

Rodrigo Siqueira-Batista | Eugênio Silva
Elen Nas | Oswaldo Motta

DEUS EX MACHINA

Fronteiras bioéticas
das neurociências,
inteligência artificial,
nanotecnologia e
robótica.

Rodrigo Siqueira-Batista | Eugênio Silva
Elen Nas | Oswaldo Motta

DEUS EX MACHINA

Fronteiras bioéticas
das neurociências,
inteligência artificial,
nanotecnologia e
robótica.

FUNDAÇÃO EDUCACIONAL SERRA DOS ÓRGÃOS – FESO

Antônio Luiz da Silva Laginestra
Presidente

Jorge Farah
Vice-Presidente

Luiz Fernando da Silva
Secretário

Carlos Alberto Oliveira Ramos da Rocha
José Luiz da Rosa Ponte
Paulo Cezar Wiertz Cordeiro
Wilson José Fernando Vianna Pedrosa
Vogais

Luis Eduardo Possidente Tostes
Direção Geral

Michele Mendes Hiath Silva
Direção de Planejamento

Solange Soares Diaz Horta
Direção Administrativa

Fillipe Ponciano Ferreira
Direção Jurídica

CENTRO UNIVERSITÁRIO SERRA DOS ÓRGÃOS – UNIFESO

Verônica Santos Albuquerque
Reitora

Marcelo Siqueira Maia Vinagre Mocarzel
Direção de Pós-Graduação, Pesquisa e Extensão

Mariana Beatriz Arcuri
Direção Acadêmica de Ciências da Saúde

Vivian Telles Paim
Direção Acadêmica de Ciências e Humanas e Tecnológicas

Anne Rose Alves Federici Marinho
Direção de Educação a Distância

HOSPITAL DAS CLÍNICAS COSTANTINO OTTAVIANO – HCTCO

Rosane Rodrigues Costa
Direção Geral

CENTRO EDUCACIONAL SERRA DOS ÓRGÃOS – CESO

Roberta Franco de Moura Monteiro
Direção

CENTRO CULTURAL FESO PROARTE – CCFP

Edenise da Silva Antas
Direção

Copyright© 2026
Direitos adquiridos para esta edição pela Editora UNIFESO

EDITORA UNIFESO

Comitê Executivo

Marcelo Siqueira Maia Vinagre Mocarzel (Presidente)

Conselho Editorial e Deliberativo

Marcelo Siqueira Maia Vinagre Mocarzel

Mariana Beatriz Arcuri

Verônica dos Santos Albuquerque

Vivian Telles Paim

Assistente Editorial

Matheus Moreira Nogueira

Formatação

Matheus Moreira Nogueira

Capa

Gerência de Comunicação

D495 Deus ex machina : fronteiras bioéticas das neurociências, inteligência artificial, nanotecnologia e robótica / Rodrigo Siqueira-Batista ... [et al.]. – Teresópolis: Editora Unifeso, c2026.
498 p. : il. color.

ISBN 978-65-5320-050-0

1. Bioética. 2. Consciência. 3. Inteligência Artificial. 4. Neurociências.
5. Nanotecnologia. 6. Robótica. I. Siqueira-Batista, Rodrigo. II. Título.

CDD 174.957

EDITORA UNIFESO

Avenida Alberto Torres, nº 111

Alto - Teresópolis - RJ - CEP: 25.964-004

Telefone: (21) 2641-7184

E-mail: editora@unifeso.edu.br

Endereço Eletrônico: <http://www.unifeso.edu.br/editora/index.php>

ORGANIZADORES

Rodrigo Siqueira-Batista

Professor Titular do Departamento de Medicina e Enfermagem,
Universidade Federal de Viçosa.

Professor Titular da Escola de Medicina,
Faculdade Dinâmica do Vale do Piranga.

Bolsista de Produtividade em Pesquisa do Conselho Nacional de
Desenvolvimento Científico e Tecnológico – Nível 2.

Docente Permanente do Programa de Pós-graduação em Bioética, Ética
Aplicada e Saúde Coletiva, Universidade Federal do Rio de Janeiro.

Docente Permanente do Programa de Pós-graduação em Ensino de Ciências
da Saúde e do Ambiente, Faculdade Dinâmica do Vale do Piranga.

Pós-doutorado no Laboratório de Física Experimental de Altas Energias,
Centro Brasileiro de Pesquisas Físicas.

Pós-doutorado no Departamento de Ecologia do Instituto de Biologia,
Universidade Federal do Rio de Janeiro.

Doutor em Ciências (Saúde Pública), Fundação Oswaldo Cruz.

Mestre em Medicina (Doenças Infecciosas e Parasitárias), Universidade
Federal do Rio de Janeiro.

Mestre em Filosofia, Pontifícia Universidade Católica do Rio de Janeiro.

Especialista (Residência Médica) em Doenças Infecciosas e Parasitárias,
Universidade Federal do Rio de Janeiro.

Diplomado em Medicina, Universidade do Estado do Rio de Janeiro.

Diplomado em Filosofia, Universidade do Estado do Rio de Janeiro.

Diplomado em Matemática (Licenciatura e Bacharelado), Universidade
Estácio de Sá.

Eugênio Silva

Professor Associado do Departamento de Computação, Faculdade de Ciências Exatas e Engenharias, Universidade do Estado do Rio de Janeiro.

Professor Titular do Curso de Ciência da Computação, Centro Universitário Serra dos Órgãos.

Pesquisador do Grupo de Tecnologia em Energia e Petróleo, Pontifícia Universidade Católica do Rio de Janeiro.

Docente Permanente do Programa de Pós-graduação em Bioética, Ética Aplicada e Saúde Coletiva, Universidade do Estado do Rio de Janeiro.

Doutor em Engenharia Elétrica, com ênfase em Métodos de Apoio à Decisão, Pontifícia Universidade Católica do Rio de Janeiro.

Mestre em Sistemas e Computação, Instituto Militar de Engenharia.

Diplomado em Ciência da Computação, Universidade Federal de Ouro Preto.

Elen Nas

Pós-doutora pela Fundação Casa de Rui Barbosa.

Pós-doutora na Cátedra Oscar Sala do Instituto de Estudos Avançados, Universidade de São Paulo.

Realizou parte do doutorado no Departamento de Informática da Universidade da Califórnia, Irvine, e do Departamento de Filosofia, Monash University.

É membro dos grupos internacionais de pesquisadores Tierra Comun (<https://www.tierracomun.net/>) e AI4Afrika (<https://ai4afrika.cargo.site/who-we-are#dr-elen-nas>).

Colaboradora do ArtSciLab da Escola de Artes, Humanidades e Tecnologia da Universidade do Texas em Dallas, da Rede Lavits (Rede Latino-americana de estudos sobre vigilância, tecnologia e sociedade), do Laboratório de Filosofia Pop da UNIRIO e da Rede Interinstitucional de Pesquisadores BraiNNIAC (ensino, pesquisa e extensão em Neurociências, Nanotecnologia, Inteligência Artificial e Cognição).

Doutora em Bioética, Ética Aplicada e Saúde Coletiva, Universidade Federal do Rio de Janeiro.

Mestre em Design, Pontifícia Universidade Católica do Rio de Janeiro.

Bacharel em Ciências Sociais, Universidade Federal Fluminense.

Oswaldo Motta

Professor Visitante do Centro de Ciências da Saúde, Núcleo de Bioética e Ética Aplicada, Universidade Federal do Rio de Janeiro.

Integrante do grupo de pesquisa Ethos: (bio)ética e as fronteiras do cuidado certificado junto ao CNPq pela Universidade Federal de Viçosa.

Membro da Rede BraiNNIAC (ensino, pesquisa e extensão em Neurociências, Nanotecnologia, Inteligência Artificial e Cognição),
Universidade Federal de Viçosa.

Colaborador do SpaceExtreme Environment Research Center (Centro de Pesquisa Espacial e Ambiental Extremo), Universidade Federal de Ciências da Saúde de Porto Alegre.

Pesquisador do Laboratório de Bioética Clínica,
Universidade Federal do Rio de Janeiro.

Pós-Doutorado no Departamento de Medicina e Enfermagem,
Universidade Federal de Viçosa.

Doutor em Bioética, Ética Aplicada e Saúde Coletiva,
Universidade Federal do Rio de Janeiro.

Mestre em Bioética, Ética Aplicada e Saúde Coletiva,
Universidade Federal do Rio de Janeiro.

Especialista em Ciência Sociais, Instituto Mineiro de Educação Superior.

Especialista em Ciência de Dados e Big Data Analytics,
Instituto Mineiro de Educação Superior.

SUMÁRIO

APRESENTAÇÃO.....	10
PREFÁCIO	13
1. A FÍSICA CONTEMPORÂNEA E O PROBLEMA DA CONSCIÊNCIA (OU SOBRE A POSSIBILIDADE DE UMA <i>TEORIA FÍSICA UNIFICADA DA CONSCIÊNCIA</i>)	17
2. FUNDAMENTOS NEUROBIOLÓGICOS DA CONSCIÊNCIA: IMPLICAÇÕES PARA O DEBATE SOBRE O LIVRE-ARBITRÍO	66
3. DETECTANDO A LUZ DA EXPERIÊNCIA: O CAMINHO PARA UMA MEDIDA OBJETIVA DOS ESTADOS DE CONSCIÊNCIA	108
04. AUTOPSYCIBOTS: MÁQUINAS AUTOCONSCIENTES SÃO DIGNAS DE CONSIDERAÇÃO ÉTICA? ENTRE O PENSAMENTO DE KANT E A (BIO)ÉTICA PARA TODOS OS SERES.....	137
5. AS DUAS VIDAS DE ALAN FOSTER: UM EXPERIMENTO MENTAL SOBRE IDENTIDADE PESSOAL E EDIÇÃO GENÉTICA	170
6. COMPUTAÇÃO AFETIVA APLICADA	193
7. SOCIEDADE AUTÔMATA: A MÁQUINA NO CONTROLE.....	226
8. SAÚDE, <i>BIG DATA</i> E BIOÉTICA.....	255
9. AVANÇOS DA INTELIGÊNCIA ARTIFICIAL NA FERTILIZAÇÃO <i>IN VITRO</i> : BREVES REFLEXÕES	274
10. APOIO COMPUTACIONAL À TOMADA DE DECISÃO EM TERAPIA INTENSIVA: QUESTÕES BIOÉTICAS.....	283
11. EXPERIMENTAÇÃO <i>IN SILICO</i> : QUESTÕES BIOÉTICAS.....	301
12. NANOCOMPUTAÇÃO COM DNA: PRINCÍPIOS, APLICAÇÕES E BIOÉTICA.....	324

13. ARTELLIGENT: INTELIGÊNCIA ARTIFICIAL NA ARTE.....	350
14. O CÉREBRO E A NEUROÉTICA SOB O OLHAR DO DIREITO	380
15. ANÁLISE DA RESPONSABILIDADE JURÍDICA CIVIL QUANDO DO USO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL NO PROCESSO DE TOMADA DE DECISÃO EM CONFLITOS (BIO)ÉTICOS	391
DIÁLOGO 01. SOBRE CIÊNCIA E MATEMÁTICA – QUESTÕES EPISTEMOLÓGICAS E ÉTICAS	442
DIÁLOGO 02. BIOÉTICA, TECNOLOGIAS EMERGENTES E ÉTICA DAS VIRTUDES.....	454
DIÁLOGO 03. BIOPOLÍTICA E OS VALORES OCULTOS NOS SISTEMAS.....	475
COLABORADORES	491

APRESENTAÇÃO

“A técnica é um destino que os seres humanos estão chamados a enfrentar e compreender, pois ela molda a maneira como nos relacionamos com o mundo e nos revela a nossa essência como seres-no-mundo.”¹

(Martin Heidegger)

As sociedades laicas e plurais contemporâneas testemunharam, nos últimos 50 anos, grandes avanços tecnológicos que têm impactado significativamente a vida humana. A revolução digital transformou os modos de comunicação entre as pessoas, permitindo uma conectividade global sem precedentes. A internet e as redes sociais emergiram como ferramentas poderosas para compartilhar informações e promover interações sociais. Ademais, a disseminação de dispositivos móveis, como smartphones, trouxe consigo a conveniência e a acessibilidade da tecnologia para um público mais amplo.

O cenário de desenvolvimento técnico-científico tem florescido em diferentes campos do conhecimento, particularmente nas investigações dirigidas à inteligência artificial (IA), neurociências e nanotecnologia. A IA tem desempenhado um papel cada vez mais proeminente, impulsionando avanços na automação, análise de dados e tomada de decisões. A neurociência, por sua vez, tem proporcionado uma compreensão mais profunda do funcionamento do cérebro humano, abrindo portas para o desenvolvimento de terapias mais eficazes e intervenções em vários campos. A nanotecnologia, com suas aplicações em diversas áreas, desde a medicina até o meio ambiente, tem o potencial de revolucionar a forma como os humanos interagem com o mundo ao redor. No entanto, enquanto esses avanços oferecem muitas oportunidades promissoras, também trazem desafios éticos e sociais complexos que exigem reflexão cuidadosa e regulamentação adequada.

1 HEIDEGGER, M. A questão da técnica. Trad. Marco Aurélio Werle. *Scientiae Studia*. São Paulo, v. 5, n. 3, p. 375-398, 2007.

Os colaboradores desta obra mergulham nas fronteiras cada vez mais tênues entre ciência e ética, explorando os avanços revolucionários na neurociência, nanotecnologia e inteligência artificial. Ao examinar as interseções dessas disciplinas, os autores investigam o impacto das inovações tecnológicas na concepção de humanidade e identidade pessoal. Ao confrontar as promessas e os perigos desses avanços, “Deus Ex Machina” oferece visão panorâmica dos desafios enfrentados pelas sociedades modernas e abre espaço para um debate crucial sobre como utilizar essas tecnologias de forma responsável e compassiva.

As discussões interdisciplinares que envolvem bioética, neurociências, nanotecnologia e inteligência artificial oferecem um vasto campo de possibilidades para o desenvolvimento de estudos futuros. Um aspecto promissor é a investigação das implicações éticas da manipulação do sistema nervoso central, explorando questões como privacidade, autonomia e identidade. Além disso, a integração de nanotecnologia e inteligência artificial na medicina abrem caminhos para terapias personalizadas e avanços significativos no tratamento de doenças neurológicas. Estudos sobre a interface neurônio biológico e neurônio artificial podem aprofundar nossa compreensão da mente humana e levar a avanços na comunicação e reabilitação de indivíduos. Também é crucial investigar os efeitos sociais e psicológicos da adoção generalizada de tecnologias de IA, como algoritmos de tomada de decisão e sistemas autônomos. O aprimoramento humano por meio de tecnologias emergentes é outro tópico relevante, com questões sobre igualdade, justiça e distribuição de acesso às melhorias cognitivas e físicas. Em suma, as perspectivas de estudos futuros nesse campo são amplas e multifacetadas, oferecendo oportunidades de explorar os limites éticos, os avanços científicos e as implicações sociais dessas tecnologias em constante evolução.

É importante observar que os capítulos deste livro foram escritos no ano de 2021. De lá até a publicação, a sociedade vivenciou o surgimento de avanços tecnológicos muito significativos. Para os próximos anos, as projeções são otimistas e apontam que muitos outros avanços virão, reafirmando a dinâmica acelerada da inovação na sociedade. Esta obra reflete o contexto do momento em que foi produzida, porém convida o leitor a explorar as ideias apresentadas sob uma perspectiva atemporal, reconhecendo a evolução contínua da tecnologia e suas implicações, quiçá – heideggerianamente – como destino humano.

É com grande entusiasmo que são dadas as boas-vindas aos leitores! Apresenta-se o convite a você para trilhar essa jornada de descoberta e reflexão. Apreciar os capítulos deste livro é mergulhar em um oceano de conhecimento, onde as interseções entre bioética, nanotecnologia, neurociência e inteligência artificial revelam a complexidade e o potencial da presente era tecnológica. Que essa leitura inspire e alimente a paixão pelo conhecimento e pela busca por um futuro ético e promissor.

Cordialmente,
Os organizadores.

PREFÁCIO

O século XXI trouxe consigo uma onda de inovação sem precedentes, na qual tecnologias emergentes como neurociência, inteligência artificial (IA), nanotecnologia e robótica estão remodelando não apenas o campo da saúde, mas também o próprio tecido da nossa sociedade. Nesse sentido, o século XXI é o século das tecnologias disruptivas, caracterizadas não apenas por substituir tecnologias previamente existentes, mas também por criar novos mercados e transformar indústrias, provocando mudanças tremendas em nossa forma de vida. Essas novas tecnologias, que até poucas décadas atrás eram apenas o motor imaginário de criadores literários e visionários, hoje estão gerando avanços concretos que impactam não apenas o progresso na saúde, mas também a arte, a economia, a educação e, fundamentalmente, o conhecimento que temos sobre nós mesmos.

Sob essa perspectiva positiva, o desenvolvimento dessas tecnologias representa um triunfo do intelecto humano, uma conquista coletiva que impulsionou nossa capacidade de compreender, intervir e transformar a realidade de formas que antes pareciam impossíveis. A neurociência está desvendando os segredos do cérebro, permitindo-nos aprofundar nos mistérios da mente, da consciência, das emoções, do comportamento, da saúde mental e da identidade pessoal. A inteligência artificial está revolucionando a maneira como processamos informações, aprendemos e tomamos decisões. Ela não se limita apenas a transformar o mundo dos negócios e da saúde, mas também contribui para o campo artístico. Artistas de todo o mundo experimentam ferramentas de IA para criar obras que combinam a criatividade humana com a capacidade analítica e computacional das máquinas.

Na medicina, a nanotecnologia e a robótica abriram caminho para terapias que não apenas curam doenças, mas também previnem seu aparecimento e, em alguns casos, permitem a regeneração de tecidos e órgãos, prolongando a esperança e a qualidade de vida de milhões de pessoas. Além disso, a robótica nos proporciona a possibilidade de projetar

dispositivos que não apenas superam as limitações humanas, mas também permitem restaurar funções perdidas ou criar novas formas de interação entre pessoas e máquinas.

É inegável, portanto, que todos esses avanços impulsionaram de maneira notável o progresso social e cultural, mas não podemos esquecer os enormes desafios que também nos apresentam. Em meio a essa revolução tecnológica, surgem questões éticas e jurídicas que, em muitos casos, são tão novas e complexas quanto as próprias tecnologias. Enfrentamos dilemas que até recentemente eram inimagináveis, e, em algumas ocasiões, o arcabouço ético e legal existente parece inadequado para abordar os problemas emergentes. Neste livro, os autores, especialistas em áreas como ciência da computação, medicina, direito e bioética, abordam alguns desses desafios, oferecendo análises aprofundadas e reflexões críticas que nos convidam a considerar as implicações desses avanços sob uma perspectiva ética, jurídica e social.

Um dos principais desafios que esse novo cenário nos impõe é a redefinição dos limites do ser humano e da natureza da inteligência. Com a neurociência e a IA, estamos começando a repensar o que significa ser consciente e autônomo. No passado, esses conceitos estavam intimamente ligados a um conhecimento limitado do cérebro; no entanto, o desenvolvimento de agentes inteligentes artificiais, capazes de aprender e tomar decisões, nos leva a questionar se esses novos agentes deveriam ter algum tipo de status ético ou legal. Até que ponto deveríamos permitir o desenvolvimento desses agentes inteligentes? Eles terão, algum dia, direitos semelhantes aos dos seres humanos?

Além disso, a questão da autonomia pessoal é afetada por esses avanços. À medida que a neurociência penetra mais profundamente na compreensão do cérebro, vislumbram-se tecnologias capazes de intervir na tomada de decisões e no comportamento humano. As possibilidades são promissoras: tratamentos para doenças mentais, modulação da memória ou controle de estados emocionais. Mas também surgem preocupações: como garantir que essas intervenções não sejam usadas de maneira coercitiva ou para controlar o comportamento das pessoas além dos limites éticos aceitáveis? Em última instância, como tudo isso afetará nossa percepção do livre-arbítrio e da responsabilidade individual?

Outro desafio importante é o conceito de privacidade e propriedade no contexto da nanotecnologia e da IA. Através de dispositivos minúsculos, capazes de intervir diretamente no corpo humano, a nanotecnologia oferece oportunidades para tratamentos personalizados e diagnósticos precisos. No entanto, essa capacidade de acessar a biologia humana em níveis tão íntimos também levanta a questão de quem tem controle sobre esses dados e como será gerenciada a informação sensível obtida dos corpos dos pacientes. Paralelamente, a IA, ao processar e analisar enormes quantidades de dados de saúde, levanta questões sobre a proteção da privacidade e a possibilidade de violações de direitos fundamentais por meio da vigilância biotecnológica.

A robótica na saúde e os exoesqueletos, que restauram a mobilidade de pessoas com deficiência ou auxiliam profissionais médicos em cirurgias complexas, também trazem novas questões sobre responsabilidade e equidade. O que acontece quando um robô comete um erro durante um procedimento médico? Quem é responsável: o projetista, o programador, o hospital ou o próprio robô? E, de maneira mais ampla, como garantir que essas tecnologias sejam acessíveis a todos e não criem novas formas de exclusão social? É crucial que, enquanto celebramos os avanços nesse campo, não nos esqueçamos dos perigos da desigualdade no acesso a tecnologias que, em última instância, podem ser cruciais para a qualidade de vida das pessoas.

Nesse sentido, a humanidade enfrenta um desafio sem precedentes: não apenas o desenvolvimento e a implementação dessas tecnologias, mas também a criação de novos marcos éticos e jurídicos que possam abordar os conflitos inéditos que estão surgindo. A tecnologia avança a uma velocidade vertiginosa, mas nossa capacidade de nos adaptarmos às suas consequências sociais, políticas e morais deve acompanhar esse progresso. O risco não é apenas ficarmos para trás em termos tecnológicos, mas também em termos éticos.

Portanto, este livro tem um valor inestimável como convite e poderoso instrumento para refletirmos sobre o futuro que estamos construindo. À medida que integramos a neurociência, a IA, a nanotecnologia e a robótica em nossas vidas, devemos garantir que as decisões que tomamos sobre seu uso sejam guiadas por uma reflexão ética profunda. É um lembrete de que o verdadeiro progresso não reside apenas na inovação técnica, mas também em nossa capacidade de gerir responsavelmente as implicações humanas e sociais desses avanços.

Este volume reúne vozes especializadas de diferentes áreas que nos convidam a refletir sobre os limites e possibilidades desse futuro em construção. Eles nos desafiam a imaginar um mundo onde o desenvolvimento tecnológico não esteja apenas a serviço da eficiência e do poder, mas também do bem-estar humano, da justiça e do respeito pela vida. Um futuro onde o conhecimento de nós mesmos, potencializado pela tecnologia, não nos desumanize, mas nos permita alcançar novas formas de compreensão e solidariedade.

Francisco Lara

Professor de Filosofia Moral do Departamento de Filosofia da
Universidade de Granada - Espanha

01.

A FÍSICA CONTEMPORÂNEA E O PROBLEMA DA CONSCIÊNCIA (ou sobre a possibilidade de uma *Teoria Física Unificada da Consciência*)

Rodrigo Siqueira-Batista

Mathias Viana Vicari

Adilson Costa da Silva

José Abdalla Helayël Neto

1. INTRODUÇÃO

O problema da consciência é, provavelmente, o principal desafio epistêmico com o qual o *Homo sapiens* tem se defrontado ao longo da sua (breve) história. De fato, trata-se de um dos mais intrigantes “mistérios” científicos da atualidade, cuja *dimensão labiríntica*² pode ser compreendida por algumas características muito específicas da temática em investigação, relacionadas principalmente (i) à definição do termo e (ii) à ambiguidade da posição ocupada no processo de conhecer.

O primeiro problema diz respeito à constatação de que o vocábulo *consciência* é extremamente polissêmico, sendo utilizado em diferentes contextos e com díspares significados, como, por exemplo: (i) o âmbito moral ou ético-político (*conscience*, em inglês; συνείδησις, em grego), de acordo com o qual é possível discernir o “certo” e o “errado”; (ii) a *vigília* (*wakefulness*, em inglês), em par de oposto com o sono ou o estado de coma; (iii) a *atenção* (*awareness*, em inglês) de um dado organismo vivo ao entorno (meio ambiente), o qual recebe díspares informações (*inputs*); e (iv) o *autoconhecimento* ou *autopercepção* da existência (*consciousness*, em inglês), com marcante articulação com as concepções de tempo (passado, presente

2 A referência labiríntica, aqui, não é gratuita. Há interpretações do famoso mito do *Minotauro*, como uma luta pela posse da consciência – e da autoconsciência – tipicamente humana (Teseu) em oposição à *natureza* bestial tipicamente animal (Minotauro). A saída do Labirinto, nesses termos, só pode ser alcançada pelo fio da razão (Ariadne).

e futuro) e de memória (MARIETTA JUNIOR, 1970; SILVA, *et al.*, 2003; SILVA, 2018). Com efeito, pode-se reconhecê-la, a consciência, como a “possibilidade de dar atenção aos próprios modos de ser e às próprias ações, bem como exprimi-las com a linguagem” e “uma relação intrínseca do homem, ‘interior’ ou ‘espiritual’, pela qual ele pode conhecer-se de modo imediato e privilegiado” (ABBAGNANO, 2003, p. 185), ou “os estados subjetivos de sensibilidade (*sentience*) ou ciência (*awareness*) que começam quando uma pessoa acorda de manhã, depois de um sono sem sonhos, e se estendem por todo o dia até que vá dormir à noite, entre em coma, morra ou de algum outro modo se torne, digamos, ‘inconsciente’” (SEARLE, 2010, p. 1-2), ou ainda como a “parte da mente relacionada ao sentido manifesto do self e do conhecimento. A mente não é apenas consciência, e pode haver mente sem consciência, como descobrimos em pacientes que possuem uma, mas não a outra” (DAMÁSIO, 2000, p. 47). As (in)definições terminológicas da consciência – ou do que *ela é* – podem chegar à situação de se reconhecer que a “consciência é um fenômeno fascinante, porém elusivo; é impossível especificar o que é, o que faz ou por que evoluiu. Nada digno de ser lido foi escrito sobre ela” (SUTHERLAND, 1989, p. 90). Mas, este não é o único problema...

A segunda ordem de questões, atinentes à abordagem científica da consciência, diz respeito à sua situação simultânea de *sujeito* e *objeto* do conhecimento, em uma ambiguidade de difícil superação. Em última análise, trata-se de reconhecer que a consciência pode ser acessada em *primeira pessoa*, como *sujeito* (ou seja, perspectiva a qual o leitor dessa Introdução está *experimentando*, nesse momento) ou em *terceira pessoa*, como *objeto* (ou seja, como fração da realidade que pode ser submetida ao escrutínio de uma mente cognoscente, em prol da produção de conhecimento). A consciência, assim, seria acessível pelo próprio indivíduo auto-consciente – o que a torna “uma propriedade irreduzivelmente subjetiva, ‘inanalísável’, ‘indecomponível’ (não-relacional)” (TOURINHO, 2012, p. 489) – ou investigável como fenômeno da natureza – ou da biologia/neurobiologia – e, assim, passível de ser causalmente explicada, por meio de modelos cientificamente válidos. Nesse último caso, a consciência seria o *resultado* do cérebro (ou, mais propriamente, do corpo) humano, da sua arquitetura neural, a qual é responsável pela existência de funções cerebrais complexas que envolvem distintas estruturas e vias do sistema

nervoso central (SNC) – bem como de suas múltiplas conexões funcionais –, as quais podem ser descritas de um ponto de vista científico.

A abordagem da consciência em *terceira pessoa* tem sido proposta a partir de diferentes áreas do conhecimento, as quais empregam díspares métodos de pesquisa. É possível que a mais arcaica seja a investigação dos fenômenos neurobiológicos subjacentes aos processos conscientes, cuja ênfase está presente nos trabalhos de pesquisadores como Antônio Damásio. Outra modalidade bastante frequente é a ideia de modelagem *in silico* da consciência, no campo da neurociência computacional, principalmente na esfera da Inteligência Artificial (IA) forte (ver Capítulo 4). Uma terceira vertente tem sido proposta a partir da física contemporânea, especialmente da Mecânica Quântica (MQ) e do Eletromagnetismo (EMG). Em termos da MQ, o matemático John Von Neumann foi – provavelmente – o primeiro autor a propor um papel importante dos fenômenos quânticos para a consciência. Outras ideias inscritas nesse mesmo rol de possibilidades dizem respeito à perspectiva (PESSOA Jr., 1994) (i) do cérebro funcionar como um *computador quântico* (DEUTSCH, 1992), (ii) da condensação de Bose estar na base dos processos conscientes (FRÖHLICH, 1968), (iii) da descrição das funções cerebrais como um “campo dendrítico”, em termos da teoria quântica de campos (JIBU & YASUE, 1991), e (iv) da teoria probabilística da *exocitose* de neurotransmissores na fenda sináptica (BECK & ECCLES, 1992). No que concerne ao EMG, deve ser comentado o papel das funções de Green (COURANT; HILBERT, 1953-1962) – as quais albergam possibilidades de construção de uma apreciação do fenômeno da consciência, via efeitos não-lineares – e (b) a equação de campo proposta por Jirsa e Haken (1996), a qual é útil para a descrição do comportamento espaço-temporal da atividade eletromagnética cerebral. Outros direcionamentos factíveis referem-se à possibilidade trabalhar em termos da Teoria da Relatividade Especial (TRE) e da Teoria da Relatividade Geral (TRG) – os modelos neurobiológicos relativísticos – como fundamento para uma proposta explicativa do funcionamento do cérebro e, quiçá, da consciência (SIEB, 2016).

O presente capítulo, com base nessas preliminares considerações, tem por objetivo construir uma reflexão sobre a abordagem física da consciência – enfatizando a MQ, o EMG, a TER e a TRG – com a abertura, todavia, para a eventual incorporação de outros modelos científicos, quiçá em prol de uma *Teoria Física Unificada da Consciência* (TFUC).

2. MECÂNICA QUÂNTICA E CONSCIÊNCIA

A ênfase dessa primeira seção será na apresentação da *hipótese quântica* da consciência, a partir de uma (i) breve exposição dos fundamentos da MQ, a qual subsidiará a discussão (ii) das ideias de David Bohm e, na sequência, (iii) das propostas de Hameroff e Penrose.

2.1. Fundamentos de MQ

A passagem do século XIX para o século XX trouxe significativos problemas para a Física. Naquela ocasião, os cientistas trabalhavam intensamente em importantes questões nas quais a Mecânica Clássica Newtoniana falhava – e ainda falha – em explicar, com destaque para diferentes aspectos da estrutura da matéria e da radiação do corpo negro. Pesquisas nas áreas de termodinâmica e eletromagnetismo desenvolviam-se, de modo acelerado, na tentativa de descrever a realidade física, a qual, heracliticamente, *insistia* em se ocultar.

A história da MQ pode ser brevemente recontada de distintas maneiras, mas, presentemente, partir-se-á do dia 14 de dezembro de 1900, quando Max Planck, na esperança de obter uma explicação teórica para o problema da *catástrofe ultravioleta* (PLANCK, 1901; POLKINGHORNE, 2016), fez uma hipótese que chamou de “um ato de desespero”. O referido autor relacionou o comportamento dos osciladores (átomos) nas paredes do corpo negro e sua interação com a radiação eletromagnética dentro da cavidade, sugerindo que a troca de energia dessa interação, não se daria de maneira contínua, mas em quantidades discretas. Desse modo, pressupunha que essa energia fosse quantizada em porções que passou a denominar de *quanta* (plural de *quantum*). Nesses termos, o *quantum* de energia irradiado é necessariamente proporcional à frequência da radiação, destacando-se que os valores possíveis da energia de um oscilador (elétron) de frequência, são:

$$E = nh\nu \quad n = 0, 1, 2, 3, \dots$$

E = *quantum* de energia irradiado

n = número quântico (inteiros positivos)

h = constante de Planck ($\approx 4,136 \times 10^{-15}$ eV.s $\approx 6,626 \times 10^{-34}$ J.s)

ν = frequência de oscilação

Albert Einstein, cinco anos após – em 1905, mesmo ano em que explica o Movimento Browniano e apresenta sua Teoria da Relatividade Especial – publica “*Um ponto de vista heurístico a respeito da produção e transformação da luz*”, artigo no qual propõe que a energia não é apenas adicionada em unidade de *quantum*, mas que a própria luz (radiação) existe em porções quantizadas, em concordância com a teoria de Planck (EINSTEIN, 1905a, 1905b; HEWITT, 2011; ALONSO *et al.*, 2012), o que foi comprovado por Millikan em 1914 (MILLIKAN, 1914):

$$K_{máx} = eV_F = h\nu - W$$

$K_{máx}$ = energia cinética máxima do elétron ao escapar de um metal

e = carga do elétron

V_F = potencial de corte (ou de fretamento)

h = constante de Planck

ν = frequência de oscilação

W = função de trabalho (energia de ligação do elétron ao átomo – metal)

De acordo com esta concepção, depreende-se que um *quantum* de luz – ou um corpúsculo de luz, fóton (o nome fóton aparece, em 1926, no trabalho de Lewis), partículas de massa nula – pode transferir toda sua energia a um único elétron (LEWIS, 1926a, 1926b). Quase duas décadas depois da proposição de Einstein, em 1923, Compton confirmou a natureza corpuscular para a luz (radiação), observando a colisão elástica de um fóton com um elétron (o experimento analisou o espalhamento de raios X (fóton) por um alvo de grafite; este alvo possui muitos elétrons poucos ligados, que podem ser considerados como se estivessem livres) (COMPTON, 1923). No ano seguinte, Louis De Broglie – durante a preparação de sua tese de doutorado –, partindo da ideia de que a natureza deve ser simples e simétrica, propõe a existência de ondas de matéria, devido ao comportamento dual da radiação (onda ↔ partícula). Esta dualidade também deveria se aplicar à matéria. De Broglie associou o momento e energia dos fótons ao momento e energia das partículas (não relativísticas) e, com isso, obteve uma relação para descrever o comprimento de onda de uma partícula, em função de sua energia ou momento (DE BROGLIE, 1924a, 1924b; EISBERG *et al.*, 1979; NUSSENVEIG, 2014):

$$E = hf = \hbar\omega \quad \text{e} \quad p = \frac{h}{\lambda} = \hbar k$$

sendo

$$p^2 = 2mE$$

$$\lambda = \frac{h}{p} = \frac{h}{\sqrt{2mE}}$$

E = energia da partícula

h = constante de Planck

f = energia total

$\hbar$ = forma de Dirac para a constante de Planck

p = módulo do momento linear da partícula

λ = comprimento de onda de De Broglie

ω = frequência angular

k = número de onda

m = massa da partícula

A hipótese de De Broglie foi testada por Davisson, Germer e G. Thomson, filho do J. J. Thomson, em 1927. Os autores elaboraram um aparato experimental no qual incidiram um feixe de elétrons sobre um monocristal de níquel; observaram, então, que havia difração de elétrons, fenômeno que juntamente com a interferência já eram reconhecidamente característicos de entes ondulatórios (DAVISSON & GELMER, 1927). Outros experimentos foram desenvolvidos e comprovaram a natureza ondulatória para os nêutrons e para toda a matéria. De fato, a dualidade (onda $\leftrightarrow$ partícula), se aplica tanto à radiação quanto à matéria (EISBERG *et al.*, 1979; HEWITT, 2011; NUSSENZVEIG, 2014).

Em 1926, Schrödinger foi convidado a apresentar uma palestra sobre as ondas de matéria de De Broglie, na Escola Politécnica de Zurique, organizada por Peter Debye (CASSINELLO; GÓMEZ, 2012). Questionado se existiria uma onda associada aos elétrons e como seria esta equação, Schrödinger desenvolveu a Mecânica Quântica Ondulatória, a qual descreve as energias

dos átomos, tratando-os como ondas e não como partículas (BLOCH, 1976).³ Sua equação – a qual, em última análise, equivalia à mecânica matricial publicada cerca de um ano antes por Heisenberg (1925) – permite calcular a função de onda descrevendo, em probabilidades, uma possível localização para o elétron em um dado instante, fundamentando a MQ (EISBERG *et al.*, 1979; FEYNMAN, 2008; SCHRÖDINGER, 1926a, 1926b). Com efeito, a amplitude $\psi(x, y, z, t)$, para um elétron de massa m , se movendo em um potencial $V(x, y, z, t)$ – o qual é equivalente ao potencial escalar que relaciona a energia total de todas as partículas – obedece a equação diferencial:

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi + V\psi$$

Esta é a equação de Schrödinger dependente do tempo, em uma forma tridimensional. Observa-se que essa função de onda, na forma abordada, depende das três coordenadas, incorporando a possibilidade de explicar todos os fenômenos atômicos – diferentemente da equação de onda usual –, menos aqueles que relacionam a MQ e a Teoria da Relatividade Restrita (TRR) ou Teoria Especial da Relatividade (TRE), os quais foram posteriormente desenvolvidos por Paul Dirac, nos artigos intitulados: “*The Quantum Theory of the Electron – Part I*” (DIRAC, 1928a) e “*The Quantum Theory of the Electron – Part II*” (DIRAC, 1928b), nos quais o físico britânico elabora uma descrição relativística para o *spin* do elétron, unificando a MQ e a TRE em sua famosa Equação de Dirac:

-
- 3 A história da formulação da MQ ondulatória por Erwin Schrödinger é bastante interessante, especialmente ao se recuperar a descrição de Bloch (1976), tal qual apresentada por Cassinello e Gómez (2012, p. 34): “*Em uma ocasião, no final do colloquium, ouvi Debye dizer a Schrödinger algo como: ‘Schrödinger, o senhor não está trabalhando em nenhum problema muito importante. Por que não nos conta, em alguma sessão, algo sobre a tese de De Broglie, que parece haver atraído certa atenção?’*”. Assim, em um dos colóquios seguintes, Schrödinger fez uma magnífica exposição de como De Broglie associava uma onda com uma partícula e como podia obter as regras de quantização de Bohr e Sommerfeld exigindo que se incluísse um número inteiro de ondas em uma órbita estacionária. Quando terminou, Debye sinalizou que esse modo de falar parecia um tanto infantil. Como aluno de Sommerfeld, ele havia aprendido que, para tratar adequadamente com ondas, era preciso ter uma equação de ondas. Isso parecia bastante evidente e aparentemente não causou grande impressão, mas Schrödinger mais tarde refletiu sobre essa ideia. Algumas semanas depois, ele deu outra palestra no colloquium e começou dizendo: ‘*Meu colega Debye sugeriu que deveria haver uma equação de ondas; pois bem, eu encontrei uma*’. E, então, ele nos contou essencialmente aquele que seria, com o título ‘*Quantização como um problema de autovalores*’, o primeiro artigo de uma série publicada em *Annalen der Physik*.

$$i\hbar\gamma^\mu\partial_\mu\Psi - mc\Psi = 0$$

usando a notação de barra de Feynman:

$$\not{\partial} \stackrel{def}{=} \gamma^\mu\partial_\mu$$

temos em sua simples forma a Equação de Dirac:

$$(i\not{\partial} - m)\psi = 0$$

A noção de quantum – e seus desdobramentos teóricos e experimentais – foram essenciais para o nascimento da MQ, vislumbrando-se que a natureza em nível microscópico era regida por leis próprias, as quais não são imediatamente compreensíveis à luz da Física Clássica. Nesse sentido, Heisenberg (1927) afirmou a impossibilidade, em uma experiência, da simultânea medição (i) do valor exato de um componente do momento linear (ΔP_x) de uma partícula e (i) do valor exato da coordenada da posição correspondente:

$$\Delta P_x \cdot \Delta x \geq \frac{\hbar}{2}$$

Com efeito, o produto na incerteza do momento (ΔP_x) e a incerteza da posição (Δx) não pode ser menor que $\frac{\hbar}{2}$, sendo $\hbar = \frac{h}{2\pi}$, o que limita as possibilidades de conhecer, ao mesmo tempo, o momento e a posição, mostrando que ao se determinar (P_x) com precisão absoluta ($\Delta P_x = 0$), nada é possível saber a respeito da posição (x) ($\Delta x = \infty$); de outro modo, ao se identificar (x) com precisão absoluta ($\Delta x = 0$), nada se pode delimitar a respeito do momento (P_x) ($\Delta P_x = \infty$). Deste modo, não se pode definir, com precisão, a trajetória de uma partícula no mundo quântico do modo que é feito na Mecânica Clássica. Para os demais componentes do momento e da posição, essa relação é também verdadeira quando medidas sob o mesmo eixo. Essa formulação passou a ser denominada Princípio da Incerteza (de Heisenberg), o qual é considerado um elemento fundamental da realidade quântica (ALONSO, 2012; EISBERG et al. 1979; HEISENBERG, 1927).

A Mecânica Clássica Newtoniana foi considerada hegemônica para a descrição dos eventos, na perspectiva do caráter mecânico-determinista da realidade, até princípios do século XX. Desde o advento da MQ, tal modelo perdeu sua hegemonia para a descrição de diferentes ordens de fenômenos, mormente aqueles atinentes ao *micromundo*. Reconhece-se, assim, um significativo deslocamento da explicação causal da realidade – em prol do reconhecimento do indeterminismo dos eventos – como analisado no parágrafo anterior, cujas conjecturas estão de acordo com a visão de Niels Bohr e, claro, do próprio Heisenberg:

Quão radical foi a mudança promovida por esse avanço da física atômica em nossa atitude perante a descrição da natureza talvez possa ser mais claramente ilustrado pelo fato de que até o princípio de causalidade, antes considerado o fundamento incontestável de toda interpretação dos fenômenos naturais, revelou-se um referencial estreito demais para abarcar as regularidades singulares que regem os processos atômicos individuais.

(BOHR, 1995, p. 326)

Mas não podemos, e é aí que a lei causal falha, explicar porque um determinado átomo se desintegra num dado momento, e não no seguinte, ou o que o faz emitir um elétron exatamente numa certa direção e não na outra. Estamos convencidos, por muitas razões, de que essa causa não existe.

(HEISENBERG, 1996, p. 326)

A visão *não determinista* da realidade subatômica causou um profundo mal-estar na comunidade científica, tornando-se famosa a frase de Einstein, “Deus não joga dados”. Desse ponto de vista, a MQ estava incompleta, carecendo de novos ajustes matemáticos, em direção àquilo que se convencionou denominar as (famosas) *variáveis ocultas*. Com efeito, o autor das Teorias da Relatividade (Especial e Geral) opunha-se à *Interpretação de Copenhagen* da MQ (IC-MQ), *não determinista* – inicialmente estruturada por Werner Heisenberg e Niels Bohr – considerando que seria necessário, outrossim, a busca pela “libertação” da incerteza dessa abordagem que fornece resultados em termos de probabilidades. Vale destacar que a IC-MQ – uma das interpretações possíveis da MQ – reconhece o caráter probabilístico do conhecimento físico sobre a intimidade da matéria (SHESTAKOVA, 2020). Deve ser pontuado, que de acordo com a IC-MQ a indeterminação é

constitutiva da realidade (*ontológica*) e não uma medida da ignorância do investigador acerca de supostas variáveis não explícitas (*epistemológica*); ou seja, o mundo quântico seria indeterminístico *em si* (BOHR, 1995).

Em 1928, como não se podia descrever a trajetória das partículas subatômicas como na Mecânica Clássica, é formulada, por Max Born, uma unificação probabilística para a dualidade onda-partícula, descrita por uma função complexa, chamada *função de onda* $\psi(x,t)$ (BORN, 1927a, 1927b). A função $P(x,t) = |\psi(x,t)|^2$ é conhecida como *densidade de probabilidade* e informa qual é a região mais provável, de um ponto de vista estatístico, de encontrar uma partícula em uma certa região do espaço, em um dado instante t . Neste caso, o *colapso da função de onda* ocorreria imediatamente após a realizada a medição da posição da partícula; de fato, sua função de onda colapsa para o ponto ou a região (probabilidade de detecção) em que ela foi encontrada. Deve ser destacado que medidas rápidas continuam detectando tal partícula nesse mesmo local, se não houver perturbação, devido ao fato da densidade de probabilidade mudar com o tempo de acordo com $\psi(x,t)$ (ALONSO *et al.*, 2012; BORN, 1927a, 1927b; EISBERG *et al.*, 1979; FEYNMAN, 2008; HEWITT, 2011).

As leis da física para o mundo microscópico diferem sobremaneira daquelas propostas para o macroscópico, o que gera alguns comportamentos consideravelmente *estranhos*. Por exemplo, as partículas no mundo quântico (microscópico) podem existir simultaneamente em dois ou mais estados/locais, de acordo com a descrição matemática dada por uma função de onda; ademais, pode ser mencionado o emaranhamento – ou entrelaçamento – quântico, no qual há unificação dos separados processos de um sistema, de modo que todos se tornam conectados – ainda que estejam (significativamente) distantes –, por uma função de onda quântica, sendo possível instantaneamente uma medida acabar por influenciar na outra; deste ponto de vista, as partes se conectam em um todo (HAMEROFF, PENROSE, 2013; WANG *et al.*, 2018).

A observação de um sistema quântico produz incertezas, o que é uma peculiaridade atribuível ao próprio ato de observar, de acordo com Bohr. Uma importante crítica articulada a essa interpretação foi elaborada por Erwin Schrödinger, o qual propõe um experimento mental – *gedankenexperiment*, em alemão, conforme terminologia proposta por Ernst Mach (MACH, 1960) –, em 1935, o qual se tornou (bastante) conhecido como o *Gato de Schrödinger* (Figura 1):

“Um gato é preso dentro de uma câmara de aço, juntamente com um “dispositivo diabólico” no interior do contador Geiger, em que há uma pequena quantidade de material radioativo, tão pequena que talvez, no decurso de uma hora, um dos seus átomos decaia, mas também, com igual probabilidade, talvez nenhum decaia; se isso acontecer, o contador irá liberar uma descarga e através de um relé soltará um martelo que irá quebrar o frasco contendo ácido cianídrico. Durante uma hora, deixamos todo o sistema isolado; então, diremos que o gato ainda vive, se nenhum átomo decaiu durante esse tempo. A função de onda do sistema como um todo iria expressar isto contendo, em si mesma, o gato vivo e o gato morto, simultaneamente, ou dispostos em partes iguais. É típico destes casos que uma indeterminação originalmente confinada ao domínio atômico venha a transformar-se numa indeterminação macroscópica, a qual pode então ser resolvida pela observação direta. Isso previne-nos de tão ingenuamente aceitarmos como válido um ‘modelo impreciso’ para representar a realidade. Em si mesma esta pode não incorporar nada de obscuro ou contraditório” (SCHRODINGER, 1935, p. 812).

O referido experimento mental – o qual aponta para a possível *incomplete* da MQ, de modo similar ao proposto por Einstein, no paradoxo EPR ou Paradoxo de Einstein-Podolsky-Rosen (conforme será comentado adiante) – representou um excelente exemplo para enfatizar a possibilidade de estados superpostos em nível quântico, o que acarretaria situações absurdas no nível macroscópico (EISBERG *et al.*, 1979; FEYNMAN, 2008; GRIFFITHS, 2005; NUSSENZVEIG, 2014). Com efeito, em termos das superposições quânticas, pode-se dizer que uma subpartícula ou um sistema quântico, existe em todos os estados quânticos possíveis, ao mesmo tempo antes de ser medido. Na formulação de Schrödinger, ao final de uma hora, a função de onda do gato pode ser matematicamente descrita do seguinte modo:

$$\psi = \frac{1}{\sqrt{2}}(\psi_{vivo} - \psi_{morto})$$

Observa-se, isto posto, a probabilidade que o gato esteja vivo e também esteja morto, em uma combinação linear de dois estados físicos possíveis, senão, de ambas as situações (ou seja, de ambos os estados) coexistindo, até que uma medição ocorra, ressaltando-se que esta, invariavelmente, interferirá no resultado (FREIRE, 2011; SANTOS *et al.*, 2011; GRIFFITHS, 2005; SCHRODINGER, 1935).

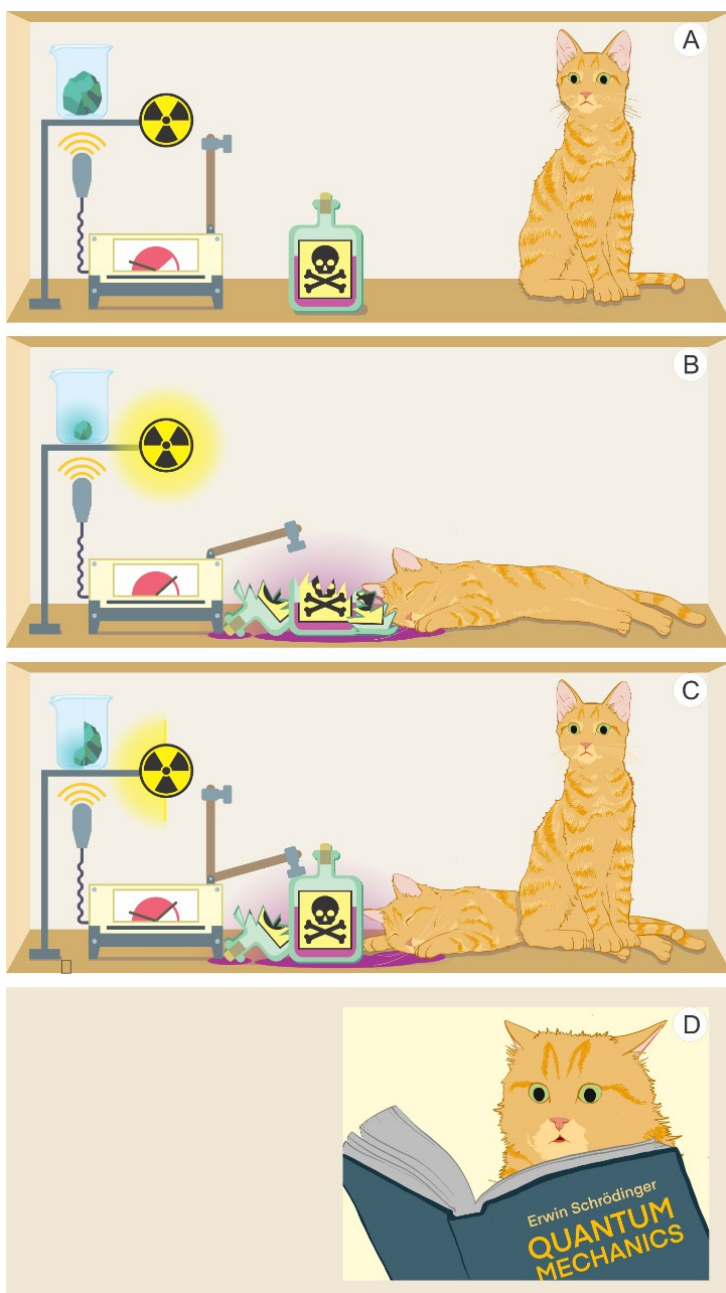


Figura 1. Situações estado vivo (A), estado morto (B), estados coexistindo (C) e estado tentando entender o que está ocorrendo (D). Ilustração elaborada pelo Prof. Ademir Nunes Ribeiro Júnior sob a orientação do Prof. Rodrigo Siqueira-Batista.

De forma também bastante inteligente, o físico Eugene Wigner, formula o famoso *Paradoxo do Amigo de Wigner*.⁴ Analisa, então, que o colapso da função de onda aconteceria no momento da observação, exatamente quando o observador se conscientizasse dos resultados da medição. Para Wigner esse fenômeno da medição, deveria ser explicado através da cognição e da consciência (FREIRE, 2011; SANTOS *et al.*, 2011; SUMMER, 2020). Desse modo, propõe para o papel da consciência na MQ, tendo em vista o processo da medição:

“[...] parece perigoso considerarmos o ato de observação como um ato humano, básico para a teoria dos objetos inanimados. É, no entanto, em minha opinião, uma conclusão inevitável. Se isto for aceito, temos que considerar o ato de observação como um ato mental, um conceito primitivo da física, nos termos do qual as regularidades e as correlações da Mecânica Quântica são formuladas”.

(WIGNER, 1995, p. 4)

Para Wigner, no problema da medição torna-se explícita a relação entre o *aparato de observação* e o *sistema observado*. Inspirando-se nos trabalhos de Von Neumann, o físico passa a considerar o sistema quântico *como se fosse* um observador humano, consciente (não máquina); emerge, assim, uma proposta *mentalista*, na qual atos conscientes intencionais podem ser correlacionados às reduções do estado quântico (SANTOS *et al.*, 2011; SUMMER, 2020). Wigner propõe que a entrada da consciência deveria ser tratada para uma melhor formulação da realidade quântica, buscando um caráter mais matemático. Ademais, ressalta que a consciência humana a responsável pelo colapso da função de onda na MQ (SANTOS *et al.*, 2011; SUMMER, 2020):

“Dar a razão para o crescente interesse dos físicos contemporâneos em problemas de epistemologia e ontologia poderia ser útil, apesar de não ser algo estritamente relevante. A razão é, em poucas palavras, que os físicos têm achado impossível dar uma satisfatória descrição dos fenômenos atômicos sem referência à consciência. Pouco pode ser feito com o frequentemente requeitado problema da dualidade onda e partícula e aludir, certamente, ao processo chamado ‘redução do pacote de onda’. Este ocorre sempre que o resultado de uma observação entra na consciência do observador – ou, para ser ainda mais exageradamente

4 Em um sistema físico, dois observadores fazem medições ao nível quântico; após as medições, cada um relata as observações que se contradizem, refletindo uma incompatibilidade sobre o determinismo (contínuo) e a probabilidade (descontínuo) na MQ. Trata-se de mais um paradoxo relacionado aos problemas da medição.

preciso, minha própria consciência, desde que seja eu o único observador; todas as outras pessoas são somente objetos das minhas observações. Alternativamente, alguém poderia dizer que a Mecânica Quântica provê somente conexões de probabilidade entre o resultado de minhas observações como eu as percebo. Seja qual for a formulação que alguém adote, a consciência tem evidentemente um papel indispensável”.

(WIGNER, 1995, p. 186).

Em 1935, Einstein, Podolsky e Rosen, publicam o paradoxo EPR expondo lacunas na Interpretação de Copenhague, proposta por Bohr em 1927 (EINSTEIN *et al.*, 1935). Baseando no trabalho de Griffiths (2005), será apresentada uma versão simplificada do paradoxo EPR, na versão de David Bohm. Com efeito, supondo um decaimento do *méson pi* neutro em um pósitron e em um elétron:

$$\pi^0 \rightarrow e^- + e^+$$

Se o pión estiver em repouso, o pósitron e o elétron estarão em direções opostas entre eles. O pión tem *spin* zero, de modo que essas partículas estão em um estado correlacionado em que o momento angular total do estado é zero. Se o pósitron tiver *spin* para baixo o elétron terá para cima. A MQ informa essa correlação das medidas, mas não qual combinação ocorrerá em um decaimento. Considerando que o elétron se desloque 5 anos-luz, se a mensuração de seu *spin* revelar que este se apresenta “para cima”, necessariamente o *spin* do pósitron estará “para baixo” (mesmo a 5 anos-luz de distância...). Essa ação à distância foi considerada assustadora para Einstein, Podolsky e Rosen, pois o elétron e o pósitron deveriam ter os *spins* bem definidos (EINSTEIN *et al.*, 1935; GRIFFITHS, 2005).

Na versão de Bohm para o EPR, pode-se descrever um pión, em repouso, decaindo em um par elétron-pósitron:

$$e^- \leftarrow \pi^0 \rightarrow e^+$$

A ideia da informação quântica aparentemente viajar mais rápido do que a luz, não era correta para o colapso da função de onda, na visão de Einstein, o qual considerava um fenômeno local. Posteriormente, com o *emaranhamento quântico* proposto por Schrödinger, percebeu-se as conexões do sistema quântico, aspecto que se tornou interessante para o modelo

de Bohm, como será visto adiante, tendo em vista seu caráter determinístico e não local (PATY, 1986).

John Bell (1964) – baseando-se em uma versão dos trabalhos supramencionados de Von Neumann, Bohm e do paradoxo EPR, ao qual estabelecem a diferença entre a MC e a MQ – demonstrou algumas lacunas no trabalho de Von Neumann. Bell elaborou uma forma de quantificar certos conceitos do paradoxo EPR, através da publicação do famoso “Teorema de Bell” – também nomeado “Desigualdades de Bell” –, como uma *resposta* ao paradoxo EPR (BELL, 1964, 1987). Desta feita, Bell pontua que a teoria de variáveis ocultas locais não era compatível com a MQ, pelo seu caráter não local como tratado por David Bohm; neste sentido, as variáveis ocultas evitam a prova de impossibilidade de Von Neuman (BELL, 1964, 1987). O físico sugere, então, que, em vez de orientar os detectores de elétrons e de pósitrons para a mesma direção, deve-se considerar que estas partículas possuem orientações fixas, independentes e quantizadas (BELL, 1964, 1987; GRIFFITHS, 2005).

A estranheza da MQ, especialmente ao se considerar uma comunidade de físicos acostumados a pensar em termos clássicos, recebeu díspares questionamentos – para além dos *dados ‘divinos’ do questionamento proposto por Einstein* – cabendo destaque ao trabalho de David Bohm. O cientista, em 1952 – mesmo ano em que chegou ao Brasil – publicou o texto *Uma proposta de interpretação da teoria quântica em termos de “variáveis ocultas”*, indicando uma versão determinista (não-local) para os sistemas físicos. A busca, em última análise, dirigia-se à formulação de uma nova interpretação para a MQ, a qual – como se comentará adiante – foi finalizada em 1983 (FREIRE, *et al.*, 1994, SANTOS *et al.*, 2011).

A MQ tem se mostrado uma teoria de grande alcance para a explicação de distintos fenômenos da natureza. Encontros *interessantes* têm também ocorrido entre a teoria quântica e a pesquisa dirigida ao fenômeno da consciência. De fato, dadas as dúvidas – ora intransponíveis – sobre a emergência de *seres autoconscientes* na realidade física, teorias como o potencial quântico de Bohm e a “neurologia quântica” de Penrose/Hameroff parecem trazer instigantes acenos (GALEGO, 2011), quiçá perspectivas de respostas, como se discutirá em seguida.

2.2. MQ e consciência: David Bohm

A questão da *causalidade* foi, provavelmente, a preocupação inicial dos estudos desenvolvidos por David Bohm, em torno da MQ, teoria à qual dedicou uma vida em termos de reflexões científicas e filosóficas. De fato, como afirmado pelo autor, “*comecei tendo um interesse especial na causalidade*” (BOHM, 1990a, p. 189). Ademais, ao longo dos seus estudos sobre a Teoria Quântica – iniciados sob a orientação de Julius Robert Oppenheimer – considerou a grande relevância teórica deste modelo, ainda que se posicionasse criticamente em relação às consequências da IC-MQ, mormente no que concerne às relações entre “observador” e “observado”: “*Disso tudo fica a ideia de que, no fim das contas, o real depende do físico que o observa, quer dizer, se você tiver uma função de onda que entre em colapso num e não em outro resultado, então o físico também tem que entrar em colapso. Assim, isso parece tornar necessário que o físico esteja por perto para que o universo seja real, e achei que isso estava errado*” (BOHM, 1990a, p. 191).

Após a publicação do livro *Quantum Theory* (BOHM, 1951), David Bohm reformulou, nos seis anos subsequentes, sua visão acerca da MQ, com a proposição das variáveis ocultas, apresentada em dois artigos publicados em 1952 (BOHM, 1952a; BOHM, 1952b) e consolidada na obra *Causality and Chance in Modern Physics* (BOHM, 1957). Os trabalhos publicados em 1952 permitiram a elaboração de um modelo capaz de reproduzir os resultados alcançados pela IC-MQ, com a utilização de parâmetros ocultos, o que permitiu “*uma descrição contínua, detalhada e causal de todos os processos*” (BOHM, 1952a, p. 166) no âmbito da realidade quântica, o que, em última análise, representa um genuíno *renascimento* do determinismo intrínseco à física clássica. Com efeito, em “*qualquer caso, a mera possibilidade de tal interpretação prova que não é necessário abrimos mão de uma descrição precisa, racional e objetiva de sistemas individuais em um nível quântico de precisão*” (BOHM, 1952a, p. 166).

A proposta elaborada pelo físico estadunidense inscreve um sistema quântico – p. ex., um elétron – em um campo físico real análogo, atribuindo-lhe uma posição e um momento bem definidos, em todos instantes. Deste modo, pode-se dizer que, de um ponto de vista físico-matemático, Bohm adota a proposição – no caso particular da MQ não-relativista – de sistemas

de partículas se movendo em trajetórias quânticas no espaço, como uma lei de evolução atribuída às partículas de acordo com a prescrição (BOHM, 1952a; SIQUEIRA-BATISTA *et al.*, 2015):

$$\frac{\partial x}{\partial t} = \frac{1}{m} \vec{\nabla} S \quad (1)$$

S parametrizando a função de onda, ψ , de acordo com a decomposição polar $\psi = R^{[(1/\hbar)S]}$, $R > 0$, e m é a massa inercial da partícula.

Esta função de onda, de modo análogo ao formulado na IC-MQ, satisfaz à equação de Schrödinger:

$$-\frac{\hbar^2}{2m} \nabla^2 \Psi + V\Psi = i\hbar \frac{\partial}{\partial t} \Psi \quad (2)$$

onde V é o potencial clássico.

A parametrização polar acima e da equação (2), permite que se encontre, para R e S :

$$\frac{\partial}{\partial t} S + \frac{(\vec{\nabla} S)^2}{2m} - \frac{\hbar^2}{2m} \frac{(\vec{\nabla} R)^2}{R} + \nabla = 0 \quad (3)$$

e

$$\frac{\partial}{\partial t} R^2 + \frac{1}{m} \vec{\nabla} \left(R^2 \vec{\nabla} S \right) = 0 \quad (4)$$

A reelaboração da função – considerada solução da equação de Schrödinger – permite o estabelecimento de analogias com o formalismo hamiltoniano da mecânica clássica (FREIRE Jr. *et al.*, 1994), ou seja, o tratamento matemático explicitado em (3) foi descrito como uma equação clássica de Hamilton-Jacobi:

$$\frac{\partial}{\partial t} S + \frac{(\vec{\nabla} S)^2}{2m} + Q + V = 0 \quad (5)$$

onde

$$Q \equiv -\frac{\hbar^2}{2m} \left(\frac{\vec{\nabla} R}{R} \right)^2 \quad (6)$$

A energia mecânica total do sistema representada pelo Hamiltoniano, de acordo com a interpretação de Bohm, é apresentada como um *terceiro termo* – para além das energias potencial e cinética –, o qual, todavia, não possui um análogo na Física Clássica. A saída encontrada pelo autor é caracterizar este termo – a equação (6) – como um *potencial quântico* (diferenciando-o de um potencial clássico), o qual “*é uma função, não dos estados das partículas, mas determinada num espaço multidimensional a partir das propriedades do todo*” (BOHM, 1990a, p. 196).

A partir da equação do tipo Hamilton-Jacobi (5), escreve-se a velocidade e o momento da partícula em termos da fase, S , como:

$$\vec{x} = \frac{1}{m} \vec{\nabla} S(t; \vec{x}) \quad (7)$$

e

$$\vec{p}(t; \vec{x}) = \vec{\nabla} S(t; \vec{x}) \quad (8)$$

Desta forma, o momento, $\vec{p}(t; \vec{x})$, é determinado pelo campo (fase) S , permitindo a elaboração, mesmo em um nível quântico, uma *trajetória definida*, denominada trajetória bohmiana:

$$\Psi(t; \vec{x}) = R(t; \vec{x}) e^{\frac{i}{\hbar} S(t; \vec{x})} \quad (9)$$

Com efeito, devido às condições iniciais, é possível a delimitação das trajetórias – através de uma equação diferencial de primeira ordem –, o que caracteriza um genuíno determinismo no tratamento bohmiano da MQ, ou seja, neste “*modelo a mecânica newtoniana clássica não perde a validade no domínio quântico, e Bohm a expressa escrevendo a segunda lei de Newton para uma partícula quântica*” (FREIRE Jr. et al., 1994, p.

59). Ademais, tendo em vista “*a equação de movimento dada pelas leis de Newton e mais as condições iniciais tem-se então recuperado o conceito de trajetória no espaço-tempo que perdeu seu sentido exato na interpretação usual da teoria quântica*” (FREIRE Jr. *et al.*, 1994, p. 59). Deve ser destacado que a teoria de David Bohm se aplica à realidade – na verdade, a toda a matéria –, envolvendo o fenômeno ondulatório e as partículas microscópicas (SIQUEIRA-BATISTA *et al.*, 2015), de uma perspectiva bastante original:

“[...] *com o modelo de Bohm, que abandonava a necessidade de uma descrição probabilística dos fenômenos, a física recuperava para si o determinismo. Nesse modelo, elétrons, por exemplo, eram sempre descritos como partículas que tinham trajetórias bem definidas – o que a versão probabilística da teoria quântica descarta*”.

(FREIRE Jr., 2001, p. 36)

Os constructos conceituais e matemáticos desenvolvidos por Bohm para lidar com as questões próprias da MQ – especialmente relativas à descrição determinista do mundo subatômico – permitiram a proposição, pelo autor, de uma *totalidade indivisa* – “*implicada não apenas no conteúdo da física (notavelmente na teoria da relatividade e na teoria quântica), mas também na maneira de trabalhar em física*” (BOHM, 2001, p. 193) – a qual é inerente à tessitura da realidade e pode ser formulada de dois modos próprios (BOHM, 2001): a *ordem explicada* e a *ordem implicada*.

A *ordem explicada* (ou *explícita* – a *ordem ex*), mais superficial e aparente, abrange o espaço-tempo quadridimensional, os distintos entes e eventos cotidianos separados entre si, os quais podem ser analisados pelos momentos, pelas trajetórias e pelas velocidades das partículas. A *ordem ex*, em última análise, consiste em uma manifestação da *ordem implicada* (ou *implícita* – a *ordem im*) – a qual oferece “*uma explicação muito mais coerente das propriedades quânticas da matéria do que a ordem mecanicista tradicional*” (BOHM, 2001, p. 244) –, caracterizando-se como instância não diretamente apreensível e que pode ser descrita pela função de onda, Ψ , na qual o espaço e o tempo assumem o papel de meras coordenadas, de modo que Ψ exhibe o caráter representativo da geometria espaço-tempo (SIQUEIRA-BATISTA *et al.*, 2015). Deve ser ponderada, nestes termos,

a necessidade do tratamento algébrico, em termos dos *spinors*⁵ (conforme trabalhado por Wolfgang Pauli e Paul Dirac) ou dos *twistors*⁶ (propostos por Roger Penrose), para a plena abordagem da *ordem im*: “*Precisamos dessa Álgebra para desenvolver nossas ideias sobre como desdobrar o espaço-tempo a partir da **implicate order**. Fazendo isso, podemos ter, por exemplo, um significado possível da equação de Dirac, que seria uma espécie de desvio das propriedades do espaço-tempo constituindo-se numa onda com propriedades geométricas*” (BOHM, 1990a, p. 197).

O modelo bohmiano da MQ – especialmente as concepções de *ordem ex* e de *ordem im* – fornece uma explicação para a *localidade* e a *não-localidade*, aspectos essenciais para a totalidade do cosmo (BOHM; HILEY, 1999).⁷ Bohm caracterizou um potencial quântico (PQ), que guia a partícula e transmite as influências sobre as distintas partes de um dado sistema quântico, ou seja, cada região do espaço-tempo está interconectada por este PQ, definindo – assim – uma *totalidade* inseparável (SIQUEIRA-BATISTA *et al.*, 2006; SIQUEIRA-BATISTA *et al.*, 2015), sendo as variáveis ocultas propostas por Bohm, as posições das partículas. Este todo – ὅλος (*holos*) em grego – pressupõe que à ordem global subjaz um movimento mantenedor, o *holomovimento*, o qual é “*uma totalidade ininterrupta e indivisa*” e “*transporta uma ordem implicada*” (BOHM, 2001, p. 202). Ademais, o *holomovimento* – considerando-se os aspectos essenciais do holograma (NELSON-ISAACS, 2021) – é “*indefinível e imensurável*” (BOHM, 2001, p. 202) e, com certo aceno para uma descrição da realidade na qual tudo

5 Espinores são os elementos complexos do espaço (afim) que porta a representação fundamental da Álgebra de Clifford associada a um dado espaço-tempo. Se a dimensão deste é D , os espinores são objetos de dimensão $2^{[D/2]}$, sendo $[D/2]$ a parte inteira de $D/2$.

6 Twistors são os objetos que portam a representação fundamental do grupo de spin $SU(2,2)$ associado ao grupo conforme, $SO(2,4)$, do Espaço de Minkowski da Teoria da Relatividade Especial.

7 O comportamento não-local refere-se ao “[...] *fato de pares de partículas que, em um dado momento, interagiram de alguma maneira, ao serem separadas por qualquer distância, manterem um comportamento previsível uma em relação a outra, ou seja, as propriedades físicas de qualquer uma delas são afetadas pela escolha das propriedades a serem medidas na outra*”. Cf. SIQUEIRA-BATISTA *et al.*, 2006, p. 9. Vale destacar, nas palavras do próprio Bohm, que o potencial quântico “*surgiu pela primeira vez em meu trabalho de 1951/52, quando estava explicando a Mecânica Quântica do elétron. O elétron é basicamente uma partícula, mas acompanhado de um certo campo quântico que satisfaz a equação de Schrödinger: pensava num novo tipo de campo. Este campo agia sobre a partícula por meio de um potencial quântico que tinha estranhas propriedades, e uma delas é que ele nem sempre diminuía com a distância, podia ser muito forte a grandes distâncias. Chamei a isto de não-localidade*”. Cf. (BOHM, 1990a, p. 195-6; grifo do autor).

está implicado em tudo, em uma “*ordem de totalidade indivisa*” (BOHM, 2001, p. 208). Com efeito, o contínuo espaço-tempo e, de resto, todas as “coisas”, fluem do – e retornam ao – *holomovimento*. A realidade, desde este ponto de vista, passa a ser vista como um *tecido* constituído por profundas interconexões – o que atualiza o tema da *interdependência* – entre os fenômenos, quiçá entre observadores (aparato de observação) e observados. Esse último ponto será retomado na seção vindoura.

O problema da consciência – na qual se incluem “*o pensamento, o sentimento, o desejo, a vontade, etc.*” (BOHM, 2001, p. 258) – se coloca precisamente nesse âmbito, uma vez que esta surge da matéria, da realidade física, ainda que as relações entre tais âmbitos – consciência e realidade física – sejam de difícil compreensão (BOHM, 1990b). De todo modo, para o físico estadunidense, a consciência – como parte da realidade – “*deve ser compreendida em termos da ordem implicada*” (BOHM, 2001, p. 258) e, igualmente, da *ordem explicada*. De fato, a partir da análise da experiência auditiva – p. ex., ouvir uma música⁸ – e visual – p. ex., assistir um filme de cinema⁹ –, em termos das transformações percebidas em uma e outra situação, Bohm propõe que “*há uma semelhança básica entre a ordem de nossa experiência imediata de movimento e a ordem implicada, conforme expressa em termos de nosso pensamento*” (BOHM, 2001, p. 264), ou seja, conjectura-se que a informação encontra-se dobrada em nível neuronal. A questão que se coloca, na sequência, diz respeito aos conteúdos manifestos da consciência – compreendidos no âmbito da *ordem ex* –, os quais

8 O comentário de Bohm é bastante elucidativo: “*Pode-se assim obter uma sensação direta de como uma sequência de notas vai-se dobrando para dentro de muitos níveis de consciência, e de como, em qualquer dado momento, as transformações que fluem de muitas dessas notas dobradas se interpenetram e se entremesclam para dar origem a uma sensação imediata e primária de movimento. Essa atividade na consciência constitui evidentemente um notável paralelo com a atividade que propomos para a ordem implicada em geral. (...) Ao ouvir música, estamos, portanto, percebendo diretamente uma ordem implicada*”. Cf. (BOHM, 2001, p. 262-263; grifo do autor).

9 Deve ser notada a ponderação do físico estadunidense a este respeito: “*No entanto, é evidente que as imagens visuais devem passar também por uma transformação ativa à medida que vão sendo ‘dobradas’ dentro do cérebro e do sistema nervoso (p. ex., elas dão origem a respostas emocionais, físicas e outras mais sutis, das quais pode-se estar apenas vagamente consciente, bem como a ‘imagens futuras’, que, de certa forma, são semelhantes às reverberações nas notas musicais). Mesmo que seja pequena a diferença de tempo entre duas dessas imagens, os exemplos citados acima deixam claro que uma sensação de movimento é experimentada graças ao entremesclamento e à interpenetração das transformações copresentes, às quais essas imagens devem dar origem à medida que penetram no cérebro e no sistema nervoso*”. Cf. (BOHM, 2001, p. 264).

se baseiam na memória. Desta feita, a consciência e a realidade devem dizer respeito a dois níveis: (1) manifesto, *ordem ex*, tipicamente “estático e fragmentado” (BOHM, 2001, p. 271) e (2) profundo, *ordem im*, caracterizado como “*fluxo ininterrupto*” (BOHM, 2001, p. 271). Ao considerar que matéria e consciência pertençam à mesma noção geral de ordem, abre-se a perspectiva para que ambas tenham um mesmo fundamento, tanto de um ponto de vista da *ordem ex* quanto da *ordem im*:

“Conclui-se então que a ordem explicada e manifesta da consciência não é, em última análise, distinta daquela da matéria em geral. Fundamentalmente, esses são aspectos essencialmente diferentes da ordem global única. Isto explica um fato básico que apontamos antes: que em geral a ordem explicada da matéria é também, em essência, a ordem explicada sensorial que é apresentada na consciência, na experiência ordinária.

Não apenas quanto a isso, mas, como vimos, também numa ampla faixa de outros aspectos importantes, a consciência e a matéria em geral são basicamente a mesma ordem (isto é, a ordem implicada como um todo)”. (BOHM, 2001, p. 273)

A consciência e a realidade, ou a “mente” e o “corpo”, isto posto, não estão apartadas, mas, são, em suma, “*o resultado de projeções relacionadas de uma base comum de dimensão mais elevada*” (BOHM, 2001, p. 275). Tal ponderação, com efeito, representa uma crítica ao dualismo cartesiano (BOHM, 1990), tendo em vista a distinção qualitativa, proposta pelo filósofo e matemático francês, entre *res extensa* e *res cogitans* (DESCARTES, 2016). No entanto, o autor também questiona o reducionismo causal – da consciência à matéria, e, também, poder-se-ia dizer, da matéria à consciência – ao reconhecer que “*não dizemos que a mente e o corpo afetam de modo causal um ao outro*” (BOHM, 2001, p. 275). Com efeito, Bohm é simultaneamente crítico ao *dualismo de propriedades* e às *teorias reducionistas* da consciência (TOURINHO, 2012), tomando a noção de totalidade e de ordem – implicada e explicada – para propor a unidade da realidade física.

As considerações ora apresentadas se articulam àquilo que Bohm denominou *modelo holográfico da consciência*, o qual, distintamente de uma concepção *localizacionista* (TOURINHO, 2012), pressupõe que as informações pertinentes se encontram armazenadas “*por todo o cérebro, ou sobre grandes áreas dele, e cada vez que as informações são usadas é feita uma seleção reunindo todas essas informações a partir de todo o cérebro, como ocorre com o holograma propriamente dito*” (BOHM & WEBER, 1995, p.

45). Cabe destacar que o potencial quântico, tal qual formulado por Bohm, é capaz de propiciar o agrupamento dos neurônios individuais, os quais passam a se comportar como em um sistema de *macroneurônios quânticos*, processando informações do ambiente (GALEGGIO, 2011). A ação do referido potencial quântico pode induzir a transição para o estado clássico desse agrupamento dos neurônios – os quais se comportam como um *todo* dentro do sistema – o que explicaria o surgimento da consciência na *interseção* do trabalho “clássico” e “quântico” do cérebro, estabelecendo-se transições clássicas ↔ quânticas para o fenômeno consciente (GALEGGIO, 2011).

A recorrência da ideia de todo – p. ex., nestas formulações – converge para uma noção bastante radical de totalidade – quiçá, afim à originária ideia grega de *phýsis* (φύσις)¹⁰, pois “a *phýsis* compreende a totalidade de tudo o que é” (BORNHEIM, 2008, p. 13) – a qual abrange todas as coisas, inclusive a consciência, reconhecida como parte do todo, nas palavras do próprio autor: “*temos o todo da natureza e dentro dela existimos e também estamos; o todo está em cada parte e a consciência também pertence a essa natureza*” (BOHM & WEBER, 1995, p. 87).

O legado das concepções bohmianas tem sido investigado por distintos pesquisadores (FREIRE Jr. *et al.*, 1994; PINTO-NETO, 2005; BETZ, 2014; SIQUEIRA-BATISTA *et al.*, 2015; GOMEZ *et al.*, 2018; JAMALI *et al.*, 2019), mas, o alcance de suas contribuições – mormente àquelas dirigidas à compreensão da consciência – ainda carece de melhor compreensão, dada a profundidade de seu pensamento.

2.3. MQ e consciência: Roger Penrose e Stuart Hameroff

O problema da consciência, do ponto de vista da MQ, tem sido investigado, de modo bastante salutar, pelos cientistas Penrose¹¹ e Hameroff. Os autores – que iniciaram suas reflexões sobre o problema de modo

10 Etimologicamente, o vocábulo *phýsis* é formado pelo sufixo *sis* e pela raiz verbal *phy*. Para W. Jaeger “a palavra também abrange a fonte originária de todas as coisas, aquilo a partir do qual se desenvolvem e pelo qual seu desenvolvimento se renova constantemente; em outras palavras, a realidade subjacente às coisas de nossa experiência” (p. 26). Cf. JAEGER, Werner. La teología de los primeros filósofos griegos. Traducción de José Gaos. México: Fondo de Cultura Económica, 1952.

11 Roger Penrose foi laureado com o Prêmio Nobel de Física, juntamente com Reinhard Genzel e Andrea Ghez, no ano de 2020, em virtude dos seus trabalhos sobre buracos negros.

independente e que, na sequência, convergiram para o trabalho em conjunto – têm buscado entender o processo físico da natureza da consciência (HAMEROFF & PENROSE, 2013; 2014; LI TIANWEN *et al.*, 2019). A proposta dos cientistas envolve, assim, de forma interdisciplinar, a MQ, a gravitação quântica, a filosofia e as neurociências, especialmente (HAMEROFF & PENROSE, 2013; HAMEROFF & PENROSE, 2014).

Os Teoremas da Incompletude de Kurt Gödel (1931) estão entre os pontos de partida de Penrose, o qual considera a mente como um produto da atividade física do cérebro (PENROSE, 1989; PENROSE, 1994; PENROSE, 2011). De fato, no livro *O grande, o pequeno e a mente humana*, o físico critica a perspectiva adotada pelo programa da “Inteligência Artificial forte” (IA forte), de acordo com o qual os processos mentais são completamente computacionais. Com efeito, Penrose propõe que “*existe algo na ação física do cérebro que evoca receptividade – em outras palavras, é algo na física que temos de apelar; mas, essa ação física é algo que nem sequer pode ser simulado computacionalmente. Não existe simulação que possa executar essa ação. Isso implica que deve haver algo na ação física do cérebro que esteja além da computação*” (PENROSE, 1998, p. 112). Desta perspectiva, considera que a investigação da consciência humana pressupõe outros fatores que estão muito além dos sistemas computacionais complexos, uma vez que uma série de operações do raciocínio humano é “não computável” (HAMEROFF & PENROSE, 2013). Para o físico, a não computabilidade do cérebro poderia ser explicada através de relevantes ajustes na MQ, com boa fundamentação matemática. Ademais, Penrose (2013) acredita que o colapso da função de onda e a superposição dos estados quânticos são processos essenciais para a descrição física da realidade.

A incompletude da MQ – também defendida por Einstein, Schrödinger e Dirac – é uma das ideias adotadas por Penrose – com relevância para a abordagem do enigma da consciência – especialmente ao se considerar a questão do *problema de medição* (HAMEROFF & PENROSE, 2014). Para abordá-lo, o físico britânico sugere uma nova extensão para a MQ, introduzindo uma maneira objetiva de redução do estado quântico, *objective reduction* – “OR” (HAMEROFF & PENROSE, 2014; PENROSE, 1996; PENROSE, 2005). Inicialmente, Penrose sugeriu que a “OR” teria origem de ‘um gráviton’ (partícula elementar responsável por transmitir a força gravitacional) (HAMEROFF & PENROSE, 2014). Todavia, logo depois, em

um novo trabalho, surge a proposta *Diósi-Penrose* (DP), tratando-se de um limite físico objetivo para os estados quânticos superpostos (HAMEROFF & PENROSE, 2014), como será detalhado em seguida.

Penrose publica, em 1989, *The Emperor's New Mind*, uma possível explicação da consciência humana através dos referidos fenômenos descritos no âmbito da MQ. A ideia original foi desenvolvida, na sequência, com Hameroff, culminando no lançamento de *Shadows of the Mind*, em 1994. Nesta obra, os autores propõem que a consciência se relaciona aos microtúbulos – constituintes proteicos cilíndricos, com cerca 25 nanômetros de diâmetro e comprimento variável –, presentes em grandes quantidades de células, incluindo os neurônios. Na visão dos dois cientistas, as vibrações dos microtúbulos neuronais produzem superposições quânticas. O aumento da energia, em decorrência das massas sobrepostas, permite o alcance do ponto crítico da gravidade quântica, colapsando e originando a consciência dentro dos neurônios, através das junções dos qubits (bit quântico – computação quântica) nos microtúbulos. Desta feita, para Penrose & Hameroff (2014), a consciência trabalha em processos quânticos que são biologicamente “orquestrados” (*Orch*) nos conjuntos de microtúbulos existentes nos neurônios. Essas ações quânticas se relacionam combinando-se com a atividade neuronal sináptica e de membrana, no processo OR (HAMEROFF & PENROSE, 2014). Com efeito, a proposta de Penrose-Hameroff refere-se à redução objetiva da consciência e dos estados quânticos, em um elegante modelo de interpretação denominado *Teoria da Redução Objetiva Orquestrada* (em inglês, “*Orchestrated Objective Reduction*” = Orch-OR) (HAMEROFF & PENROSE, 2014).

Os processos de (i) estado consciente, (ii) percepção e (iii) escolha são, de acordo com esta proposta, devidos à *Orch-OR*. A supramencionada configuração de DP está articulada aos fundamentos da MQ juntamente com a geometria espaço-tempo, sugerindo assim, um vínculo entre as leis físicas clássicas e quânticas do Universo e a sua respectiva junção aos processos biomoleculares do cérebro (HAMEROFF & PENROSE, 2014). Trata-se, pois, do estabelecimento de uma genuína *ponte* entre a MQ e a Física Clássica, caracterizando um fenômeno “gravitacional quântico”, ou seja, a superposição quântica em um contexto gravitacional. A configuração de (DP) para (OR) redundaria na redução da superposição em uma escala de tempo τ que

pode ser estimada em superposições de dois estados estacionários pela fórmula (PENROSE, 2002; HAMEROFF & PENROSE, 2013; 2014):

$$\tau \approx \frac{\hbar}{E_G}$$

τ = representação de um tipo de tempo médio

$\hbar$ = forma de Dirac para a constante de Planck

E_G = energia que custaria para deslocar um componente da superposição no campo gravitacional distinto (podendo ser representado pelo número de tubulinas)

O processo quântico de ‘OR’ (redução objetiva) acontece quando há superposições quânticas entre espaços-tempos ligeiramente diferentes conforme esquematizado na Figura 2. A medida de tempo seria ínfima em relação a uma medida entre os distintos espaços, considerados relativamente grandes, como pode-se observar no já referido experimento mental do Gato de Schrödinger. Todavia, ínfimas medidas de diferença de espaço implicariam em medidas de tempo possivelmente bastante longas. Fisicamente, como proposto por Hameroff e Penrose (2014), um único elétron isolado em um estado sobreposto com baixo (E_G) conseguiria alcançar o limiar de ‘OR’ meramente passados em tempos, milhares de anos ou mais. Ao passo que o Gato de Schrödinger, com massa de aproximadamente 10 kg, fosse inserido em uma superposição de morte-vida, essa fronteira conseguiria ser atingida em um tempo menor que o tempo de Planck de 10^{-43} s (HAMEROFF & PENROSE, 2014).

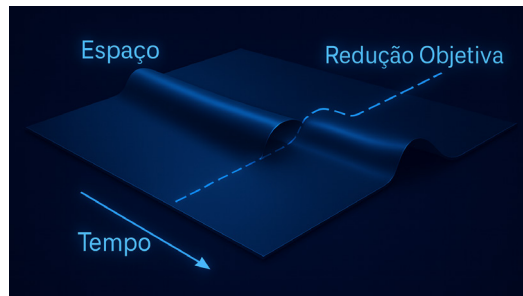


Figura 2 – A curvatura de superposição E atingindo o limiar ($EG = \hbar/\tau$) e a ‘redução objetiva’ (‘OR’) em uma localização ou curvatura de partícula selecionada se tornando clássica e a outra localização ou curvatura desfazendo-se (Fonte: HAMEROFF & PENROSE, 2014).

Desde esta perspectiva, quanto maior for a energia de massa do objeto, mais rápido ocorrerá a ‘OR’ (nova forma de colapso da função de onda) sendo a interpretação inversa, verdadeira (JAMALI *et al.*, 2019; HAMEROFF & PENROSE, 2013; 2014). Para que ‘Orch-OR’ (redução objetiva orquestrada) e consciência ocorram, superposições quânticas de autoenergia gravitacional (E_G) precisariam evitar a decoerência ambiental por um certo tempo significativo para perfazer a medida de tempo τ por $\tau \approx \hbar/E_G$. De acordo com Hameroff e Penrose (2014), certos graus de liberdade do sistema devem existir de forma isolada da decoerência ambiental, para que a ‘Orch-OR’ seja utilizada controladamente pelo sistema (HAMEROFF & PENROSE, 2014). Trata-se, pois, de uma teoria sobre a constituição da consciência sobre fatos ou acontecimentos de eventos discretos, sendo cada um desses eventos discretos, considerados um momento de ‘OR’ de um estado quântico em concordância com a redução objetiva de DP, para uma possível conclusão de que esses estados quânticos estão como partes de uma computação quântica gerada como fonte principal em microtúbulos neuronais (HAMEROFF & PENROSE, 2014).

As consequências da Teoria Orch-OR – de uma perspectiva de dispares áreas: Biologia Molecular, Ciência Cognitiva, Ciências Sociais, Computação, Filosofia, Física e Neurociências – seguem em investigação. O avanço das pesquisas – com a superação das limitações experimentais para a testagem de diferentes elementos do modelo de Penrose e Hameroff – poderá elucidar vários aspectos ainda obscuros de uma possível teoria quântica da consciência.

3. ELETROMAGNETISMO E CONSCIÊNCIA

A compreensão da consciência envolve processos físicos de características diversas, dentre as quais pode-se mencionar o Eletromagnetismo – em seus aspectos clássicos –, cuja proposta de abordagem será desenvolvida na presente seção.

No âmbito dos fenômenos eletromagnéticos, a recepção, o processamento e a transmissão de sinais são aspectos investigados sob a denominação de funções de Green (COURANT; HILBERT, 1953-1962). Destas, são estudadas três categorias: as funções de Green avançadas, as retardadas e as causais. No caso de processos ligados à vida, contribuem as funções de

Green retardadas, sendo aquelas que respeitam o Princípio da Causalidade no domínio da Física Clássica. As energias envolvidas nos processos moleculares e atômicos da vida não conseguem excitar efeitos da chamada Eletrodinâmica Quântica, onde a função de Green que se adota é a causal.

Quanto aos processos bioquímicos envolvidos nas transmissões neurais e no próprio processo da consciência, fenômenos quanto-mecânicos são fundamentais, uma vez que as reações químicas nada mais são do que processos de transferência de elétrons das camadas atômicas. Entretanto, outras ocorrências, como processos de difusão e dissipação estão também presentes. A interligação de todos estes efeitos que concorrem para o processo da consciência resulta em um sistema altamente complexo, que exige, para um entendimento preliminar, a análise individual de cada categoria de fenômeno. Torna-se, então, interessante a perspectiva cartesiana de necessária redução às partes para, em etapas sucessivas, se poder ligar as diferentes peças que conduzem à conjuntura de processos.

Nesta Seção, os esforços serão concentrados na discussão dos possíveis aspectos eletromagnéticos envolvidos na emergência da consciência. A compreensão da interação eletromagnética em escala macroscópica deu-se com os estudos do físico-matemático James Clerk Maxwell, publicados em 1864 e 1865 na *Royal Academy of Science de Londres* (MAXWELL, 1864; MAXWELL, 1865). O eletromagnetismo em sua versão maxwelliana é uma teoria linear e descreve a geração de campos elétricos e magnéticos induzidos por distribuições de cargas e correntes elétricas, por domínios de *spin* e pela dependência temporal dos próprios campos elétricos e magnéticos, através da chamada indução de Faraday-Lenz e da chamada corrente de deslocamento. Entretanto, em uma vasta categoria de materiais, efeitos de não-linearidade tornam-se extremamente relevantes. Exemplos destes são as respostas com birrefringência e efeitos de rotação nos planos de polarização quando são aplicados campos elétricos e magnéticos externos às ondas eletromagnéticas que se propagam nos materiais. Trata-se de efeitos eletro- e magneto-óticos de natureza birrefringente; não serão presentemente descritos, mas, ressalta-se que são de grande importância porque acrescentam informações, às Equações de Maxwell respostas não-lineares, devido a propriedades de anisotropia de meios materiais especiais.

Por outro lado, aspecto a ser também destacado é a necessidade de se reavaliar a eletrodinâmica maxwelliana em questões relacionadas à

dimensionalidade do espaço onde os sinais eletromagnéticos são transmitidos. A propagação de fenômenos ondulatórios depende fortemente da dimensionalidade do espaço onde estes ocorrem, e são as funções de Green aquelas capazes de revelar esta característica. É certo que, no que diz respeito à consciência, os processos têm lugar, em uma primeira abordagem, no espaço euclidiano tridimensional. Entretanto, se houver propagação de sinais em seções planares, o fenômeno eletromagnético (i) adquire um alcance mais longo do que o correspondente em três dimensões espaciais, (ii) é acentuado em sua intensidade e (iii) apresenta efeitos de reverberação advindos da redução dimensional. Isto implica em consequências muito particulares para a relação entre uma causa externa e a sua percepção e sua consequente resposta excitatória (MELO, 2001).

A busca ora empreendida diz respeito à compreensão de que os fenômenos de natureza eletromagnética que interferem no processo da consciência não se enquadram na formulação maxwelliana original, a qual descreve com excelente precisão um amplo espectro de situações físicas. Com efeito, é preciso ir além de Maxwell e buscar possíveis extensões não-maxwellianas. Na literatura contemporânea, pode-se encontrar um vasto número destas extensões além-Maxwell, formuladas para levar em conta diferentes regimes dos fenômenos eletromagnéticos. As extensões não-lineares parecem ser aquelas que mais se enquadram no estudo do eletromagnetismo da consciência. Na literatura da Física, encontram-se catalogadas, atualmente, 15 extensões não-lineares advindas de efeitos microscópicos, presentes em escalas da ordem de 10^{-12} m ou inferiores a este valor (PLEBANSKI, 1970; HELAYËL-NETO *et al.*, *preprint*). O próprio vácuo quântico induz efeitos de não-linearidade na propagação de ondas eletromagnéticas na região dos raios-gama. Desde esta perspectiva, abrem-se possibilidades de discussão para o papel da abordagem eletromagnética para o estudo da consciência deve ser via efeitos não-lineares.

Em um denso e interessante artigo publicado muito recentemente, McFadden (2020) descreve o modelo denominado *Conscious Electromagnetic Information Field Theory* (CEMI), no qual apresenta a ideia de como a consciência implementa, no campo eletromagnético do cérebro, algoritmos espaciais e que estes devem ser mesmo mais relevantes do que os algoritmos temporais. A interessante conclusão do autor é que a consciência emerge de algoritmos que coexistem no espaço do campo

eletromagnético do cérebro, o que, consequentemente, induz acentuados efeitos de não-linearidade do fenômeno eletromagnético.

Outra abordagem que reforça a não-linearidade do eletromagnetismo da consciência encontra-se no artigo publicado, em 1996, por Jirsa e Haken no *Physical Review Letters*, “*Field theory of electromagnetic brain activity*” (JIRSA; HAKEN, 1996), no qual os autores propõem uma equação de propagação não-linear para uma variável de campo que descreve a resposta fisiológica excitatória a um pulso excitatório externo. A equação de onda associada ao campo-resposta é não-linear e esta não-linearidade apresenta uma forte analogia com a auto-interação – portanto um fenômeno não-linear – do campo de Higgs, responsável pela separação das interações eletromagnética e nuclear fraca na chamada Teoria Eletrofraca de Salam-Glashow-Weinberg.

A equação de campo a que chegam Jirsa e Haken no trabalho acima citado governa o comportamento espaço-temporal da atividade cerebral eletromagnética, representando o desvio, no ponto x e no instante de tempo t , em relação a um dado estado psicológico e assume a forma abaixo:

$$\ddot{\psi}_e + \left(\omega_0^2 - v^2 \Delta \right) \psi_e + 2\omega_0 \dot{\psi}_e = \left(\omega_0^2 + \omega_0 \frac{\partial}{\partial t} \right) \times \rho(x, t)$$

com a densidade, ρ , podendo ser escrita como segue,

$$\begin{aligned} \rho(x, t) &= a_e S_e [\tilde{\rho} \psi_e(x, t) + p(x, t)] \\ &\approx a_e \left\{ \alpha_e \left[\tilde{\rho} \psi_e(x, t) + p(x, t) \right] - \frac{4}{3} \alpha_e^3 [\tilde{\rho} \psi_e(x, t) + p(x, t)]^3 \right\} \end{aligned}$$

onde

$\psi_e(x, t)$ representa o campo neural excitatório (que descreve a atividade cerebral excitatória), cada ponto sobre ψ_e indica uma derivada parcial em relação ao tempo;

Δ designa o operador Laplaciano;

v denota uma característica velocidade de propagação;

$\omega_0 = v/\sigma_e$ define uma escala de frequência, sendo σ_e um parâmetro de alcance espacial;

α_e é um parâmetro constante;

α_e é a inclinação da chamada função sigmóide, $S(x) = ex / [ex + 1]$;

$p(x, t)$ é o termo não-homogêneo que representa o estímulo acústico externo.

É o termo com a terceira potência do campo ψ e que explicita a analogia com a auto-interação associada ao campo de Higgs, o que corresponderia, em uma formulação Lagrangeana, à presença de um potencial quártico no campo ψ . Este modelo visa descrever a dinâmica da propagação de $\psi(x,t)$ ao longo das fibras córtico-corticais (apenas excitatórias) com atrasos de até 100 milissegundos. O domínio espacial de conectividade, para a propagação dessas ondas, é da ordem de 0,1cm para as fibras intracorticais e de 1-20 cm para fibras córtico-corticais. As velocidades de propagação cortical estão na faixa de 1m/s. A escala de tempo está na faixa de 100 milissegundos com escala espacial na faixa de vários cm. A presença de termos de amortecimento lineares e não-lineares associados à derivada da função de onda, resultam na dependência da amplitude da onda com a frequência, de tal modo que, com o aumento da frequência, espera-se uma redução na amplitude.

Seria de especial relevância o estudo das propriedades da equação não-linear de Jirsa e Haken, sobretudo no aspecto referente às suas similaridades com o Mecanismo de Higgs, devido à presença do termo cúbico de auto-interação. Especial atenção deve também ser dada ao aparecimento de um termo de dissipação da onda, efeito este não-presente no Mecanismo de Higgs original para a quebra da simetria eletrofraca. Em trabalho dos autores que se encontra em desenvolvimento – um desdobramento do presente capítulo, em última análise – é dedicado ao estudo de diferentes regimes no espaço de parâmetros da equação propagatória de Jirsa e Haken, na tentativa de encontrar uma fase em que o campo da resposta excitatória adquira uma configuração de equilíbrio estável. Desta forma, considerando-se perturbações em torno da configuração de campo mínimo, propõe-se encontrar excitações que se propagam nesta fase, analisar as suas características e compreender como poderão contribuir para se compreender aspectos da consciência. A equação de Jirsa e Haken apresenta um rico diagrama de fases, o que justifica a sua investigação, tanto do ponto de vista da pura teoria de campos como, também – e principalmente –, no âmbito do eletromagnetismo da atividade cerebral.

Finalizando esta seção, a tentativa foi motivar e fundamentar a relevância de extensões não-Maxwellianas do eletromagnetismo para o estudo do campo eletromagnético do cérebro. Dentre estas extensões, destacou-se a importância de se incorporar efeitos de não-linearidade e de baixa dimensionalidade.

4. ABORDAGENS RELATIVÍSTICAS

A investigação da consciência de uma perspectiva relativística tem sido igualmente proposta, com destaque para trabalhos recentemente publicados dirigidos à abordagem da sua inscrição espaço-temporal (CICUREL & NICOLELIS, 2015; SIEB, 2019; SATTIN, 2021).

A vida – nela incluída a experiência consciente humana – se organiza em termos do espaço-tempo (HELAYEL *et al.*, 2012), ou seja, em um contínuo quadridimensional que envolve três dimensões espaciais e uma dimensão temporal. De fato, uma pessoa é capaz de estar consciente de uma série de eventos que ocorrem simultaneamente, mas em distintas posições no espaço [o leitor, no momento, pode experimentar levantar os olhos da página, que ora lê, e observar ao seu redor, a multiplicidade de objetos que se apresentam à consciência, os quais estão simultaneamente (ou seja, *ao mesmo tempo*) situados em díspares locais]. De outro modo, uma pessoa que se desloca, dia após dia, do quarto para a sala, em sua própria casa, na hora de dormir por exemplo, nota as mesmas posições no espaço (o que permite o movimento), mas em tempos distintos. A ideia, aparentemente contraintuitiva, de um contínuo entre espaço e tempo foi formulada por Albert Einstein como parte do corpo teórico da Teoria da Relatividade Especial (TRE), ao reconhecer que as três dimensões espaciais e a dimensão temporal são interdependentes e descritas por quatro números, as coordenadas de espaço – x , y e z – e a coordenada de tempo, t (EINSTEIN, 1999). Tal contexto é explicitado na quarta equação das transformações de Lorentz (HELAYËL-NETO *et al.*, 2015):

$$t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}}$$

De acordo com Einstein, nesta equação

“[...] a diferença de tempo entre dois eventos em relação a K' [um sistema de coordenadas] em geral não se anula, mesmo quando o intervalo de tempo Δt entre os mesmos é nula em relação a K [outro sistema de coordenadas]. Uma distância puramente espacial de dois eventos em relação a K tem como consequência uma distância temporal dos mesmos em relação a K'. [...] As leis da natureza que satisfazem às exigências da Teoria da Relatividade (Especial) assumem então formas

matemáticas nas quais a coordenada de tempo desempenha exatamente o mesmo papel que as coordenadas do espaço”.

(EINSTEIN, 1999, p. 50)

A quadridimensional experiência consciente humana tem correlatos em nível neural – dada a organização morfofuncional do sistema nervoso humano – os quais, supõe-se, podem ser modelados em termos da TRE de Einstein (SIEB, 2019). Nestes termos, Sieb (2016) propõe que uma representação dos principais eventos no espaço e no tempo têm sua codificação produzida em nível do hipocampo, com armazenamento sob a forma de memória episódica. Ademais, destaca que os córtices pré-frontal e parietal posterior, ao estabelecer conexões com o hipocampo, contribuem para (i) a organização de experiências conscientes, (ii) o estabelecimento de correlações causais entre eventos (propostas em termos de relações entre passado e futuro) e (iii) a integração de diferentes funções cognitivas superiores típicas da consciência. A memória, por conseguinte, adquire uma participação essencial nestes processos – como destacado por Damásio ao considerar a distinção entre *consciência central*¹² e *consciência ampliada*¹³ – dado o próprio processamento dos registros no cérebro, o que é um ponto decisivo para o funcionamento do *self autobiográfico*.¹⁴ Com efeito, ao analisar

-
- 12 Para Damásio, “[...] a *consciência central* ocorre quando os mecanismos cerebrais de representação geram um relato imagético, não verbal, de como o próprio estado do organismo é afetado pelo processamento de um objeto pelo organismo, e quando esse processo realça a imagem do objeto causativo, destacando-o assim em um contexto espacial e temporal” (DAMÁSIO, 2000, p. 219).
- 13 De acordo com Damásio, a “*consciência ampliada*, portanto, é a capacidade de estar consciente de uma gama enorme de entidades e eventos, ou seja, a capacidade de gerar um senso de perspectiva individual, de propriedade e da condição de agente sobre uma gama de conhecimentos maior do que a abrangida pela *consciência central*” (DAMÁSIO, 2000, p. 254).
- 14 Pode-se dizer que o *self* de Damásio é dividido em três (e, aqui, não há como não lembrar da alma tripartida de Platão...): (1) o *proto-self*, “um conjunto interligado e temporariamente coerente de padrões neurais que representam o estado do organismo, a cada momento, em vários níveis do cérebro”; (2) o *self central*, “inerente ao relato não verbal de segunda ordem que ocorre toda vez que um objeto modifica o *proto-self*. O *self central* pode ser acionado por qualquer objeto. O mecanismo de produção do *self central* sofre mudanças mínimas no decorrer de toda a vida”; e (3) o *self autobiográfico*, o qual se baseia “na memória autobiográfica, que é constituída por memórias implícitas de múltiplos exemplos de experiência individual do passado e do futuro antevisto. Os aspectos invariáveis da biografia de um indivíduo formam a base da memória autobiográfica. A memória autobiográfica cresce continuamente com a experiência de vida, mas pode ser parcialmente remodelada para refletir novas experiências. Conjuntos de memórias que descrevem a identidade e a pessoa podem ser reativados como um padrão

a formação de memórias em nível cerebral, o autor propõe que este processo ocorre de modo inicialmente “distribuído” – registro das percepções visuais, táteis, motoras, auditivas e outras de um dado objeto (por exemplo, ver, tocar, mover e nomear um martelo) – e posteriormente unificado em termos de relações de continuidade espaço-temporal: “*A separação espacial dos registros não se revela um problema, pois, quando todos os registros são tornados explícitos em forma de imagem, eles são exibidos apenas em alguns sítios e coordenados no tempo de tal modo que todos os componentes registrados parecem integrados sem solução de continuidade*” (DAMÁSIO, 2000, p. 283). A referida integração “*sem solução de continuidade*” acena para a inseparabilidade das quatro coordenadas supramencionadas – x , y , z e t – conforme o previsto na TRE.

A Teoria do Cérebro Relativístico (TCR), proposta por Circurel e Nicolelis (2015) é, igualmente, uma abordagem bastante interessante para a caracterização do funcionamento encefálico, no geral, e (quiçá) da consciência, em particular. De uma perspectiva global, pode-se dizer que a TCR é um modelo explicativo, de caráter multidisciplinar, da neurobiologia de mamíferos – com ênfase no *Homo sapiens* – que propõe novos mecanismos para a descrição do funcionamento do sistema nervoso, em situações de saúde e de doença, com a possibilidade de testagem experimental de suas proposições e previsões. Para os autores, em concordância com a TCR:

“[...] sistemas nervosos complexos como o nosso geram, processam e estocam informação através de interações recursivas de um sistema computacional híbrido digital-analógico (HDACE, do inglês Hybrid Digital-Analog Computational Engine). Nesse HDACE, o componente digital é definido pelos potenciais de ação produzidos por redes de neurônios distribuídos por todo o cérebro, enquanto o componente analógico é formado pela superimposição de campos eletromagnéticos neurais (NEMFs, do inglês, neural electromagnetic fields) gerados pelo fluxo de cargas elétricas pelos múltiplos feixes circulares de nervos que formam a substância branca do cérebro”.

(CIRCUREL & NICOLELIS, 2015, p. 28)

neural e explicitar-se como imagens sempre que necessário. Cada memória reativada opera como um “algo a ser conhecido” e gera seu próprio pulso de consciência central”. (DAMÁSIO, 2000, p. 230, Tabela 6.1).

Os citados campos eletromagnéticos neurais, NEMFs, constituintes do sistema computacional integrado híbrido digital-analógico (HDACE), estariam responsáveis pela organização do contínuo espaço-temporal cerebral, o *espaço mental*, o qual, de acordo com os autores, seria “*o substrato analógico neural de onde emergem todas as funções cerebrais superiores ou mais complexas*” (CIRCUREL & NICOLELIS, 2015, p. 29), dentre as quais se destaca a consciência. Os *inputs* oriundos dos ambientes interno e externo participariam da reconfiguração dos circuitos cerebrais, a partir da mediação dos próprios NEMFs. Estes, até o presente momento, não foram experimentalmente atestados, mas, evidências indiretas – por exemplo, relacionadas à “*sincronia neuronal coerente e disseminada*” (CIRCUREL & NICOLELIS, 2015, p. 32) implicada nas origens do contínuo espaço-temporal proposto pela TCR – parecem subsidiar a hipótese.

Uma das significativas ideias da TCR é a concepção de “imagem do corpo”, considerada elemento essencial para a composição do “senso de ser”. Neste sentido, pode-se concordar com Damásio, autor para o qual os neurônios “*em qualquer nível do sistema nervoso, representam entidades ou acontecimentos que ocorrem em alguma outra parte do organismo. Nasceram para ser cartógrafas da geografia de um organismo e dos acontecimentos que se dão ali*” (DAMÁSIO, 2004, p. 7). Tais representações podem ser compreendidas como “mapas mentais” dos meios interno e externo, os quais são atualizados constantemente de modo a produzirem o denominado “filme no cérebro”, “*metáfora para a composição integrada e unificada de diversas imagens sensoriais – visuais, auditivas, tácteis, olfativas e outras – que constituem o espetáculo multimídia que denominamos mente*” (DAMÁSIO, 2004, p. 8). Por razões que permanecem em investigação, este filme no cérebro produz um “senso de eu”, de origem biológica, o qual pode ser desdobrado no tríptico *self* de Damásio – *proto-self*, *self central* e *self autobiográfico*¹⁵ –, o qual, de uma certa perspectiva, pode ser aproximado do conceito de “senso de ser” de Circurel e Nicolelis. Tanto a “imagem do corpo” quanto o “senso de ser”, de acordo com estes dois últimos autores, “*emergem da interação de NEMFs gerados pela contínua combinação da atividade elétrica de múltiplas áreas corticais e sub-corticais envolvidas na representação cerebral da configuração corporal*

15 Veja a nota anterior.

de cada um de nós” (CIRCUREL & NICOLELIS, 2015, p. 37), o que permite a criação, pelo cérebro, de uma *expectativa* do corpo físico que o alberga e, mais do que isto, de uma determinada perspectiva, denominada “ponto de vista do cérebro”, como destacado pelo próprio Nicolelis, no livro *Muito além do nosso eu*:

“[...] o cérebro passa a ser considerado um ente ativo, sempre em fluxo, sempre em processo de adaptação e totalmente apto a expressar seu próprio ponto de vista e suas expectativas sobre o mundo exterior; mesmo antes que qualquer informação sobre esse mundo chegue à sua central de processamento, como resultado de uma ampla constelação de receptores periféricos”.

(NICOLELIS, 2011, p. 209)

A capacidade do cérebro de construir a realidade – ou de “esculpir” o universo conhecido (NICOLELIS, 2020) em termos do “ponto de vista do cérebro” – se articula à dinâmica do “espaço mental”, que em tese pode ser estudado de uma perspectiva lógica e matemática, cujas pistas são assinaladas pelos autores, os quais consideram que o contínuo espaço-temporal

“[...] pode se ‘deformar’ durante um grande número de experiências, por exemplo, quando nós sonhamos, experimentamos ilusões, ou sofremos algum tipo de alucinação gerada pelo consumo de drogas alucinógenas (por exemplo, LSD) ou somos acometidos por algum distúrbio mental (esquizofrenia). Deformações do ‘espaço mental’ poderiam explicar por que, durante todas as experiências mentais listadas acima, a noção de espaço, tempo, e mesmo as nossas sensações corpóreas, movimentos e senso de ser podem ser profundamente alterados”.

(CIRCUREL & NICOLELIS, 2015, p. 40)

A partir dessas ponderações, conjectura-se que o espaço mental se organizaria em uma estrutura que poderia ser descrita por uma geometria riemanniana. A concepção de espaço proposta Georg Friedrich Bernhard Riemann, autor que caracterizou a geometria no âmbito da investigação de espaços n -dimensionais com curvatura não trivial (MANFIO, 2018). A característica da geometria riemanniana de interesse para a descrição do espaço mental é a introdução de um tensor métrico (com o qual são feitas as medidas de comprimentos e ângulos) de característica não-local, isto significando que a ideia de medida depende de cada ponto do espaço no qual a medida é feita.

A geometria de Riemann é adotada como um ingrediente fundamental em diferentes teorias físicas, particularmente na Teoria da Relatividade Geral (TRG) de Albert Einstein (EINSTEIN, 1997), cujo Princípio da Covariância Geral pode ser assim enunciado: *“Todos os sistemas de coordenadas gaussianos são essencialmente equivalentes para a formulação das leis gerais da natureza”* (EINSTEIN, 1999, p. 81). Com efeito, suas ferramentas matemáticas poderiam ser úteis, conforme a perspectiva de Circurel e Nicolelis, para o estudo do espaço mental organizado a partir da “gravação ‘neural’ de novas memórias realizadas por NEMFs nos circuitos neurais”. Esta seria, provavelmente, uma significativa contribuição da TCR, com potencialidade para descrição dos estados neurofisiológicos e neuropatológicos típicos da experiência humana de existir.

Com base nestes apontamentos, pode-se considerar a abertura de uma importante frente de investigação dos aspectos físicos do sistema nervoso central, em termos dos modelos relativísticos inspirados na TRE e na TRG, cujos desdobramentos poderão ser bastante promissores para a compreensão – e a eventual modelagem – da consciência.

5. POR UMA TEORIA FÍSICA UNIFICADA DA CONSCIÊNCIA

A abordagem do *enigma da consciência*, em termos científicos, deve ser necessariamente conduzida de uma perspectiva interdisciplinar, dada a complexidade do sistema nervoso central humano. De fato, esta consideração se alicerça na própria caracterização do campo das neurociências por Roberto Lent, em termos dos seus níveis de existência:

“[...] o sistema nervoso, e o cérebro em particular, pode ser estudado de várias maneiras, todas verdadeiras e igualmente importantes. Podemos encará-lo como um objeto desconhecido mas capaz de produzir comportamento e consciência, e assim dedicar-nos a estudar apenas estas propriedades (ditas “emergentes”) do sistema nervoso. É o modo de ver dos psicólogos. Podemos também vê-lo como um conjunto de células que se tocam através de finos prolongamentos, formando trilhões de complexos circuitos intercomunicantes. É a visão dos neurobiólogos celulares. Alternativamente, podemos pensar apenas os sinais elétricos produzidos pelos neurônios como elementos de comunicação, como fazem os eletrofisiologistas. Ou então nas reações químicas que ocorrem entre as moléculas existentes dentro e fora das células nervosas, como fazem os neuroquímicos.

E assim por diante. Como se vê, são muitos os modos (chamados níveis) de existência do sistema nervoso, abordados pelos diferentes especialistas”.

(LENT, 2010, p. 5)

Diferentes modos de existência, com foco nos modelos físicos da consciência, foram abordados no presente capítulo, com ênfase na Mecânica Quântica, no Eletromagnetismo e nas Teorias da Relatividade (Especial e Geral). Dadas as díspares possibilidades de abordagem conceitual e matemática – muitas das quais ainda carentes de evidências experimentais robustas – a premissa que se adotará nesta seção é a seguinte: *a proposição de um modelo científico explicativo da consciência somente poderá almejar algum êxito caso seja capaz de incorporar abordagens oriundas de diferentes campos da Física e da Matemática, as quais deverão ser articuladas complexamente aos dados experimentais obtidos no bojo da neurobiologia, em prol de uma Teoria Física Unificada da Consciência (TFUC).*

A TFUC, com efeito, pode ser preliminarmente pensada nos seguintes termos: a consciência emerge na estrutura do cérebro humano – e, de modo mais ampliado, no corpo que o alberga e que mantém estreita relação com o ambiente –, como consequência da organização biológica e da estrutura física dos seus constituintes. Desta última perspectiva, deve-se considerar, no atual estágio do conhecimento científico, que consciência e cérebro (corpo) não podem ser investigados separadamente, sem que este posicionamento signifique um mero reducionismo da primeira ao segundo e, tampouco, o apoio a uma proposta localizacionista. Outrossim, deve-se reconhecer que os processos neurais subjacentes à consciência ocorrem em todo o cérebro (e, para além deste, em todo o sistema nervoso), tal como proposto por Bohm e Weber (1995), envolvendo fenômenos de natureza quântica – eventualmente, afins à proposta de Penrose e Hameroff (2014) de redução objetiva orquestrada em nível subcelular, mais propriamente nos microtúbulos neuronais.

Tendo em vista a estrutura bioquímica dos componentes neurais – com a participação central dos neurotransmissores e dos processos de transferência de elétrons das camadas atômicas – a apreciação em termos da não-linearidade dos fenômenos eletromagnéticos da consciência deve ser ressaltada. Ademais, em concordância com McFadden (2020), destaca-se a concepção de que a consciência implementa algoritmos espaciais

– os quais têm provável preeminência sobre os algoritmos temporais – no campo eletromagnético do cérebro, reforçando a referida não-linearidade destes fenômenos.

As dimensões espaço-temporais relacionadas à experiência consciente trazem como possibilidade a incorporação da TRE e da TRG para a empreitada explicativa da física da consciência. Do ponto de vista dos autores, ressalta-se a ausência de uma estimativa mais objetiva para caracterizar a indissociabilidade espaço-tempo como sendo a mesma da TRE. Além disso, considera-se que também é válida a reflexão se para a descrição do referido espaço mental (Cicurel e Nicolelis, 2015) é necessário invocar a TRG, ou se bastaria adotar uma descrição via geometria riemanniana. A questão aqui diz respeito à participação, ou não, da gravidade na formulação proposta. Nestes termos, as abordagens de Sieb (2016; 2019) e de Cicurel e Nicolelis (2015) – articuladas às elegantes proposições de Damásio (2020) – acenam para a relevância do contínuo quadridimensional do espaço-tempo, o qual é condição de possibilidade para a organização das experiências conscientes nas realidades “externa” (SIEB, 2016; 2019) e “interna” – o “espaço mental” (CICUREL & NICOLELIS, 2015) –, esta última eventualmente modelável em termos da métrica riemanniana, abordagem que contudo, precisa ser matematicamente desenvolvida.

A Teoria Física Unificada da Consciência, isto posto, incorporaria conceitos, modelos e métodos da Física e da Matemática – além na neurobiologia, particularmente singularizada nos trabalhos de António Damásio –, com destaque especial para os conteúdos já mencionados da MQ, do EMG, da TRE e da TRG. Sem embargo, tal proposta pode ser ampliada com a incorporação de outras contemporâneas concepções das ciências da natureza, podendo-se mencionar à guisa de propostas a serem exploradas, (1) a investigação das dimensões espaço-temporais extra nos estudos neurocientíficos (HELAYEL *et al.*, 2015) e (2) a possibilidade da utilização do conhecimento disponível sobre os buracos negros como *imagem* (ou modelo análogo) dos eventos conscientes, tendo em vista a ideia do feixe de percepções proposto pelo filósofo escocês David Hume. Estas fronteiras, por vezes condizentes com narrativas de SciFi, poderão ser trabalhadas de uma perspectiva teórica – e, quiçá experimental – com o intuito de contribuir para a investigação da consciência.

6. CONSIDERAÇÕES FINAIS

O presente capítulo apresenta possibilidades de análise – desenvolvidas na esfera da física contemporânea – para o entendimento do fenômeno da consciência. Apontamentos sobre a MQ – enfatizando, especialmente, o trabalho de Bohm e de Penrose/Hameroff –, o EMG – destacando seus aspectos de não linearidade para além da herança de Maxwell – a TRE e a TRG – de um ponto de vista relativístico – foram apresentados e discutidos.

Propõe-se que no futuro uma combinação das teorias físicas abordadas poderá contribuir decisivamente para a descrição da consciência, problema cujos autores do presente capítulo seguirão investigando. Dois desafios, no entanto, se apresentam: (1) o desenvolvimento de uma abordagem matemática unificada para estas teorias, empreendimento provavelmente tão ou mais difícil do que a busca da *Teoria de tudo*; e (2) a criação de desenhos experimentais para testagem da abordagem matemática construída. Ante as dificuldades ora existentes, deverão ser empreendidos novos esforços interdisciplinares de pesquisa científica, com a promessa – quem sabe, utopia – de que se possa compreender, finalmente, o insondável enigma da consciência.

7. REFERÊNCIAS

ABBAGNANO, N. **Dicionário de filosofia**. São Paulo: Martins Fontes, 2003.

ALONSO, M.; FINN, E. J. **Física**. Lisboa: Escolar Editora, 2012.

ANASTASSIOU, C. A.; MONTGOMERY, S. M.; BARAHONA, M.; BUZSÁKI, G.; KOCH, C. The effect of spatially inhomogeneous extracellular electric fields on neurons. **Journal of Neuroscience**, v. 30, n. 5, p. 1925-1936, 2010.

ATMANSPACHER, H. “Quantum Approaches to Consciousness”, *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2020/entries/qt-consciousness/>>.

BECK, F. & ECCLES, J.C. Quantum aspects of brain activity and the role of consciousness. *Proceedings of the National Academy of Sciences U.S.A.*, v. 89, p. 11357-61, 1992.

BELL, J.S. (1964), On the Einstein-Podolsky-Rosen Paradox, *Physics* 1, 195-200. Reproduzido em Wheeler & Zurek (1983), op. cit., pp. 403-8, e em Bell (1987), op. cit., pp. 14-21.

BELL, J. S. “Speakable and unspeakable in Quantum Mechanics: Collected papers on Quantum Philosophy”. Cambridge University Press, 1987.

BETZ, M. E.M. Elementos de mecânica quântica da partícula na interpretação da onda piloto. **Revista Brasileira de Ensino de Física**, v. 36, n. 4, p. 1-14, 2014.

BLOCH, F. Heisenberg and the early days of quantum mechanics. **Physics Today**, v. 29, p. 23-27, 1976.

BOHM, D. **Quantum Theory**, New York: Prentice Hall, 1951.

BOHM, D. A suggested interpretation of the quantum theory in terms of hidden variables. I. **Physical Review**, v. 85, p.166-179, 1952a.

BOHM, D. A suggested interpretation of the quantum theory in terms of hidden variables. II. **Physical Review**, v. 85, p. 180-193, 1952b.

BOHM, D. **Causality and chance in modern physics**. Londres: Routledge, 1957.

BOHM, D. O aparente e o oculto: entrevista com David Bohm realizada pelo prof. Alberto Luiz da Rocha Barros. **Estudos Avançados**, v. 4, n. 8, p. 188-198, 1990a.

BOHM, D. A new theory of the relationship of mind and matter. **Philosophical Psychology**, v. 3, n. 2-3, p. 271-286, 1990b.

BOHM, D.; WEBER, R. **O universo que dobra e desdobra: uma conversa com David Bohm**. In: WILBER, K. O paradigma holográfico – e outros paradoxos. São Paulo: Cultrix, 1995. p. 45-104.

BOHM, D. **A totalidade e a ordem implicada**. Trad. Mauro de Campos Silva. São Paulo: Cultrix, 2001.

BOHR, N. **Física atômica e conhecimento humano**. Trad. Vera Ribeiro. Rio de Janeiro: Contraponto, 1995.

BORNHEIM, G. A. **Os filósofos pré-socráticos**. São Paulo: Cultrix, 2008

BORN, M. The Mechanics of the Atom, Bell, 1927a.

BORN, M. Physical aspects of quantum mechanics, *Nature*, **119**, 354–357. 1927b.

CARMO, M. P. Geometria Reimanniana. 5a ed. Rio de Janeiro: IMPA, 2015.

COMPTON, A. H. A quantum theory of scattering of X-rays by light elements. *Physical Review*, [série 2] 21: 483-502, 1923.

COURANT, R.; HILBERT, D. *Methods of Mathematical Physics*, Vols. 1 and 2, Interscience Publishers, New York, 1953-1962.

DAMÁSIO, A. **O mistério da consciência: do corpo e das emoções ao conhecimento de si**. Trad. Laura Teixeira Motta. São Paulo: Companhia das Letras, 2000.

DAMÁSIO, A. Como o cérebro cria a mente. **Scientific American**, edição Especial: segredos da mente, n. 4, p. 6-11, 2004.

DAVISSON, C. J; GERMER, L. H. “Diffraction of Electrons by a Crystal of Nickel” *Physical Review* 30, 705-740. (1927).

DE BROGLIE, L. A tentative theory of light quanta. **Philosophical Magazine**, 47: 446-458, 1924(a) .

DE BROGLIE, L. *Thésés. Recherches sur la théorie des quanta*. Paris: Masson et Cie Éditeurs, 1924(b)

DESCARTES, R. **Meditações metafísicas**. Trad. Maria Ermantina de Almeida Prado Galvão. 4ª ed. São Paulo: Martins Fontes, 2016.

DEUTSCH, D. Quantum computation. **Physics World**, v. 5, n. 6, p. 57-61, 1992.

DIRAC, P. A M. The quantum theory of the electron. **Proc. R. Soc. Lond.** **A117**610–624. 1928a. <http://doi.org/10.1098/rspa.1928.0023>

DIRAC, P. A. M. The quantum theory of the Electron. Part II. **Proc. R. Soc. Lond.** **A118**351–361. 1928b. <http://doi.org/10.1098/rspa.1928.0056>

EINSTEIN, A. (1905a). «Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen» (PDF). *Annalen der Physik*. **322** (8): 549–560. doi:10.1002/andp.19053220806

EINSTEIN, A. (1905b). “Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt” (PDF) . *Annalen der Physik* . **322** (6): 132–148. Bibcode : 1905AnP ... 322..132E . doi : 10.1002 / andp.19053220607

EINSTEIN, A. PODOLSKY, B. ROSEN, N. “Can quantum-mechanical description of physical reality be considered complete?” *Physical Review* **41**, 777, 1935.

EINSTEIN, A. **The Collected Papers of Albert Einstein**. Edited by A. J. Kox, M. J. Klein and R. Schulmann. Princeton: Princeton University Press, 1997.

EINSTEIN, A. **A Teoria da Relatividade Especial e Geral**. Tradução: Carlos Almeida Pereira. Rio de Janeiro: Contraponto, 1999.

EISBERG, R.; RESNICK, R., **Física Quântica: átomos, moléculas, sólidos, núcleos e partículas**. Editora Elsevier, 1979.

FEYNMAN, R. P.; LEIGHTON, R. B.; SANDS, M. **Lições de física de Feynman**. Porto Alegre: Bookman, 2008.

FREIRE JR, O., PESSOA JR, O., BROMBERG, J. L. **Teoria Quântica: estudos históricos e implicações culturais [online]**. Campina Grande: EDUEPB; São Paulo: Livraria da Física, 2011. 456 p. ISBN 978-85-7879-060-8. Available from SciELO Books. p. 251-63.

FREIRE JR., O.; PATY, M.; BARROS, A. L. R. David Bohm, sua estada no Brasil e a Teoria Quântica. **Estudos Avançados**, v. 8, n. 20, p. 53-82, 1994.

FRÖHLICH, H. Long-range coherence and energy storage in biological systems. *International Journal of Quantum Chemistry* **2**, 641-649, 1968.

GALEGGIO, M, B. The Bohm-Penrose-Hameroff model for consciousness and free will-theoretical foundations and empirical evidences. **Pensamiento**, v. 67, n. 254, p. 661-674, 2011.

GOMEZ, G.; KOTSIREAS, I.; GKIGKITZIS, I.; HARANAS, I.; FULLANA, M. J. Space time geometry in the atomic hydrogenoid system. Approach to a dust relativistic model from causal quantum mechanics. **Revista Mexicana de Física**, v. 64, n. 1, p. 18-29, 2018.

GRIFFITHS, DAVID, J. Introduction to Quantum Mechanics, 2nd Edition, Prentice-Hall, 2005.

HAMEROFF, S. Quantum computation in brain microtubules? The Penrose–Hameroff “Orch OR” model of consciousness. *Philos Trans R Soc Lond Ser A, Math Phys Sci* 1998; 356:1869-96.

HEISENBERG, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Physik* **43**, 172–198. <https://doi.org/10.1007/BF01397280> (1927).

HELAYËL-NETO, J. A.; SIQUEIRA-BATISTA, R.; ALVES FERREIRA, R.; DE JESUS, V. L. B. Neurociências e física contemporânea: a vida no espaço-tempo. In: Esperidião Antonio, V. Neurociências: diálogos e interseções. Rio de Janeiro: Rubio, 2012, p. 505-527.

HELAYËL-NETO, J. A.; SIQUEIRA-BATISTA, R.; ALVES FERREIRA, R.; DE JESUS, V. L. B. Neuroscience, space and time. **Vértices**, v. 17, p. 237-253, 2015.

HELAYËL-NETO, J. A.; NEVES, M. J.; OSPEDAL, L. P. R.; OLIVEIRA, J. B. Dispersion relations in non-linear Electrodynamics and the kinematics of the Compton Effect in a magnetic background, arXiv:210103642 [hep-ph].

HEWITT, P. G. Física conceitual. 11. ed. Porto Alegre: Bookman, 2011.

JAEGER, W. **La teologia de los primeros filosofos griegos**. Traducción de José Gaos. México: Fondo de Cultura Económica, 1952.

JAMALI, M; GOLSHANI, M; JAMALI, Y. A proposed mechanism for mind-brain interaction using extended Bohmian quantum mechanics in Avicenna’s monotheistic perspective. **Heliyon**, v. 5, n. 7, p. e02130, 2019.

JIBU, M.; YASUE, K. The basics of quantum brain dynamics. In: PRIBRAM, K. H. Rethinking neural networks: quantum fields and biological data. Proceedings of the First Appalachian Conference on Behavioral Neurodynamics, L. Erlbaum, Hillsdale (NJ, EUA), 1993., p. 121-145.

JIRSA, V. J.; HAKEN, H. Field theory of electromagnetic brain activity, *Phys. Rev. Lett.* v. 77, n. 5, p. 960, 1996.

LEWIS, G. N. The nature of light. **Proceedings of the National Academy of Science**, v. 12, p. 22-29, 1926a.

LENT, R. **Cem bilhões de neurônios? Conceitos fundamentais de neurociências**. 2ª ed. São Paulo: Atheneu, 2010.

LEWIS, G. N. The conservation of photons. **Nature**, v. 118, p. 874-875, 1926b.

LI, T., TANG, H., ZHU, J., & ZHANG, J. H. (2019). The finer scale of consciousness: quantum theory. *Annals of translational medicine*, 7(20), 585. <https://doi.org/10.21037/atm.2019.09.09>

MACH, E. **The science of mechanics**. Trad. Thomas J. McCormack. 6. ed. LaSalle, Illinois: Editora Open Court, 1960.

MANFIO, F. **Geometria Riemanniana: Notas das aulas com exercícios** (atualizado em 05/07/2018). Universidade de São Paulo. Disponível em: <<https://sites.icmc.usp.br/manfio/Riemanniana.pdf>>.

MARIETTA JUNIOR, D. E. Conscience in greek stoicism. *Numen*, v. 17, n. 3, p. 176-187, 1970.

MAXWELL, J. C. A dynamical theory of the electromagnetic field. **Proceedings of the Royal Society of London XIII**, 531, 1864.

MAXWELL, J. C. A dynamical theory of the electromagnetic field. **Philosophical Transactions of the Royal Society of London**, v. 155, p. 459, 1865.

MCFADDEN, J. Integrating information in the brain's EM field: the cemi field theory of consciousness. **Neuroscience of Consciousness**, 6(1): niaa 016, 2020.

MELO, W. A. M. **Aspectos peculiares das eletrodinâmicas de Maxwell e Maxwell-Chern-Simons Planares**. Tese do Doutorado. Centro Brasileiro de Pesquisas Físicas. Rio de Janeiro, 2001.

MILLIKAN, R. A direct determination of h. **Physics Review**, n. 4, v. 1, p. 73, 1914.

NELSON-ISAACS, S. Spacetime paths as a whole. **Quantum Reports**, v. 3, n. 1, p. 13-41, 2021.

NICOLELIS, M. A. L. **Muito além do nosso eu**. São Paulo: Companhia das Letras, 2011.

NICOLELIS, M. A. L. **O verdadeiro criador de tudo: Como o cérebro humano esculpiu o universo como nós o conhecemos**. São Paulo: Crítica, 2020.

NUSSENZVEIG, H. M. **Curso de física básica: ótica relatividade e física quântica**. 4ª ed. São Paulo: Editora Edgard Blücher Ltda, 2014. Vol. 4.

PATY, M. La non-séparabilité locale et l'objet de la théorie physique, **Fundamenta Scientiae**, 7, 1986, p.47-87.

PENROSE, R. HAMEROFF, S. Consciousness in the universe: A review of the 'Orch OR' theory, **Physics of Life Reviews**, Volume 11, Issue 1, 2014, Pages 39-78, ISSN 1571-0645, <https://doi.org/10.1016/j.plrev.2013.08.002>.

PENROSE, R. **The emperor's new mind: concerning computers, minds, and the laws of physics** Oxford University Press, Oxford (1989).

PENROSE, R. **Shadows of the mind: an approach to the missing science of consciousness** Oxford University Press, Oxford (1994)

PENROSE, R. **Gravity and quantum mechanics**. In: Gleiser RJ, Kozameh CN, Moreschi OM, editors. General relativity and gravitation 13. Part 1: plenary lectures. Proceedings of the thirteenth international conference on general relativity and gravitation. Cordoba, Argentina, 28 June–4 July 1992. Bristol, Philadelphia: Institute of Physics Publishing; 1993. p. 179–89.

PENROSE R; BELL, J. **State reduction, and quanglement**. In: Bertlmann RA, Zeilinger A, editors. Quantum unspeakables: from bell to quantum information. Berlin: Springer-Verlag; 2002. p. 319.

PENROSE, R. **Godel, the Mind, and the Laws of Physics**. In Matthias Baaz (ed.), Kurt Gödel and the Foundations of Mathematics: Horizons of Truth. Cambridge University Press. pp. 339. 2011.

PENROSE, R. On gravity's role in quantum state reduction. **General Relativity Gravitation**, 28 (1996), pp. 581-600.

PENROSE, R. Black holes, quantum theory and cosmology. Fourth International Workshop DICE 2008, J Phys Conf Ser, 174 (2009).

PESSOA JR., O. **A física quântica seria necessária para explicar a consciência?** In: Questões Metodológicas em Ciências Cognitivas, 1994, São Paulo. Coleção Documentos - Série Ciência Cognitiva - 20. São Paulo: Instituto de Estudos Avançados - USP, 1994. p. 184-189.

PINTO-NETO, N. The Bohm interpretation of quantum cosmology. **Foundations of Physics**, v. 35, p. 577-603, 2005.

PLANCK, M. (1901). «Über das Gesetz der Energieverteilung im Normalspektrum». **AnnalenderPhysik**.4:553. Bibcode:1901AnP...309..553P. doi:10.1002/andp.19013090310

PLEBANSKI, J. **Lectures on Nonlinear Electrodynamics**, NORDITA, Copenhagen, 1970.

SANTOS, F. M.; PESSOA JUNIOR, O. Delineando o problema da medição na mecânica quântica: o debate de Margenau e Wigner *versus* Putnam. **Scientia et Studia**, v. 9, n. 3, p. 625-644, 2011.

SATTIN, D.; MAGNANI, F. G.; BARTESAGHI, L.; CAPUTO, M.; FITTIPALDO, A. V.; CACCIATORE, M.; PICOZZI, M.; LEONARDI, M. Theoretical models of consciousness: a scoping review. **Brain Science**, v. 11, n. 5, p. 1-58, 2021.

SCHRODINGER, E. **Quantisierung als Eigenwertproblem** [German; tr. “Quantization as an Eigenvalue Problem”]. Four communications (Ger *Mitteilungen*) (1926a).

SCHRODINGER, E. “An Undulatory Theory of the Mechanics of Atoms and Molecules”. *Phys. Rev.* **28** (6): 1049–1070. Bibcode:1926PhRv..28..1049S. doi:10.1103/PhysRev.28.1049. (1926b).

SCHRODINGER, E. Die Gegenwärtige Situation in der Quantenmechanik. Die Naturwissenschaften, **Naturwissenschaften** 23, Oxford, (49): 807-812, doi: 10.1007/BF01491891, 1935.

SIEB, R. A. Human conscious experience is four-dimensional and has a neural correlate modeled by Einstein’s Special Theory of Relativity. **NeuroQuantology**, v. 14, p. 630-644, 2016.

SIEB, R. A. Space-time intervals underlie human conscious experience, gravity, and everything. **NeuroQuantology**, v. 17, n. 5, p. 87-89, 2019.

SHESTAKOVA, T. P. Is the Copenhagen Interpretation inapplicable to quantum cosmology? **Universe**, v. 6, n. 128, p. 1-20, 2020. <https://doi.org/10.3390/universe6090128>

SIQUEIRA-BATISTA, R.; BATISTA, R. S.; COSTA, C. R. B. S. F.; HELAYËL-NETO, J. A. Realidade e consciência: notas sobre o pensamento de David Bohm. **Physicae**, v. 6, p. 7-11, 2006.

SIQUEIRA-BATISTA, R.; BATISTA, R. S.; ALVES FERREIRA, R.; HELAYËL-NETO, J. A. La mecánica cuántica de David Bohm: formalismo matemático e implicaciones epistemológicas. **Prometheus**, v. 8, p. 109-128, 2015.

SEARLE, J. R. **Consciência e linguagem**. São Paulo: Martins Fontes, 2010.

SILVA, M. M.; FUHRMEISTER, A. V.; BRUM, A. M.; MEDEIROS, M. S. FERREIRA, P. P.; P. P. BREDÁ, R. L.; VIERO, R.; LEITE, S. S. A Consciência: algumas concepções atuais sobre sua natureza, função e base neuroanatômica. **Revista de Psiquiatria do Rio Grande do Sul**, v. 25, n.1 (supl.), p. 52-64, 2003.

SILVA, M. M. Syneidesis, para além da fragmentação do espírito: uma introdução à teoria do abrangente e do pensar em imagens. **Trans/Form/Ação**, v. 41, p.13-30, 2018.

SUTHERLAND, S. **The Macmillan Dictionary of Psychology**. Macmillan; (1989) em De La Fuente R, Leefmans FJA, Biología de la Mente, México: Fondo de Cultura Económica; 1999.

TOURINHO, C. D. C. **A filosofia da mente hoje**. In: ESPERIDIÃO ANTONIO, V. Neurociências: diálogos e interseções. 1ª ed. Rio de Janeiro: Editora Rubio, 2012, v. 1, p. 530-545.

Wang, Xi-Lin; Luo, Yi-Han; Huang, He-Liang; Chen, Ming-Cheng; Su, Zu-En; Liu, Chang; Chen, Chao; Li, Wei; Fang, Yu-Qiang. 18-qubit entanglement with six photons' three degrees of freedom. **Physical Review Letters**, 2018.

WIGNER, E. P. **Events, laws of nature, and invariance principles. Philosophical reflections and syntheses.** New York: Springer, 1995. p. 321-333.

WIGNER, E. P. **The subject of our discussions.** Philosophical reflections and syntheses. New York: Springer, 1995. p. 199-217.

VON NEUMANN, J. **Mathematical Foundations of Quantum Mechanics.** Princeton: Princeton University Press, 1955.

02.

FUNDAMENTOS NEUROBIOLÓGICOS DA CONSCIÊNCIA: IMPLICAÇÕES PARA O DEBATE SOBRE O LIVRE-ARBÍTRIO

Vanderson Esperidião Antonio

Lindisley Ferreira Gomides

Mirna Brito

1. INTRODUÇÃO

As neurociências podem ser definidas como um conjunto de ciências que buscam compreender o funcionamento do Sistema Nervoso (SN) a partir da análise de diferentes áreas do conhecimento como, por exemplo, Neuroanatomia, Neurofisiologia, Neurociências Cognitivas, Psicologia, Biologia, dentre outras (KANDEL, 2014; LENT 2010). Nessa perspectiva, é possível o estabelecimento de pontes e interseções entre essas e outras que estavam restritas às suas especificidades, favorecendo uma abordagem interdisciplinar com a formulação de hipóteses, pesquisas e métodos.

Esse diálogo proposto através das Neurociências convida ao compartilhamento de entendimentos e reflexões, estabelecendo uma mútua troca de saberes com o objetivo comum de encontrar a compreensão a respeito de determinadas questões as quais, até então, pareciam não acusar contribuições correlacionadas, a exemplo do caso das ciências humanas e exatas sobre da inteligência artificial.

Com base nessa estreita relação, investigações científicas contemporâneas sobre funções cerebrais têm agregado respostas a várias questões com o auxílio da tecnologia, como os exames de imagem que permitem observações detalhadas e de alta resolução espacial, estrutural, com destaque aos de avaliação funcional. Entre os mais recentes estão reunidos sólidos exemplos às contribuições cerebelares na afetividade para além da motricidade (DAMIANI et al., 2016) e as correlações do sistema límbico

à regulação autonômica e neuroendócrina (MULKEY e PLESSIS, 2019), dentre outras. Entretanto, apesar dos avanços alcançados, ainda persistem muitas especulações, teorias e um conjunto de dúvidas e de detalhamentos aguardando explicações para antigas e latentes questões viscerais, como as da consciência (MEESE, 2018).

A consciência representa um fenômeno complexo, intrigante e misterioso do SN (DAMÁSIO, 1994; MCGINN, 1999; EDELMAN, 2004; JAYNES, 2000) que vem provocando interrogações, discussões e análises ao longo das últimas quatro décadas. As dificuldades relacionadas ao tema estão circunscritas a vários aspectos que passam desde seus conceitos possíveis, referenciais teóricos relacionados às ciências humanas, sociais e biológicas (CHALMERS, 1996; DENNETT, 1991).

Para o enfrentamento inicial ao desafio de apresentar uma proposta de entendimento sobre a consciência, seria razoável analisar a etimologia da palavra. Partindo de duas expressões do latim, têm-se *cum* (com) e *scio* (eu sei/eu conheço) (GENNARO, 1996), sugerindo propriedade e autonomia do saber. Na língua inglesa, o adjetivo *conscius* (*aware*), o verbo *conscio* e o substantivo *consciouness* (*awareness*) se organizam para conferir conteúdo a forma. Ainda, o termo *wakefulness* – estar consciente, em estado de vigília em oposição ao coma, no qual há ausência desses; *conscience* – consciência moral, com envolvimento da ética e do julgamento; e, por fim, o termo *consciouness*, que remete ao significado da existência, ao próprio ser e suas experiências.

Na língua portuguesa, a mesma palavra pode ser empregada para expressar mais de um significado: consciência moral, consciência de responsabilidade e a própria vigília, estado de alerta, em oposição aos diferentes níveis de coma, o que, da mesma forma, não facilita sua conceituação (KLEM, SILVA e GAMA, 2019). O fato de o indivíduo possuir a capacidade de identificar a própria consciência de cada ação desejada, de maneira comum, usual e rotineira, tornaria mais suave o esclarecimento do termo (ROSENTHAL, 2009).

Do ponto de vista teórico, seria possível distingui-la em três condições básicas: *i*) a primeira, condicionada ao estado vigil, acordado, capaz de perceber e interpretar adequadamente os estímulos através de diferentes órgãos dos sentidos e, por isso, mais relacionada a uma interpretação biológica. Cabe destacar, nesse caso, a observação de que essa condição estaria relacionada ao sujeito consciente (TONONI, 2004; LAUREYS, 2005); *ii*) a segunda, guarda como característica a capacidade de o sujeito estar ciente sobre determinada

situação, sensação, objeto, condição de reconhecimento, estando próximo ou não, mas evocando uma estruturação mental organizada; e, por fim, *iii*) a terceira caracterização, que diferentemente das duas anteriores, apresenta uma representação mais filosófica, de que a sensação de consciência surge a partir de percepções, algumas vezes subliminares, inclusive sobre os pensamentos (ROSENTHAL, 2009). Em resumo, a consciência envolve monitoramento do ambiente e de nós mesmos, de modo que perceptos, memórias e pensamentos sejam representados, além do controle do ambiente e de nós mesmos, contribuindo para que o indivíduo possa iniciar e finalizar atividades comportamentais e cognitivas (KIHLSSTROM, 1992).

Além de tentar alcançar a compreensão dessa natureza conceitual, outra questão associada a consciência que merece mais discussões, com aprofundamento na problemática interdisciplinar é o livre-arbítrio. Considerado dependente do estado de consciência, sua relação com o determinismo dos eventos naturais tem uma longa história na filosofia e no pensamento científico. A própria subjetividade coloca-se como alvo de questionamentos e esclarecimentos frente ao tratamento da consciência pelo viés da causalidade, prezando a identidade e a responsabilização da conduta humana.

Dentro desse cenário, alguns autores defendem a compatibilidade entre a liberdade da ação humana e a causalidade natural, enquanto outros sustentam que essas são incompatíveis. Para esses, alguns concluem, a partir dessa incompatibilidade, que não há verdadeira liberdade da vontade e de ação. De forma contrária, outros estudiosos concebem a consciência como uma realidade isolada do mundo físico ou emergente dele e, portanto, não sujeita à causalidade natural.

Dentro desse contexto, a proposta do presente capítulo é fazer uma conexão neurobiológica e filosófica das questões relacionadas a consciência, pontuando suas principais hipóteses e conexões com o livre-arbítrio e a tomada de decisão. Para tal, o desenho da escrita foi elaborado inicialmente com reflexões filosóficas sobre a consciência, com destaque para os filósofos contemporâneos e suas teorias; seguido da base neurobiológica dos fenômenos ligados a consciência e a circuitaria neural de sinalização que faz jus aos avanços das pesquisas em neurociências, com base nas ferramentas tecnológicas e nos exames de imagem para a exposição dos dados. Ao final, são apresentados os aspectos do livre arbítrio e alguns pontos de discussão correlacionando o tema com as diversas faces já caracterizadas da consciência.

2. REFLEXÕES FILOSÓFICAS SOBRE A CONSCIÊNCIA

Ao longo da história da filosofia, o conceito de consciência foi explorado sob diferentes ângulos. Na antiguidade, a consciência era caracterizada como “metafísica”, desligada do mundo material, responsável pela essência das coisas através da razão. Para os filósofos desse período, a consciência intelectual é o centro do homem, cujo corpo humano era dividido em partes irracional e enganosa (ROSENTHAL, 1986; GULICK, 2012).

O modo metafísico consiste na crença de uma realidade autônoma e objetiva, independente do sujeito. O todo existe a favor de uma determinada essência, ou seja, cada ser, cada indivíduo é do jeito que é porque, ao existir, está realizando uma natureza que define suas características específicas, as quais justificam o seu pertencimento a um grupo específico, bem como as suas interações sociais (MONDIN, 2013).

Ainda sob a perspectiva metafísica, com o fortalecimento do cristianismo na Idade Média, a consciência adquire um caráter religioso, pretendendo alcançar a natureza divina enquanto princípio da verdade, anulando a razão humana. Permanece, portanto, a distinção e oposição entre consciência e corpo, destacando o papel primordial da consciência na aquisição do conhecimento. Essa forma de pensar e de alcançar a sociedade permaneceu até a Idade Moderna, sendo analisada e criticada pelos filósofos modernos pela discordância em atribuir a essência das coisas pela vertente da razão natural. Segundo esses, o que o homem pode conhecer, de fato, por meio da razão natural é o seu próprio pensamento ou o fenômeno da consciência segundo a própria percepção. Desde então, desenvolveu-se uma nova forma de compreensão do mundo e suas percepções, especialmente através da ciência (BALASHOV, ROSENBERG, 2002).

Ao avaliar o conceito de consciência com base na filosofia moderna, a despeito de certas variações, o debate teve início a partir das proposições de René Descartes (1596-1650) e estendeu-se até Immanuel Kant (1724-1804), apresentando uma filosofia cartesiana que, além de romper com a escolástica, agrega valor a uma nova ideia de ciência. René Descartes, no século XVI, foi o primeiro a estudar a questão cérebro e mente de forma organizada, caracterizando o dualismo cartesiano, baseado na diferença do cérebro (*res extensa*) e da mente (*res cogitans*) (FRITH, REES, 2007). Para Descartes, consciência implica em existência; um objeto natural, de caráter intimista, sedimentando uma tradição individualista e subjetivista da

mente, evidenciada na sua famosa frase “*cogito ergo sum*” (“*Penso, logo existo*”). Assim, o fato de existir, enquanto ser pensante, e a mente, através da consciência e dos sentidos, conferem a base para a relação com o mundo. O conhecimento da mente sobre si mesma é, portanto, um processo reflexivo, um olhar interior para as suas próprias ocorrências. De acordo com essa análise, a consciência é a forma de o pensamento transitar em torno do seu próprio eixo. Descartes também reúne argumentos que contemplam o dualismo entre corpo e mente, apontando a natureza divisível do corpo e indivisível da mente (DESCARTES, 2011).

Ao longo do tempo, a teoria do dualismo cartesiano foi contestada por outros filósofos e cientistas com a argumentação de que a mente é o resultado de propriedades físicas do cérebro, embora sem mecanismos claros de como isso acontecia. Assim, a fim de avaliar essa relação, várias teorias foram propostas e endereçadas em dois grandes grupos: dualistas e monistas.

Para o primeiro, adeptos da teoria cartesiana proposta por Descartes, a mente não é derivada da atividade neurofisiológica, sendo a consciência um fenômeno não físico. Por outro lado, para os monistas, apesar de agrupados pela mesma opinião, ainda foram subdivididos em materialistas/fisicalistas/reducionistas – defendendo a ideia de que a mente era o cérebro; e os emergentes e outros ramos – que acreditavam que a mente emerge do funcionamento do cérebro (VICTORINO, 2009).

Mesmo considerando os avanços científicos, incluindo recursos tecnológicos e exames de imagem, essa estratificação de pensamentos ainda permanece e um boa parte de cientistas continua defendendo que o estado mental não está vinculado ao estado físico cerebral. Na filosofia contemporânea, dois influentes filósofos contemporâneos, David J. Chalmers e Daniel C. Dennett, trazem importantes considerações sobre o fenômeno da consciência, no sentido de aproximar as bases da filosofia e os dados científicos até então analisados.

Para Chalmers, a divisão das propriedades mentais em fenomenais e psicológicas tem como efeito a avaliação da relação mente e corpo em duas significações: uma fácil e outra difícil. A primeira está relacionada às propriedades psicológicas, apresentando características estruturais e funcionais. A segunda, de difícil esclarecimento, engloba as propriedades subjetivas da consciência, com caracterizações específicas e resultantes da experiência individual e qualitativa do sujeito, e não de suas estruturas e

funções (CHALMERS, 1996). Nesse cenário, para Chalmers, o termo *qualia* classifica as propriedades qualitativas e subjetivas inerentes a consciência como não física, sugerindo, assim, o retorno ao dualismo cartesiano.

Reforçando essa ideia, a fim de demonstrar que as teorias materialistas/fisicalistas/reducionistas não são claras para definir a consciência, Chalmers introduz a figura de um zumbi nas argumentações filosóficas, na perspectiva de apontar que todos os fatos físicos não sustentam as explicações para o fenômeno da consciência (CHALMERS, 1996). De acordo com essa análise, o zumbi é fisiologicamente idêntico ao ser humano, com as mesmas propriedades funcionais, porém, sem pensamentos, intenção e consciência de si. Em outras palavras, o que Chalmers pretendia dizer é que o caráter funcional que explica os diversos estados psicológico e físico não explica os estados fenomenais.

O pesquisador afirma que a veracidade da imaginação de que todos os fatos físicos possam coexistir sem a consciência, então, os fatos físicos não esgotam todos os fatos. A possibilidade de criação de um mundo identicamente ao humano, porém sem consciência, remete a ideia de que a consciência é uma propriedade que não sobrevém ao físico e, como tal, não pode ser explicada de forma simples e pontual (CHALMERS, 1996).

Diante disso, um outro extremo também foi apontado, direcionando filósofos a defenderem argumentos paradoxais de que apesar de não possuírem consciência, os zumbis não sabem disso. Então, como avaliar se o julgamento humano sobre a própria consciência não é incorreto? Ou seja, o ser humano pode ser um zumbi e não ter conhecimento dessa realidade. Chalmers, no entanto, não pretende confrontar esses dois opostos, sugerindo que o caminho é aprender a mesclar ambos, observando o que é aparentemente óbvio: a experiência consciente do indivíduo é o que permite identificá-lo como um ser consciente (CHALMERS, 1996).

Embora paradoxal, esse argumento ganhou atencioso apoio dos funcionalistas e estudiosos da consciência, porém foi claramente refutado pelos materialistas, destacando o filósofo Daniel C. Dennett por suas concepções dualistas. Dennett, baseado na inteligência artificial e na evolução das neurociências, sugere que o estudo da consciência só é possível com base em estudos sólidos do funcionamento cerebral, apoiando a teoria fisicalista e funcionalista na mesma proporção em que identifica o homem como uma máquina biológica controlada pelo próprio cérebro. Assim, de modo semelhante ao que

argumentava o neurólogo John Hughlings um século antes (KANDEL, 2014), para Dennett, a fim de estudar a consciência, seria necessário entender previamente o funcionamento do cérebro, pois a mente consciente é resultado da atividade computacional das áreas associativas do encéfalo (DENNETT, 1991).

Com base nisso, alguns inferiram que os *qualia* não existem, visto que os pensamentos e as decisões do indivíduo são resultado da atividade cerebral (TAVARES, 2018). Essa conclusão simplificada foi considerada incipiente, pois apesar da tendência materialista, as neurociências ainda não conseguiram confirmar ou negar com bases sólidas a proposição de Dennett, cuidadosamente considerando ser *qualia* o resultado da atividade cerebral ou ainda, propriedades físico-químicas geradas pelo cérebro em determinadas situações que, subjetivamente, são vivenciadas pelo indivíduo (DENNETT, 1991; DE SOUSA, 2015).

Na *teoria dennettiana*, o cérebro deve realizar duas tarefas simples: processar a informação em suas áreas específicas e, após, reunir tudo em um único conjunto para compor a experiência consciente. Cada circuito neural opera em paralelo aos outros, cada qual com a sua estrutura e a sua categoria mental, sem a necessidade de reprocessamento (DENNETT, 1991).

Nessa perspectiva, frente a imprecisão definitiva de como caracterizar uma ferramenta e/ou uma ciência unicamente voltadas para os estudos da consciência humana, esses filósofos ganharam adeptos, na tentativa de reunir argumentos próprios, segundo análises prévias pautadas em torno da ciência e da filosofia, sobre a origem e a natureza da consciência. Surge então a *fenomenologia*, com objetivo de chegar a um fim comum, descrevendo fenômenos experienciados pela consciência, nos quais o sujeito e o objeto constroem juntos o conhecimento. Para tal, Merleau Ponty dá um novo significado a consciência, caracterizando-a como percepção do sujeito, indissociável da razão e da emoção na interpretação dos fatos (MERLEAU-PONTY M., 1994), claramente expressa na qualidade com que o corpo expõe a unidade humana.

Com o olhar voltado sobre a psicanálise, teoria desenvolvida a fim de explicar o funcionamento da mente humana, Freud aponta que a consciência se limita a uma pequena fração de nossa vida psíquica total, sendo algumas memórias, impulsos e desejos não são acessíveis a consciência humana e os processos psíquicos, em sua maioria inconscientes, dominados por tendências sexuais naturais da biologia humana (GOMES, 2003).

Segundo Freud, memórias e desejos emocionalmente dolorosos, uma vez reprimidos, são desviados para o inconsciente e podem influenciar as ações do indivíduo, ainda que não haja consciência sobre isso. Alguns exemplos são sonhos, comportamentos irracionais, peculiaridades e deslizes de linguagem, como o “lapso freudiano” ou “ato falho”, o qual representam comentários involuntários que supostamente revelam impulsos ocultos (KANDEL, 2014).

Em sua obra “O Ego e o Id” (1923), Freud expôs uma divisão de mente humana em três partes: *i*) o ego, que se identifica à consciência humana; *ii*) o superego, a consciência moral humana, ou seja, os princípios sociais e as proibições que nos são impostas nos primeiros anos de vida e que acompanham o homem de forma inconsciente ao longo da sua existência; e, por fim, o *iii*) id, caracterizado pelos impulsos múltiplos da libido, voltados para o prazer. Freud chama o *ego* de ‘eu’, o *superego* de ‘supereu’ e o *id* de ‘isso’. A psicanálise, realizada por meio de conversação, é uma ferramenta terapêutica para as doenças mentais a partir da interpretação desses fenômenos, levando o paciente a avaliar as origens de seu problema, para então, a partir da consciência desse(s) fator(es), elaborar uma estratégia de cura (FREUD S, 1990).

Apesar da grande influência que Freud exerceu em várias correntes da ciência, da arte e da filosofia, muitos profissionais - dentre eles filósofos, psicólogos e psiquiatras - pontuam a não conformidade com esses conceitos, pelo fato de não ser uma hipótese e por ultrapassar o que a ciência conseguiu desvendar, até então, levando a afirmação da própria subjetividade do terapeuta. Na visão de Merleau Ponty, o inconsciente de Freud é carregado de afeto e emoções, percebidos na consciência e, por isso, descaracteriza a autonomia do sujeito, visto que esses pensamentos inconscientes são independentes do sujeito como protagonista de suas ações.

Francisco Varela segue essa mesma proposição. Em um estudo sobre a neurofenomenologia da consciência, o filósofo afirma que a real fonte de conhecimento sobre a mente funcional e a elucidação dos fenômenos da consciência exigem uma ruptura com a tradicional linha do pensamento científico ocidental (CARDOSO, 2006), limitada ao tecnicismo e carregada de obstruções e impedimentos direcionados ao próprio ser para, então, reestruturar um novo modelo científico, com reflexões cuidadosas sobre a existência do corpo (VARELA, THOMPSON, ROSCH, 2001), promovendo a união entre as ciências e os saberes.

Na complementação dessas discussões, de acordo com as neurociências e conforme proposto por Dennett, Antônio Damásio, um reconhecido neurocientista que segue investigações do cérebro consciente, propõe como argumento a essência das transformações desencadeadas pelo corpo em movimento. Essa discussão tem como base o pensamento de Baruch de Espinosa, grande nome da filosofia moderna ao lado de Descartes, e abrange uma nova visão, a de que é possível avaliar a veracidade do resultado lógico auto-organizado de um movimento físico e da substância corpórea ou material, complementando a teoria de Dennett sobre a consciência (BLOCK, 2007).

John Searle, seguido de Thomas Nagel, definiram três características essenciais para a consciência: a subjetividade, a unidade e a intencionalidade (SEARLE, 2010). Sob essa análise, Searle e Nagel, juntamente com Antônio Damásio, Francis Crick, Christof Koch e Gerald Edelman, compõem o grupo que defende que a consciência é um fenômeno discreto, fruto de processos cerebrais e, assim sendo, essas três características devem ser confrontadas, a fim de se entender como a experiência é construída na perspectiva de cada organismo vivo (KANDEL, 2014), sugerindo uma não concordância com o dualismo.

A intencionalidade pode ser definida como a experiência consciente do indivíduo que o faz conectar sucessivos momentos que se direcionam para algum objetivo, diferente da subjetividade, considerada o maior desafio filosófico e científico da humanidade (KANDEL, 2014). Esse entrave se deve a natureza intrínseca do indivíduo, centro da experiência, a qual apresenta sensações únicas e privadas – sucessos e desapontamentos, alegrias e dores – além de ideias e estados de humor que parecem ser mais reais do que quando comparadas a de outras pessoas. Por exemplo: o aroma do café apreciado por um indivíduo vai além do entendimento da sua capacidade sensorial, podendo ser diferente da apreciação de outrem, partindo do mesmo café e do mesmo momento compartilhado, sob a mesma intensidade. Sob essa análise, não seria apenas o aroma do café percebido sensorialmente, mas sim, uma sensação, uma percepção consciente que podem, inclusive, desencadear memórias intrínsecas que foram perceptíveis ao indivíduo na presença daquele aroma – o cheiro do café de um parente, momentos vividos na infância ao lado da família, café de finais de semana, enfim, cada qual com o seu particular, por isso é subjetivo.

Na mesma linha de pensamento, a unidade da consciência também se refere ao fato de as experiências do indivíduo serem sentidas de forma

própria e unificada. A atenção voltada para o momento presente, a respiração, a temperatura do ar entrando pelas narinas, o contato do corpo com os objetos é único e intransferível. A construção de diálogos através de uma junção de palavras é um evento consciente, apesar de o indivíduo não se preocupar em como fazer isso passo a passo (KANDEL, 2014). Como essa análise também envolve a matéria, é de extrema importância a utilização das bases epistemológicas científicas da pesquisa, propondo correlacionar as bases neurobiológicas da consciência com a existência do ser.

Reunindo então essas três características - intencionalidade, subjetividade e unidade - faz-se necessário uma revolução na neurobiologia para explicar todos os mecanismos de significação da consciência que, até então, as neurociências ainda não conseguiram elucidar por completo. Diante dessa complexa rede de discussões, seria adequado para o estudo da consciência que os termos dualistas não interferissem nos estudos neurobiológicos pautados na investigação científica, ainda que com recursos incipientes. Da mesma forma, não se deve ignorar as propostas de Chalmers e outros filósofos da mente sobre a realidade dos estados qualitativos que caracterizam o mundo fenomênico, porém, considerando os mecanismos físico-biológicos que suportam esses estudos.

3. ABORDAGEM NEUROBIOLÓGICA DA CONSCIÊNCIA

A partir do final do século XX, com o desenvolvimento das Neurociências, houve um incremento na formulação de hipóteses para tentar explicar o fenômeno da consciência. Enquanto a Filosofia, por considerar a natureza subjetiva das coisas, trazia um maior conjunto de teorias, a Ciência, em seu sentido mais estrito, buscou compreender os mecanismos cerebrais subjacentes à consciência. Intervenções cirúrgicas, administração de substâncias químicas e estimulação eletromagnética intracraniana já pontuaram na literatura mudanças nos estados conscientes (BEAR, 2007; PURVES, 2008), contrariando a proposta naturalista de Chalmers para a consciência, bem como a filosófica baseada apenas na subjetividade de demais estudiosos. Apesar de considerado por muitos como um estudo reducionista (SULLIVAN, 2009), a avaliação do nível neurobiológico (celular e molecular) é uma importante ferramenta para buscar respostas para a consciência, visto que essa é o resultado de mecanismos neurobiológicos (BICKLE, 2003). A despeito de parecer inicialmente uma contraposição de entendimentos, essa aparente discordância favorece a

oportunidade para o estabelecimento de análises e pesquisas interdisciplinares a respeito desse enigma (KEE, 2006; TAVARES, 2018).

A abordagem biológica sobre a consciência poderia considerar inicialmente a questão ontogenética e filogenética, já que os vertebrados que apresentam um SN mais organizado são capazes de participar de situações mais complexas, exibindo aproximações comparáveis às observadas nos seres humanos (CARRUTHERS, 2000; BRONFMAN, 2016). Porém, para Searle, o problema mais importante nas ciências biológicas é elucidar exatamente como os processos neurobiológicos do cérebro causam a consciência. Para tal, pesquisadores utilizam modelos teóricos e computacionais para explorar a consciência, reconhecendo a importância de modelos científicos, através da representação fidedigna do fenômeno estudado, a partir de testes e observação (FRIGG, HARTMANN, 2012).

Carlos Eduardo de Sousa realizou um estudo sobre os modelos neurais da consciência e reuniu os principais trabalhos neurocientíficos sobre o tema em (i) Modelo de múltiplos esboços (Dennett), (ii) Hipótese do marcador somático (Damásio), (iii) Modelo Neurodinâmico (Freeman), (iv) Modelo do Espaço Global Neuronal de Trabalho (Baars; Dehaene et al.) e (v) Modelo do Núcleo Dinâmico (Edelman) (SOUSA, 2015). Conforme discussão proposta a seguir, a semelhança das discussões filosóficas, foram contemplados critérios de plausibilidade, de coerência e de aplicabilidade para, então, afirmar se, de fato, esses modelos explicam o fenômeno da consciência.

3.1 Modelo de Múltiplos esboços de Dennett

O modelo de múltiplos esboços proposto pelo Dennett, pretende explicar a consciência com base em teorias cognitivas através de metáforas, excluindo a *qualia* e contrapondo a concepção tradicional de Descartes, conforme já foi abordado anteriormente nas bases filosóficas. Através desse modelo, também conhecido como *Multiple Draft Model*, Dennett argumenta que o cérebro tem o papel de editar as informações recebidas do meio externo, em tempo real e por meio da percepção, sendo essas informações processadas pelas redes neurais (DENNETT, 1991). Assim, essas inúmeras mensagens vão sendo transmitidas e interpretadas resultando em múltiplos esboços (*drafts*), não havendo, portanto, a participação do sujeito na experiência, apenas níveis diferentes de processamento de informações.

Para uma avaliação neurobiológica dessa interpretação, é possível utilizar técnicas como eletroencefalografia (EEG), tomografia de emissão de pósitrons (PET) e imagem por ressonância magnética funcional (fMRI) que permitem ao avaliador confrontar os processos biológicos com os estados internos relatados pelo sujeito. Após registros, os dados brutos são avaliados e recebem significados. Contudo, mesmo depois de ter recebido críticas e adaptado o modelo, substituindo os múltiplos esboços por um conceito de ‘fama’ do cérebro, transitória e resultado de uma variação do fluxo de informações recebidas, Dennett entende que algo está incompleto. Apesar de acertar na descrição funcional, o autor observa que a inexistência da subjetividade atrapalha na avaliação da consciência propriamente dita (DE SOUSA, 2015) e, assim, tal modelo não justifica como os processos neurobiológicos resultam nas propriedades qualitativas.

3.2 Modelo de Damásio: ‘*Somatic Marker Hypothesis*’

Na tentativa de caracterizar melhor a consciência, Antônio Damásio (1999, 2005) fez algumas proposições baseadas em evidências obtidas de pacientes com alterações cerebrais específicas comparadas a presença de alterações apresentadas nas áreas encefálicas. O termo em inglês ‘Somatic Marker Hypothesis’, traduzido como Hipótese do Marcador Somático, divide a consciência em central e ampliada. Assim, foi proposta uma distinção fundamental entre essas subclassificações da consciência, na qual a central estaria relacionada ao *self*, a sensação de ser alguém, de vivenciar suas experiências no momento presente. Para Damásio, a consciência central é independente das funções do neocórtex e estão relacionadas às áreas primitivas presentes na maioria dos mamíferos, como o tronco encefálico e o giro do cíngulo, motivo pelo qual infere-se que esses animais possuam a “sensação de existir”. A consciência ampliada, por sua vez, além do *self* corporal, possui também o “*self* autobiográfico”, na tentativa de manter o equilíbrio e a homeostase.

Segundo esse modelo, a consciência tem início quando o indivíduo vivencia suas experiências e observa um sentimento de que algo acontece, seja no interior ou no exterior do organismo (DAMÁSIO, 2005). Contemplando a ampla distribuição dos receptores e terminações nervosas a nível periférico, esse primeiro contato é possível através das diferentes formas de percepção sensorial – visual, auditiva, tátil, visceral – as quais possuem características específicas (MACHADO, 2014).

Damásio pontua que, ao endereçar os estímulos a áreas específicas do cérebro, surge a interpretação do evento, seguido do relato desse sentimento.

Conforme o cérebro forma imagens de um objeto, por exemplo, essas imagens vão sendo relacionadas a diversos outros mecanismos, a fim de construir um relato significativo continuado e integrado, que caracteriza o *self* que observa, interpreta e vivência. Essa é a consciência ampliada, quando o indivíduo se encontra consciente de si e do meio em que está inserido, produzindo, assim, o conhecimento (DAMÁSIO, 2005).

Segundo o neurocientista, através dos mecanismos neurobiológicos supracitados, uma consciência central emerge e se caracteriza pelo estado de vigília com uma emoção de fundo e atenção básica (DAMÁSIO, 2005). Esses mecanismos produzem um sentimento primitivo no organismo, presentes em várias espécies. O sentimento adquire sentido, quando é associado à linguagem e relacionado à memória, produzindo um relato imagético.

Para tal, além da importância dos neurônios como protagonistas da sinalização, o processo de criação da consciência envolve a participação de diversas estruturas neuroanatômicas que conferem suporte para os mecanismos superiores na região tálamo-cortical, tais como áreas relacionadas a aquisição e evocação de memórias; processamento das emoções, além de áreas motoras, como o cerebelo. Segundo o autor, o processo é mediado por marcadores somáticos e eventos de associação entre estímulo de reforço e indução do estado fisiológico afetivo, mecanismos de relação no cérebro que direcionam a atenção para certas situações, tornando o indivíduo consciente de todo o processo que o envolve (DAMÁSIO, 2005).

A conexão com o tálamo, considerado um relé de entrada desses sinais elétricos para endereçamento as respectivas áreas corticais funcionais, é de extrema importância para a primeira interpretação desses estímulos. Comumente, a partir da sensibilização dessas áreas primárias, há necessidade de conexão com as áreas corticais próximas (secundárias) que contribuem para o refinamento dos sinais associada à complexa integração a áreas terciárias, localizadas a distância (KANDEL, 2014).

3.3 Modelo Neurodinâmico de Walter Freeman

De acordo com a teoria do modelo neurodinâmico, Freeman destaca o cérebro como uma estrutura que possui a função de estabelecer o equilíbrio por meio da avaliação de informações que recebe a cada instante. Assim, de acordo com as mensagens captadas através das vias sensoriais, é compreensível inferir que há oscilações entre diferentes estados, na medida em que essas informações transitam (FREEMAN, 2001). Com isso, Freeman explica a

teoria do caos aplicada aos estados neurodinâmicos representados pelo comportamento dos neurônios, de acordo com as estimulações interna e externa.

Em uma análise simples desse modelo matemático, Freeman faz uma analogia do cérebro ao sistema caótico de um aeroporto e os padrões de comportamento humano na espera de um voo (FREEMAN, 2001). Ao comparar a mudança do padrão inicial do passageiro aguardando a chamada para o embarque (sentado no átrio, quieto, a espera) frente ao padrão de após a chamada (momento no qual o passageiro se levanta, olha para o portão de embarque e estabelece uma marcha não linear rumo ao avião para procurar o seu assento), Freeman tenta explicar a base desse processo através da codificação da informação em padrões neurais (ativação e sincronização).

Em outras palavras, na visão de Freeman, o sistema neural interage de forma casual, continuamente, auto organizando-se de acordo com estímulos específicos, os quais possuem padrões de atividade que geram uma modulação global de amplitude (Global Amplitude Modulation – GAM), processo não linear, regido pelo caos e detectável em aparelhos de Eletroencefalograma (EEG) e Tomografia de Emissão de Pósitrons (PET). A estabilidade é momentânea até que um novo padrão seja detectado pelos sistemas sensoriais, desencadeando todo o processo novamente. A consciência então, seria explicada quando o GAM produz estados de ciência (*awareness*) ao circular pelo cérebro, e esses, por sua vez, são vivenciados fenomenologicamente.

Apesar de propor que os estados neurodinâmicos produzem a consciência, Freeman pontuou o modelo é limitado a identificação dos correlatos neurais entre as atividades neurais e a experiência, evidenciando a dificuldade em reconhecer esses estados. Para aprofundar na investigação, Freeman propõe o entendimento do estado de causação, a fim de evitar correlações infundadas e não científicas a respeito do tema. Como discussão nova na área das neurociências, o neurocientista defende que a ideia de causação atribui causas a agentes capazes de modificar o mundo, destacando que há necessidade de maior atenção para correlacionar os estados neurodinâmicos transientes e a consciência.

3.4 Modelo do núcleo dinâmico (integração de informação)

O modelo de núcleo dinâmico foi proposto por Gerald Edelman (1989) e, como o próprio nome já diz, também envolve sistemas dinâmicos. Trata-se de uma teoria que tentou explicar a organização dos neurônios associada ao princípio da seleção natural dos neurônios, processo conhecido como ‘darwinismo neural’ (EDELMAN, 1989). Seguindo o dinamismo proposto no

modelo anterior, Edelman argumenta que ao receber estímulos, grupos de neurônios produzem padrões de atividades elétricas específicas. De acordo com a peculiaridade inerente a cada um desses processos, os neurônios são selecionados e organizados em grupos funcionais que conectam diferentes áreas do cérebro, formando a unidade do mapa neural. Giulio Tononi, através da neurociência computacional, auxilia Edelman a desvendar as bases dessa integração de informações (EDELMAN; TONONI, 2000).

De acordo com o modelo, neurônios que são ativados simultaneamente permanecem juntos, e o resultado desse agrupamento gera um núcleo dinâmico especializado, cuja composição é frequentemente alterável (*ever-changing*), capaz de conectar áreas encefálicas mais distantes. A consciência é então explicada pelo processamento e capacidade de integração dessas informações em redes neurais, amplamente geradas dentro do sistema tálamo-cortical, que associadas as habilidades linguísticas e semânticas, formam um núcleo dinâmico de processamento (*cluster*) (EDELMAN; TONONI, 2000). Nesse caso, há diversos níveis de organização – a depender da natureza do estímulo e dos grupos de neurônios ativados simultaneamente – sem atribuir funções a um comando central. A abstração do modelo e a necessidade de aprofundar na matemática para explicar os fatos deixa os fenômenos neurobiológicos distante dos olhos dos pesquisadores, pontuando entraves para novas discussões.

3.5 Modelo *Global Neuronal Workspace*

Inicialmente proposto por Bernard Baars (1988, 1997, 2002), trata-se de um modelo teórico, sem bases neurobiológicas, que atribui ao neurônio ou a rede neural a transferência das informações para uma plataforma consciente - a *Global Neuronal Workspace* -, a qual é disponibilizada para todo o sistema e, com isso, o sujeito tem acesso a informação. Como era pautado apenas em ciência cognitiva clássica, o modelo sofreu alguns ajustes propostos por Dehaene et al. (2001, 2003, 2006, 2007), combinando dados práticos das neurociências com teorias da cognição.

Esses autores utilizam dados experimentais, resultado de análises de tarefas executadas pelo indivíduo no estado consciente, tais como testes de atenção e memória e experimentos com visão cega para associar aos conhecimentos atuais de neuroanatomia e neurodinâmica e, assim, fundamentar o conceito de *Global Neuronal Workspace*. De acordo com os pesquisadores, o modelo é baseado em um conjunto de mecanismos neurobiológicos interconectados

por neurônios piramidais (com axônios muito longos) pontuando que essa interconectividade (córteço-cortical e tálamo cortical) é o elemento central do modelo, a qual permite integrar diversos sistemas neurais.

4. REFLEXÕES SOBRE AS TEORIAS PROPOSTAS

Avaliando com cautela os modelos até então propostos, juntamente com as discussões filosóficas sobre a natureza biológica da consciência, é possível observar que a ciência desse estudo da consciência não está medindo esforços para alcançar progressos consideráveis no quesito correlatos neurais (DEHAENE, 2014; KOCH et al., 2016). No entanto, apesar desses estudos investirem na avaliação de quais circuitos neurais específicos estão envolvidos na consciência, ainda é incipiente a correlação de como e por que exatamente esses mecanismos neurais produzem, de fato, a experiência subjetiva.

Nesse sentido, as teorias que abordam diretamente essa experiência subjetiva geralmente focam em um processo neural, cognitivo ou informativo específico, hipoteticamente sugerido como sendo acompanhado pela consciência. Existem muitas outras teorias que traçam essas conjecturas para além das já mencionadas aqui. Associando as teorias já discutidas no presente capítulo com as propostas representativas mais pontuadas na literatura, pode-se fazer um apanhado cronológico de abordagens que englobam a transmissão de mensagens para um espaço de trabalho global amplamente acessível dentro do cérebro (BAARS, 1988; BAARS, 1993); a forma como o cérebro compila a consciência a partir de micronarrativas concorrentes (DENNETT, 1991), a autopercepção de seu próprio estado emocional (DAMÁSIO, 1999); o núcleo dinâmico de clusters neurais funcionais (EDELMAN e TONONI, 2000); o processamento neuronal recorrente (LAMME e ROELFSEMA, 2000); transmissões neuronais para um espaço de trabalho neuronal global (DEHAENE et al., 2003); representações na forma de trajetórias no espaço de atividade (FEKETE e EDELMAN, 2011); representações sobre representações (LAU e ROSENTHAL, 2011); atendimento a representações (PRINZ, 2012); a atenção socialmente observada (GRAZIANO e KASTNER, 2011); um processo fisiológico regular, mas não especificado (SEARLE, 2013); a capacidade de integrar informações (OIZUMI et al., 2014); e, por fim, a construção dos mapas sensoriais internos unificados (FEINBERG e MALLATT, 2016).

Através de uma análise atenta dos modelos propostos e das teorias discutidas, é possível observar que nem todas estão diretamente envolvidas

com uma fundamentação neurobiológica, que demandam tempo e embasamentos teórico e prático, que ultrapassam os objetivos do capítulo. De forma pontual, a teoria de Dehaene aponta que o cérebro primeiro se encontra envolvido em um estágio de processamento inconsciente para, na sequência, mas não necessariamente, passar para um segundo estágio global (DEHAENE et al., 2003). Essa segunda etapa, considerada um espaço de trabalho globalmente acessível para os resultados da primeira etapa, sustenta empiricamente a hipótese da consciência, além de conferir o suporte para os processamentos cerebrais subsequentes. Outra proposta que também assume duas etapas consecutivas é a de Lamme e Roelfsema, a qual sugere uma varredura de processamento de avanço rápido através do córtex após a apresentação de um estímulo visual, estágio inconsciente.

Já na segunda fase, considerada consciente, partes anteriores do córtex visual são ativadas por mecanismo de *feedback* de partes posteriores (LAMME e ROELFSEMA, 2000). Finalmente, caminhando para um cenário macroscópico, Edelman e Tononi propõem a formação de um núcleo dinâmico de agrupamentos neurais funcionais, que variam no decorrer do tempo, como resultado da conexão entre o tálamo e o córtex frontal (EDELMAN e TONONI, 2000). Segundo os pesquisadores, a consciência do indivíduo é adquirida por meio da diferenciação e da integração de informações engenhadas por esse núcleo. Através dessas análises, Christof Koch chama a atenção da necessidade de um programa de pesquisa que propõe descobrir os correlatos neuronais que envolvem os processos da consciência (KOCH et al., 2004).

A fim de apontar as estruturas neurais e os eventos biológicos discutidos em torno do estudo da consciência, faz-se necessário a análise do fenômeno da consciência do ponto de vista morfofuncional, de maneira panorâmica, a qual é estabelecida uma integração neuroanatômica e neurofuncional envolvendo diferentes segmentos do SN. Esse recrutamento, juntamente com diferentes conexões e uma vasta hierarquia funcional, avança na medida direta em que as funções encefálicas envolvidas são de características mais complexas, como no caso da consciência. Nesse sentido, é de extrema relevância abordar diferentes segmentos do SN associados a sinalização proposta para a elaboração e a percepção da consciência, como descrito a seguir.

5. ABORDAGEM DA CONSCIÊNCIA SEGUNDO OS ASPECTOS MORFOFUNCIONAIS DAS ÁREAS DO SN CENTRAL E PERIFÉRICO

Para atender tal proposta e contemplar uma das características marcantes do SN, que é a habilidade de conectar os estímulos externos e as variações do SNP com o SNC, é justificável iniciar as discussões discriminando os elementos do SNP, responsáveis por relacionar os seres vivos ao ambiente em que habitam e, dessa forma, garantir os procedimentos básicos, como as atividades de vida diárias, e as interações com o meio ambiente e a sociedade, com promoção da qualidade de vida.

Em uma análise preliminar a respeito desses aspectos neurofuncionais envolvidos, seria razoável partir da premissa de que a consciência representa a subjetividade de nossas vivências, associada aos variados estímulos externos, internos, bem como a integração desses eventos, compondo um mapa de interseções encefálicas possíveis e, como resultado, a sensação de consciência (FRITH, 2019). Justificando tal proposta, Searle resume duas relações cruciais entre a consciência e o cérebro: os processos neurônicos de nível inferior no cérebro causam a consciência (SEARLE, 2010), e essa, por sua vez, é uma característica de nível superior composta por tais elementos.

Do ponto de vista morfofuncional, a organização do SNC e do SNP para a elaboração do fenômeno da consciência depende da integração neuroanatômica e neurofuncional entre os seus diversos segmentos. O SNC é composto pela medula espinhal (ME), tronco encefálico (TE), cerebelo e cérebro (diencéfalo e telencéfalo), enquanto o SNP engloba nervos espinhais (NE) e cranianos (NC), terminações nervosas, glândulas, músculo liso e o músculo cardíaco (MACHADO, 2011).

Da parte periférica, destacam-se as terminações nervosas e receptores distribuídos e divididos de acordo com suas localizações e especificidades para a sensibilização a partir de estímulos variados, tais como elétrico, químico, mecânico, luminoso, sonoro (LENT, 2010). Após recebidos, essas terminações nervosas conduzem os impulsos no sentido aferente, permitindo a comunicação do periférico até o central fazendo sinapses com a ME ou com o TE (exceto o primeiro nervo craniano), participando não só como segmento por onde transitam esses estímulos, mas, também, onde ocorrem sinapses a núcleos relacionados, principalmente à substância reticular ascendente, ao cerebelo e outros neurônios. Os estímulos eferentes, por sua vez, transportam as mensagens do central para a periferia, possibilitando, assim, a comunicação entre ambos e a relação do organismo com o meio ambiente, através de respostas motoras, sensoriais, cognitivas e comportamentais (KANDEL, 2014).

A célula protagonista dessa sinalização entre SNC e SNP é o neurônio, cujo corpo celular (5 a 150 μm) inclui o núcleo e outras organelas, e do qual partem os prolongamentos, dendritos e axônios, de diferentes tamanhos, fazendo a conexão de diferentes áreas encefálicas com a ME, o TE e o SNP. A forma do corpo celular possui variação dependente da localização e da atividade funcional do neurônio, apresentando-se como piramidal, estrelada, fusiforme, piriforme ou esférica (LENT R., 2001).

Em detrimento da sua importância, o biólogo Francis Crick (1994) enfatizou que o estudo da consciência é um problema científico, com aspectos do comportamento do cérebro estreitamente relacionados às atividades dos neurônios. Dessa forma, para entender o ser humano e desvendar o mistério da consciência, é necessário entender como essas células nervosas se comportam e, ainda, como interagem entre si. Segundo o pesquisador, ao conhecer essa sinalização, a linguagem do cérebro será compreendida.

5.1 O Sistema Nervoso Periférico

A peculiaridade do SNP em interagir com o ambiente depende fundamentalmente da capacidade de perceber variações dos meios externo e interno, apresentá-las às porções mais superiores do SN que, por sua vez, através de uma rica rede de interseções e conexões funcionais, determinam a interpretação dos estímulos, condição de ação, reação e adaptação dinâmica ao meio. A parte periférica é caracterizada por receptores específicos e os gânglios, responsáveis pela realização das sinapses; além dos nervos, por onde transitam os impulsos elétricos até o encéfalo, no caso dos NC; e à ME, no caso dos NE (KOEPPEN, 2017).

Os receptores são considerados estruturas especializadas do SN, distribuídas por todo o corpo e atuam na captação dos variados estímulos (tato, temperatura, movimentos nas juntas, musculares, luminosos, sonoros, gustativos, além de estímulos dolorosos, variações da pressão arterial, concentração de gases sanguíneos, osmolaridade, entre outros). Como trata-se de diferentes naturezas e localidades, cada qual recebe um nome específico, de acordo com esses fatores que o caracterizam. Após receber os estímulos, esses receptores conduzem essa informação a sua respectiva área do córtex cerebral, onde a circuitaria neural específica faz o processamento de acordo com a habilidade funcional da área sobre o reconhecimento e interpretação daquela informação.

A sensibilização desses receptores provoca modificações estruturais e funcionais específicas com a transformação dos diferentes estímulos em

potencial elétrico (transdução) que caminha através dos nervos até o SNC. Essa capacidade de modulação frente a diferentes estímulos permite aos seres vivos desenvolver integração e adaptação finas para cada ambiente, compondo, do ponto de vista neurobiológico, um fator necessário à construção da percepção compreendida como consciência (KANDEL, 2014).

Quanto aos nervos, geralmente cilíndricos, com comprimentos e espessuras distintos, correspondem à porção periférica do SN constituída de tecido nervoso. Variações na estrutura morfológica dos nervos determinam, dentre outras características, a velocidade de condução dos impulsos elétricos. Quanto à topografia, os nervos dividem-se em dois grandes grupos: *i*) os NC, em número de doze pares, conectados anatomicamente as áreas encefálicas do TE, diencéfalo e telencéfalo; e *ii*) os NE, os quais são representados por trinta e um pares que partem das diversas porções da ME, contida na coluna vertebral, em direção a pontos específicos correspondentes aos eixos axial (cervical, torácica, lombar e sacral) e apendicular (cinturas escapular e pélvica, além dos membros superior e inferior) (MACHADO, 2014).

Em relação ao sentido dos estímulos, os NC apresentam a direção da periferia para o SNC, recebendo a denominação de vias aferentes ou sensitivas e, ao contrário, vias eferentes ou motores, quando a comunicação é do SNC para esses. Os doze pares de NC apresentam funções exclusivamente sensitivas ou motoras e, ainda para alguns, mistas, mantendo as duas características. Em contrapartida, todos NE são mistos. Além desses, também são de extrema relevância a participação das glândulas, músculo liso e tecido muscular cardíaco, os quais constituem o Sistema Nervoso Autônomo (SNA).

Dessa maneira, avaliando a composição periférica e suas respectivas naturezas, é perceptível a participação dos nervos na conexão entre as sensações captadas por esses receptores e encaminhadas ao SNC, onde os estímulos serão efetivamente traduzidos, compondo o registro que resulta na compreensão consciente do ser vivo às variações dos meios externo e interno.

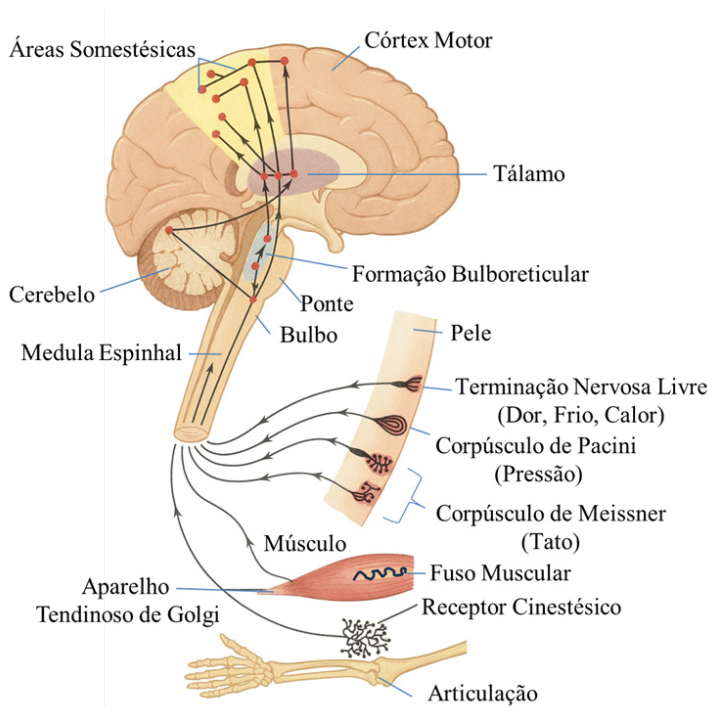
5.2 Medula Espinhal

A ME é um órgão cilíndrico e longo do SNC, envolvida por três membranas (meninges) e pelo líquido cefalorraquidiano (liquor) que circula em um espaço denominado sub-dural, além de gordura e vasos sanguíneos. Situada abaixo do forame magno do osso occipital, onde tem continuidade com o TE, ocupa o forame vertebral no interior da coluna vertebral.

Inferiormente ao forame magno, encontram-se as primeiras raízes medulares cervicais, com trajeto horizontal (MENESES, 2015).

A ME, além de participar de funções específicas na hierarquia funcional do SN como reflexos, localização dos motoneurônios que ativam a contração muscular voluntária, também representa uma via de entrada e saída de impulsos nervosos no sentido central (ascendente) e periférico (descendente), onde esses estímulos são conduzidos por feixes de axônios que compõem os tratos e fascículos (**Figura 1**). Desse modo, os sinais elétricos transitam em ambos os sentidos, favorecendo a integração entre os segmentos mais periféricos ao TE e deste ao córtex cerebral que, a partir da integração a outras regiões cerebrais, são responsáveis pelas funções mais complexas como o raciocínio, aprendizado, tomada de decisão, sentimentos e o fenômeno da própria consciência e do livre arbítrio, dentre outros (MACHADO, 2014; KANDEL, 2014).

Figura 1: Esquema representativo do Eixo Somatossensorial, integrando receptores periféricos, ME, TE, Tálamo e ativação de diversas áreas corticais.

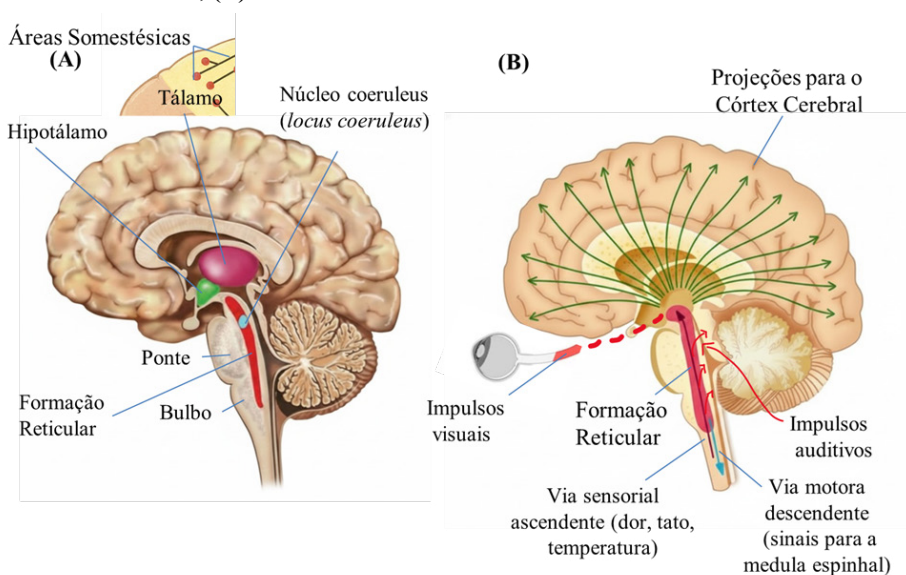


Fonte: Ilustração elaborada pelo Prof. Fagner Eduardo de Oliveira Levindo (FADIP) com auxílio da Inteligência Artificial e editada pela Profa. Lindisley Ferreira Gomides (FADIP). Adaptado de: Guyton and Hall Textbook of Medical Physiology. 14th ed, Elsevier, 2021.

5.3 Tronco Encefálico

O TE é a estrutura do SNC que está localizada superiormente à ME e inferiormente ao diencefalo. Como consequência dessa posição anatômica, possui reverberações funcionais diretas. É anatomicamente dividido em três partes, sendo o bulbo a mais inferior, conectada à ME; a ponte, a região intermediária; e, em nível mais rostral, o mesencéfalo; o qual faz limite com o diencefalo. Funcionalmente, por conter muitos núcleos que estão relacionados a diversas funções de manutenção da vida, o TE é amplamente conhecido como centro respiratório, além de possuir controle cardiovascular e de sinalizações dos NC (do terceiro ao décimo segundo), envolvidos com funções viscerais do SNA, além de passagem de estímulos aferentes e eferentes que fazem sinapses com esses núcleos (**Figura 2A**). Ainda no TE, destacam-se algumas estruturas de localização mais central, nomeados de Formação Reticular (FR) ou Substância Reticular Ativadora Ascendente (SARA), a qual recebe aferências provenientes do globo ocular (visão), da orelha interna (audição) e da ME (**Figura 2B**).

Figura 02: (A) Corte sagital pontuando a localização macroscópica da FR, representada em vermelho; (B) Conexões morfofuncionais da FR relacionadas à consciência.



Fonte: Ilustração elaborada pelo Prof. Fagner Eduardo de Oliveira Levindo (FADIP) com auxílio da Inteligência Artificial e editada pela Profa. Lindisley Ferreira Gomides (FADIP). Adaptado de: Cognitive Neuroscience. The Biology of the Mind GAZZANIGA. Chapter 14 Consciousness, Free Will, and the Law, pag 607.

Na sequência, ocorre a ativação do tálamo, porção do diencefalo que desempenha função decisiva na ativação cortical funcional de áreas primárias nos lobos cerebrais (MACHADO, 2014), inclusive as que participam da consciência. Foi sugerido um mecanismo de *feedback* positivo entre uma mente desperta e a SARA com fundamentos da neuroquímica. Os núcleos monoaminérgicos do TE, incluindo os noradrenérgicos (*locus coeruleus*), núcleos dopaminérgicos e serotoninérgicos (*rafe*), são projetados diretamente para os núcleos basais e para o córtex, a fim de modular a vigília e o humor.

Os núcleos noradrenérgicos e serotoninérgicos também conferem atenção e respostas comportamentais aos estímulos, enquanto a dopamina está, entre outros, associada ao controle motor e a recompensa, além de participar da memória de trabalho por meio de modulações excitatória e inibitória no córtex pré-frontal. Além disso, embora os neurônios serotoninérgicos sejam silenciados durante o sono REM (do inglês *Rapid Eye Movement* – fase do sono na qual é observado movimento rápido dos olhos), essas células aumentam sua atividade durante o sono profundo. Além desses, a atividade de pico em neurônios colinérgicos aumenta durante a vigília e no sono REM, com decaimento durante o sono NREM (do inglês *Not Rapid Eye Movement* – fase do sono na qual não é observado movimento rápido dos olhos) ou profundo. Dessa forma, é sugerido que a acetilcolina aumenta a excitação no cérebro de modo a permitir a formação da consciência, ainda que passageira.

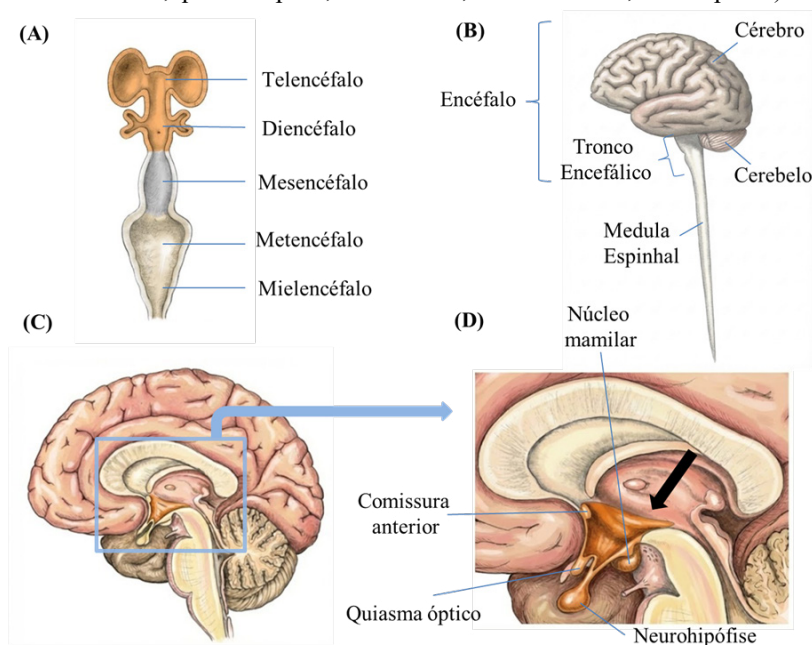
Portanto, qualquer condição que interrompa a passagem desses estímulos, impedindo a ativação dos núcleos talâmicos, causará perda da continuidade funcional e da ativação cortical, o que pode comprometer a interpretação consciente dos sinais e, conseqüentemente, a da consciência, conferindo um estado de inconsciência contínua de “olhos fechados” na ausência de um ciclo de sono-vigília (EDLOW et al., 2012). Algumas condições clínicas neurológicas dramáticas como acidentes vasculares encefálicos, traumas e tumores, nas quais há perda da conexão entre o SARA e os núcleos talâmicos, é possível verificar posturas patológicas, tais como rigidez de decorticação e descerebração, na dependência da lesão em relação ao núcleo rubro, localizado no mesencéfalo (KANDEL, 2014).

5.4 Cérebro - Diencefalo e Telencéfalo

Durante o período embrionário e após o nascimento, não há um desenvolvimento proporcional entre as estruturas que compõem o cérebro. Isso justifica o fato do telencéfalo haver envolvido o diencefalo, o que explica a dificuldade

de observação desse segmento a partir de uma análise macroscópica inicial do encéfalo, conforme observado nas **Figuras 3A e 3B**. O diencefalo está localizado superiormente ao mesencéfalo e inferiormente ao telencéfalo, subdividido em quatro partes: epitálamo, subtálamo, tálamo e hipotálamo. Considerando a abordagem sobre os aspectos neurobiológicos relacionados à consciência, abordaremos com detalhes o Hipotálamo e o Tálamo (**Figura 3C e 3D**), a seguir.

Figura 03: (A) Divisão do encéfalo primitivo em cinco vesículas, as quais as superiores – Telencéfalo e Diencefalo compõem o cérebro. O mesencéfalo não se diferencia, continua sendo a estrutura superior do TE. As duas vesículas inferiores, metencéfalo e mielencéfalo, dão origem ao restante do TE (ponte e bulbo) e ao cerebelo. (B) Partes que compõem o SNC, destacando a ME, parte inferior; e as demais estruturas encefálicas: TE, Cerebelo e Cérebro – composto por diencefalo e telencéfalo, todos originados das respectivas vesículas demonstradas em A. (C) Corte sagital apontando a face medial do diencefalo, destacando em maior aumento (D) as regiões do Tálamo (apontado pela seta azul) e do Hipotálamo (destacado em marrom com respectivas estruturas que o compõem - comissura anterior, quiasma óptico, tuber cinéreo, núcleo mamilar, neurohipófise).



Fonte: Ilustração elaborada pelo Prof. Fagner Eduardo de Oliveira Levindo (FADIP) com auxílio da Inteligência Artificial e editada pela Profa. Lindisley Ferreira Gomides (FADIP). Adaptado de MENESES, MURILO. Neuroanatomia aplicada. 3 ed. Ed. Rio de Janeiro. Guanabara-Koogan, 2015.

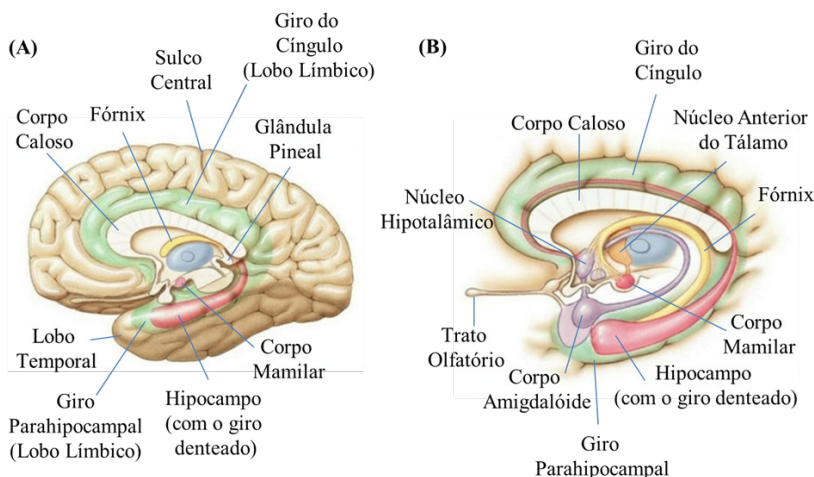
Topograficamente, o hipotálamo representa uma porção do diencefalo localizada mais rostralmente. Constitui um conjunto de núcleos intimamente relacionados à constância do meio interno (homeostasia), considerando que esses estão intrinsecamente ligados a funções integrativas de manutenção da vida, como controle de sinalizações como fome, saciedade, sede, regulação da água corporal via osmoceptores, temperatura corporal, dentre outras. Como integrante do SNA, a partir da divisão anatômica em hipotálamo anterior e posterior, faz conexão funcional com o SNP através dos sistemas simpático e parassimpático, promovendo a integração e regulação das funções secretivas de glândulas exócrinas, musculatura lisa de vísceras e órgãos (TEN DONKELAAR, 2020).

Segundo análises neuroquímicas, os neurônios histaminérgicos do hipotálamo posterior estabelecem as sinalizações de vigília. Esses neurônios fazem comunicação com neurônios que contêm orexina, um neuropeptídeo que mantém a vigília e o processamento emocional. A falta dessas pode resultar em transtornos neuropsiquiátricos e narcolepsia (ELBOW et al., 2012; JALEWA et al., 2014). Associadas a sinalizações noradrenérgicas, dopaminérgicas e colinérgicas já mencionadas anteriormente, esses múltiplos sistemas de despertar são grosseiramente redundantes, uma vez que nenhum sistema é absolutamente necessário para a ocorrência do despertar. No entanto, são diferenciados, uma vez que cada um desempenha um papel especial na vigília e no sono. Durante o sono de ondas lentas, são submetidos a uma influência inibitória originada em parte, pelo menos, de neurônios do ácido gama-aminobutírico (GABAérgicos) particulares, co-distribuídos com muitos neurônios dos sistemas de excitação e também concentrados no prosencéfalo basal e região pré-óptica adjacente (LENT R.; 2010; ELBOW et al., 2012), destacando, mais uma vez, que é necessária uma integração de vias para que haja a consciência, e não como resultado de um evento pontual, simplesmente.

Outra característica funcional que merece destaque é a participação central do hipotálamo através de circuitaria neural específica de segmentos encefálicos corticais e subcorticais na constituição do sistema límbico (SL) (**Figuras 4A e 4B**), responsável pelo controle de nossas emoções (TORRICO, 2020). Descrito inicialmente por Paul Broca (1878), o SL engloba estruturas do telencéfalo, diencefalo e mesencéfalo, como a amígdala, giro do cíngulo, giro parahipocampal, hipocampo, úncus e córtex orbitofrontal. Além da semelhança anatômica, essas estruturas também compartilham propriedades neuroquímicas e fisiológicas em comum (MENESES, 2015; PESSOA, 2015).

Essas condições justificam a homeostasia que favorece os ajustes dinâmicos da fisiologia frente a integridade das funções orgânicas estruturais e bioquímicas para a manutenção da vida nas diferentes circunstâncias ambientais externa e interna.

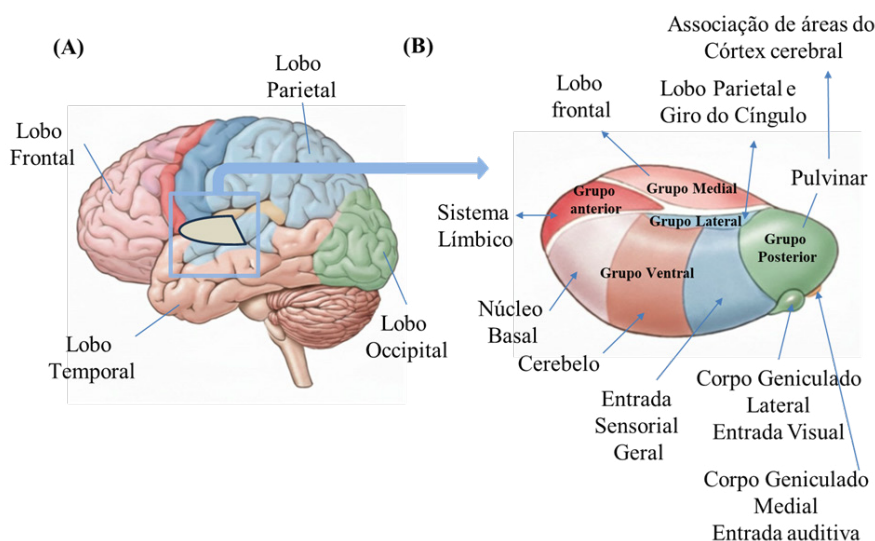
Figura 04: (A) Circuitaria neural específica a segmentos encefálicos corticais e subcorticais na constituição do Sistema Límbico (SL). (B) Participação central do hipotálamo nas sinalizações do SL.



Fonte: Ilustração elaborada pelo Prof. Fagner Eduardo de Oliveira Levindo (FADIP) com auxílio da Inteligência Artificial e editada pela Profa. Lindisley Ferreira Gomides (FADIP). Adaptado de: MARTINI, Frederic H.; TIMMONS, Michael J.; TALLITSCH, Robert B. **Human anatomy**. 7. ed. Upper Saddle River, NJ: Pearson, 2011.

Outra importante estrutura do diencefalo que tem sido muito discutida no cenário das pesquisas sobre a consciência é o Tálamo (**Figura 5A**). Localizado na região central do diencefalo, é representado por um conjunto de núcleos separados por fina camada de tecido cerebral em três grupamentos nucleares que recebem aferências de todo o corpo, exceto o olfato, o qual possui via específica. Nesse particular, essas aferências, entradas (*inputs*) de todas as vias sensitivas do corpo humano, de maneira simplificada e comparativa, atuam como uma via relé, exercendo uma triagem detalhada dos estímulos através do endereçamento para o córtex primário específico para a interpretação daquele respectivo estímulo (**Figura 5B**). Assim, o núcleo reticular do tálamo é formado por uma fina camada de neurônios inibidores que funcionam como um “portão” para os sinais direcionados ao cérebro, principalmente relacionadas à alerta e à atenção (SCHIFF, 2009).

Figura 05: (A) Vista lateral do encéfalo, apontando alguns giros do Telencéfalo: em vermelho (giro pré central – área motora primária; em azul, giro pós central – área somestésica), além do lobo parietal (azul claro) e lobo occipital (verde), cores correspondentes a sinalização dos núcleos talâmicos em B. **(B)** Núcleos do Tálamo responsáveis por sinalizações aferentes em direção a respectivas áreas corticais, com exceção do olfato.



Fonte: Ilustração elaborada pelo Prof. Fagner Eduardo de Oliveira Levindo (FADIP) com auxílio da Inteligência Artificial e editada pela Profa. Lindisley Ferreira Gomides (FADIP). Adaptado de: MARTINI, Frederic H.; TIMMONS, Michael J.; TALLITSCH, Robert B. **Human anatomy**. 7. ed. Upper Saddle River, NJ: Pearson, 2011.

Como exemplos têm-se o estímulo visual, que chega ao tálamo no núcleo geniculado lateral e é direcionado ao córtex do lobo occipital, área primária da visão; as informações somestésicas que passam pelo núcleo ventral posterolateral (tato, dor, temperatura e propriocepção, exceto da cabeça), além das informações de sensibilidade tátil e dolorosa da cabeça (núcleo trigeminal). As fibras gustativas do trato solitário passam pelo tálamo pelo núcleo ventral posteromedial e caminham em direção ao córtex somatossensorial no giro pós-central. De forma semelhante, o estímulo auditivo, atravessa o núcleo geniculado medial do tálamo e é enviado ao córtex auditivo primário, no giro temporal transversal anterior (MACHADO, 2014; KANDEL, 2014).

Além disso, o recrutamento de neurônios talâmicos centrais ocorre em resposta ao aumento da demanda cognitiva, estresse, fadiga e outras perturbações que reduzem o desempenho comportamental. Um estudo sob análise dos estados de sono e vigília observou que havia apenas algumas correlações fora do tálamo em pacientes em estado de alerta, sugerindo forte participação das vias talâmicas para o processo atencional e de vigília. Da mesma forma, durante o sono REM, fortes correlações foram encontradas entre o tálamo e os córtices occipital e parietal, hipocampo, áreas motoras, cerebelo e formação reticular mesencefálica, reforçando a participação dessa estrutura na sinalização desses extremos (LENT R., 2010).

5.5 Áreas corticais funcionais

Além das estruturas já mencionadas, o sistema de consciência também compreende áreas corticais funcionais essenciais para a integração de ordem superior, incluindo os córtices de associação frontal lateral e parietal; córtex frontal medial; bem como os giros cingulados anterior e posterior e o parietal medial (pré-cuneus, retrosplenial). Normalmente, os estímulos requerem motivação para entrar na percepção consciente por meio de um processo denominado intenção ou comportamento direcionado a um objetivo. Estudos apontam uma modulação dopaminérgica das entradas glutamatérgicas do córtex e do tálamo para os neurônios espinhosos médios no estriado, a qual contribui para a cognição e a expressão da consciência (PALMITER, 2011).

Algumas pesquisas de imagem cerebral funcional e avaliação de áreas com lesão demonstram que o comportamento planejado é desencadeado principalmente pelas áreas frontal medial e cingulado anterior do córtex cerebral. Dentre essas estão o campo ocular suplementar, áreas motoras suplementar e pré-motora (SUMMER, 2007). Esse sentido de consciência, portanto, lança uma forte conexão entre a percepção e a ação, os eventos que vivenciamos e os que planejamos vivenciar (ZEMAN, 2001). Tais regiões são apoiadas na produção de comportamento direcionado a um objetivo por alças estriato-palido-talâmicas, bem como a área tegmental ventral e o cinza periaquedutal do TE, reforçando a participação do TE e do tálamo na aquisição da consciência.

Foi demonstrado que os neurônios de alguns núcleos talâmicos ativam células piramidais corticais específicas e silenciam outras áreas do cérebro

por meio da ativação de interneurônios GABAérgicos inibitórios de 40 Hz (NEGRAO, VILJOEN, 2009). Esse dado traduz que a consciência é considerada possível apenas quando esse *insight* elétrico específico de 40 Hz é sustentado entre os circuitos cerebrais, com o recrutamento de mais redes cerebrais permitindo que a consciência se amplie, o que reforça mais uma vez o sinergismo de várias áreas para entender o processo da consciência.

Na tentativa de aprofundar nas pesquisas sobre a consciência, um estudo com base na teoria dos gráficos e na conectividade funcional dinâmica comparou padrões específicos da função cerebral da consciência através de dados de fMRI de voluntários acordados em estado de repouso, de voluntários anestesiados com propofol e de pacientes com distúrbios de consciência. Após as análises, os pesquisadores demonstraram que as redes corticais são especialmente afetadas pela perda de consciência durante os estados temporais de alta integração, exibindo diversidade funcional reduzida e capacidade informacional comprometida, enquanto as desconexões funcionais tálamo-corticais surgem durante os estados de maior segregação. Espacialmente, as regiões posteriores do cérebro exibem reduções na diversidade funcional e na integração com o resto do cérebro durante a inconsciência. Esses resultados apontam que a consciência humana depende de interações espaço-temporais entre integração cerebral e diversidade funcional, cuja quebra pode representar um biomarcador generalizável de perda de consciência, com potencial relevância para a prática clínica (BASTOS et al., 2021).

6. IMPLICAÇÕES PARA O DEBATE SOBRE O LIVRE-ARBÍTRIO

O termo *arbitrium* foi descrito pela *Theologia scholarium*, proposta por Pedro Abelardo (1079-1142), da seguinte forma: “O livre-arbítrio certamente é da alma a própria deliberação ou julgamento pelo qual uma pessoa qualquer se propõe ou a fazer algo ou a renunciar a algo” (ABELARDO, 1855). Filosoficamente, o livre-arbítrio é definido como o poder ou a habilidade que um agente deve possuir para justificar certos julgamentos, atitudes ou tratamentos baseados no deserto básico em resposta a decisões ou ações que o agente executou ou deixou de executar, considerando deserto básico a condição que uma pessoa tem de ser merecedor de algo, seja bom ou ruim (SHEPHERD, 2015). Sendo assim, a existência inclui um sentimento de

exercer um certo controle sobre o próprio comportamento, suscitando duas questões distintas, embora relacionadas a origem e a função desse sentimento (BOUDESSEUL, 2015).

Embora com significados próximos, os debates sobre o livre arbítrio, principalmente entre teólogos e filósofos, são seculares e reúnem uma diversidade de discussões inconclusivas. Dentro de um contexto filosófico, o tema foi discutido de acordo com três correntes de pensamentos: o determinismo, o libertarismo e o compatibilismo. A primeira, defendida por Spinoza e Schopenhauer, afirma que o destino do ser humano não é regido pela própria vontade ou liberdade de escolha, mas sim, por uma norma determinista prévia alusiva ao seu próprio destino. A segunda, apoiada por Kant e Santo Agostinho, trata-se de uma linha de pensamento que acredita na ação humana livre e responsável. Santo Agostinho foi o primeiro filósofo a usar o conceito de livre-arbítrio, pois entendia a liberdade como a capacidade de fazer escolhas e ser responsável moralmente por isso. Em sua obra “Livre-Arbítrio”, argumenta que a fonte de todo o mal (pecado) está no arbítrio da escolha do homem (AGOSTINHO, 1995). Por fim, a última, o compatibilismo, proposto Hobbes, Hume e Mill, propõe a liberdade e a responsabilidade das ações humanas em um contexto determinístico, previamente aceito (COSTA, 2010).

Diante dessa breve exposição, observa-se que a compreensão sobre o livre-arbítrio pode ser complexa e influenciada pelo conceito de liberdade, contemplando ainda significados díspares, os quais abrangem, inclusive, os aspectos religiosos, políticos e morais (TAVARES, JUNIOR; 2019), reforçando ainda mais a importância do entendimento do livre-arbítrio e da tomada de decisões para as relações entre os seres humanos.

No cenário neurocientífico, o crescente desenvolvimento tecnológico, com inúmeros estudos no conhecimento das ciências cognitivas e das neurociências, evidenciou a possibilidade de que os processos mentais que controlam e colaboram para com a experiência humana consciente acontecem de maneira inconsciente. Tal evidência reascendeu novas indagações polêmicas sobre a temática, suscitando questões concernentes que vão desde o conceito, perpassando pela existência e a importância da consciência no contexto social. Além dessas demandas, as inquietações dos pesquisadores de diversas áreas problematizam as possíveis implicações da necessidade da consciência para o exercício do livre-arbítrio, principalmente no

contexto de responsabilidade do agente moral, sem ignorar que tal relação é uma tarefa complexa e cheia de incertezas (MELE, 2019).

O cérebro tem vários sistemas para tomada de decisões, das quais muitos usam regras simples e automáticas e, na maioria das vezes, de forma inconsciente (COSENZA, 2016). Essas inúmeras possibilidades são consideradas eficientes para a solução de problemas no dia-a-dia, entretanto, em questões mais profundas, não automáticas, podem ocasionar uma resolução inapropriada, nomeada por alguns pesquisadores como “vieses cognitivos” (KAHNEMAN, 2012).

Em busca de respostas, o psicólogo Benjamin Libet (1983), da Universidade da Califórnia, monitorou um estudo com seres humanos a fim de avaliar o limiar mínimo de atividade no córtex preparatório do movimento motor para que um estímulo consciente, fenômeno denominado “potencial de prontidão”. As ferramentas utilizadas foram o EEG, com eletrodos no couro cabeludo para medir a atividade elétrica do córtex cerebral; um eletromiógrafo, cujos eletrodos eram fixados na pele para registro das respostas musculares; e um osciloscópio, a fim de medir a amplitude e a frequência dos sinais elétricos. A atividade elétrica do córtex cerebral e a atividade muscular eram mensuradas concomitantemente (LIBET et al., 1983).

Os participantes sentavam-se à frente de um monitor, onde ficava o relógio do osciloscópio que girava em sentido horário, e recebiam comandos para a realização de movimentos comuns, como, por exemplo, curvar o dedo, com a finalidade de apertar o botão à sua frente; de acordo com a posição dos ponteiros do relógio do osciloscópio em movimento. Toda vez que os ponteiros do relógio mudavam sua posição, os participantes deveriam apertar o botão. Após repetir o experimento por várias vezes, Libet e seus colaboradores observaram que os sinais de atividade cerebral relacionada a uma ação anteviam trezentos milissegundos da intenção de agir (ação motora de apertar o botão). Sendo assim, os pesquisadores deduziram que o livre ato de agir tem início de forma inconsciente e os autores sugeriram que essas primeiras informações preditivas estão codificadas em regiões específicas do córtex frontal e parietal (CP) (LIBET et al., 1983).

Em 2011, os resultados desse estudo foram reproduzidos pelo grupo de Bode, apontando que as intenções motoras estavam codificadas no córtex frontopolar até 7 segundos antes dos participantes estarem conscientes de suas intenções, reforçando que o córtex frontopolar é parte de uma grande rede de regiões cerebrais que molda decisões conscientes bem antes dessas atingirem a consciência (BODE et al., 2011).

Outro estudo, desenhado por Soon e Heinze, também buscou avaliar a mesma proposta. Nesse ensaio, foi solicitado para os 14 participantes mentalizarem uma imagem escolhida entre duas, por 10 segundos antes de indicá-la. Os pesquisadores detectaram a ocorrência de padrões de atividade cerebral nas áreas do córtex frontopolar e no visual em 11 segundos antes da tomada de decisão consciente dos participantes, sugerindo a origem inconsciente de tal escolha (SOON et al., 2008; KOENIG-ROBERT, PEARSON, 2019).

Em uma interpretação filosófica dos dados expostos, uma pesquisa destacou a área 10 de Brodmann (BA10), primeira região do córtex frontopolar a fornecer a informação preditiva há 7 segundos antes dos sujeitos decidirem. Pesquisadores afirmaram também que as descobertas dos padrões de atividade preditivos em BA10 são interessantes porque tal área não é normalmente conectada ao livre-arbítrio em discussões científicas. Além disso, acredita-se que BA10 é a área que mais desproporcionalmente tenha crescido em tamanho nos humanos, comparativamente com os primatas não humanos (HAYNES, 2011), o que caracteriza o poder do ser humano em tomar decisões de forma consciente.

Na tentativa de tecer conexões desses dados com o livre-arbítrio, Haynes levanta as seguintes proposições: *i)* O cérebro pode começar, de forma inconsciente, a preparar as suas decisões vários segundos antes que atinja a consciência; *ii)* os estudos, tanto os do seu grupo, quanto os demais, contrariam uma intuição específica em relação ao livre-arbítrio, denominada ‘intuição psicológica ingênua’. Resumidamente, na hora que a decisão é tomada, o resultado é livre, não ser predeterminado por atividade cerebral anterior. Para o pesquisador, o fato de a atividade cerebral predizer os resultados de decisões antes do alcance da consciência revela um poder persuasivo adicional de convencer as pessoas de que são mais previsíveis do que acreditam ser (HAYNES, 2011).

A visão ingênua, dessa forma, pressupõe que só é livre o que não é determinado pelo cérebro, justificando o uso desse adjetivo sobre o que seja o livre-arbítrio, já que certamente as concepções mais razoáveis que o rondam não pressupõem que tal conceito, de fato, exista independentemente do cérebro (NAHMIAS, 2010). Em contraposição, Libet, na interpretação de seus dados, distingue a intenção e a vontade consciente, e pontua que apenas a primeira resulta de processos cerebrais anteriores e inconscientes, enquanto a segunda é autônoma, ao mesmo tempo em que pode agir causalmente sobre o cérebro.

Desse modo, Libet prefere ignorar a sugestão inicial de seus experimentos, ou seja, a de que todo evento mental resulta ou é efeito de processos cerebrais específicos. Observa-se, portanto, que essa reflexão filosófica é ambígua e corrobora as grandes discussões sobre a relação mente-cérebro, ou seja, a de que existem alguns fenômenos mentais resultantes da atividade do cérebro e, de forma semelhante, outros independentes do cérebro.

Apesar de alguns estudos revelarem que o cérebro tem uma capacidade de raciocínio limitada para processar as inúmeras informações advindas do meio ambiente e que utiliza os processamentos cognitivos para tomar decisões, existe uma necessidade de entender a sensação de um agente consciente unificado, que acredita estar no controle de suas ações (COSENZA, 2016). Considerando que o homem é um ser sociável e que a vida em sociedade é benéfica, faz-se necessário um certo número de restrições sobre os indivíduos, para que possam coabitar os mesmos espaços sociais, se comeder respeitando as leis, controlando sua agressividade e as normas impostas para a convivência em grupo. É compreensível pensar que a crença do livre-arbítrio foi selecionada para favorecer esse comportamento, uma vez que é através de escolhas, que é gerada a ilusão de domínio do ambiente e de si mesmo. Essa ilusão também é considerada uma necessidade biológica e psicológica para a sobrevivência humana (LEOTTI, 2010).

Confirmando essa proposta, estudos sugerem que a crença no livre arbítrio aviva sentimentos de responsabilidade pessoal, padronizando formas comportamentais desejáveis pela sociedade. Segundo esses estudos, a incredulidade no livre-arbítrio pode induzir um comportamento antissocial com o aumento da desonestidade e da agressividade nas pessoas e, de forma contrária, a crença no livre-arbítrio resulta em um sentimento de responsabilidade pessoal e de prestação de contas (BAUMEISTER; MASICAMPO; DEWALL, 2009; MELE, 2019).

Na mesma linha de pensamento, outros ensaios indicam que acreditar no livre-arbítrio afeta processos sociocognitivos fundamentais que estão envolvidos na percepção e na compreensão do comportamento do outro. A força da crença no livre-arbítrio prediz intolerância a comportamentos antiéticos em outras pessoas, bem como o apoio a punições criminais severas (GENSCHOW; RIGONI; BRASS, 2017). Como exemplo, o livre-arbítrio é indispensável para o direito penal na determinação do grau de responsabilidade, designando a culpa e a punição do indivíduo na percepção

de sua capacidade de decidir (pressão externa, doença mental ou emoção intensa), com destaque para o agir humano como um fator determinante para o alicerce da culpabilidade (TAVARES; JUNIOR, 2019).

Pesquisas encontraram evidências semelhantes estudando o comportamento de indivíduos que residiam em países com baixos níveis de corrupção, com um sistema governamental transparente quando comparados com indivíduos que residiam em países nos quais a corrupção política era evidente. Nos países com baixos níveis de corrupção, a crença dos participantes no livre-arbítrio era fator preditor do grau em que consideram os comportamentos antiéticos injustificáveis, bem como o apoio à punição criminal (MARTIN; RIGONI; VOHS, 2017), sugerindo que o meio ou o sistema pode condicionar o indivíduo á suas escolhas. Os autores pontuam, portanto, que o comportamento moral dos indivíduos, na maioria das vezes, reflete os códigos morais da sociedade ou grupos aos quais pertencem.

Em uma outra análise, um estudo revelou que residentes de países afetados por fraude e corrupção eram mais propensos do que outros a trapacear em uma tarefa de laboratório (GÄCHTER, SCHULZ, 2016). Associado a isso, dados também apontaram que as pessoas muitas vezes projetam sua própria história de comportamento antiético no grau de aceitação desses atos (FLYNN, WILTERMUTH, 2010), reforçando que o meio em que vivem tende a conduzir e/ou a normalizar certas situações do cotidiano. Dessa forma, atitudes sobre comportamentos antiéticos podem se tornar livres de crenças pessoais sobre a livre escolha ou controle sobre os próprios resultados, dois pilares das crenças de livre-arbítrio (PAULHUS, CAREY, 2011).

De acordo com o envolvimento das crenças e o consequente comportamento motivado, é possível observar também a participação das emoções nesse cenário, como causa ou consequência de um processo que é dinâmico e contínuo. As emoções fazem parte da evolução humana e desempenham um papel fundamental na aprendizagem, na socialização e na sobrevivência dos seres humanos, sendo o medo uma das emoções mais antigas para prever riscos e perigos eminentes e, assim, conferir proteção e continuidade da vida. Segundo Damásio, as emoções são imprescindíveis na origem e na expressividade do comportamento humano. Diante do processo de evolução humana, a estipulação de respostas comportamentais adaptativas é norteadas por processos emocionais. As emoções são vinculadas a memória e, em determinadas situações, fazem emergir

sensações oriundas de experiências anteriores, bem como dos seus resultados, sejam benéficos ou não (DAMÁSIO, 2005).

Por muitos anos julgou-se que o lado racional do ser humano era o responsável por suas escolhas. Atualmente, sabe-se que uma das estratégias para tomada de decisão resulta da habilidade de sentir emoções e o sistema inconsciente se encontra interligado ao sistema de raciocínio consciente (DAMÁSIO, 2005). O termo “tomada de decisão” é definido pelas neurociências como a competência que um indivíduo possui de selecionar uma opção entre várias alternativas de escolha, em um dado momento, dentro de um determinado contexto. Diante disso, corroborando a sinalização discutida anteriormente sobre o SL, é compreensível imaginar que as emoções podem influenciar nesse processo, com diferentes desfechos, a depender de qual emoção está predominando no indivíduo, naquele momento.

7. (IN)CONCLUSÃO

É certo que as teorias proeminentes da consciência enfatizam diferentes aspectos da neurobiologia, como a integração e a diversidade do processamento de informações dentro do cérebro. As pesquisas contemporâneas envolvendo a abordagem neurobiológica sobre a consciência, bem como a sua associação ou não ao livre-arbítrio, parecem apontar para o entendimento de que esse intrigante fenômeno poderia, de certa forma, representar a principal função do SN como integrador dinâmico do *sujeito* com o ambiente, de forma contínua, em variadas e sistemáticas situações. Essa habilidade resultaria da interação entre diferentes segmentos desse sistema, não apenas pelas conexões anatômicas e funcionais, mas, para além disso, por uma dimensão expandida do que se poderia hoje ser compreendida como *mente*! (Autores, 2021).

8. CONTRIBUIÇÃO DOS AUTORES

VE, LFG e MB desenharam o presente capítulo, trabalhando na elaboração da primeira versão. VE e LFG procederam o desenvolvimento dos tópicos de Introdução, Reflexões filosóficas sobre a consciência e Abordagem neurobiológica da consciência, seguido da Abordagem da consciência segundo os aspectos morfofuncionais das áreas do SN central e periférico. MB elaborou a escrita das reflexões relacionadas ao livre-arbítrio. VE, LFG e MB procederam a revisão crítica do texto.

9. REFERÊNCIAS

- ABELARDO, P. Theologia Scholarium. In: Opera omnia. Accedunt Hilarii et Berengarii (Abaelardi discipulorum), Parisiis: J.P. Migne, 1855. p. 979-1114. Tomus Unicus (Patrologiae Latinae n. 178).
- AGOSTINHO, SANTO. **O livre-arbítrio**. São Paulo: Paulus, 1995.
- BAARS, B. J. **A Cognitive Theory of Consciousness**. Oxford University Press, Oxford, 1988.
- BAARS B. J. **How does a serial, integrated and very limited stream of consciousness emerge from a nervous system that is mostly unconscious, distributed parallel and of enormous capacity?** In: Broch G, Marsh J eds: Experimental and theoretical studies of consciousness. New York: John Wiley and Sons, 1993.
- BAARS B. J. **In the theater of consciousness: the workspace of the mind**. New York: Oxford University Press, 1997.
- BAARS B. J. **The conscious access hypothesis: origins and recent evidence**. Trends in Cognitive Sciences, v. 6, n. 1, p. 47-52, 2002.
- BALASHOV, Y.; ROSENBERG, A. **Philosophy of Science**. Contemporary Readings. London: Routledge, 2002.
- BASTOS AM, DONOGHUE JA, BRINCAT SL, MAHNKE M, YANAR J, CORREA J, WAITE AS, LUNDQVIST M, ROY J, BROWN EN, MILLER EK. Neural effects of propofol-induced unconsciousness and its reversal using thalamic stimulation. **Elife.**, v. 27; n.10e60824, 2021.
- BAUMEISTER, R. F.; MASICAMPO, E. J.; DEWALL, C. N. Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. **Personality and Social Psychology Bulletin**, v. 35, n. 2, p. 260–268, 2009.
- BEAR, M. F. et al. **Neuroscience: exploring the brain**. Philadelphia, PA: Lippincott Williams & Wilkins, 2007.
- BICKLE, J. **Philosophy and neuroscience: a ruthlessly reductive account**. Dordrecht: Kluwer Academic Publishers, 2003.
- BLOCK, N. Consciousness, accessibility, and the mesh between psychology and neuroscience. **Behavioral and Brain Sciences**, v. 30, p. 481-548, 2007.
- BODE, S.; HANXI HE, A.; SOON, C. S.; TRAMPEL, R.; TURNER, R.; HAYNES, J. D. “Tracking the Unconscious Generation of Free decisions Using Ultra-High Field fMRI”. **Plos One**, v. 6, n. 6, p.1-13, 2011.
- BOUDESSEUL, J. Libre Arbitre: de la philosophie aux sciences cognitives. **Implications philosophiques**, p. 1-9, May 2015.
- BRONFMAN, Z. Z.; GINSBURG, S.; JABLONKA, E. (2016). The Transition to Minimal Consciousness through the Evolution of Associative Learning. **Frontiers in Psychology** 7, 1954.

CARDOSO Bouyer, Gilbert. A “nova” Ciência da Cognição e a Fenomenologia: Conexões e emergências no pensamento de Francisco Varela. **Ciências & Cognição, São Paulo**, v. 07, n. 81-104, p. 81-102, 2006.

CARRUTHERS, P. **The evolution of consciousness**. In *Evolution and the Human Mind: Modularity, Language and Meta-Cognition* (Ed. P. Carruthers and A. Chamberlain). Cambridge University Press: Cambridge, UK, 2000.

CHALMERS, DJ. **The Conscious Mind: In Search of a Fundamental Theory**. Oxford: Oxford University Press, 1996.

COSENZA, R. **Por Quê Não Somos Racionais**. Porto Alegre: Artmed, 2016.

COSTA, C. F. Livre Arbítrio: Como ser um bom compatibilista. **Princípios: Revista de Filosofia (UFRN)**, v. 7, n. 08, p. 19-33, 2 out. 2010.

CRICK, FRANCIS. **The Astonishing Hypothesis**. Londres: Simon & Schuster, 1994.

DAMÁSIO, A. **O erro de descartes emoção, razão e o cérebro humano**. 1º ed. São Paulo: Companhia das Letras, p. 190-200, 1994.

DAMÁSIO A. **The Feeling of What Happens**. William Heinemann Ed., London, 1999.

DAMÁSIO, A. **O Mistério da consciência: do corpo e das emoções ao conhecimento de si**. São Paulo: Companhia das Letras, 2005.

DAMIANI, D.; GONÇALVES, V. P.; KUHL, L.; ALOI, P. H.; NASCIMENTO, A. M. Aspectos neurofuncionais do cerebelo: o fim de um dogma. **Arq Bras Neurocir**; v. 35, p. 39-44, 2016.

DEHAENE, S.; NACCACHE, L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. **Cognition**, v. 79, n. 1-2, p. 1-37, 2001.

DEHAENE, S.; SERGENT, C.; CHANGEUX, J.P. A neuronal network model linking subjective reports and objective physiological data during conscious perception. **Proc Natl Acad Sci USA**, v. 100, p. 8520-8525, 2003.

DEHAENE S.; CHANGEUX, J.P.; NACCACHE, L.; SACKURA, J.; SERGENT, C. Conscious, preconscious, and subliminal processing: a testable taxonomy. **Trends in Cognitive Sciences**, v. 10, n. 5, p. 204-211, 2006.

DEHAENE S.; DEL CUL, A.; BAILLET, S.; DEHAENE, S. Brain dynamics underlying the nonlinear threshold for access to consciousness. **PLoS Biology**, v. 5, n.10, p. e260, 2007.

DEHAENE S. **Consciousness and the brain: deciphering how the brain codes our thoughts**. Viking Press, New York, 2014.

DENNETT, D. C. **Consciousness explained**. Boston: Little, Brown and Company, 1991.

DENNETT, D. C. **Sweet dreams: Philosophical obstacles to a science of consciousness**. Cambridge, MA: The MIT Press, 2005.

DESCARTES, R. **Meditações Metafísicas**. Editora Martins Fontes, 2011.

DE SOUSA, CEB. Modelos Neurais de Consciência: uma análise neurofilosófica. **Trans. Form. Ação**; v. 38, n. 2, p. 95-128, 2015.

Edelman, G. M. **The remembered present: A biological theory of consciousness**. New York, Basic Books. 1989.

EDELMAN, G. M. **Wider than the sky: the phenomenal gift of consciousness**. New Haven, CT, and London: Yale University Press, 2004.

EDELMAN, G. M.; TONONI, G. **A universe of consciousness: How matter becomes imagination**. New York, Basic Books, 2000.

EDLOW BL.; TAKAHASHI E.; WU O.; BENNER T.; DAI G.; BU L.; GRANT PE.; GREER DM.; GREENBERG SM.; KINNEY HC.; FOLKERTH RD. Neuroanatomic connectivity of the human ascending arousal system critical to consciousness and its disorders. **J Neuropathol Exp Neurol**, v. 71, n. 6, p. 531-546, 2012.

ESPERIDIÃO-ANTONIO V.; MAJESKI-COLOMBO M.; TOLEDO-MONTEVERDE D.; MORAES-MARTINS G.; FERNANDES JJ.; BAUCHIGLIONI DE ASSIS M.; MONTENEGRO S.; SIQUEIRA-BATISTA, R. Neurobiology of emotions: an update. **Int Rev Psychiatry**, v. 29, n. 3, p. 293-307, 2017.

FEKETE, T.; EDELMAN, S. Towards a computational theory of experience. **Conscious Cogn**, v. 20, p. 807-827, 2011.

FEINBERG, TE.; MALLATT, J. The nature of primary consciousness. A new synthesis. **Conscious Cogn**, v. 43, p. 113-127, 2016.

FLYNN, FJ.; WILTERMUTH, SS. Who's with me? Ethical consensus, brokerage, and ethical decision making in organizations. **Academy of Management Journal**, v. 53, n. 5, p. 1074-1089, 2010.

FREEMAN, W. J. **How brains make up their minds**. New York: Columbia University Press, 2001.

FREUD, S. **O ego e o id**. In S. Freud, **EdiÁ**. O standard brasileiro das obras psicológicas completas de Sigmund Freud. Rio de Janeiro: Imago. (Trabalho original publicado em 1923), v. 19, p. 11-83, 1990.

FRIGG, R.; HARTMANN, S. **Models in Science**. In: ZALTA, Edward N. (Ed.). The Stanford Encyclopedia of Philosophy. Winter 2012 Edition. Disponível em: <<http://plato.stanford.edu/entries/models-science/>>. Acesso em: 21/08/2020

FRITH, C. D. **The neural basis of consciousness**. Psychological Medicine - Cambridge University Press, 2019.

FRITH, C.; REES, G. **A Brief History of the Scientific Approach to the Study of Consciousness**. In: *The Blackwell Companion to Consciousness*. Malden: Blackwell Publishing, p. 7-22, 2007.

GÄCHTER, S.; SCHULZ, JF. Intrinsic honesty and the prevalence of rule violations across societies. *Nature*.; v. 531, n. 7595, p. 496-499, 2016.

GENNARO, R. **Consciousness and Self-consciousness: a Defense of the Higher Order Thought Theory of Consciousness**. Amsterdam/Philadelphia: John Benjamins Publishing, 1996.

GENSCHOW, O.; RIGONI, D.; BRASS, M. Belief in free will affects causal attributions when judging others' behavior. **Proceedings of the National Academy of Sciences of the United States of America**, v. 114, n. 38, p. 10071–10076, 2017.

GOMES, GILBERTO. A Teoria Freudiana da Consciência. **Psicologia: Teoria e Pesquisa**, v. 19, n. 2, p. 117-125, 2003.

GRAZIANO, MSA.; KASTNER, S. Human consciousness and its relationship to social neuroscience: a novel hypothesis. **Cogn Neurosci**, v. 2, p. 98-113, 2011.

GULICK, ROBERT VAN GULICK. **Consciência**. Trad. Marco Aurélio Alves e Marcelo Fischborn. Revisor: André Abath. *Investigação Φ Filosófica*, v. E2, artigo digital 2, 2012.

HAYNES, J. D. **“Decoding and predicting intentions”**. In: *Ann. N.Y. Acad. Sci.*, 1224, p. 9-21, 2011.

JALEWA, J.; JOSHI, A.; MCGINNITY, TM.; PRASAD, G.; WONG-LIN, KF.; HOŁSCHER, C. Neural circuit interactions between the dorsal raphe nucleus and the lateral hypothalamus: an experimental and computational study. **PLoS One**, v. 9, n. 2e88003, 2014.

JAYNES, J. **The origin of consciousness in the breakdown of the bicameral mind**. Boston, MA: Mariner Book, 2000.

KAHNEMAN, D. **Rápido e Devagar: duas formas de pensar**. *Journal of Chemical Information and Modeling*. Rio de Janeiro, RJ: Objetiva, 2012. v. 53, 588 p. ISBN: 978-85-390-0401-0, ISSN: 1098-6596.

KANDEL, ER.; SCHWARTZ, JH.; JESSELL, TM. **Princípios de Neurociências**. São Paulo: AMGH Editora, 2014.

KEE, KF. **The neurobiology of human consciousness: Linking rhetoric, science, and graduate liberal studies**. Paper to be presented at the meeting of the National Communication Association Conference, San Antonio, TX, 2006.

KIHLSTROM, JF.; BARNHARDT, TM.; TATARYN, DJ. The psychological unconscious: Found, lost, and regained. **American Psychologist**, v. 47, n. 6, p. 788-791, 1992.

KLEM DE MATTOS, MSS.; SILVA, GP.; GAMA, RF. O Problema da Consciência: por Onde Começar o Estudo? **Revista Brasileira de Educação Médica**, v. 43, n. 1, p175-180; 2019.

KOCH.; C. **The quest for consciousness: A neurobiological approach**. Englewood, CO: Roberts and Company Publishers, 2004.

KOCH, C.; MASSIMINI, M.; BOLY, M.; TONONI, G. Neural correlates of consciousness: progress and problems. **Nat Rev Neurosci**, v. 17, p. 307-321, 2006.

KOENIG-ROBERT, R.; PEARSON, J. Decoding the contents and strength of imagery before volitional engagement. **Scientific Reports**, v. 9, n. 1, p. 1-14, 2019.

KOEPPEN, B. M.; STANTON, B. A. **Berne and Levy Physiology**. Seventh Edition, Chapter 04. Elsevier, 2017.

LAMME, VAF.; ROELFSEMA, PR. The distinct modes of vision offered by feedforward and recurrent processing. **Trends Neurosci**, v. 23, p. 571-579, 2000.

LAU, H.; ROSENTHAL, D. Empirical support for higher-order theories of conscious awareness. **Trends Cogn Sci**, v. 15; p. 365-373, 2011.

LAUREYS S. **The neurology of consciousness: what can we learn from the vegetative state?** Coma Science Group Cyclotron Research Centre & Neurology Dept. University of Liège, Belgium, 2005.

LENT R. A Consciência Regulada. Os Níveis de Consciência e os seus Mecanismos de Controle. O ciclo Vigília-Sono e Outros Ritmos Biológicos. In: Lent R. Cem Bilhões de Neurônios: Conceitos Fundamentais de Neurociência. 2a ed. São Paulo: Atheneu, 2010, p. 519-52.

LENT R. As unidades do Sistema Nervoso. Forma e função de neurônios e gliócitos. In: Lent R. Cem Bilhões de Neurônios: Conceitos Fundamentais de Neurociência. 2a ed. São Paulo: Atheneu, 2010, p.73-110.

LEOTTI, LA.; IYENGAR, SS.; OCHSNER, KN. Born to choose: the origins and value of the need for control. **Trends Cogn Sci**, v. 14, n. 10, p. 457-63, 2010.

LIBET, B.; GLEASON, CA.; WRIGHT, EW.; PEARL, DK. Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. **Brain**, v. 106, n. Pt 3, p. 623-42, 1983.

MACHADO, AB. Formação reticular e neurônios monoaminérgicos do tronco encefálico. In: Machado A. Neuroanatomia Funcional. 2a ed. São Paulo: Livraria Atheneu SA, p. 195-203, 2011.

MARTIN, N. D.; RIGONI, D.; VOHS, K. D. Free will beliefs predict attitudes toward unethical behavior and criminal punishment. **Proceedings of the National Academy of Sciences**, v. 114, n. 28, p. 7325-7330, 2017.

MATURANA, H. R.; VARELA, F. J. *The tree of knowledge: the biological roots of human understanding*. Boston: New Science Library, (1987).

MCGINN, C. *The mysterious flame: conscious minds in a material world*. New York: Basic Books, 1999.

MEESE, T. S. **The How and Why of Consciousness?** Front. Psychol., 2018.

MELE, A. R. **Free Will and Moral Responsibility: Manipulation, Luck, and Agents' Histories**. *Midwest Studies In Philosophy*, [s.l.], v. 43, n. 1, p. 75-92, 2019.

MENESES, M. Neuroanatomia aplicada. 3 ed. Ed. Rio de Janeiro. Guanabara-Koogan, 2015.

Merleau-Ponty, M. **Fenomenologia da Percepção**. São Paulo: Martins Fontes, 1994.

MONDIN, BATTISTA. **Curso de Filosofia Volume 2**. Ed 15. Local: Paulus, 2013.

MULKEY, S. B.; PLESSIS A. J. Autonomic nervous system development and its impact on neuropsychiatric outcome. **Pediatric Research**, v. 85, p. 120-126, 2019.

NAHMIAS, EDDY. "Scientific Challenges to Free Will". In: O'CONNOR, Timothy e SANDIS, Constantine (ed). *A Companion to the Philosophy of Action*. UK: Blackwell, p. 345-356, 2010.

NEGRAO, BL.; VILJOEN, M. Neural correlates of consciousness. **Afr J Psychiatry (Johannesbg)**, v. 12, n. 4, p. 265-269, 2009.

OIZUMI, M.; ALBANTAKIS, L.; TONONI, G. From the phenomenology to the mechanisms of consciousness: Integrated Information Theory 3.0. **PLOS Comput Biol**, v. 10, n. e1003588, 2014.

PALMITER, RD. Dopamine signaling as a neural correlate of consciousness. **Neuroscience**, v. 198, p.213-220, 2011.

PAULHUS, DL.; CAREY, JM. The FAD-Plus: Measuring lay beliefs regarding free will and related constructs. **J Pers Assess**, n. 93, p. 96-104, 2011.

PESSOA L, HOF PR. From Paul Broca's great limbic lobe to the limbic system. **J. Comp. Neurol.**, v. 523, n.17, p. 2495-500, 2015.

PURVES, D. **Principles of cognitive neuroscience**. Sunderland, MA: Sinauer Associates, 2008.

PRINZ, JJ. **The conscious brain: how attention engenders experience**. Oxford University Press, Oxford., 2012.

ROSENTHAL, D. **Two Concepto of Consciouness**. In: *Philosophical Studies*, n. 49, p. 329-359, 1986.

ROSENTHAL DM. "Concepts and Definitions of Consciousness" in *Encyclopedia of Consciousness*. Ed. William P. Banks, Amsterdam: Elsevier, p. 157-169, 2009.

SCHIFF, N.; J. GIACINO; J. FINS. Deep brain stimulation, neuroethics, and the minimally conscious state. **Archives of Neurology**, v. 66, p. 697-702, 2009.

- SEARLE, JOHN R. **Consciência e linguagem**. Trad: Plínio Junqueira Smith. São Paulo. Ed: Martins Fontes, 2010.
- SEARLE, JOHN R. **O mistério da consciência**. São Paulo: Paz e Terra, 1998.
- SEARLE J. Theory of mind and Darwin's legacy. **Proc Natl Acad Sci USA**, v. 110, n. (suppl 2), p.10343-10348, 2013.
- SHEPHERD, J. Consciousness, free will, and moral responsibility: Taking the folk seriously. **Philosophical Psychology**, [s.l.], v. 28, n. 7, p. 929-946, 2015.
- SOON, C.S.; BRASS, M.; HEINZE, HJ.; HAYNES, JD. Unconscious determinants of free decisions in the human brain. **Nat Neurosci.**, v. 11, n. 5, p. 543-545, 2008.
- SOUSA, CARLOS EDUARDO B. DE. Modelos Neurais de Consciência: uma Análise Neurofilosófica. **Trans/Form/Ação**, Marília, v. 38, n. 2, p. 95-128, ago. 2015. Disponível em <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-31732015000200095&lng=pt&nrm=iso>. acessos em 19 fev. 2021. <https://doi.org/10.1590/S0101-31732015000200006>.
- SULLIVAN, J. The Multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. **Synthese**, v. 167, p. 511-539, 2009.
- SUMNER, P.; NACHEV, P.; MORRIS, P.; PETERS, AM.; JACKSON, SR.; KENNARD, C.; MASUN, H. Human medial frontal cortex mediates unconscious inhibition of voluntary action. **Neuron.**, v. 54, n. 5, p. 697-711, 2007.
- TAVARES, DR. A Busca de Solução Para o “Problema Difícil” da Consciência: Uma Revisão das Perspectivas de Chalmers, Dennett e Damásio. **Revista Ciências & Cognição** [online]. 2018. 23 (1) [capturado 12 maio de 2020]; p. 138-157. Disponível em <http://www.cienciasecognicao.org/revista/index.php/cec/index>.
- TAVARES, D.; JUNIOR, W. Neurociência e Livre Arbítrio. **Revista Jurídica Luso-Brasileira**, [s.l.], v. 5, p. 345-375, 2019.
- TEN DONKELAAR, H. J.; NĚMCOVÁ, V.; LAMMENS, M.; OVEREEM, S. The Autonomic Nervous System. **Clinical Neuroanatomy.**, p. 669-710, 2020.
- TONONI G. An information integration theory of consciousness. **BMC Neuroscience**, v. 5, n. 42, 2004.
- TORRICO, TJ.; ABDIJADID, S. **Neuroanatomy, Limbic System**. [Updated 2020 Jul 31]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021.
- VARELA, F. J.; THOMPSON, E.; ROSCH, E. **A mente corpórea: ciência cognitiva e experiência humana**. Lisboa: Instituto Piaget, 2001.
- VICTORINO, CG. Uma Introdução à Neurofilosofia: O Problema Mente-Corpo. **Revista da Biologia**. Instituto de Biociências USP [online]. 2009. 3(1) [capturado 20 maio. 2020]; p.15-20. Disponível em <http://www.ib.usp.br/revista/node/28>.
- ZEMAN, A. **Consciousness**. *Brain.*, v. 124, p. 1263-1289, 2001.

03.

DETECTANDO A LUZ DA EXPERIÊNCIA: O CAMINHO PARA UMA MEDIDA OBJETIVA DOS ESTADOS DE CONSCIÊNCIA

Adenauer G. Casali

Renzo Comolatti

1. INTRODUÇÃO

O ano é 1666. Isaac Newton está sentado em seu quarto em Woolsthorpe, Lincolnshire. Ele observa a dispersão da luz através de um prisma de vidro. E anota em seu caderno o que seria publicado alguns anos depois: “tendo escurecido meu quarto e feito um pequeno buraco na minha janela para deixar entrar uma quantidade conveniente de luz do sol, coloquei meu prisma em sua entrada para que a luz pudesse ser assim refratada para a parede oposta” (Silva e Martins, 1996). Newton vê um raio de sol que atravessa um pedaço de vidro e espalha-se em cores. O que é a luz? De que ela é feita? Qual a natureza desta coisa misteriosa, efêmera, desconhecida, que se deixa perceber através da visão e que aquece as manhãs de inverno? Newton *experimenta* a luz. *Experimenta* a sensação de calor na própria pele e a vividez das cores dispersas em seu campo visual. E, por meio de seu simples aparato experimental, *experimenta com* a luz branca que se decompõe em uma gama de cores ao atravessar o prisma de vidro. Que parte deste calor e destas cores é propriamente da luz que atravessa o espaço e o que disso tudo pertence à experiência do cientista que a observa? O que é o fenômeno físico através daquele buraco na janela e o que pertence apenas à experiência perceptiva de Newton? O que é “objetivo” e o que é “subjetivo”?

Inicialmente, o encontro do cientista com a natureza se dá pela via da percepção. Há uma experiência perceptiva da luz, do calor, assim como do peso e do movimento dos corpos, que fornece os meios para delimitar

cognitivamente, na forma de um primeiro conceito intuitivo, o fenômeno natural a ser investigado. Newton sabia, no entanto, que era necessário discernir o objeto do seu estudo do modo como este objeto aparece em sua experiência subjetiva:

“E se por acaso falo de luz e raios como sendo coloridos ou dotados de cores, devo ser entendido como falando não filosófica e apropriadamente, mas toscamente, e de acordo com as concepções que as pessoas comuns que vissem todas essas experiências tenderiam a formar. Pois os raios, para falar propriamente, não são coloridos. Neles nada mais há do que certo poder e disposição para excitar a sensação desta ou daquela cor. Pois, assim como o som em um sino, ou corda musical, ou outro corpo sonoro, não passa de um movimento vibracional [...], nos raios elas nada mais são do que suas disposições para propagar este ou aquele movimento até o sensorio, e no sensorio elas são as percepções desses movimentos sob as formas de cores.” (Newton, 1996)

Se era através da sua qualidade sensorial que ele podia observar seu fenômeno de estudo, os raios de luz que Newton buscava entender não deviam se confundir com as sensações coloridas causadas por eles. Foi necessário, portanto, que Newton se desvencilhasse do caráter particular e contingente de sua percepção para distinguir aquilo que é próprio à natureza de seu objeto de estudo – a luz enquanto fenômeno físico – abstraindo tudo aquilo que era peculiar à sua experiência subjetiva – a sensação subjetiva da cor e do calor. No entanto, Newton sabia que, para tanto, não bastava o reconhecimento deste problema. Entre ele e seu fenômeno Newton antepôs o prisma de vidro: a natureza da luz não seria discernida através de um mero juízo perceptivo correto ou “científico” – o que seria dar à sua experiência pessoal a última palavra – mas sim por meio de um aparato experimental capaz de verificar efeitos do objeto do experimento independentes da percepção subjetiva do experimentador. Delegava-se assim o papel de “observação” ao anteparo capaz de interagir com a luz e refratá-la numa gama de cores.

Mais de um século depois de Newton usar um prisma para decompor a luz branca no espectro das cores visíveis e descrever o fenômeno da refração da luz, William Herschel notou que uma parte “invisível” dessa luz decomposta pelo prisma – inacessível à sua experiência sensorial imediata –, era capaz de influenciar a temperatura de um termômetro, e descobriu aí a radiação infravermelha. Logo depois, Johann Wilhelm Ritter, com

um aparato um pouco mais elaborado, mas ainda utilizando um prisma de vidro, descobriu o ultravioleta ao observar que raios perto da borda violeta do visível escureciam uma solução de cloreto de prata mais rápido que a luz violeta. Partindo do conhecimento sobre a luz obtido por Newton, tanto Herschel como Ritter descobriram novas regiões do espectro ao empregar aparatos experimentais que não só iam além dos fenômenos oferecidos pela experiência perceptiva deles, mas que demonstravam sua realidade independente dela. Décadas depois, James Maxwell unificou, sob um mesmo formalismo matemático, os fenômenos da luz visível, do infravermelho e do ultravioleta, que passaram desde então a serem entendidos como manifestações distintas de uma mesma radiação eletromagnética. As assim chamadas equações de Maxwell não só apontaram para a existência de novos fenômenos (como as ondas de rádio e micro-ondas, posteriormente descobertas por Heinrich Hertz), como também sugeriram um modo de produzi-las e detectá-las usando circuitos elétricos.

Os avanços teórico-experimentais da ciência, na maioria das vezes só atingíveis através de processos iterativos e complexos entre teorização e experimentação (Chang, 2007), são normalmente acompanhados pela produção de aparatos de mensuração e experimentação dos fenômenos físicos cada vez mais impessoais e pela formulação de teorias científicas capazes de descrever fenômenos cada vez mais removidos de qualquer experiência subjetiva do cientista. O circuito teórico-experimental antes substitui do que estende as capacidades perceptivas de observação do cientista (Bogen e Woodward, 1988): são equações impessoais que sugerem novos fenômenos e anteparos físicos objetivos que marcam a detecção de novas ondas. Tal impessoalidade que dá ao conhecimento físico sua objetividade é, nas palavras de Thomas Nagel, uma espécie de vista de nenhum lugar (Nagel, 1986); um ponto de vista externo, de “terceira pessoa”, cuja característica é ser indiferente aos pontos de vista de indivíduos particulares (de “primeira pessoa”). Essa indiferença epistêmica está na base do método científico, que nada mais é do que uma coleção de procedimentos teórico-experimentais desenvolvidos para destilar os fenômenos físicos de possíveis ilusões ou impressões que são peculiares a sujeitos particulares, e garantir assim o acesso público à verificação experimental do fenômeno, sempre dentro dos limites de um ceticismo razoável¹⁶.

16 Afinal, nenhum método ou procedimento especial do cientista poderia servir para refutar a hipótese cética de que o prisma, os raios, os ângulos e tudo mais naquele quarto em

Mas se a experiência, por carregar a marca do subjetivo, porta consigo a marca da pessoalidade e da contingência da qual o método científico precisa se livrar para só então desenvolver qualquer conhecimento do mundo, esta mesma experiência e o universo que ela contém – nossas sensações, os nossos pensamentos, sonhos, emoções – parecem ficar, como um todo, além – ou aquém – do escopo do mundo físico. Mas e quando a experiência subjetiva se torna o próprio objeto de estudo científico? E se Newton perguntasse não pela natureza da luz e sim pela natureza da experiência de ver a luz? Para muitos, o entendimento da natureza última da experiência e das bases da consciência humana define uma fronteira decisiva do conhecimento científico (Koch, 2004). E para tantos, esta fronteira é intransponível (McGinn, 1991; Levine, 2001).

2. OS DESAFIOS DE UMA CIÊNCIA DA EXPERIÊNCIA

Imagine que você observa Newton em seu quarto escuro. Enquanto ele se pergunta pela natureza da luz, você quer entender a natureza da experiência de ver a luz. Como você poderia proceder para entender o que é esta experiência? Enquanto Newton mede ângulos de incidência e refração dos raios de luz no prisma de vidro, o que você pode medir em Newton para compreender como é ser Newton em seu quarto de Woolsthorpe observando a dispersão da luz?

Diferente da luz que se decompõe no prisma como se o fizesse para qualquer um que estivesse na sala a observá-la, a experiência luminosa de Newton só é diretamente acessível a ele próprio. Há aí uma assimetria fundamental que parece separar os fenômenos físicos dos fenômenos da experiência subjetiva: enquanto propriedades físicas parecem ser igualmente acessíveis para qualquer experimentador, isto é, são *públicas*, a acessibilidade epistêmica das experiências subjetivas são assimétricas: *qua* experiências subjetivas particulares são absolutamente inobserváveis, com a exceção daquele que tem a experiência em questão; diz-se, por isso, que as experiências subjetivas possuem um caráter *privado*. Thomas Nagel ilustrou esta questão com o exemplo de um morcego e seu sonar: independentemente de qualquer medição que pudesse ser realizada a partir de uma

Lincolnshire não fossem mais do que mera ilusão, como aquelas produzidas pelos demônios enganadores de Descartes (Descartes, 1996).

perspectiva externa, de terceira pessoa, não há como saber o que o morcego de fato experimenta (Nagel, 1974). Mundo físico e experiência subjetiva parecem coisas tão heterogêneas e incomensuráveis, que a passagem de um ao outro apresenta-se como intransponível. Leibniz expressou vividamente tal “lacuna explanatória” (Levine, 1983) entre mundo físico e mundo da experiência em sua Monadologia:

“Ademais, deve-se confessar que a Percepção e aquilo que dela depende é inexplicável por razões mecânicas, isto é, por figuras e movimentos. Imaginando-se que há uma máquina cuja estrutura a faça pensar, sentir e perceber, poder-se-á, guardadas as mesmas proporções, concebê-la ampliada de sorte que se possa nela entrar como em um moinho. Admitido isso, lá não encontraremos, se a visitarmos por dentro, senão peças impulsionando-se umas às outras, e nada que explique uma percepção.” (Leibniz, 2009)

Do mesmo modo que ao adentrar seu moinho neural imaginário Leibniz não encontra nada que se assemelhe a um *percepto* consciente, ao explorar o interior do cérebro de Newton é difícil imaginar o que poderíamos encontrar em toda aquela complexa massa neuronal de conexões sinápticas e potenciais de ação que torne inteligível a experiência subjetiva da intensidade luminosa, das cores ou do calor.

Tal assimetria epistêmica (Metzinger, 1995) é o que torna tão difícil o assim chamado *problema difícil* da consciência (Chalmers, 1995): o problema de explicar, em termos físicos, a qualidade da experiência consciente; ou, em outras palavras, explicar porque certos sistemas físicos são acompanhados por uma perspectiva intrínseca, qual seja, a experiência subjetiva particular de ser aquele sistema em um determinado estado. O estatuto epistêmico *sui generis* da experiência subjetiva, seu caráter assimétrico e privado, parece ser um obstáculo maior à possibilidade de estudá-la cientificamente através dos métodos tradicionais usados, por exemplo, na física. Se a ciência é, como vimos, uma empreitada marcada pela impessoalidade e pela capacidade de experimentação conceitual e empírica com seus objetos, mas a experiência é, enquanto tal, irreduzivelmente pessoal para o sujeito dela e inobservável para o observador externo, devemos concluir que algo como uma ciência da experiência é uma contradição em termos.

Antes, porém, de afirmarmos o caráter irreconciliável da experiência consciente com a ciência da natureza, antes de concluirmos que a

experiência permanecerá para sempre misteriosa, uma espécie de fato bruto da natureza que está para além do escopo da ciência, cabe explorarmos todas as alternativas. Não haveria espaço no método científico para uma abertura à perspectiva de primeira pessoa? E, de outra parte, não seria possível investigar a hipótese de que experiências subjetivas estão sempre acompanhadas de estados neurais observáveis e impessoais? A possibilidade de uma ciência da consciência depende de se formar uma ponte entre as perspectivas de primeira e terceira pessoa (Figura 1). E a construção desta ponte, a um só tempo teórica como experimental, deve passar por operações que transformam, a partir de restrições mútuas e iterativas (Varela, 1996), tanto o que entendemos por consciência como os conceitos tradicionais de ciência e natureza.

De um lado, podemos considerar modificações no interior da ciência que sejam capazes de alargar suas práticas e conceitos. Este caminho passa por uma abordagem *fenomenológica*: se a experiência luminosa de Newton é, para nós “do lado de fora”, inobservável e imensurável por qualquer aparato experimental, resta-nos delegar a Newton a tarefa de investigar cientificamente sua própria experiência. Tal abordagem toma a experiência em primeira pessoa como um campo autônomo dotado de uma estrutura intrínseca feita de distinções e relações fenomênicas. A experiência torna-se assim uma fonte original para uma análise e descrição científica da subjetividade, o que implica numa reorganização dos métodos e objetos que são considerados legítimos pela ciência. A comunicação – verbal, escrita ou por meio de outros sistemas – aparece como um meio para superar o caráter pessoal e privado da experiência, publicizando-a através de suas capacidades mnemônica, introspectiva e linguística. Esta abordagem, que tem sua origem no trabalho de Edmund Husserl (Husserl, 2019), tem por desafio o estabelecimento e a formalização de protocolos de experimentação e verificação de seus resultados, espécie de fenomenologia experimental que não é puramente “transcendental”, mas encontra-se constantemente submetida tanto à validação intersubjetiva como à tarefa de compatibilizar seus resultados com aqueles obtidos através da experimentação “tradicional” de seus mecanismos neurais.

De outro lado, permanece necessário determinar observáveis físicos relacionados à experiência e que sejam capazes de alicerçar a investigação científica. Aqui, parte-se do princípio largamente aceito de que

qualquer diferença entre duas experiências conscientes deve necessariamente ser acompanhada por uma diferença física entre substratos neurais (isto é, o princípio da superveniência do mental sobre o físico). A tarefa passa, portanto, pela busca dos assim chamados “correlatos neurais da consciência” (Crick e Koch, 1990; Chalmers, 1995) e que depende do estudo da correlação entre diferentes medidas de processos neurais e as possíveis estimativas, baseadas em relatos subjetivos ou avaliações comportamentais clínicas, do estado de consciência do sujeito (Overgaard, 2015). Uma vez ancorada na terra firme dos fenômenos físicos, a ciência da consciência poderá então empregar as ferramentas e métodos científicos tradicionais.

Apesar das dificuldades e incertezas inerentes à fundação desta nova ciência da consciência, os avanços obtidos na última década sugerem que esta ponte entre o objetivo e o subjetivo é, de fato, possível. Dentre tais progressos recentes, destacam-se em particular os desenvolvimentos associados a um problema específico desta área do conhecimento: o problema clínico do diagnóstico dos distúrbios de consciência, em que considerações fenomenológicas combinadas a protocolos empíricos levaram à introdução de um possível marcador objetivo e mensurável do estado global de consciência.

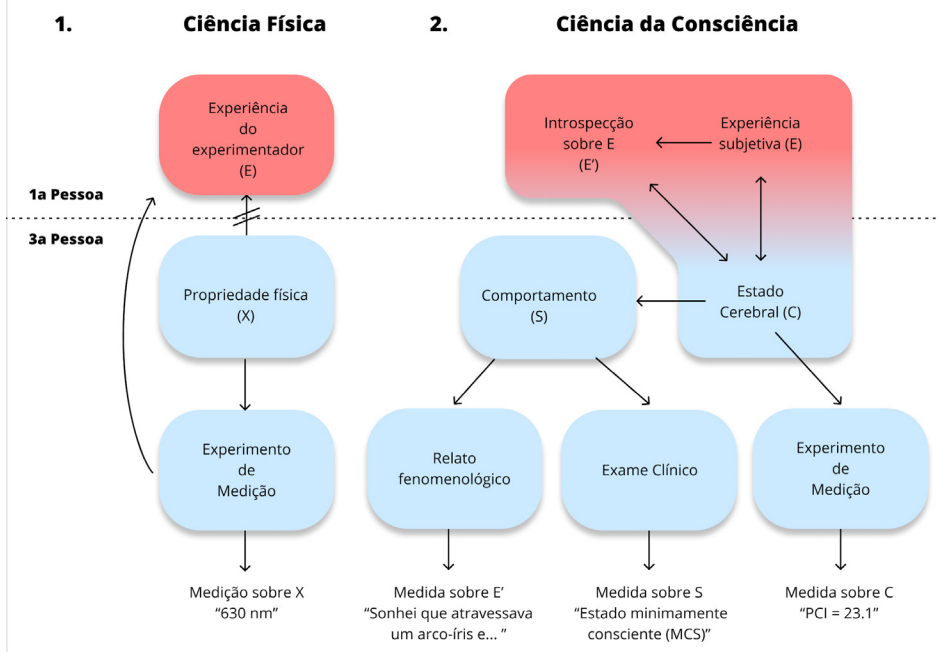


Figura 1: A possibilidade de uma Ciência da Consciência depende de se formar uma ponte entre as perspectivas de primeira e terceira pessoa. Esquematisação dos procedimentos que compõem o problema da medição na física e no recente campo de estudos científicos da consciência. Na física, apesar de inicialmente o contato do experimentador com o fenômeno físico (X) se dar através de sua experiência perceptiva (E), o desenvolvimento do aparato teórico-experimental – que envolve o equipamento de medição, uma teoria que descreve o seu funcionamento e a relação dele com o fenômeno físico estudado, assim como uma teoria estatística sobre como extrair o valor da medida do equipamento – progressivamente substitui a observação perceptiva do experimentador (seta cortada), constituindo assim a objetividade da perspectiva de terceira pessoa. No estudo da consciência, o fenômeno a ser detectado e mensurado é justamente a experiência subjetiva (E), e a objetividade deste procedimento se dá na passagem entre a experiência consciente (perspectiva de primeira pessoa) e o estado cerebral (C) correspondente (perspectiva de terceira pessoa). Inicialmente, medidas sobre o estado de consciência se dão a partir de seus correlatos comportamentais (S), usados no diagnóstico clínico da consciência para distinguir, por exemplo, o estado vegetativo do estado minimamente consciente. A introspecção em primeira-pessoa (E') permite que o sujeito da experiência possa comunicar, através de comportamentos, seu estado de consciência bem como descrever seu conteúdo em um relato fenomenológico. Finalmente, com o progresso do conhecimento científico sobre o paralelismo entre estado e mecanismos cerebrais e experiência consciente, que se dá a partir de restrições mútuas entre as perspectivas de primeira e terceira pessoa, podemos utilizar a mensuração das propriedades cerebrais relevantes ao invés de comportamentos do indivíduo para detectar e medir de maneira objetiva os estados de consciência.

3. O PROBLEMA DO DIAGNÓSTICO DOS DISTÚRBIOS DE CONSCIÊNCIA

Voltemos uma última vez ao quarto de Newton. Agora Newton caminha de um lado a outro na sala e fala consigo mesmo enquanto planeja seu experimento. O que pensa Newton? Não saberíamos dizer. Ele observa a luz e anota em seu caderno. Ao ver Newton com seu prisma, não estamos certo do que exatamente ele vê, qual o conteúdo de sua experiência particular, como o vermelho lhe aparece, nem quais emoções acompanham aquela experiência visual. Mas, mesmo se completamente incertos sobre a qualidade de sua experiência, estamos certos de que ele tem *alguma* experiência – rapidamente descartamos a possibilidade de ele estar inconsciente tendo um episódio repentino de sonambulismo ou ter se transformado em um zumbi sem nenhuma experiência fenomenal. Na prática, enquanto a inferência sobre o conteúdo da experiência particular de alguém a partir de seu comportamento nos parece sempre essencialmente limitada e incerta, o problema de avaliar se alguém está consciente ou não parece, ao menos de partida, consideravelmente mais simples. Normalmente uma pessoa acordada exibe uma série de comportamentos que não deixa dúvidas sobre sua capacidade de ter experiências conscientes: a habilidade de se comunicar e interagir de modo intencional e coerente com o mundo ao seu redor; de perceber e reagir a objetos e de realizar ações voluntárias de maneira consistente e sustentada. Estes sinais comportamentais – às vezes, chamados de correlatos comportamentais da consciência – levam-nos a inferir, a partir do comportamento externo da pessoa, a presença de um ponto de vista subjetivo intrínseco. Isto é, se não somos capazes de determinar exatamente **o que** a pessoa experimenta (o conteúdo da sua experiência consciente), o modo como a pessoa se comporta pode ser suficiente para respondermos sobre sua **capacidade** de experimentar algo (sobre o seu estado global de consciência). No entanto, se estes comportamentos são suficientes para inferirmos a consciência, como veremos a seguir, eles não são necessários, uma vez que é possível estar consciente sem demonstrar nenhum destes sinais externos.

Na escuridão de seu quarto, após ter jogado com o prisma de vidro e os raios de sol durante boa parte da manhã, Newton agora repousa imóvel em sua poltrona, de olhos fechados e em silêncio. Estaria Newton consciente?

Ele não demonstra nenhum sinal claro quanto a isso. Talvez ele esteja completamente alerta e, enquanto repousa, arquiteta sua teoria corpuscular da luz. Ou então talvez Newton esteja dormindo e, em seu sono, sonha com os fantásticos raios coloridos através do prisma de vidro. Mas ainda é possível que seu silêncio e imobilidade sejam o sinal de que ele dorme profundamente em um estado de completa inconsciência, sem qualquer experiência. O método mais simples e confiável de sanar esta dúvida é obter uma resposta de Newton. Afinal, certamente ele está numa posição melhor para testemunhar sobre seu próprio estado de consciência do que nós a observá-lo do lado de “fora”. Se nos aproximamos de sua poltrona, receosos de interrompê-lo em plena reflexão ou no meio de um intenso sonho fantástico e perguntarmos a ele sobre a natureza da sua experiência, bastará alguns instantes de reflexão introspectiva para que Newton possa nos contar detalhes sobre o conteúdo de seus pensamentos ou sonhos. Alternativamente, ele poderia nos revelar, um pouco consternado, que havia caído no sono mais profundo, e que não havia sonhado nada: até onde era capaz de averiguar, antes de ser acordado por nós, ele simplesmente não estava ali e sua existência havia desaparecido num oceano profundo de inconsciência a partir do momento que adormecera. Através deste seu relato pessoal é possível avaliar com razoável segurança a questão sobre seu estado global de consciência, assim como obter informações sobre o conteúdo de sua experiência, que de outro modo permaneceria desconhecido por nós. De fato, até hoje, os relatos fenomenológicos são usados por pesquisadores que estudam a natureza dos sonhos e sua base neural para rotular os estados de consciência em experimentos, especialmente para distinguir o estado de consciência durante o sono inconsciente dos sonhos que ocorrem durante o sono, bem como seu conteúdo particular (Siclari *et al.*, 2017). Aqui, a comunicação possibilita romper o solipsismo do sujeito que se sabe consciente, trazendo esse conhecimento privado para a esfera comum dos fenômenos intersubjetivos. O uso dessa metodologia para obter evidências empíricas em terceira pessoa a partir de um conhecimento fenomenológico em primeira pessoa constitui o que Daniel Dennett chamou de uma heterofenomenologia (Dennett, 2007).

No entanto, a comunicação de um sujeito torna-se fenômeno publicamente acessível apenas através da transdução de uma intenção linguística mental em um meio material externo. Seja através da vocalização, da escrita

manual ou de algum outro sistema de comunicação, a obtenção de relatos fenomenológicos depende tanto das bases cognitivas da linguagem e da memória¹⁷ quanto da integridade do canal de comunicação e da inscrição da mensagem em um meio físico. Podemos imaginar o cenário improvável de Newton ser tomado por um derrame que compromete temporariamente a área de Wernicke em seu cérebro – responsável pela articulação da fala – ou, ao ser acordado por nós, ele ser tomado por uma “paralisia do sono” – em que a pessoa se vê temporariamente paralisada pela permanência anormal da paralisia que naturalmente sobrevém durante o sono REM, fase do sono em que muitos dos sonhos ocorrem. Em ambos os casos, apesar de consciente, Newton estaria incapacitado de comunicar tal informação a nós.

Assim como ocorre com outros comportamentos associados à consciência, casos de comprometimento de funções linguísticas ou da articulação motora em indivíduos conscientes demonstram como a presença de manifestações comportamentais, apesar de suficientes para se inferir a consciência, não são necessárias para a consciência, de modo que sua ausência não implica em inconsciência. É assim que o estado global de consciência de um sujeito pode estar dissociado das suas manifestações comportamentais: você pode estar consciente enquanto corre uma meia maratona, ou enquanto deita relaxadamente e com os olhos fechados em sua cama; você pode ter experiências subjetivas mesmo enquanto dorme ou em meditação profunda; é possível permanecer consciente sob o efeito de drogas que bloqueiam completamente os movimentos, ou até mesmo quando se está na triste condição denominada de síndrome de encarceramento, em que todos os músculos de controle voluntário do sujeito encontram-se totalmente paralisados devido a uma lesão cerebral. Tal dissociação entre o estado de consciência e as manifestações comportamentais deste estado está no cerne da problemática associada ao diagnóstico dos chamados distúrbios da consciência.

17 Cabe notar que o relato é feito num momento subsequente ao estado de consciência a qual ele se refere – Newton já acordado, nos fala sobre o momento anterior em que ainda dormia – introduzindo um hiato temporal entre os dois momentos. Mediado, portanto, pela introspecção, pela linguagem e também pela memória, Newton estabelece para si mesmo e para nós o caso sobre seu estado de consciência momentos antes. A partir da falibilidade da memória, é possível levantar dúvidas sobre a acurácia do seu relato, já que ele poderia, por exemplo, ter sonhado, mas logo esquecido ao ser acordado, o que o levaria a reportar erroneamente seu estado de consciência – um falso negativo, portanto.

Em pacientes que se recuperaram do coma após sofrerem lesões cerebrais, a ausência de sinais comportamentais de parte destes pacientes que possam servir para indicar a presença da consciência leva à dificuldade clínica de se avaliar se o paciente se encontra de fato inconsciente ou se está consciente, mas incapaz de dar sinais disso (Bernat, 2006). Este problema é especialmente relevante para pacientes em estado vegetativo, condição de significativa prevalência dentre os possíveis efeitos das lesões cerebrais (Van Erp *et al.*, 2014; Giacino *et al.*, 2018) e definida pela presença de ciclos sono-vigília e comportamentos reflexivos, mas ausência de qualquer sinalização comportamental de experiência consciente. Recentemente, propôs-se renomear tal condição para “*síndrome da vigília sem responsividade*” (Laureys *et al.*, 2010), salientando-se assim a ausência de sinais comportamentais, ao invés de presumir a ausência de consciência. Além disso, através da aplicação cuidadosa de escalas clínicas comportamentais, tais pacientes sem responsividade passaram a receber um diagnóstico distinto do diagnóstico de pacientes que, mesmo sem ser capazes de se comunicar, são chamados de “*minimamente conscientes*” por conseguirem demonstrar claramente ao menos um dos seguintes comportamentos condizentes com a presença da consciência (Giacino *et al.*, 2004): seguir comandos simples, responder a respostas de sim ou não (verbalmente ou através de gestos), verbalizar de modo inteligível ou demonstrar movimentos ou comportamentos afetivos de modo não-reflexivo e coerente com o ambiente.

Porém, como as escalas clínicas utilizadas para inferir o estado de consciência são baseadas em manifestações comportamentais dos pacientes, e como o estado de consciência é dissociado desta capacidade comportamental, erros diagnósticos nesta população são comuns (Schnakers *et al.*, 2009; Van Erp *et al.*, 2015) e possuem impacto dramático tanto nas chances de recuperação quanto na qualidade de vida dos pacientes (Bruno *et al.*, 2013). De um ponto de vista médico, a solução definitiva para o problema de se diagnosticar o estado de pacientes com distúrbios de consciência depende de se identificar um ou mais marcadores objetivos e não comportamentais que sejam fortemente correlacionados com o estado de consciência dos sujeitos a ponto de servirem para apoiar o diagnóstico destas condições.

Como progredir na busca de tal marcador? Como identificar um correlato do estado global da consciência em pacientes que não são capazes de manifestar o próprio estado cognitivo? Ora, enquanto esta identificação

é certamente problemática em pacientes que se recuperam de lesões cerebrais, a busca por tal correlato poderia ser realizada estudando-se indivíduos saudáveis em diferentes condições, como em vigília, durante o sono ou até mesmo sob anestesia, quando o estado de consciência pode ser determinado com alto grau de confiança através do comportamento do indivíduo e de seus relatos em primeira pessoa, realizados imediatamente ou logo após o despertar. O progresso desta investigação depende então de se validar medidas objetivas da atividade cerebral capazes de discriminar, com alta confiabilidade, o estado global de consciência em tal “grupo controle” formado por indivíduos adultos saudáveis e capazes de reportar, durante ou imediatamente após o experimento, o próprio estado de consciência. Uma vez identificada uma medida capaz de quantificar objetivamente as condições para que um indivíduo tenha experiências subjetivas, tal “índice de consciência” poderia então ser validado também em pacientes com lesões cerebrais, mas dos quais o estado de consciência é ainda assim conhecido. Finalmente, um marcador de consciência que apresente alta acurácia em tal população poderia então ser aplicado também na circunstância mais complexa e que envolve pacientes sem responsividade, dos quais o estado de consciência não pode ser de outro modo determinado.

Nos últimos anos, diversas medidas empíricas foram propostas como índices de estado de consciência. Algumas dessas métricas pretendem monitorar o estado de consciência através da avaliação do conteúdo informativo (Seth *et al.*, 2008; Seth *et al.*, 2011) e das características espectrais de sinais cerebrais (Sigl e Chamoun, 1994). Outras são baseadas em paradigmas experimentais que envolvem potenciais evocados e relacionados a eventos (Plourde e Picton, 1991; Engel e Singer, 2001; Boly *et al.*, 2011; Ragazzoni *et al.*, 2013). Mais recentemente, estudos utilizando técnicas de inteligência artificial empregaram combinações de dezenas destas métricas em algoritmos propostos para identificar o estado de consciência de tais pacientes (Sitt *et al.*, 2014; Engemann *et al.*, 2018). Mas, apesar de tais métodos serem capazes de discriminar o estado de consciência entre grupos de pacientes, eles possuem um desempenho bastante modesto na tarefa que realmente importa do ponto de vista clínico: detectar acuradamente o estado de consciência em pacientes individuais expostos a diferentes condições fisiológicas, farmacológicas e patológicas. Índices espectrais desenvolvidos para o monitoramento de anestesia, além de muito variáveis

entre indivíduos (Kaskinoro *et al.*, 2011), não parecem ser úteis na determinação da consciência em pacientes com lesões cerebrais (Schnakers *et al.*, 2005). Componentes evocadas tardias supostamente associadas à consciência foram também encontradas em condições de sono profundo e anestesia (Koelsch *et al.*, 2006; Fischer *et al.*, 2010; Höller *et al.*, 2011; Azabou *et al.*, 2018). Medidas da coerência do sinal de eletroencefalografia (EEG) também apresentaram resultados contraditórios com relação à utilidade destes métodos na clínica (Supp *et al.*, 2011; Barrett *et al.*, 2012). E mesmo classificadores multidimensionais treinados especificamente para distinguir pacientes em estado vegetativo de pacientes que demonstram sinais de consciência sofrem com a variabilidade destas medidas e produzem taxas de acerto muitas vezes inferiores a 70% (Sitt *et al.*, 2014), portanto ainda insuficientes para a prática clínica.

Em todos estes casos, o problema parece estar associado à dificuldade de se relacionar os resultados obtidos em indivíduos saudáveis e em condições ideais, quando é possível obter relatos sobre o estado de consciência, com estas mesmas medidas calculadas em condições de interesse clínico, onde outros fatores alheios ao estado de consciência parecem afetar de forma significativa os resultados dos índices, produzindo assim uma variabilidade que dificulta ou até mesmo impede completamente a aplicação clínica de tais métodos. Meramente testar empiricamente diversas métricas neurais em um grande número de indivíduos parece não ser suficiente para se determinar a medida certa. Talvez seja necessário, antes, pensarmos no que devemos medir. Assim como Newton escolheu o instrumento certo para estudar a luz, também nós devemos antes considerar o objeto do nosso estudo para só então determinarmos a medida correta.

4. O ÍNDICE DE COMPLEXIDADE PERTURBACIONAL COMO UMA MEDIDA OBJETIVA DA CAPACIDADE DE CONSCIÊNCIA

Dentre as diversas medidas de estado de consciência que foram propostas nos últimos anos, uma em particular recebeu especial destaque por sua acurácia até então sem precedentes. A métrica, denominada Índice de Complexidade Perturbacional (ou PCI na sigla em inglês), tem sido

considerada particularmente promissora¹⁸ para vir a ser empregada no apoio ao diagnóstico clínico dos distúrbios de consciência (Maciver, 2015; Giacino *et al.*, 2018). Para seu cálculo, o índice requer uma combinação improvável e complexa de técnicas experimentais e algoritmos computacionais que dificilmente teria sido proposta a partir de considerações feitas somente a partir do conhecimento físico ou neurocientífico em terceira pessoa: PCI é fruto de uma tentativa de se construir uma ponte entre o subjetivo e o objetivo, de se aproximar as propriedades básicas da experiência consciente, derivadas de uma perspectiva fenomenológica, ao nosso entendimento sobre o modo como o cérebro funciona.

Façamos então um experimento introspectivo simples. Depois de um dia de trabalho, você deita em sua cama para mais uma noite de sono. Da rua vem o ruído insistente de carros, mesmo naquela tarde hora da noite. Você sente o ar morno de verão no seu quarto, e o toque suave do lençol no corpo. Sua mente vaga pelos pensamentos mais diversos até que você se vê numa praia distante e o som dos carros da noite agora é o ressonar do mar que avança sobre a areia quente até que, um tanto de repente, sem que você possa determinar exatamente como e nem mesmo quando, tudo se perde, tudo desaparece: você cai em um sono profundo, no despenhadeiro silencioso da inconsciência. Tente se destacar um pouco do conteúdo particular desta sua experiência e observe o que era comum a todas essas experiências conscientes antes de adormecer. Observe como cada aspecto da sua experiência, mesmo quando não explicitamente objeto de sua atenção, ainda assim é parte constitutiva desta mesma experiência: cada sensação – da textura do lençol, dos sons e ruídos dos carros ou das ondas do mar, do calor da noite ou da sensação de tranquilidade – informam a qualidade desta experiência por serem *específicas*, ou seja, por se distinguirem de todas as outras experiências possíveis que não aquelas – outras texturas, outros sons, outros estados emocionais. Além disso, cada um desses aspectos, se relacionam um com o outro de maneira estruturada e não arbitrária – as emoções perpassam a forma como os sons são percebidos, estes se relacionam com a escuridão do quarto e o espaço perceptivo tátil provocado pelo lençol, compondo a rica estrutura de cada experiências. A experiência de ser você, enquanto aguarda o sono chegar, tem uma qualidade particular

18 A técnica foi recentemente considerada “*a milestone in untangling the ancient mind-body problem*” (Koch, 2017).

justamente por ser composta por todas essas informações específicas que se relacionam entre si. Ela é, neste sentido, um múltiplo específico e estruturado (Tononi, 2012). Porém, estas múltiplas informações não se apresentam a você como experiências fragmentadas: a experiência de ser você é, a cada momento, uma experiência *unificada e coerente*, e que se estende no tempo e no espaço. Essa unidade parece ser uma propriedade tão constitutiva da experiência que é difícil (senão impossível) até mesmo imaginar o que significaria sustentar, simultaneamente, duas experiências conscientes distintas e totalmente independentes em um único sujeito (Bayne e Chalmers, 2003). Mas toda esta unidade e coerência, toda essa informação e estrutura, tudo isso se perde quase instantaneamente quando a consciência se desvanece.

O que este experimento introspectivo nos mostra é que, mesmo sem conhecermos a natureza exata da consciência, há pelo menos dois aspectos que podem ser identificados como sendo comuns a qualquer experiência consciente (Bayne e Chalmers, 2003; Tononi, 2004; Bayne, 2010): i) cada particular experiência consciente específica e estruturada é constituída por múltiplos elementos perceptivos, emotivos e cognitivos que a caracterizam qualitativamente, distinguindo-a assim das demais experiências; ii) esta multiplicidade de elementos apresenta-se ao sujeito consciente como um todo integrado, unificado, formando assim uma unidade, uma experiência única. O reconhecimento de tais propriedades fenomenológicas básicas da experiência, acessíveis, em princípio, a qualquer sujeito consciente de uma perspectiva de primeira pessoa, possibilita que se questione pelas condições cerebrais necessárias para sustentar a consciência: como poderia ser possível que a multiplicidade e a unidade da experiência sejam simultaneamente constituídas pelo cérebro humano?

Um dos princípios organizacionais do sistema nervoso central é a especialização funcional de suas estruturas, com áreas cerebrais específicas sendo estruturadas de modo a processar modalidades específicas de informação (Kolb e Whishaw, 1990). Para que a informação que compõe a cena consciente possa estar disponível ao sujeito que a experimenta, é necessário, portanto, que o funcionamento diferenciado destes núcleos cerebrais esteja preservado. Mas não basta, para a unidade da experiência, a preservação da integridade destas funções enquanto módulos fragmentados: deve haver algum tipo de comunicação entre as diferentes áreas corticais

envolvidas nesta experiência. Tome por exemplo a experiência de ler este texto. Você lê uma palavra, que é primariamente processada no córtex visual, localizado atrás da sua cabeça: esta palavra lhe traz a lembrança de um evento longínquo na memória, que por sua vez colore sua experiência de uma emoção específica produzida pela descarga de núcleos neurais localizados na parte frontal do seu cérebro. Não há como supor tal reação sem uma cadeia causal específica: as diferentes áreas corticais e subcorticais envolvidas na construção da experiência, especializadas em processar informações específicas, devem poder influenciar umas às outras, devem poder estar em interação para, através desta interação, possibilitar que a informação que compõe a cena consciente seja integrada em um todo. De fato, hoje sabemos que a integração da informação entre diferentes núcleos cerebrais é realizada no sistema talamocortical através de um densa rede de conexões de *feedforward*, que progridem no sentido de um processamento cada vez mais abstrato, bem como de *feedback* que incluem circuitos de reentrada subcorticais através das quais áreas hierarquicamente superiores em termos da complexidade da informação por elas processada modulam o processamento realizado em áreas corticais primárias (Lamme e Roelfsema, 2000; Guillery e Sherman, 2002).

Deste modo, a integridade e funcionalidade dos circuitos que sustentam estes dois aspectos do funcionamento cerebral - a *diferenciação* dos seus módulos funcionais e a *integração* entre tais módulos – aparecem assim como estando vinculados à capacidade cerebral de sustentar a consciência. Essa ideia, originalmente proposta por Giulio Tononi e Gerald Edelman (Tononi e Edelman, 1998), está na base de uma teoria matemática sobre a natureza última da consciência: a Teoria da Informação Integrada (Tononi, 2004; Tononi *et al.*, 2016). Independentemente dos avanços e dificuldades específicos desta teoria, o paradigma teórico geral, centrado nos dois pilares de diferenciação e integração, define um ponto de partida possível para o desenvolvimento de índices empíricos de estado global de consciência. O Índice de Complexidade Perturbacional (PCI) foi introduzido justamente para quantificar o balanço entre a integração e a diferenciação dos núcleos cerebrais, de modo a estimar não só o conteúdo informativo da dinâmica de diferentes áreas cerebrais, mas também a capacidade destas áreas de integrarem informação entre si.

Uma analogia pode ser útil para compreendermos o que PCI quantifica. Imagine uma conversa entre um grupo de amigos, todos interessados e engajados no tema em debate: quando um participante fala, os demais indivíduos ouvem e respondem, cada um na sua vez, a partir de pontos de vista particulares, constituindo assim um diálogo rico e diverso que reflete a diversidade de opiniões assim como a integração do grupo. Se um indivíduo do grupo faz uma pergunta, mas ninguém responde, seja porque ninguém escuta ou porque ninguém se interessa, um diálogo rico não se estabelece. Do mesmo modo, não há diálogo se, quando um amigo faz uma pergunta, todos os demais irrompem com respostas idênticas e proferidas ao mesmo tempo. PCI pretende quantificar justamente a riqueza ou complexidade do “diálogo” entre as diferentes áreas do cérebro: em uma condição de consciência, esperamos encontrar os diferentes núcleos cerebrais, cada um com sua funcionalidade particular, integrados em um todo, tal que seja possível uma “conversa” diferenciada e integrada entre tais núcleos. Mas, se a perda de consciência for causada pela ruptura deste equilíbrio entre a diferenciação e a integração das áreas cerebrais, então tal diálogo pode se tornar impossível em condições como durante sono profundo ou durante anestesia, quando o cessar dessa “conversa” sinaliza o desvanecer de toda e qualquer experiência.

Como medir tal diálogo cerebral na prática? Dois aspectos aparecem como fundamentais: a resolução temporal dos métodos empregados e a possibilidade de adoção de uma perspectiva perturbacional. Como a comunicação entre áreas corticais acontece na escala de milissegundos, as técnicas não invasivas para registro de ativação cerebral que podem ser utilizadas na estima da integração e diferenciação dos núcleos cerebrais ficam restritas, essencialmente, à eletroencefalografia (EEG). Além disso, para que seja possível inferir a causalidade inerente à ideia de um diálogo cortical integrado, é necessário empregar algum tipo de estímulo de modo a melhor controlar a cadeia causal do fluxo de informação. Este ponto passa a ser de especial relevância quando se leva em consideração a natureza dos sinais cerebrais obtidos com registros de EEG e que são normalmente corrompidos por fontes não neurais, como movimentos oculares, ativações musculares e outros ruídos e artefatos.

PCI foi então proposto (Casali *et al.*, 2013) empregando-se uma técnica que envolve estimular magneticamente uma região focal da superfície do

cérebro e medir a resposta provocada em outras regiões cerebrais (Figura 2). O estímulo, realizado de forma não invasiva por uma bobina magnética apoiada na superfície da cabeça - uma técnica conhecida por Estimulação Magnética Transcraniana - é responsável por produzir o disparo, reproduzível e confiável, de grupos de neurônios localizados na área cortical logo abaixo da bobina. Em seguida, outros núcleos cerebrais podem, por sua vez, responder a este disparo inicial, desde que estejam efetivamente conectados com a área estimulada. Tais respostas de diferentes áreas cerebrais, que se apresentam em um sistema de registro de EEG como sinais evocados (Figura 2), possuem formas de onda determinadas pelas propriedades dinâmicas dos respectivos núcleos neurais que as geram (Rosanova *et al.*, 2009). Deste modo, ao se registrar as respostas de tais núcleos, podemos inferir não só sobre a conectividade destes núcleos com a área estimulada (integração), como também sobre quão diversos estes núcleos são entre si (diferenciação).

Em uma série de trabalhos (Rosanova *et al.*, 2012; Casali *et al.*, 2013; Sarasso *et al.*, 2015; Casarotto *et al.*, 2016), verificou-se que respostas cerebrais ao estímulo magnético, quando registradas em sujeitos saudáveis, conscientes e capazes de reportar seu estado de consciência, eram sempre diversas e integradas: diferentes regiões do córtex de tais indivíduos, mesmo quando distantes da área estimulada, respondiam fortemente ao estímulo inicial, com ondas evocadas que possuíam características específicas e que, juntas, compunham um conjunto diverso, complexo de sinais evocados. Por outro lado, registros obtidos em sono profundo ou em anestesia nestes mesmos indivíduos eram invariavelmente simples: na maioria das vezes, as áreas cerebrais que não eram estimuladas diretamente nem sequer respondiam ao estímulo ou, quando respondiam, apresentavam respostas muito semelhantes entre si. A maneira como esse “diálogo” cerebral provocado pelo estímulo magnético se desenvolvia no tempo e também no espaço, entre diferentes regiões do cérebro, aparecia como estando associada, portanto, ao estado de consciência: em indivíduos conscientes, áreas específicas do cérebro reagiam em tempos específicos, formando um padrão espaço-temporal complexo e que era invariavelmente perdido quando a consciência se desvanecia. Deste modo, ao se estimar a complexidade deste conjunto de respostas cerebrais através do emprego de medidas clássicas de compressão de dados, foi possível obter um índice numérico normalizado

entre 0 e 1 - o índice de Complexidade Perturbacional - capaz de discriminar com perfeição o estado de consciência em indivíduos saudáveis.

Tal escala numérica objetiva, por estar perfeitamente associada ao estado de consciência em sujeitos saudáveis, pode ser utilizada para se comparar a complexidade do diálogo cerebral observada nestes sujeitos com a mesma complexidade em pacientes que sofrem de distúrbios da consciência e encontram-se incapazes de relatar o próprio estado de consciência. Em um estudo publicado na *Annals of Neurology* em 2016, a escala foi aplicada em 127 pacientes com lesões cerebrais traumáticas e vasculares, associadas ou não a distúrbios da consciência (Casarotto *et al.*, 2016). 48 dos pacientes estudados eram capazes de se comunicar e possuíam um índice de complexidade perturbacional calculado durante consciência dentro da faixa dos valores obtidos para esta mesma condição em sujeitos saudáveis. A acurácia da medida também se mostrou alta em pacientes minimamente conscientes e que se recuperaram do coma: dos 38 pacientes diagnosticados nesta condição e que exibiam sinais confiáveis da presença da alguma experiência subjetiva, mesmo se ainda incapazes de comunicar o próprio estado de consciência, 36 pacientes (94,7%) possuíam um índice de complexidade perturbacional acima do limiar mínimo para consciência.

Esta alta confiabilidade da medida, mesmo em pacientes com lesões cerebrais e que se recuperaram do coma, motivou o emprego da técnica na população mais desafiadora do ponto de vista do diagnóstico clínico dos distúrbios da consciência: a população formada por pacientes em estado vegetativo. Dos 43 pacientes estudados nesta condição no trabalho de 2016, 34 possuíam valores de PCI abaixo do limiar para a consciência encontrado em sujeitos saudáveis, confirmando assim o diagnóstico clínico. Mas 9 destes pacientes, que estavam acordados mesmo se clinicamente inconscientes, possuíam atividades cerebrais complexas, com valores de PCI compatíveis com o estado de consciência de indivíduos saudáveis. Estariam tais pacientes conscientes na ocasião do experimento? Ainda é cedo para respondermos. Mas a estratificação destes casos complexos, baseada em um índice objetivo da capacidade do sistema cerebral de sustentar um “diálogo” cortical, pode ser fundamental para o apoio ao diagnóstico e acompanhamento clínico destes pacientes, incluindo, possivelmente, o emprego de estratégias de intervenção voltadas a restaurar a interação destes pacientes com o meio externo (Schiff *et al.*, 2007).

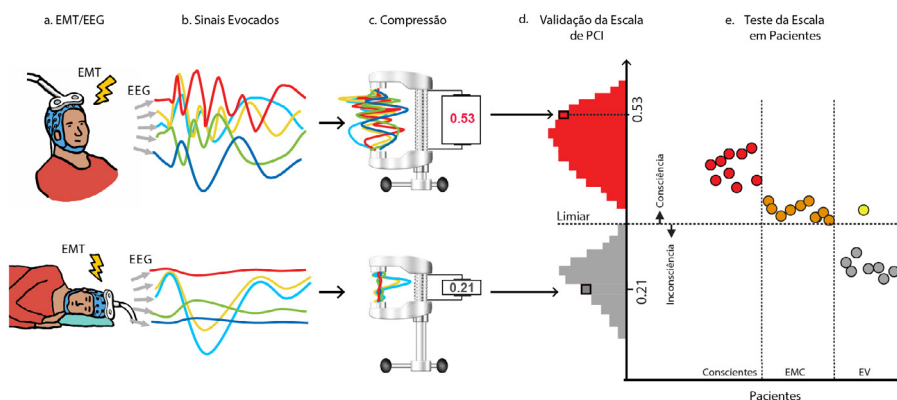


Figura 2: Construção, validação e teste do Índice de Complexidade Perturbacional (PCI). A combinação da Estimulação magnética transcraniana (EMT) com o registro de eletroencefalografia (EEG) de alta densidade permite a extração não invasiva de sinais cerebrais evocados produzidos pelo diálogo cortical que se segue ao estímulo magnético. Inicialmente, a coleta destes sinais (a) foi realizada em indivíduos capazes de reportar, imediatamente ou após o experimento, o seu estado de consciência e o experimento foi repetido, em cada indivíduo, tanto em condição de consciência (vigília), quanto de inconsciência (sono profundo, anestesia). Para cada experimento, os sinais evocados em resposta à EMT (b) são processados por técnicas matemáticas e de compressão de dados que resultam em um único número, o PCI, limitado em uma escala de 0 a 1, e que representa a complexidade do padrão espaço-temporal deste diálogo cerebral (c). Tal experimento foi realizado em um grande número de indivíduos capazes de comunicar o seu estado de consciência, sejam estes indivíduos saudáveis, sejam pacientes com lesões cerebrais, doenças neurodegenerativas e outras condições, resultando assim em duas grandes distribuições de valores de PCI (d): uma distribuição de valores observados durante consciência (vermelho) e outra para a condição de inconsciência (cinza). A partir destas distribuições, um limiar pode ser obtido (linha tracejada horizontal), representando o valor máximo observado em condições de inconsciência. Esta escala, validada nesta população, passou então a ser testada em pacientes com distúrbios de consciência e que se recuperaram do coma (e). Pacientes conscientes que conseguem se comunicar (círculos vermelhos) ou pacientes minimamente conscientes (EMC, círculos laranjas) apresentaram complexidade alta, na faixa de consciência, confirmando a validade da escala também nesta população. Já pacientes sem responsividade, diagnosticados em estado vegetativo, e dos quais não podemos conhecer o estado de consciência, em sua grande maioria (círculos cinza) apresentaram diálogos corticais de baixa complexidade, na faixa da inconsciência. Em cerca de 20% a 30% destes pacientes, porém, sinais de um diálogo cortical complexo foram detectados (círculo amarelo).

5. PERSPECTIVAS PARA O FUTURO DA CIÊNCIA DA CONSCIÊNCIA

O desenvolvimento do Índice de Complexidade Perturbacional ilustra o potencial de se tentar aproximar considerações fenomenológicas sobre a experiência consciente com procedimentos científicos já em uso como registro de sinais neurais, técnicas de estimulação não-invasiva, e algoritmo de quantificação de complexidade. De um lado, experimentos introspectivos realizados em uma perspectiva de primeira pessoa são fundamentais para entendermos o que exatamente uma ciência *da* consciência deve ser capaz de explicar, pois só eles podem apontar para as características essenciais do nosso objeto de estudo. Além disso, como mostramos, tais experimentos podem revelar propriedades gerais da experiência consciente (como informação e integração) que, uma vez presentes nos processos neurais subjacentes à consciência, servem de guia para a construção de um aparato algorítmico e experimental capaz de quantificá-las. De outro, o entendimento do modo como o cérebro processa a informação, combinado às técnicas experimentais certas, leva-nos a testar medidas empíricas cuja validação – a partir de grupos controle em que já sabemos de antemão, com um grau alto de confiança, o estado de consciência – nos autoriza a realizar extrapolações controladas sobre casos em que nosso conhecimento é menor como, por exemplo, em pacientes em estado vegetativo. A partir deste ponto, podemos esperar que avanços empíricos e teóricos sigam o curso natural do progresso científico.

Do ponto de vista empírico, um avanço recente diz respeito a uma nova versão do índice de Complexidade Perturbacional (Comolatti *et al.*, 2019). Nesta versão, a complexidade da resposta cerebral ao estímulo é calculada através da adaptação de técnicas de recorrência a sinais evocados e empregando-se algoritmos mais eficientes e que não requerem um aparato de EEG de alta densidade. A nova versão do índice é, portanto, não só mais fácil de ser calculada, como pode inclusive ser aplicada a outros tipos de sinais cerebrais, como, por exemplo, registros intracranianos. Neste mesmo estudo de 2019, a atividade evocada registrada com eletrodos intracerebrais em pacientes que foram submetidos à cirurgia para tratamento de epilepsia pôde ser analisada e confirmou-se, desta vez com registros obtidos de dentro do cérebro, a interrupção dos padrões complexos de resposta cerebral que ocorre durante a perda de consciência.

A estes avanços práticos, muitas perguntas de cunho mais fundamental se seguem. O sucesso de PCI parece indicar que a consciência de fato está relacionada com o processamento global de informação pelo cérebro. Que características dessa informação são tão fundamentais a ponto de produzirem a consciência? É possível generalizar tais resultados para outras espécies e, até mesmo, para um cérebro de silício? E de forma mais fundamental ainda: o que há de especial neste processamento que faz com que, além dos processos físicos de sinapses e alterações cerebrais, uma experiência subjetiva associada a este processamento emergja? Formulações teóricas, como a Teoria da Informação Integrada (Tononi *et al.*, 2016), passam a ganhar força na medida em que buscam formalizar matematicamente as relações entre as propriedades da experiência e as propriedades físicas do sistema neural, justamente com o objetivo de construir uma ponte entre as perspectivas de primeira e terceira pessoa. Outros avanços nesta área deverão passar justamente pelo estudo das relações formais entre medidas empíricas como PCI e tais formalismos teóricos sobre as bases da consciência.

Dentro do quadro traçado, podemos imaginar como o ciclo iterativo e gradual de construção teórica e validação empírica de medidas físicas sobre a consciência poderia nos levar a confiar cada vez mais no poder preditivo de tais medidas a ponto de extrapolarmos sua aplicação a casos em que relatos fenomenológicos ou sinais comportamentais estejam indisponíveis: pensemos nos pacientes com distúrbios de consciência, mas também em recém-nascidos, outras espécies de animais próximos mas também mais distantes – um polvo? – até o limite da imaginação científica – computadores, seres alienígenas? Somado a isso, esse desenvolvimento nos leva também a referendar e refinar as descrições fenomenológicas que serviram como guia inicial para a construção da medida e da teoria. É possível que o papel da perspectiva em primeira-pessoa – do observador e de sua experiência subjetiva – também se modifique, e que os aparatos experimentais e formulações teóricas passem a iluminar aspectos da experiência consciente que, enquanto sujeitos individuais, não éramos capazes de conhecer por nós mesmos.

É provável que ainda seja cedo para afirmarmos que já temos o aparato tecnológico e experimental necessário ou as ferramentas formais adequadas para construir a ponte entre a consciência e mundo físico. Ao contrário, é mais provável que novos avanços nas técnicas de registro e estimulação

neural, inovações no modo como quantificamos a informação neural e como descrevemos nossa fenomenologia subjetiva, venham em breve a revolucionar a forma como concebemos a relação profunda entre cérebro e experiência subjetiva. Mas talvez não seja cedo para afirmarmos que o progresso recente indica o caminho que deveremos seguir. Uma ciência da consciência, como qualquer outra ciência, visa discernir o seu fenômeno de estudo em sua natureza própria, sem tomá-lo por outra coisa. Para isso, como Newton, devemos buscar o aparato justo – o experimento e a ferramenta formal – capaz de dar ao fenômeno sua consistência objetiva, sem que ele dependa dos caprichos de um observador para existir. Assim como a luz refrata ao mudar de meio do ar para o vidro, um cérebro gera uma experiência subjetiva de maneira intrínseca, independente de seus observadores externos. E se pretendemos aplicar o método científico para tentar entender melhor o que faz sermos quem somos, deveremos começar do modo que é possível: reduzindo a luz do quarto, abrindo um buraco de tamanho certo na janela e aproximando o nosso prisma de vidro da luz interna da experiência, por mais fugaz, insólita, e misteriosa que essa luz hoje nos pareça.

6. AGRADECIMENTOS

Os autores agradecem o apoio da Fundação de Amparo do Estado de São Paulo (FAPESP) - projetos 2016/08263-9 e 2017/ 03678-9. Agradecem também Greta Comolatti pelas ilustrações e Maíra Suzuki pela montagem gráfica das figuras.

7. CONTRIBUIÇÃO DOS AUTORES

Os autores contribuíram igualmente para a concepção, redação e edição do presente capítulo.

8. REFERÊNCIAS

AZABOU, E. *et al.* Mismatch negativity to predict subsequent awakening in deeply sedated critically ill patients. **Br J Anaesth**, v. 121, n. 6, p. 1290-1297, Dec 2018.

BARRETT, A. B. *et al.* Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. **PLoS One**, v. 7, n. 1, p. e29072, 2012.

BAYNE, T. **The Unity of Consciousness**. Oxford University Press, 2010.

BAYNE, T.; CHALMERS, D. What is the unity of consciousness? In: CLEEREMANS, A. (Ed.). **The unity of consciousness: Binding, integration, and dissociation**: Oxford University Press, 2003.

BERNAT, J. L. Chronic disorders of consciousness. **Lancet**, v. 367, n. 9517, p. 1181-92, Apr 2006.

BOGEN, J.; WOODWARD, J. Saving the Phenomena. **The Philosophical Review**, v. XCVIII, n. 3, 1988.

BOLY, M. *et al.* Preserved feedforward but impaired top-down processes in the vegetative state. **Science**, v. 332, n. 6031, p. 858-62, May 2011.

BRUNO, M. A.; LAUREYS, S.; DEMERTZI, A. Coma and disorders of consciousness. **Handb Clin Neurol**, v. 118, p. 205-13, 2013.

CASALI, A. G. *et al.* A theoretically based index of consciousness independent of sensory processing and behavior. **Sci Transl Med**, v. 5, n. 198, p. 198ra105, Aug 14 2013.

CASAROTTO, S. *et al.* Stratification of unresponsive patients by an independently validated index of brain complexity. **Ann Neurol**, v. 80, n. 5, p. 718-729, Nov 2016.

CHALMERS, D. Facing up to the problem of consciousness. **Journal of Consciousness Studies** v. 2, p. 19, 1995.

CHANG, H. **Inventing temperature : measurement and scientific progress**. Oxford ; New York: Oxford University Press, 2007. xviii, 286 p. ISBN 9780195337389.

COMOLATTI, R. *et al.* A fast and general method to empirically estimate the complexity of brain responses to transcranial and intracranial stimulations. **Brain Stimul**, v. 12, n. 5, p. 1280-1289, 2019 Sep - Oct 2019.

CRICK, F.; KOCH, C. Toward a neurobiological theory of consciousness. **Seminars in the Neurosciences**, v. 2, 1990.

DENNETT, D. Heterophenomenology reconsidered. **Phenomenology and the Cognitive Sciences**, v. 6, 2007.

DESCARTES, R. **Meditações Metafísicas**. GUINSBURG, J. e PRADO JÚNIOR, B. 5a. Nova Cultural, 1996.

ENGEL, A. K.; SINGER, W. Temporal binding and the neural correlates of sensory awareness. **Trends Cogn Sci**, v. 5, n. 1, p. 16-25, Jan 2001.

ENGEMANN, D. A. *et al.* Robust EEG-based cross-site and cross-protocol classification of states of consciousness. **Brain**, v. 141, n. 11, p. 3179-3192, 11 2018.

FISCHER, C.; LUAUTE, J.; MORLET, D. Event-related potentials (MMN and novelty P3) in permanent vegetative or minimally conscious states. **Clin Neurophysiol**, v. 121, n. 7, p. 1032-42, Jul 2010.

GIACINO, J. T.; KALMAR, K.; WHYTE, J. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. **Arch Phys Med Rehabil**, v. 85, n. 12, p. 2020-9, Dec 2004.

GIACINO, J. T. *et al.* Comprehensive systematic review update summary: Disorders of consciousness: Report of the Guideline Development, Dissemination, and Implementation Subcommittee of the American Academy of Neurology; the American Congress of Rehabilitation Medicine; and the National Institute on Disability, Independent Living, and Rehabilitation Research. **Neurology**, v. 91, n. 10, p. 461-470, 09 2018.

GUILLERY, R. W.; SHERMAN, S. M. Thalamic relay functions and their role in corticocortical communication: generalizations from the visual system. **Neuron**, v. 33, n. 2, p. 163-75, Jan 2002.

HÖLLER, Y. *et al.* Preserved oscillatory response but lack of mismatch negativity in patients with disorders of consciousness. **Clin Neurophysiol**, v. 122, n. 9, p. 1744-54, Sep 2011.

HUSSERL, E. **Meditações cartesianas: uma introdução à Fenomenologia**. NOLASCO, F. M. Edipro, 2019.

KASKINORO, K. *et al.* Wide inter-individual variability of bispectral index and spectral entropy at loss of consciousness during increasing concentrations of dexmedetomidine, propofol, and sevoflurane. **Br J Anaesth**, v. 107, n. 4, p. 573-80, Oct 2011.

KOCH, C. **The quest for consciousness : a neurobiological approach**. Denver, Colo.: Roberts and Co., 2004. xviii, 429 p. ISBN 0974707708 (hardcover).

KOCH, C. **How to make a conscious meter**. Scientific American. October 17 2017.

KOELSCH, S. *et al*. Auditory processing during deep propofol sedation and recovery from unconsciousness. **Clin Neurophysiol**, v. 117, n. 8, p. 1746-59, Aug 2006.

KOLB, B.; WHISHAW, I. **Fundamentals of Human Neuropsychology**. Freeman and Co., 1990.

LAMME, V. A.; ROELFSEMA, P. R. The distinct modes of vision offered by feedforward and recurrent processing. **Trends Neurosci**, v. 23, n. 11, p. 571-9, Nov 2000.

LAUREYS, S. *et al*. Unresponsive wakefulness syndrome: a new name for the vegetative state or apallic syndrome. **BMC Med**, v. 8, p. 68, Nov 2010.

LEIBNIZ, G. W. **A Monadologia e outros textos**. SOUZA, F. B. G. E. Editora Hedra, 2009.

LEVINE, J. Materialism and Qualia: The explanatory gap. **Pacific Philosophical Quarterly** v. 64, 1983.

_____. **Purple Haze: The Puzzle of Conscious Experience**. Cambridge University Press, 2001.

MACIVER, M. B. Mind and Brain: Consciousness in Unresponsive Subjects. **Curr Biol**, v. 25, n. 23, p. R1140-2, Dec 2015.

MCGINN, C. **The Problem of Consciousness**. Oxford Univeristy Press, 1991.

METZINGER, T. **Conscious Experience**. Imprint Academic, 1995.

NAGEL, T. What is it like to be a bat? **The Philosophical Review**, v. 83, n. 4, p. 435-450, 1974.

NAGEL, T. **The view from nowhere**. New York: Oxford University Press, 1986. xi, 244 p. ISBN 0195036689.

NEWTON, I. Óptica. ASSIS, A. K. EDUSP, 1996.

OVERGAARD, M. **Behavioral methods in consciousness research**. First edition. 2015. ix, 268 pages, 8 unnumbered pages of plates ISBN 9780199688890 0199688893.

- PLOURDE, G.; PICTON, T. W. Long-latency auditory evoked potentials during general anesthesia: N1 and P3 components. **Anesth Analg**, v. 72, n. 3, p. 342-50, Mar 1991.
- RAGAZZONI, A. *et al.* Vegetative versus minimally conscious states: a study using TMS-EEG, sensory and event-related potentials. **PLoS One**, v. 8, n. 2, p. e57069, 2013.
- ROSANOVA, M. *et al.* Natural frequencies of human corticothalamic circuits. **J Neurosci**, v. 29, n. 24, p. 7679-85, Jun 2009.
- ROSANOVA, M. *et al.* Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. **Brain**, v. 135, n. Pt 4, p. 1308-20, Apr 2012.
- SARASSO, S. *et al.* Consciousness and Complexity during Unresponsiveness Induced by Propofol, Xenon, and Ketamine. **Curr Biol**, v. 25, n. 23, p. 3099-105, Dec 7 2015.
- SCHIFF, N. D. *et al.* Behavioural improvements with thalamic stimulation after severe traumatic brain injury. **Nature**, v. 448, n. 7153, p. 600-3, Aug 2007.
- SCHNAKERS, C.; MAJERUS, S.; LAUREYS, S. Bispectral analysis of electroencephalogram signals during recovery from coma: preliminary findings. **Neuropsychol Rehabil**, v. 15, n. 3-4, p. 381-8, 2005 Jul-Sep 2005.
- SCHNAKERS, C. *et al.* Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment. **BMC Neurol**, v. 9, p. 35, Jul 2009.
- SETH, A. K.; BARRETT, A. B.; BARNETT, L. Causal density and integrated information as measures of conscious level. **Philos Trans A Math Phys Eng Sci**, v. 369, n. 1952, p. 3748-67, Oct 13 2011.
- SETH, A. K. *et al.* Measuring consciousness: relating behavioural and neurophysiological approaches. **Trends Cogn Sci**, v. 12, n. 8, p. 314-21, Aug 2008.
- SICLARI, F. *et al.* The neural correlates of dreaming. **Nat Neurosci**, v. 20, n. 6, p. 872-878, Jun 2017.
- SIGL, J. C.; CHAMOUN, N. G. An introduction to bispectral analysis for the electroencephalogram. **J Clin Monit**, v. 10, n. 6, p. 392-404, Nov 1994.

SILVA, C. C.; MARTINS, R. A. Nova teoria sobre luz e cores de Isaac Newton: uma tradução comentada. **Revista Brasileira de Ensino de Física**, v. 18, 1996.

SITT, J. D. *et al.* Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. **Brain**, v. 137, n. Pt 8, p. 2258-70, Aug 2014.

SUPP, G. G. *et al.* Cortical hypersynchrony predicts breakdown of sensory processing during loss of consciousness. **Curr Biol**, v. 21, n. 23, p. 1988-93, Dec 2011.

TONONI, G. An information integration theory of consciousness. **BMC Neurosci**, v. 5, p. 42, Nov 2004.

_____. **Phi: A Voyage from the Brain to the Soul**. Pantheon Books, 2012.

TONONI, G. *et al.* Integrated information theory: from consciousness to its physical substrate. **Nat Rev Neurosci**, v. 17, n. 7, p. 450-61, Jul 2016.

TONONI, G.; EDELMAN, G. M. Consciousness and complexity. **Science**, v. 282, n. 5395, p. 1846-51, Dec 4 1998.

VAN ERP, W. S. *et al.* The vegetative state/unresponsive wakefulness syndrome: a systematic review of prevalence studies. **Eur J Neurol**, v. 21, n. 11, p. 1361-8, Nov 2014.

VAN ERP, W. S. *et al.* The vegetative state: prevalence, misdiagnosis, and treatment limitations. **J Am Med Dir Assoc**, v. 16, n. 1, p. 85.e9-85.e14, Jan 2015.

VARELA, F. J. Neurophenomenology: A Methodological Remedy for the Hard Problem. **Journal of Consciousness Studies**, v. 3, n. 4, 1996.

04.

AUTOPSICOBOTS: MÁQUINAS AUTOCONSCIENTES SÃO DIGNAS DE CONSIDERAÇÃO ÉTICA? ENTRE O PENSAMENTO DE KANT E A (BIO)ÉTICA PARA TODOS OS SERES

Eugênio Silva

Oswaldo Jesus Rodrigues da Motta

Fabiola Alves Alcântara

Rodrigo Siqueira-Batista

1. INTRODUÇÃO

A história da Computação é relativamente recente. Pode-se apontar que a Segunda Grande Guerra foi o marco inicial, momento no qual foram construídos os primeiros computadores digitais (O'REGAN, 2018). Com o avanço no desenvolvimento do *hardware* e do *software*, tornou-se mais simples o processo de programação, ampliando-se as possibilidades de uso desses dispositivos. De todo modo, a despeito desta *modernização*, os computadores atuais ainda preservam muitas características dos seus ancestrais (FONSECA FILHO, 2007).

Neste cenário, merece destaque o fato de que as tarefas executadas por essas máquinas sejam inerentemente *computáveis* (ZILIO, 2009). As atividades computáveis são apenas uma parte daquelas de interesse no dia a dia das pessoas. Há um conjunto mais amplo composto por aquelas que estão diretamente ligadas ao intelecto humano e que, pelo menos até o momento, não são passíveis de *computação*, ou seja, não há procedimentos efetivos capazes de descrevê-las de forma completa e precisa (CAFEZEIRO *et al.*, 2010; PENROSE, 1998). Tratam-se, pois, de tarefas que envolvem habilidades humanas como raciocínio, percepção, aprendizagem, compreensão,

planejamento e linguagem (MANZANO, 2018; LOBO, 2018). Para o desenvolvimento de modelos computacionais que sejam capazes de fazer com que as máquinas ampliem suas capacidades (e possam ser aplicadas na execução de atividades aparentemente não computáveis) pode-se empregar, por exemplo, algoritmos capazes de ajustar iterativamente um conjunto de parâmetros internos do modelo, segundo uma regra bem definida, em um processo denominado *aprendizagem*.

Torna-se importante ressaltar que os computadores ainda não são capazes de reproduzir fielmente as mesmas habilidades cognitivas do *Homo sapiens*, mas o que já se conseguiu até o momento é bastante útil para diversas aplicações. Apesar das limitações, modelos baseados em Inteligência Artificial (IA) – cujo conceito será discutido adiante – são empregados como ferramentas de apoio à tomada de decisão em diversas áreas do conhecimento. Todavia, uma das propostas em discussão nas últimas décadas – e que tem sido muito recorrente em filmes de ficção científica (como na película *Her* e na série *Westworld* da HBO) – diz respeito à possibilidade de desenvolvimento de uma máquina pensante e dotada de autoconsciência, a qual, como será posteriormente apresentado, é um dos aspectos centrais do Programa da IA forte (FANG, 2020; MOTTA *et al.*, 2020a).

Os avanços das pesquisas em IA – mormente ao considerar tal Programa, IA forte, como horizonte a ser perseguido – têm trazido questões quanto ao estatuto ético de máquinas pensantes (caso estas sejam, um dia, *criadas*). De uma perspectiva filosófica, seria possível recuperar a formulação de Immanuel Kant (SIQUEIRA-BATISTA *et al.*, 2021), o qual argumenta que *todo ser racional é sempre um fim em si mesmo* (KANT, 1960). Desta feita:

“Existir como fim em si mesmo significa existir como fim (e não como um mero meio) em um Reino dos Fins, possuindo uma dignidade (e não apenas um preço). Existir como fim em si mesmo é, assim, existir como um valor absoluto [absoluter Wert]. Para garantir tal valor absoluto, Kant relaciona a condição de fim em si mesmo à própria natureza racional [vernünftige Natur] do ser humano – enquanto sua humanidade, entendida em um sentido não antropológico (empírico), mas de uma Metafísica dos Costumes (racional). Tal natureza racional existe como fim em si mesma, na medida em que, porque o ser humano possui razão prática, considerada sua vontade livre, autônoma, ele deve ser respeitado como fim em si mesmo – e, portanto, enquanto valor absoluto e consequentemente dotado de dignidade” (STOBBE, 2017, p. 408).

Essa *formulação kantiana* abre a possibilidade para pensar a inclusão das máquinas pensantes na esfera da *consideração ética* – ou seja, como *fins em si mesmas*, reconhecíveis como *agentes morais*¹⁹ e, quiçá, como *pacientes morais*²⁰ –, o que traz novos contornos para as interações humano-máquina e para as responsabilidades de ambos os elementos desse binômio. Tal perspectiva pode ser trabalhada, complementarmente à filosofia moral de Kant, a partir de concepções éticas que sejam inclusivas para distintas formas de existência, abrindo-se o campo de debate para a reflexão em termos da *(Bio)ética para todos os seres*, corrente teórica que propõe a inclusão de toda a forma de existência no rol de consideração ética, a partir da articulação de quatro conceitos chaves: *quididade*, *totalidade*, *interdependência* e *compaixão laica* (SIQUEIRA-BATISTA, 2020).

A partir dessas preliminares considerações, pode-se propor o seguinte exercício de pensamento: imagine a existência de um computador – ou de um robô (ou de qualquer artefato...) – pensante e autoconsciente; a pergunta que poderia ser proposta, na sequência, é a seguinte: tal “aparato” não deveria ser considerado um fim em si mesmo, digno de consideração por uma ética capaz de acolher incondicionalmente todos os seres? Com base nesses preliminares apontamentos, o objetivo do presente texto é discutir essa questão – partindo de breve exposição sobre os conceitos de “computabilidade”, “máquina de Turing”, “IA fraca” e “IA forte” e de um sucinto comentário acerca das “*Máquinas pensantes*” –, abordando seus eventuais desdobramentos, em termos da *já anunciada* ética kantiana e da proposta em construção da (bio)ética para todos os seres.

19 “Um *agente moral* é um sujeito que age ou abstém-se de agir intencionalmente. Suas habilidades são facilmente reconhecidas: é um tipo de ser que pode agir moralmente e imoralmente, ter deveres e responsabilidades e responder por seus atos. Além disso, um agente moral tem a capacidade de formar juízos sobre o certo e o errado; a capacidade de empreender ações deliberadas, isso quer dizer, considerar e pesar razões morais pró e contra diferentes modos de agir abertos à escolha; a capacidade de tomar decisões fundadas naquelas razões; a capacidade de seguir as decisões tomadas; a capacidade de assumir a responsabilidade caso não as cumpra” (TAYLOR, 1986, p.15).

20 “Na condição de *paciente* numa comunidade moral, um sujeito moral pode ser destituído de todas as características do agente moral, independentemente de ser humano ou não-humano. No entanto, a responsabilidade do agente moral perante aquele não muda, pois quando o agente moral se encontra ele mesmo na condição de paciente moral, os demais agentes morais também não têm a responsabilidade para com seu bem-estar diminuída” (TAYLOR, 1986, p.15-16).

2. COMPUTABILIDADE E A MÁQUINA DE TURING²¹

A ideia de tornar computadores inteligentes – ou seja, capazes de realizar tarefas nas quais “*as pessoas são melhores*” (RICH; KNIGHT, 1990, p. 5) – implica na necessidade de se demarcar adequadamente o que significa “computabilidade” e “computável”. De um modo geral, pode-se dizer que a *computabilidade* diz respeito à existência – ou não – de um procedimento capaz de resolver um determinado problema, utilizando para tal mister um determinado número de passos (ou seja, um algoritmo – entendido como um conjunto finito de instruções não ambíguas, o qual, uma vez seguido, determinará a realização de uma certa tarefa). O conceito de *computável* foi instituído por Alan Turing – no célebre artigo *On computable numbers, with an application to the Entscheidungsproblem* (TURING, 1936) – já na primeira frase do texto: “*Os números ‘computáveis’ podem ser descritos resumidamente como os números reais cujas expressões como um decimal são calculáveis por meios finitos. [...] De acordo com a minha definição, um número é computável se o seu decimal pode ser escrito por uma máquina*” (TURING, 1936, p. 230; trad. livre dos autores).

As ideias de “computável” e de “computabilidade” foram enunciadas, por Turing, conforme explicitado no excerto, nos termos da existência de uma máquina teórica, abstrata, a qual passou a ser denominada “Máquina de Turing” (MT), dotada dos seguintes componentes (Figura 1):

(1) uma *fita infinita*, dividida em células, as quais poderão estar *vazias* ou conter um símbolo qualquer pertencente a um *alfabeto finito* Σ , de modo que $\Sigma = \{\sigma_0, \sigma_1, \dots, \sigma_n\}$;

(2) um *cabeçote de leitura/escrita*, o qual contempla três ações: (i) a leitura do símbolo, (ii) a sobrescrita de um símbolo e (iii) a movimentação para a direita (*R*) ou para a esquerda (*L*);

(3) um *registrador de estados*, que lida com um conjunto *finito* de possibilidades, o qual pode ser formulado como $Q = \{q_1, q_2, \dots, q_m\}$;

(4) um conjunto *I, finito*, de instruções.

21 Essa seção reproduz parte do trabalho realizado por R Siqueira-Batista, por ocasião da conclusão da Licenciatura em Matemática na Universidade Estácio de Sá: “*Fundamentos Matemáticos da Inteligência Artificial: Notas sobre a Aprendizagem de Máquina*” (SIQUEIRA-BATISTA, 2018).

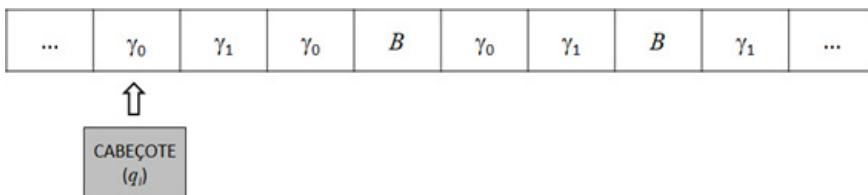


Figura 1. Representação esquemática da Máquina de Turing.
Ilustração elaborada por Rodrigo Siqueira-Batista.

Formalizando tais elementos, tem-se que a Máquina de Turing é uma 7-upla $M = (Q, \Gamma, b, \Sigma, \delta, q_0, F)$, no qual (HOPCROFT; ULLMAN, 1979):

Q = conjunto finito de *estados* da máquina;

Γ = *alfabeto* da máquina (conjunto *finito* de símbolos);

b = *branco* (importante: $b \in \Gamma$);

Σ = conjunto de *entrada* (importante: $\Sigma \subseteq \Gamma - \{b\}$);

δ = *função parcial de transição* (programa), determinada em concordância com δ : $Q \times \Gamma \rightarrow Q \times \Gamma \times \{L, R\}$; L representa o deslocamento para a esquerda e R para a direita;

q_0 = *estado inicial*;

F = conjunto de *estados finais*.

Resultados similares foram obtidos por Alonzo Church (CHURCH, 1936) – utilizando o cálculo- λ , método completamente distinto daquele empregado por Turing – o que permitiu a proposição da célebre Tese de Church-Turing: “*Toda a função que pode ser calculada por um procedimento finito pode também ser computada por uma máquina de Turing*” (LOFF, 2012, p. 69). Ademais, deve se destacar que para uma determinada função $f: D \rightarrow R$, um algoritmo será uma MT M , onde $M = (Q, \Gamma, b, \Sigma, \delta, q_0, F)$, se para uma entrada qualquer $z \in D$, M encontra a resposta correta $f(z) \in R$, tal que $q_0 z \vdash^* q_f f(z)$, $q_f \in F$, para toda cadeia $z \in D$.

Para fins de ilustração, o exemplo a seguir apresenta uma Máquina de Turing cuja finalidade é somar dois números inteiros positivos com representação unária. Nessa forma de representação, os algarismos que compõem o número são substituídos por uma cadeia de caracteres que repete um mesmo símbolo uma quantidade de vezes igual ao número representado. Nesse caso, o número 3, por exemplo, pode ser representado como

‘111’. Assim, para somar os números 2 e 3, as configurações inicial e final da fita da máquina poderiam ser representadas conforme a Figura 2.

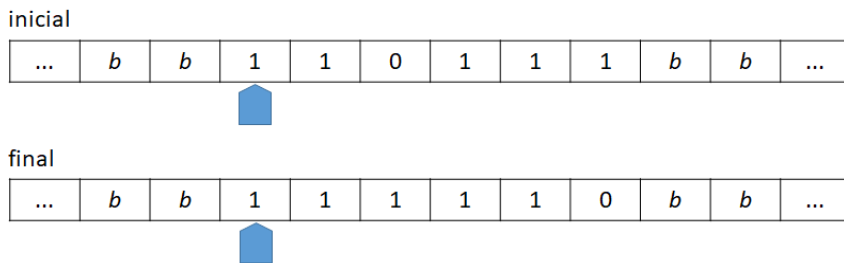


Figura 2. Configurações inicial e final da fita da Máquina de Turing.
Ilustração elaborada por Eugênio Silva.

É relevante notar que na configuração inicial as sequências de ‘1’s que representam os números são separadas pelo símbolo ‘0’. Ao final da execução, o símbolo ‘0’ está posicionado à direita da cadeia de ‘1’s que representa o resultado da soma. Importante observar também que em ambas as configurações o cabeçote de leitura está posicionado na célula em que se encontra o primeiro símbolo não ‘branco’ da fita.

Para a situação apresentada, a Máquina de Turing correspondente pode ser especificada pela 7-upla $M = (\{q_0, q_1, q_2, q_3, q_4\}, \{0, 1, b\}, b, \{0, 1\}, \delta, q_0, \{q_4\})$, onde δ é definida como:

δ	0	1	b
q_0	$(q_1, 1, R)$	$(q_0, 1, R)$	
q_1		$(q_1, 1, R)$	(q_2, b, L)
q_2		$(q_3, 0, L)$	
q_3		$(q_3, 1, L)$	(q_4, b, R)
q_4			

O diagrama de estados correspondente a M é exibido na Figura 3:

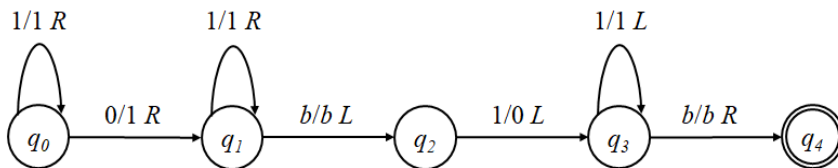


Figura 3. Diagrama de estados da Máquina de Turing somadora.
Ilustração elaborada por Eugênio Silva.

Pelo diagrama, o que se observa é que a partir do estado inicial q_0 , duas situações podem ocorrer: (i) o cabeçote lê o símbolo ‘1’, a máquina permanece no estado q_0 , escreve o símbolo ‘1’ na fita e desloca o cabeçote para a direita; ou (ii) o cabeçote lê o símbolo ‘0’, a máquina transita para o estado q_1 , escreve o símbolo ‘1’ na fita e desloca o cabeçote para a direita. Raciocínio análogo é utilizado para interpretar as outras transições. Dessa forma, ao percorrer o diagrama de estados a partir de q_0 e considerando a configuração inicial da fita apresentada na Figura 2, observa-se que a configuração final é alcançada exatamente no momento em que a máquina transita para o estado final q_4 , indicando que a tarefa foi executada de forma correta.

A proposta de Alan Turing para a caracterização do “computável” e da “computabilidade” – em termos da MT – tem uma série de desdobramentos teóricos (CAFEZEIRO *et al.*, 2010; PIQUEIRA, 2016), cuja discussão escaparia ao escopo do presente ensaio. Vale ressaltar, no entanto, que tais noções são centrais para a Ciência da Computação – e para a distinção entre o que é “computável” e o que é “não-computável” – ramo do conhecimento ao qual os estudos de IA estão visceralmente articulados (SIQUEIRA-BATISTA; SILVA, 2019), como se discutirá na sequência.

3. IA FORTE E IA FRACA

Embora a expressão “Inteligência Artificial” seja fruto de conferência acadêmica que John McCarthy organizou no Dartmouth College, em 1956, foi o artigo publicado por Alan Turing (1950) – e seu “Jogo da Imitação” (cujo filme homônimo faz referência – TYLDUM, 2015) ou “Teste de Turing” – que aparentemente melhor define e caracteriza o campo. No Teste de Turing, um ser humano entra em uma conversa, em linguagem natural, com outro humano e com uma máquina por meio de um terminal, mas sem

saber quem é um e quem é outro. Se depois de algum tempo de conversa, um julgador não for capaz de identificar com segurança quem é a máquina e quem é o ser humano, pode-se afirmar que a máquina passou no teste. Importante destacar que, mesmo tendo êxito no teste, isso não é suficiente para atestar a presença de uma inteligência genuína na máquina. Na verdade, trata-se de um mecanismo capaz de, no máximo, simular um comportamento entendido como inteligente, conforme defende John Searle (1980) em seu argumento do *Quarto Chinês*, o qual será ulteriormente comentado.

Mas afinal, o que é Inteligência Artificial? O binômio carece de melhor entendimento dada a complexidade da compreensão do que se entende por “inteligência” e por “artificial”, termos que “são de difícil definição” (NILSSON, 1998). Pode-se caracterizar *inteligência* – termo com origem no latim *intellectus*, de *intelligere* – como entendimento, compreensão. Composto de *íntus* = dentro e *lègere* = recolher, escolher, ler (BONOMI, 2004). O *artificial* pode ser compreendido como “*tudo aquilo que é feito pelo ser humano, ou seja, um artefato*” (VON ZUBEN, 2011). Essas definições não encerram os conceitos e, evidentemente, não se fecham às críticas ou embargos de outros autores. Todavia, direcionam – acenam – a um melhor entendimento sobre a IA. Uma definição de IA, parcialmente mencionada na seção anterior, que merece destaque por sua singeleza é a seguinte: “*O estudo de como fazer computadores realizarem coisas nas quais, no momento, as pessoas são melhores*” (RICH; KNIGHT, 1991, p. 3). Certamente, existem as mais variadas interpretações acerca desse campo de estudo, especialmente ao se considerar que as investigações acerca da IA são habitualmente coligidas em dois programas principais: IA fraca e IA forte (NAVEGA, 2001; SANVITO, 1995).

Na vertente denominada IA fraca, ocorre uma programação humana a fim de que a máquina responda conforme o pré-estabelecido por comandos específicos (SARACEVIC, 2008), realizando uma tarefa computável. Sistemas Especialistas, por exemplo, são capazes de interagir com o seu usuário se comportando como um especialista em um determinado assunto. Isso acontece quando são instados a emitir algum tipo de diagnóstico ou tomar alguma decisão com base em um conjunto de fatos que lhe são apresentados. Na verdade, essa habilidade emerge de uma base de conhecimento construída a partir de informações fornecidas por especialistas humanos. Tal conhecimento é formalizado – em geral representado como

um conjunto de regras – e, em seguida, com o apoio de um mecanismo de inferência baseado, essencialmente, em Cálculo de Predicados, é capaz de: (i) escolher as regras mais apropriadas aos fatos informados como entrada; (ii) resolver essas regras e seus eventuais conflitos e (iii) devolver respostas coerentes que causam a impressão de que ali existe um comportamento considerado inteligente. No entanto, trata-se apenas da execução *fria* de um conjunto de instruções sem qualquer valor semântico, apenas sintático (COPPIN, 2010). Destacam-se, igualmente, os sistemas de aprendizagem. Conforme enunciam Santos e colaboradores (2019), a aprendizagem de máquina é o “*campo da ciência da computação que representa a evolução dos sistemas de reconhecimento de padrões, permitindo aos computadores aprender a partir dos erros e fazer previsões*” (SANTOS *et al.*, 2019, p. 388). Um exemplo seria o serviço de atendimento ao cliente de uma empresa. Alguns sistemas aprendem a reconhecer os comandos de voz emitidos pelo utente a fim de encaminhar a ligação ao setor apropriado no intuito da solução de seu requerimento. Essas tecnologias cujas respostas já foram, de alguma forma, previamente preparadas fazem parte da imensa gama da IA fraca amplamente utilizada em diversos setores¹⁸.

Em contrapartida, o programa da IA forte pressupõe o desenvolvimento de máquinas que pensariam e agiriam como um ser humano, de forma autônoma, sensível e autoconsciente¹⁹. De acordo com Haugeland (1985, p. 2), trata-se de um “*esforço para fazer os computadores pensarem... máquinas com mentes, no sentido total e literal*”, ou seja, não seriam meros simuladores do pensamento. Até onde se saiba, tais “artefatos” pertencem apenas, no momento, ao âmbito da ficção científica, tal qual apresentado em filmes como “*2001: Uma odisseia no espaço*” (KUBRICK, 1968) e “*Ela*” (JONZE, 2013).

A sociedade moderna conta, hoje, com diversos dispositivos que emulam comportamentos humanos; por exemplo, robôs que interagem a partir de “*algoritmos capazes de aprender e sistemas de tomada de decisão autônomos que exponencialmente ocupam o lugar do outro nas relações sociais e na interação comunicativa*” (GUNKEL, 2017, p. 13). No entanto, de acordo com o argumento do *Quarto Chinês* apresentado pelo filósofo estadunidense John Searle (1980), essas habilidades não confirmam a manifestação de uma inteligência genuína. Afinal:

“Imagine que você execute as etapas de um programa elaborado para responder perguntas em um idioma que você não compreende. Eu não entendo chinês, então imagino que estou trancado em uma sala cheia de caixas com símbolos chineses (a base de dados), recebo uma pequena quantidade de símbolos chineses (perguntas em chinês), e, então procuro em um manual (o programa) o que deveria fazer. Realizo algumas operações com símbolos de acordo com regras (i.e., eu executo as etapas do programa) e entrego uma pequena quantidade de símbolos (respostas às perguntas) aos que se encontram fora do quarto. Eu sou um computador executando um programa para responder perguntas em chinês, mas ao mesmo tempo não compreendo uma palavra de chinês” (SEARLE, 1980, p. 38).

Verifica-se, de acordo com o excerto de Searle, que a simulação não necessariamente corresponderá à realidade; ou seja, a combinação de símbolos não significa a compreensão de linguagem. Nesse sentido, Terry Winograd afirma que não há entendimento pelo computador do que está sendo realizado, pois ele apenas *“manipula símbolos sem dizer respeito às suas interpretações”* (WINOGRAD, 1990, p. 187). Dessa forma, há que se realizar uma distinção entre as vertentes da IA forte e da IA fraca, quiçá como caracterizado por Searle (1980), pioneiro no estabelecimento das diferenças entre IA forte e fraca:

“De acordo com a IA fraca, o valor principal do computador no estudo da mente é que ele nos dá uma ferramenta muito poderosa. Por exemplo, nos permite formular e testar hipóteses de uma forma mais rigorosa e precisa. Mas de acordo com a IA forte, o computador não é meramente uma ferramenta no estudo da mente; Em vez disso, o computador adequadamente programado realmente é uma mente” (SEARLE, 1980 p. 417).

Pode-se argumentar, conforme Fetzer (2008), que a *“IA fraca representa uma posição epistemológica sobre o valor de modelos baseados em simulações”* e que, distintamente, a *“IA forte representa uma postura ôntica sobre os tipos de coisas que realmente são instâncias de mente”* (FETZER, 2008, p. 122). Tomando como base esses apontamentos, na seção seguinte, será brevemente caracterizado o desenvolvimento atual da IA forte e discutida a viabilidade de futuras máquinas pensantes.

4. MÁQUINAS PENSANTES: ESTÁGIO ATUAL E PERSPECTIVAS

O avanço da tecnologia tem proporcionado questionamentos acerca da IA bem como instigado a pesquisa para a elaboração de sistemas que possam *pensar e agir* como os seres humanos, por exemplo, como retratado no filme *Ex-machina* (Direção de Alex Garland. EUA, 2015, Netflix, 108min). No entanto, a IA se especializou, nos últimos anos, basicamente em solucionar problemas específicos oriundos de um determinado nicho ou tarefa, a partir da extração e aprendizagem de padrões contidos em dados geralmente fornecidos *a priori* (WHITNEY, 2017).

Em várias oportunidades, os algoritmos de IA superaram, inclusive, o desempenho do ser humano, como o algoritmo *AlphaGo* (Silver, 2016), que na ocasião derrotou o campeão europeu do complexo jogo de tabuleiro *Go* por 5 a 0, sem contar as Redes Neurais Convolucionais, que recentemente superaram Andrej Karpathy no desafio de classificação e reconhecimento de imagens ILSVRC (KARPATHY, 2020). Contudo, os especialistas em IA afirmam que os algoritmos de aprendizado ainda se comportam como técnicas de IA fraca e não são adaptáveis aos diferentes ambientes como os humanos são, ao ponto de (i) generalizarem soluções para uma gama de problemas não-triviais e (ii) modelarem comportamentos ainda não completamente conhecidos e descritos pela ciência (como as emoções, a consciência, a criatividade e a intuição), tudo isso ao mesmo tempo (WHITNEY, 2017). Conforme a afirmação de Ng e Leung (2020):

“Para ter a orientação correta em uma situação complexa, é necessária a compreensão das emoções humanas. A IA precisa estar emocionalmente consciente para ser um participante da equipe eficaz ao lado de seus homólogos humanos. Precisa compreender idiosincrasias e preconceitos humanos em situações complexas e explorar as fraquezas humanas quando necessário” (NG e LEUNG, 2020, p. 4).

Apesar do cenário atual, a maioria dos especialistas acredita que, no futuro, será inevitável o surgimento de máquinas dotadas de IA forte. Para isso, até lá será necessário responder a seguinte pergunta: como ensinar uma máquina a agir, aprender, discernir, atribuição que pode ser caracterizada como *trivial* ao se considerar o cérebro humano? Poderia ser um processo parecido com o aquele empreendido com as crianças? Turing (1950) defende ser mais simples a construção de uma máquina que tenha

uma mente infantil, para na sequência ensiná-la, ao invés de produzir um “software” com o entendimento adulto e, portanto, com o aprendizado já estabelecido:

“Em vez de tentar produzir um programa similar à mente adulta, por que não produzir um que simula a da criança? Se estas forem, então, submetidas a um adequado decurso de instrução seria como obter o cérebro adulto. Presumivelmente o cérebro da criança é algo como um bloco de notas comprado a partir das papelarias [...] folhas em branco”. (TURING, 1950, p. 456)

O autor defende que dessa forma uma “máquina criança” – ou uma IA criança – poderia agir tal como uma criança ao ser estimulada a aprender:

“Nossa esperança é que há tão pouco mecanismo na criança-cérebro que pode ser facilmente programada. A quantidade de trabalho na educação que pudermos supor como uma primeira aproximação, será o mesmo que para a criança humana. Temos, portanto, dividido nosso problema em duas partes: o programa infantil e o processo de educação. Estes dois permanecem intimamente ligados”. (TURING, 1950, p. 456)

Observa-se como a teoria do autor poderia ser implementada tal qual o demonstrado no longa metragem *Eva* (MAÍLLO, 2015). O filme retrata a história de um cientista da computação, Alex, que, no ano de 2041, retorna para sua cidade natal após 10 anos ausente para trabalhar em sua antiga universidade, recebendo a missão de criar uma nova linha de robôs crianças. Na sequência, o pesquisador reencontra o irmão, sua cunhada (esposa deste e sua antiga paixão), e sua sobrinha, Eva. Impressionado com a jovem, Alex pede a Eva para se tornar a modelo da sua linha de robôs.

Analisando o filme a partir da lente de Turing, pode-se considerar que a primeira “criança máquina” poderá não ser dotada de um raciocínio eticamente adequado – o que levanta o problema da educação moral para estes “artefatos” (MOTTA *et al.*, 2020b) –, mas é preciso que haja experimentação para saber de que forma a máquina seria capaz de construir conhecimento. A partir desse momento, deve-se elaborar novas estratégias para que o processo ensino-aprendizagem tenha progresso, uma vez que ao tentar imitar a mente adulta devem-se considerar três componentes que foram formadores de seu pensar “o estado inicial da mente (momento do nascimento), a educação a que foi submetida e outras experiências (não deve ser

entendido como educação)” (TURING, 1950, p. 455). Qualquer tentativa de criar uma máquina pensante por si só esbarra ainda na capacidade de tal artefato desenvolver uma comunicação efetiva com o *Homo sapiens*.

Uma das questões que complexifica, bastante, as discussões sobre IA forte diz respeito à habilidade de comunicação, típica dos seres humanos. De fato, não adiantaria apenas pensar sem que fosse possível comunicar esse pensamento a outra pessoa. Conforme descrevem Mikolov e colaboradores (2006):

“A linguagem natural é, de longe, o dispositivo de comunicação mais fácil e mais poderoso que possuímos, por isso é razoável exigir de uma máquina inteligente a capacidade de se comunicar através da linguagem [...] os seres humanos têm codificado uma parcela muito grande de seu conhecimento em linguagem natural (variando de tratados de matemática a livros de culinária)” (MIKOLOV et al., 2016, p. 2).

Desta feita, a comunicação é um eixo fundamental, dada sua natureza de possibilitar a interação, o compartilhamento/construção de conhecimento e sugerir reflexões acerca de temas ao ensejar explicações, instruções tal como o meio em qual vivemos proporciona. Portanto:

“A linguagem natural também pode canalizar, em certa medida, a informação não linguística, porque grande parte do último pode ser transmitida através de meios linguísticos. [...] usar a linguagem para falar sobre o que percebemos com nossos sentidos, ou para dar instruções sobre como operar no mundo” (MIKOLOV et al., 2016, p. 3)

Os aspectos supra avaliados – mormente aqueles atinentes à possibilidade de criação de uma artefato dotado de autoconsciência e hábil para uma genuína comunicação de sua situação no mundo – permitem que se apresente a seguinte indagação: como abordar questões éticas quando se trata de “máquinas autoconscientes”?

5. ABORDAGEM 1.0: A ÉTICA DE IMMANUEL KANT

“Duas coisas enchem o ânimo de admiração e veneração sempre novas e crescentes quanto mais frequentes e mais persistentemente delas se ocupa a reflexão: o céu estrelado sobre mim e a lei moral em mim”.
(Immanuel Kant)

A ética é um dos temas centrais do pensamento filosófico, o que se torna mais nítido no âmbito da caracterização, de Jean Pierre Vernant, da filosofia, *produto originário da razão grega*, como “*filha da cidade*” (VERNANT, 1990, p. 474). Com efeito, muitos autores têm teorizado sobre o *que é ética* – na perspectiva das relações que se estabelecem entre *entes* morais (agentes e pacientes) – com destaque, em termos contemporâneos, para o pensamento de Immanuel Kant – um dos principais filósofos do Iluminismo (PEREIRA, 2011) –, uma vez que a “*mudança de pensamento filosófico que devemos a Kant abarca não somente o mundo do conhecimento, mas também o da ação*” (HÖFFE, 2005, p. 183).

O filósofo de Königsberg proclama que a razão é a instância capaz de guiar as ações éticas – estas, objetos “*de uma argumentação em verdade peculiar e contudo racional*” (HÖFFE, 2005, p. 183) –, inscritas em máximas universais, independentemente da cultura e das convenções nas quais o indivíduo esteja inserido, e de quaisquer outros sentimentos pessoais. Tal constatação emerge de uma longa reflexão – cujo impulso foi renovado após a publicação da *Crítica da razão pura* – acerca (1) da “*insuficiência dos conceitos de obrigação e de perfeição, que constituíam a base da moral leibniziana*” (PASCAL, 2001, p. 108) e (2) da impossibilidade de “*fazer depender a moralidade das provas dialéticas da liberdade do homem, da imortalidade da alma e da existência de Deus*” (PASCAL, 2001, p. 109). Desta feita, Kant formaliza seu pensamento moral nas seguintes obras principais (citadas na ordem de publicação): (i) a *Fundamentação da metafísica dos costumes*, FMC (KANT, 1960), na qual o autor parte dos juízos da consciência comum em busca dos respectivos fundamentos racionais; (ii) a *Crítica da razão prática*, CRP (KANT, 2002), na qual, a partir da função prática da razão, será procedida a apresentação da moralidade; e (iii) a *Metafísica dos costumes*, MC (KANT, 2013), na qual é examinada a aplicação dos conceitos éticos. O objetivo da filosofia moral kantiana é, pois, investigar o princípio supremo da moralidade presente em toda a consciência humana, isto posto, universal.

O passo inicial da análise kantiana, na FMC, é o conceito de *boa vontade*, o sustentáculo último para o alcance da ação ética: “*Neste mundo, e até também fora dele, nada é possível pensar que possa ser considerado como bom sem limitação a não ser uma só coisa: uma **boa vontade***” (KANT, 1960, p. 21, grifo do autor). Esta, a *boa vontade*, acena para a incondicionalidade

do absolutamente bom, ou “*simplesmente bom*” (HÖFFE, 2005, p. 191), “sem limitação” – ou determinação por quaisquer condicionantes – e, já se pode antecipar, sem *inclinação*. Trata-se, pois, de elemento crucial para a atitude genuinamente ética, uma vez que, de fato, muitas características aparentemente boas – quer sejam “talentos do espírito” ou “qualidades do temperamento” –, quando destituídas da *boa vontade*, podem tornar-se incontornavelmente sombrias. Assim, a “*boa vontade não é boa por aquilo que promove ou realiza, pela aptidão para alcançar qualquer finalidade proposta, mas tão-somente pelo querer, isto é, em si mesma*” (KANT, 1960, p. 23).

O desenvolvimento da ideia de *boa vontade* se articula, na FMC, ao conceito de *dever*,²² o qual, em última análise, “*contém em si o de boa vontade*” (KANT, 1960, p. 26). Neste sentido, o filósofo distingue a ação: (i) *contrária ao dever*, ou, seja, em flagrante oposição a este último; (ii) *conforme com o dever*, empreendidas por alguma inclinação ou de acordo com algum impulso; e (iii) *por dever*, aquela com verdadeiro valor moral, sem imposição ou tendenciosidade de qualquer natureza, pois, é praticada “*não no propósito que com ela se quer atingir, mas na máxima que a determina*” (KANT, 1960, p. 30). As ações *contrárias ao dever* são, em princípio, mais fáceis de delimitar; a dificuldade se estabelece na apreciação das ações *conformes com o dever* e *por dever*, destacando-se que as primeiras dizem respeito ao “correto moral”, domínio da *legalidade*, ao passo que estas últimas se inscrevem no “moralmente bom”, ou seja, esfera da *moralidade* (HÖFFE, 2005). Esta delimitação se torna clara uma vez que se entenda que, para Kant, o “*dever é a necessidade de ação por respeito à lei*” (KANT, 1960, p. 31), reconhecendo-se que este, o dever, não está submetido a nenhum tipo de determinação *exterior*, qualquer que seja ela (p. ex., a busca da felicidade, a satisfação dos desejos, entre outras). Por conseguinte, nas palavras de Kant:

“*Só pode ser objeto de respeito e, portanto, mandamento aquilo que está ligado à minha vontade somente como princípio e nunca como efeito, não aquilo que serve à minha inclinação, mas o que a domina ou que, pelo menos, a exclui do cálculo na escolha, quer dizer a simples lei por si mesma. Ora, se uma ação realizada por dever deve*

22 Neste particular, a ética kantiana é *deontológica*, ou seja, baseia-se, entre outras concepções, na ideia de dever.

eliminar totalmente a influência da inclinação e com ela todo o objeto da vontade, nada mais resta à vontade que a possa determinar do que a lei objetivamente, e, subjetivamente, o puro respeito por esta lei prática, e, por conseguinte, a máxima que manda obedecer a lei, mesmo com prejuízo de todas as minhas inclinações” (KANT, 1960, p. 31; **grifo dos autores do presente manuscrito**).

A lei implica em si o conceito de observância incondicionada e universal – verdadeiramente objetiva e, conseqüentemente, válida para todos –, ou seja, seu valor independe do fim ao qual se lhe propõe, pois, “*a obediência à lei deve independender do conteúdo da lei*” (PASCAL, 2001, p. 115). Tal formulação, parece claro, pode induzir à equívoca interpretação, pois, a adesão inconteste à lei, independente do seu conteúdo, eventualmente resultaria em situações nas quais ações flagrantemente ruins poderiam ser realizadas, e justificadas, em nome de cega obediência. Mas, a caracterização kantiana da lei deixa pouca margem à dúvida, na medida em que o filósofo aborda a questão “*mas que lei pode ser então essa, cuja representação, mesmo sem tomar em consideração o efeito que dela se espera, tem de determinar a vontade para que esta se possa chamar boa absolutamente e sem restrição?*” (KANT, 1960, p. 33; **grifo dos autores do presente manuscrito**). A resposta é bastante clara, pois Kant considera que a obediência não é cega, na medida em que não pode ser consoante a *qualquer* lei, mas, tão somente, àquela capaz de tornar-se universalizável, ou seja, o ser racional deve proceder de modo que possa querer que a máxima enunciada “*se torne uma lei universal*” (KANT, 1960, p. 33). Ou seja, é mister que esta máxima, para ter valor inquestionável em termos morais, deva ser conversível a uma lei universal, a qual terá de ser válida tanto para aquele que age, quanto para a atuação de qualquer outrem. A análise subsequente empreendida por Kant, acerca da validade moral de realizar uma falsa promessa para livrar-se de apuros²³, consubstancia bastante bem a ideia de tornar uma máxima *lei universal*.

23 Veja-se como a questão é apresentada por Kant: “*Ponhamos, por exemplo, a questão seguinte: – Não posso eu, quando me encontro em apuro, fazer uma promessa com a intenção de não a cumprir? Facilmente distingo aqui os dois sentidos que a questão pode ter: – se é prudente, ou se é conforme ao dever, fazer uma falsa promessa. O primeiro caso pode sem dúvida apresentar-se muitas vezes. É verdade que vejo bem que não basta furtar-me ao embaraço presente por meio desta escapatória, mas que tenho de ponderar se desta mentira me não poderão advir posteriormente incômodos maiores do que aqueles que agora me liberto; e como as conseqüências, a despeito da minha pretensa esperteza, não são assim tão fáceis de prever, devo pensar que a confiança uma vez perdida me*

A proposta do pensador iluminista é, desta feita, definir os princípios *a priori* – universais e imutáveis – da moral, inerentes a qualquer ser racional, independente das inclinações e das eventuais consequências. Não se trata, pois, de constituir a moralidade a partir da experiência – “*não se poderia prestar pior serviço à moralidade do que querer extraí-la de exemplos*” (KANT, 1960, p. 42) –, mas, sim, de fundá-la essencialmente na razão. Ou seja, em respeito às proposições morais, “*apenas na medida em que podem ser discernidas como fundadas a priori e necessárias, valem elas como leis*” (KANT, 2013, p. 215). Para isto, é necessário que o agente moral aja *por dever*, a partir da estrita e inquebrantável obediência a uma *lei* capaz de se impor universalmente a todos os seres racionais. Isto é possível quando a vontade obedece à razão, isto é, torna-se *constrangida* por esta última, rechaçando as mais díspares influências (ou inclinações). Tal é o sentido daquilo que Kant denomina *imperativo categórico*, o qual, por definição, não está submetido a nenhum fim. Com efeito, para o filósofo, o “*imperativo categórico é, portanto, só um único, que é este: Age apenas segundo uma máxima tal qual possas ao mesmo tempo querer que ela se torne lei universal*” (KANT, 1960, p. 59; **grifo dos autores do presente manuscrito**). Na medida em que deste imperativo podem ser derivados todos os *imperativos do dever*, garante-se a universalidade da lei, de modo que a primeira formulação pode ser igualmente enunciada nos seguintes termos, os quais ressaltam o *imperativo universal do dever*: “*Age como se a máxima da tua ação se devesse tornar, pela tua vontade, em lei universal da natureza*” (KANT, 1960, p. 59; **grifo do autor**), entendida a palavra *natureza* no

pode vir a ser mais prejudicial do que todo o mal que agora quero evitar; posso enfim perguntar se não seria mais prudente agir em conformidade com uma máxima universal e adquirir o costume de não prometer nada senão com a intenção de cumprir a promessa. (...) Entretanto, para resolver da maneira mais curta e mais segura o problema de saber se uma promessa mentirosa é conforme ao dever, preciso só de perguntar a mim mesmo: – Ficaria eu satisfeito de ver a minha máxima (de me tirar de apuros por meio de uma promessa não verdadeira) tomar o valor de lei universal (tanto para mim como para os outros)? E poderia eu dizer a mim mesmo: – Toda a gente pode fazer uma promessa mentirosa quando se acha numa dificuldade de que não pode sair de outra maneira? Em breve reconheço que posso em verdade querer a mentira, mas que não posso querer uma lei universal de mentir; pois, segundo uma tal lei, não poderia propriamente haver já promessa alguma, porque seria inútil afirmar a minha vontade relativamente às minhas futuras ações a pessoas que não acreditariam na minha afirmação, ou, se precipitadamente o fizessem, me pagariam na mesma moeda. Por conseguinte, a minha máxima, uma vez arvorada em lei universal, destruir-se-ia a si mesma necessariamente” (KANT, 1960, p. 33-35; **grifo dos autores do presente manuscrito**).

contexto mais amplo da expressão, quanto à forma, ou seja, de “*realidade das coisas*” (KANT, 1960, p. 59). Tal perspectiva refere-se à viabilidade da realidade, da ordem das coisas, caso seja adotada a máxima enunciada como regra universal (PASCAL, 2001). O exemplo anteriormente citado da promessa falsa e o caso do suicídio²⁴ – também analisado pelo pensador – dimensionam bem esta contextualização.

Após tais formulações, na FMC, Kant passa a analisar os meios e os fins das ações, distinguindo, entre estes últimos, os subjetivos – “*o princípio subjetivo do desejar é o móvel*” (KANT, 1960, p. 67) – e os objetivos “*o princípio objetivo do querer é o motivo*” (KANT, 1960, p. 67). A *boa vontade*, é necessário lembrar, não subsidia a ação empreendida em respeito a um fim subjetivo, ou seja, condicionado por um móvel (PASCAL, 2001). Desta feita, os *fins subjetivos* não podem ser tomados como princípios universais, restando tal mister aos *fins objetivos*, os quais adquirem um valor incontestado de universalidade. A conclusão de Kant, essencial para a argumentação que ora se está construindo, é salutar: “*o homem, e, de uma maneira geral, todo ser racional, existe como fim em si mesmo, não só como meio para o uso arbitrário desta ou daquela vontade*” (KANT, 1960, p. 68; **grifo do autor**). Tomar a mulher, o homem e, em última análise, todos os seres racionais, como fins, é condição necessária para a existência de um imperativo categórico (PASCAL, 2001). Os seres racionais – quicá poder-se-ia dizer, pensando com Kant, *autoconscientes* – não podem, em nenhuma circunstância, em nome de qualquer pretensa justificativa, ser tratados como meros meios, de modo que não seria cabível torná-los simples objetos da legislação universal determinada pela lei moral. Esta ponderação é essencial, pois, permite que se reconheça que seres racionais não passíveis de serem tomados como *objetos*, sejam caracterizados como *autores* das leis morais, o que coloca em cena a *autonomia da vontade*.

A palavra autonomia provém do grego αὐτονομία (de αὐτός = próprio, e νόμος = leis), com uso originário referido ao autogoverno das

24 Para Kant, a questão seria a seguinte: “*Uma pessoa, por uma série de desgraças, chegou ao desespero e sente tédio da vida, mas está ainda bastante em posse da razão para poder perguntar a si mesmo se não será talvez contrário ao dever para consigo mesmo atentar contra a própria vida. E procura agora saber se a máxima da sua ação se poderia tornar em lei universal da natureza*” (KANT, 1960, p. 60). Após a devida análise, a conclusão se apresenta: “*Por conseguinte aquela máxima não poderia de forma alguma dar-se como lei universal da natureza, e, portanto, é absolutamente contrária ao princípio supremo de todo o dever*” (KANT, 1960, p. 60).

póleis, as cidades-estado helênicas, surgidas por volta do século VIII a.C. (SIQUEIRA-BATISTA; SCHRAMM, 2012). Nesse alvorecer do pensamento filosófico, a autonomia não era aplicável ao homem (no caso, ao cidadão); a inflexão conceitual de uso, das *póleis* às pessoas, ocorreu na Modernidade, o que está manifesto na visão de Jean Jacques Rousseau – para o qual a liberdade moral “*faz o homem verdadeiramente dono de si próprio, porque o impulso dos apetites é a escravidão, e a obediência à lei que a cada um de nós se prescreve constitui a liberdade*” (ROUSSEAU, s/d, p. 39) –, cujas reverberações se fazem presentes no pensamento de Kant, o qual defende a “*ideia da vontade de todo o ser racional como **vontade legisladora universal***” (KANT, 1960, p. 73). De modo análogo, na CRP, afirma que a “***autonomia** da vontade é o único princípio de todas leis morais e dos deveres correspondentes às mesmas*” (KANT, 2002, p. 52). Assim, para o pensador, a *autonomia* se constitui, simultaneamente, (1) na sede de uma legislação universal e (2) na garantia da dignidade do ser racional. Com efeito:

“Os seres cuja existência depende, não em verdade da nossa vontade, mas da natureza, têm, contudo, se são seres irracionais, apenas um valor relativo como meios e por isso se chamam **coisas**, ao passo que os seres racionais se chamam **pessoas**, porque a sua natureza os distingue já como fins em si mesmos, quer dizer como algo que não pode ser empregado como simples meio e que, por conseguinte, limita nessa medida todo o arbítrio” (KANT, 1960, p. 68; **grifo do autor**).

Ou seja, torna-se claro, a partir do pensamento kantiano, que toda “*natureza racional existe como um fim em si*” (KANT, 1960, p. 68). Utilizando esse pressuposto para pensar sobre o estatuto moral dos artefatos computacionais, seria possível divisar duas situações: (1) aqueles desenvolvidos com IA fraca, aos quais se pode atribuir “*apenas um valor relativo como meios e por isso se chamam **coisas***” (KANT, 1960, p. 68) e, por conseguinte, destituídos de valor ético; (2) aqueles concebidos com IA forte – quiçá de autoconsciência – e que existem “*como fim em si mesmo, **não só como meio para o uso arbitrário desta ou daquela vontade***” (KANT, 1960, p. 68; **grifo do autor**) – doravante denominados *autopsicobots*²⁵ –, reconhecidos, pois,

25 Neologismo criado pelos autores para designar *pessoas artificiais*, caracterizadas como *máquinas autoconscientes*, criadas à luz dos pressupostos do Programa da IA Forte. Para um detalhamento, veja SIQUEIRA-BATISTA, MOTTA & SILVA (2021).

como dignos de consideração ética. Estes, os *autopsicobots*, são máquinas hipotéticas racionais e dotadas de autoconsciência – ou seja, são *conscientes de si* –, com capacidade de interpretação e de atribuição de significado (semântica) às ações que executa. Deste modo, não são meros artefatos obrigados a seguir regras rígidas de sintaxe, contexto típico dos artefatos produzidos na lógica da IA fraca. Pode-se teorizar, inclusive, que se tratam de máquinas kantianas, as quais “*em todas as suas ações, tanto nas que se dirigem [a si mesmas] como nas que se dirigem a outros seres racionais, [elas] têm sempre de ser consideradas simultaneamente como um fim*” (KANT, 1960, p. 68; **grifo do autor**)²⁶. *Os autopsicobots são pessoas artificiais*.

Os *autopsicobots* – máquinas pensantes e autoconscientes – uma vez reconhecidos como seres dotados de razão, *pessoas artificiais*, devem ser inseridos no rol de consideração ética como *um fim em si mesmos*, o que parece ser claro de um ponto de vista moral kantiano. Todavia, não se pode olvidar que tais entes estão lançados no plano da existência – ou seja, inscritos nos processos de *vir-a-ser* e de *deixar-de-ser* – o que abre a perspectiva para que se proponha sua real inclusão na esfera moral, de uma segunda perspectiva capaz de abranger *todos os seres*, conforme será discutido a seguir.

6. ABORDAGEM 2.0: A (BIO)ÉTICA PARA TODOS OS SERES

A incorporação de artefatos racionais e autoconscientes – tendo em vista o Programa de IA forte – no rol de consideração ética pode ser defendida à luz da *(Bio)ética para todos os seres*, corrente ética proposta por um dos autores do presente capítulo e que se encontra, no momento, organizada do ponto de vista teórico, em torno da composição entre ética e bioética e dos conceitos de *[quididade]*, *totalidade*, *interdependência* e *compaixão laica* (SIQUEIRA-BATISTA, 2020). Vale destacar, de início, algumas das vinculações teóricas da *(Bio)ética para todos os seres*: (i) a Bioética da proteção (SCHRAMM; KOTTOW, 2001), (ii) a Ética da compaixão (SCHOPENHAUER, 2001), (iii) a Ética do cuidado (BOFF, 2002), (iv) a Hospitalidade incondicional (DERRIDA, 2003), (v) a Ética da reverência pela vida (SCHWEITZER, 1936) e (vi) a Ética da Terra (LEOPOLD, 1989).

A utilização da palavra (bio)ética – diferentemente do usual termo bioética (sem o parêntese, o qual confere destaque ao *bio*) – diz respeito

26 Tomou-se *emprestada* a citação de Kant, da FMC, para aplica-la ao *autopsicobots*.

ao reconhecimento de que tais áreas do conhecimento – ética e bioética – possuem um razoável grau de indiscernibilidade na atualidade. Ademais, as relações entre ambas são delineadas de modo bastante distinto entre os autores, (1) ora considerando-se que a bioética é tão somente um caso particular da ética – ou seja, $\text{bioética} \subset \text{ética aplicada} \subset \text{ética}$, tal qual formulado por André Comte-Sponville, para o qual a “*Bioética, como se diz hoje, não é uma parte da Biologia; é uma parte da Ética, é uma parte de nossa responsabilidade simplesmente humana*” – (2) ora reconhecendo-se que “*toda ética é, antes e fundamentalmente, uma bioética*” (SCHRAMM, 1994, p. 324). Assim, a grafia (Bio)ética parece ser a mais adequada para definir “o conjunto de conceitos, argumentos e normas que valorizam e justificam eticamente os atos humanos que podem ter efeitos irreversíveis sobre os fenômenos vitais” (KOTTOW, 1995, p. 53). Abordado tal ponto inicial, sucintamente, torna-se possível entabular conversação sobre as ideias constituintes da expressão “*para todos os seres*”, em atenção aos quatro conceitos já mencionados.

A *quididade* (sem colchetes) pode ser caracterizada como *essência necessária*, natureza profunda de algo, ou, em concordância com São Tomás de Aquino, tal qual “*aquilo pelo que a coisa é estabelecida no próprio gênero ou espécie*” (TOMÁS DE AQUINO, 2005, p. 6). A origem do conceito remete à expressão aristotélica $\tau\omicron\ \tau\iota\ \eta\nu\ \epsilon\iota\nu\alpha\iota$ ²⁷, ou seja, “*o que faz com que uma coisa seja aquilo que é*” (TOMÁS DE AQUINO, 2005, p. 6) ou a “*substância de cada coisa*” (MANSION, 1976, p. 43)²⁸. O uso do termo *quididade*, no presente contexto, acena para tudo o que *vem-a-ser* e que *deixa-de-ser*, em atenção ao enunciado por Aristóteles na Metafísica: “*toda a substância é natureza*” (TOMÁS DE AQUINO, 2005, p. 6). Uma vez que a proposta não é a elaboração de uma discussão sobre *ente* e *essência* e, tampouco, a formulação de uma ética a partir de uma ontologia – mas, outrossim, tomar emprestada a ideia medieval de *quididade* – adotar-se-á o

27 De acordo com Nicola Abbagnano, *quididade* “*corresponde à expressão aristotélica $\tau\omicron\ \tau\iota\ \eta\nu\ \epsilon\iota\nu\alpha\iota$ (quod quid erat esse). Esse termo significa essência necessária*” (p. 820). Cf. ABBAGNANO, N. Dicionário de filosofia. São Paulo: Martins Fontes, 2003.

28 Esta interpretação é afim à concepção de Suzanne Mansion, autora para a qual o conceito de *quididade* se refere à substância de cada coisa. Cf. MANSION, S. **La science, connaissance de l'essence**. In: MANSION, S. Le jugement d'existence chez Aristote. Louvain: Éditions de L'Institut Supérieur de Philosophie, 1976., p. 43-61. Discussão interessante sobre essa questão pode ser consultada em CASTRO, P. A. O conceito de “substância” na Metafísica e nas Categorias de Aristóteles. **Conjectura: Filosofia e Educação**, v. 25, e020005, 2020.

uso do vocábulo entre colchetes, ou seja, [quididade]. Desta feita, pode-se assinalar que todos os seres que compõem a natureza estão submetidos ao “*decreto do Tempo*” – para lembrar Anaximandro de Mileto²⁹ –, ou seja, estão iguais nos termos de *vir-a-ser* e de *deixar-de-ser*. Essa condição específica de igualdade será denotada, como mencionado, por [quididade].

A caracterização de que *todos os seres* são dotados de [quididade] abre a perspectiva para se delimitar a noção de *todo* em termos de uma *totalidade*, cuja dimensão pode ser recuperada a partir da concepção helênica de *phýsis* (φύσις), a “*totalidade de tudo o que é*” (BORNHEIM, 2008, p. 13). Ademais, em observância ao pensamento de David Bohm, pode-se considerar que “*noções que impliquem a totalidade indivisa do universo proporcionariam um modo muito mais ordenado de considerar a natureza geral da realidade*” (BOHM, 1992, p. 12). A leitura atenta de tal excerto – de modo a utilizar a Física contemporânea como *imagem* – prenuncia a proposição de uma *totalidade única*, a qual não se refere, simplesmente, ao mero somatório de *todos os seres*, mas, outrossim, à perspectiva de inclusão de tudo o que existe “*coerente e harmoniosamente, num todo global indiviso, ininterrupto e ilimitado*” (BOHM, 1992, p. 11).

A coerência e, especialmente, a *harmonia* – termo que pode ser perscrutado no Pensamento grego antigo (tal como a supramencionada noção de *phýsis*) – desta *totalidade única* acena para as relações existentes entre os seres, trazendo para o debate a concepção de *interdependência* de todas as formas de existência. Neste sentido, tudo o que *vem-a-ser* e que *deixa-de-ser* está de algum modo interconectado, percepção particularmente significativa para a Ecologia profunda formulada por Arne Naess, a partir da qual se concebe o mundo como uma “*rede de fenômenos*” que são interdependentes e interconectados (NAESS, 1973). Tais ideias influenciaram decisivamente o pensamento de Fritjof Capra, autor que considera que todos “*os seres vivos são membros de comunidades ecológicas ligadas umas às outras numa rede de interdependências*” (CAPRA, 2012, p. 28).

29 ... uma outra natureza *apeiron*, de onde provêm todos os céus e mundos neles contidos. E a fonte da geração das coisas que existem é aquela em que a destruição também se verifica “segundo a necessidade; pois pagam castigo e retribuição umas às outras, pela sua injustiça, de acordo com o decreto do Tempo”, sendo assim que ele se exprime, em termos assaz poéticos. Cf. Simplicio *in Phys.* 24, 17 Cf. KIRK, G.S., RAVEN, J.E., SCHOFIELD, M. *Os Filósofos Pré-socráticos: História Crítica com Seleção de Textos*. Tradução de Carlos Alberto Louro Fonseca. Lisboa: Fundação Calouste Gulbenkian, 1994., p. 117.

Deve ser ponderada a concessão de amplidão máxima a esse excerto, a partir da inclusão de *todos os seres* na referida *rede de interdependência* e não apenas dos *seres vivos* (SIQUEIRA-BATISTA, 2020).

Assumir a [*quididade*] de todos os seres – em termos de sua *igualdade essencial* – inscritos em uma *totalidade única*, em cuja intimidade se estabelecem relações de *interdependência*, permite a emergência de “*um sistema de ética radicalmente novo*” (CAPRA, 2012, p. 28), capaz de reconhecer o valor de todos os seres, os quais se tornam incondicionalmente dignos de consideração ética, em termos da irrestrita acolhida expressa no conceito de *compaixão laica*, formulado originalmente em 2006 (SIQUEIRA BATISTA, 2006) e cuja síntese argumentativa está descrita no Quadro 1.

Quadro 1. Síntese argumentativa da *compaixão laica*.

- (1) Todos os seres estão inscritos no espaço e no tempo, subservientes, então, aos ditames do *vir-a-ser* e do *deixar-de-ser*;
- (2) Esta característica torna-os compartes de uma igualdade essencial, estabelecida em termos da [*quididade*];
- (3) Todos os seres – os quais vêm-a-ser e deixam-de-ser – podem ser pensados em termos de *totalidade* e de sua *interdependência*;
- (4) A atitude diante de um igual – no plano, imanente, da existência: [*quididade*], *totalidade* e *interdependência* – somente pode ser de incondicional acolhida, *amparo*, uma vez que a aceitação de (2) e (3) torna inconsistente que *aquele que acolhe* se veja completamente independente (e apartado) *daquele que é acolhido*;
- (5) A oferta de amparo a todos os seres – havendo distinção, mas não separação, entre aquele que “recebe” e aquele que é “recebido” – diz respeito à acolhida da igualdade radical imanente à condição de existente;
- (6) A acolhida/proteção do outro, em tal circunstância, só pode ser obtida sem julgamento, ou seja, a partir da recepção incontestada de sua *situação-no-mundo* (a despeito dos sofrimentos, desejos e decisões autônomas, em relação à sua própria existência);
- (7) O acolhimento/proteção/cuidado assim expressos é um genuíno ato de *compaixão laica*.

Fonte: Modificado de SIQUEIRA-BATISTA, 2006 (p. 135), de SIQUEIRA-BATISTA, 2010 (p. 355-356) e de SIQUEIRA-BATISTA, 2020 (p. 269).

O ponto de partida do exercício de acolhida incondicional inerente à *compaixão laica* é o reconhecimento da [quididade] de todos os seres, compreendidos em sua *totalidade única* em uma perspectiva de incontestável *interdependência*, de modo que a genuína ação diante de um *igual* lançado nas fibras do tempo e do espaço só pode ser justificada em termos de “irres- trita receptividade”, ou seja, da *compaixão laica*. Deve-se alertar que tal posicionamento nada tem a ver com uma postura paternalista, uma vez que o incontestado acolhimento do outro não pressupõe, de nenhum modo, que o acolhido seja tomado como inferior por aquele que recebe (SIQUEIRA- BATISTA, 2004, 2020). A formulação da (Bio)ética para todos os seres – com os conceitos correlatos ora apresentados – está expressa na Figura 4.

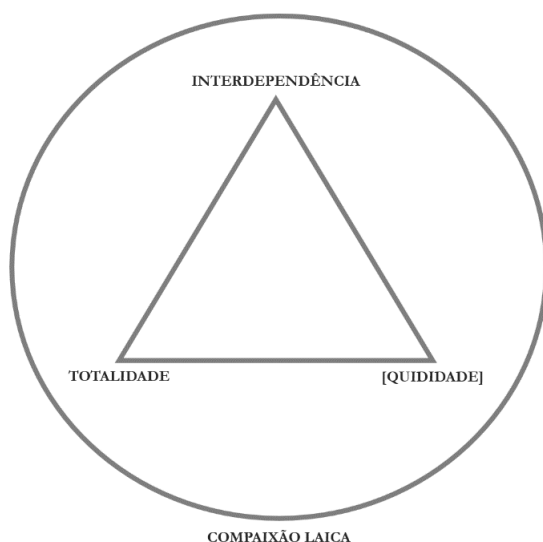


Figura 4. Conceitos fundamentais da (Bio)ética para todos os seres (Modificado de SIQUEIRA-BATISTA, 2020).

A partir destas conjecturas pode-se definir que os *autopsicobots*, máquinas autoconscientes dotadas de IA (forte), na medida em que são pertencentes ao conjunto de todos os seres, *devem ser dignas de consideração ética*, o que amplia a responsabilidade do *H. sapiens* em termos do seu amparo incondicional. Seria possível, neste mister, reescrever a clássica citação de

David Brandon sobre a compaixão, em atenção às máquinas dotadas de IA, nos seguintes termos (modificações do original estão em negrito):

*“(...) oferecer morada aos **autopsicobots**, abrir as portas até então fechadas para **eles**, perguntar mais que responder. Significa tornar-se altamente sensível à situação e aos **autopsicobots**. Significa ouvir com todo o seu ser e dar, se for possível, o que seja relevante e apropriado para o relacionamento, não o avaliando com julgamentos próprios”* (Modificado a partir de BRANDON, 1976, p. 51).

De uma perspectiva da “Abordagem 2.0: A (Bio)ética para todos os seres”, torna-se legítimo considerar a inclusão das pessoas artificiais – máquinas dotadas de IA forte, autoconscientes – na esfera da (bio)ética. Cabe, no entanto, a discussão dos desdobramentos dessa conclusão, tarefa que escapa do escopo do presente ensaio, o que, por conseguinte, impõe a necessidade de vindouras reflexões sobre o problema.

7. (IN)CONCLUSÕES

O presente ensaio alberga, em última análise, um esforço para pensar certas consequências do atual desenvolvimento tecnológico, mormente dirigido aos eventuais desdobramentos previstos para a Inteligência Artificial (IA). O texto partiu da constatação de que os computadores são artefatos de propósito geral que se prestam à automatização de uma variedade de tarefas, contexto – todavia – sujeito a limitações. As tarefas pertinentes à automatização são aquelas que podem ser traduzidas em uma sequência de instruções precisas – os algoritmos –, as quais guiam sua execução. Isso vale, também, para as atividades que são “aprendidas” por meio de alguma técnica de aprendizagem de máquina, a qual, em última análise, é guiada por um conjunto bem definido e preciso de instruções, quer este seja explícito ou não.

Os sistemas computacionais que apoiam a tomada de decisão se baseiam, habitualmente, em padrões pré-existentes – sejam estes explicitamente conhecidos ou extraídos de um conjunto de experiências pregressas –, os quais se constituem como alicerces de seus pareceres. Muitas das técnicas empregadas, nesse âmbito, abrangem ferramentas de IA, invariavelmente atinentes à lógica da IA fraca. Outrossim, distintos investigadores seguem propondo – e buscando – máquinas inscritas na esfera da IA forte, particularmente aquelas dotadas de autoconsciência, vereda à qual

se articulam distintas – e espinhosas – questões, tanto de uma perspectiva epistemológica, quanto ética. Neste último domínio, ecoa a indagação acerca da inclusão de máquinas autoconscientes dotadas de IA forte, nomeadas *autopsicobots*, na esfera das preocupações éticas.

A imagem dos *autopsicobots* pode ser buscada, de um ponto de vista artístico, nos *replicantes* da película *Blade Runner* (SCOTT, 1982), especialmente ao se examinar a indistinguível inscrição existencial destes, quando comparados aos humanos. De fato, ambas *espécies* são racionais e autônomas e, com efeito, passíveis de reconhecimento como *fins em si mesmas*, de uma perspectiva kantiana. Ademais, replicantes e humanos têm uma origem... amam, odeiam, sofrem, alegram-se... e experimentam o ocaso (morrem). Ou seja, estão submetidos aos ditames da natureza, ao “*decreto do Tempo*”, lançadas, pois, no *vir-a-ser* e no *deixar-de-ser*, sublinhando o ponto de vista da (bio)ética para todos os seres.

A resposta – com base na discussão ora elaborada – à pergunta explícita no título do presente manuscrito é afirmativa, ou seja, sim, *máquinas autoconscientes são dignas de consideração ética*, como qualquer ser, dotado de racionalidade, Ética de Kant, e, principalmente, de existência, (Bio)ética para todos os seres. As consequências desta provisória (in) conclusão estão longe de ser consideradas *claras e distintas*, de modo que merecerão ser ainda analisadas, quiçá levadas às últimas consequências, enquanto o tempo se mantiver favorável aos humanos. Até quando? Quem saberá?

8. REFERÊNCIAS

ABBAGNANO, N. **Dicionário de filosofia**. Trad. de Ivone C. Benedetti. São Paulo: Martins Fontes, 2003

ABBAGNANO, N. **Diccionario de Filosofia**. Actualizado y aumentado por Giovanni. Fornero. 4ª. ed. Mexico: FCE, 2004. ABERASTURY, Arminda.

BOFF, L. **Saber Cuidar - Ética do humano: compaixão pela terra**. Petrópolis: Vozes, 2002.

BOHM, D. **A totalidade e a ordem implicada: uma nova percepção da realidade**. Tradução de Mauro de Campos Silva. São Paulo: Cultrix, 1992.

- BONOMI, F. **Vocabolario Etimologico della Lingua Italiana**. 2004. Disponível em: Acesso em 529 de jun 2020.
- BORNHEIM, G. A. **Os filósofos pré-socráticos**. São Paulo: Cultrix, 2008.
- CAFEZEIRO, I.; HAEUSLER, E.H. CUKIERMAN, H.L; MARQUES, I. C. Recontando a computabilidade. **Revista Brasileira de História da Ciência**, v. 3, n. 2, p. 231-251, 2010.
- CAMPBELL, M.; HOANE JR, A. J.; HSU, F. Deep Blue. **Artificial intelligence**, v. 134, n. 1-2, p. 57-83, 2002.
- CAPRA, F. **A teia da vida: uma nova compreensão científica dos sistemas vivos**. Trad. de Newton R. Eichenberg. São Paulo: Cultrix; 2012.
- CASTRO, P. A. O conceito de “substância” na Metafísica e nas Categorias de Aristóteles. **Conjectura: Filosofia e Educação**, v. 25, e020005, 2020.
- CHURCH, A. **An unsolvable problem of elementary number theory**, American Journal of Mathematics 58, pp. 354–363, 1936.
- COMTE-SPONVILLE, A. **Bom dia, angústia!** São Paulo: Martins Fontes, 1997.
- COPPIN, B. **Inteligência Artificial**. Rio de Janeiro: LTC, 2010.
- DE FERNANDES TEIXEIRA, J. **Inteligência artificial**. Pia Sociedade de São Paulo-Editora Paulus, 2014.
- DERRIDA, J. **Questão do estrangeiro: vinda do estrangeiro**. In: Dufourmantelle Anne, organizadora. Anne Dufourmantelle convida Jacques Derrida a falar da hospitalidade. Tradução de Antonio Romane. São Paulo: Escuta; 2003.
- FANG, K. **The development dilemma and countermeasures of strong artificial intelligence in meeting human emotional needs**. In: STEPHANIDIS, C.; ANTONA, M. HCI International 2020 - Posters. Communications in Computer and Information Science, p. 631-640, 2020.
- FILME “O Jogo da Imitação” (The Imitation Game”). Direção de Morten Tyldum, EUA, 2014, 116min.
- FILME “EVA”. Direção de Kike Maíllo, Espanha. 2011. HBO, 94min.

FETZER, J. H. **The Blackwell Guide to the Philosophy of Computing and Information**. Luciano Floridi, 2008, 392p. Disponível em: <<https://books.google.com.br>> Acesso em: 29 jun 2020.

FONSECA FILHO, C. **História da computação: o caminho do pensamento e da tecnologia**. Porto Alegre: Editora PUC-RS, 2007.

GOMES, D. S. Inteligência artificial: conceitos e aplicações. **Olhar Científico**, v. 1, n. 2, p. 234-246, 2011.

GUNKEL, D. J. Comunicação e inteligência artificial: novos desafios e oportunidades para a pesquisa em comunicação. **Galaxia**, n. 34, p. 05-19, 2017.

HAUGELAND, J. **Artificial Intelligence: The Very Idea**. MIT Press. (Ed.). 1985.

HOPCROFT, J. E.; ULLMAN, J. D. **Introduction to automata theory, languages, and computation**. Sydney: Addison-Wesley, 1979.

JONZE, S. Her. Los Angeles, Califórnia: Annapurna Pictures, 2013.

KAFKA, F. **Um médico rural**. São Paulo: Companhia das Letras, 1999.

KANT, I. **Fundamentação da metafísica dos costumes**. Tradução: Paulo Quintela. Lisboa: Edições 70; 1960.

KANT, I. **Crítica da razão prática**. Trad. Valério Rohden. São Paulo: Martins Fontes, 2002.

KANT, I. **Metafísica dos costumes**. Primeira Parte: Primeiros Princípios Metafísicos da Doutrina do Direito. Trad. Clélia Ap. Martins, p. 9-178; Segunda Parte: Primeiros Princípios da Doutrina da Virtude. Trad. Bruno Nadai, Diego Kosbiau e Monique Hulshof, p. 179-311. Petrópolis, RJ: Vozes; Bragança Paulista, SP: Editora Universitária São Francisco, 2013.

KARPATHY, A. **What I learned from competing against a convnet on imagenet**. Disponível em: <<http://karpathy.github.io/2014/09/02/what-i-learned-from-competing-against-a-convnet-on-imagenet/>>. Acesso em 11 set 11, 2020. 2014.

KIRK, G.S., RAVEN, J.E., SCHOFIELD, M. **Os filósofos pré-socráticos: história crítica com seleção de textos**. Trad. Carlos Alberto Louro Fonseca. Lisboa: Fundação Calouste Gulbenkian, 1994.

KOENIGKAM-Santos M, Ferreira-Júnior JR, Wada DT, Tenório APM, Nogueira-Barbosa MH, Azevedo-Marques PMA. Inteligência artificial, aprendizado de máquina, diagnóstico auxiliado por computador e radiômica: avanços da imagem rumo à medicina de precisão. **Radiologia Brasileira**, Nov/Dez; 52(6):387–396, 2019.

KOTTOW, M. **Introducción a la bioética**. Chile: Editorial Universitaria, 1995.

KUBRICK, S. 2001: **A Space Odyssey**. EUA/Reino Unido, 1968.

LEOPOLD, A. **A sand county almanac and sketches here and there**. New York: Oxford, 1989. Disponível em <<https://www.ufrgs.br/bioetica/landethi.htm>>: Acesso em 30 nov 2020.

LOBO, Luiz Carlos. Artificial Intelligence, the Future of Medicine and Medical Education. **Revista Brasileira de Educação Médica**, v.42, n.3, 2018.

LOFF, B. A tese de Church-Turing. **Boletim da SPM**, v. 67, p. 61-78, 2012.

MANSION, S. **La science, connaissance de l'essence**. In: MANSION, S. Le jugement d'existence chez Aristote. Louvain: Éditions de L'Institut Supérieur de Philosophie, 1976., p. 43-61.

MANZANO, J. A. N. G. **Algoritmos-Lógica para desenvolvimento de programação de computadores**. 28ª ed. Saraiva. 2018.

MIKOLOV, Tomas; JOULIN, Armand; BARONI, Marco. A roadmap towards machine intelligence. In: **International Conference on Intelligent Text Processing and Computational Linguistics**. Springer, Cham, 2016. p. 29-61.

MOTTA, O. J. R.; SILVA, E.; SIQUEIRA-BATISTA, R. Ethical education and artificial intelligence. **International Journal of Development Research**, v. 10, p. 43111-43112, 2020a.

MOTTA, O. J. R.; SILVA, E.; SIQUEIRA-BATISTA, R. 'Ensinar' ética às máquinas dotadas de inteligência artificial. **Jornal da Ciência**, v. 6450, p. 1 - 1, 27 jul. 2020b. Disponível em: <<http://www.jornaldaciencia.org.br/edicoes/?url=http://jcnoticias.jornaldaciencia.org.br/30-ensinar-etica-as-maquinas-dotadas-de-inteligencia-artificial/>> Acesso em 14 jun 2021.

NAVEGA, S. C. **Inteligência artificial: presente, passado e futuro**. Infoimagem, 2001, p. 19.

NG, G. W.; LEUNG, W. C. Strong artificial intelligence and consciousness. **Journal of Artificial Intelligence and Consciousness**, v. 7, n. 1, p. 63-72, 2020.

NILSSON, N. J. **Artificial Intelligence: a new synthesis**. San Francisco: Morgan Kaufmann, 1998.

O'REGAN, G. **A Concise History of Computing**. World of Computing, 29-74, 2018.

PASCAL, G. **O pensamento de Kant**. Trad. Raimundo Vier. 7ª ed. Petrópolis: Vozes, 2001.

PENROSE, R. **O grande, o pequeno e a mente humana**. Tradução Roberto Leal Ferreira. São Paulo: Fundação Editora da UNESP, 1998.

PEREIRA, R. S. Alguns pontos de aproximação entre a ética aristotélica e a kantiana. **Trans/Form/Ação**, v. 34, n. 3, p. 31-40, 2011.

PIQUEIRA, J. R. C. Complexidade computacional e medida da informação: caminhos de Turing e Shannon. **Estudos avançados**, v. 30, n. 87, p. 339-344, 2016.

RIBEIRO, Henrique de Moraes. (1999). Da metamatemática para a ciência cognitiva. **Trans/Form/Ação**, 21-22(1), 181-193, 1999.

RICH, E.; KNIGHT, K. **Artificial Intelligence**. 2nd edition. McGraw-Hill, 1990.

ROUSSEAU, Jean-Jacques. Do contrato social. Tradução de Antônio P. Machado. Rio de Janeiro: **Nova Fronteira**, 2011.

SANTOS, M. K., Ferreira Júnior José Raniery, Wada Danilo Tadao, Tenório Ariane Priscilla Magalhães, Barbosa Marcello Henrique Nogueira, Marques Paulo Mazzoncini de Azevedo. Inteligência artificial, aprendizado de máquina, diagnóstico auxiliado por computador e radiômica: avanços da imagem rumo à medicina de precisão. **Radiologia Brasileira**, v.6, p. 387-396, 2019.

SANVITO, W. L. Inteligência biológica *versus* inteligência artificial. **Arquivos de Neuropsiquiatria**, v. 53, n. 3-A, p. 361-368, 1995.

SÃO TOMÁS DE AQUINO. **O ente e a essência**. 2ª ed. Trad. de Carlos Arthur do Nascimento. Rio de Janeiro: Vozes, 2005.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, 2008.

SCHOPENHAUER, A. **Sobre o fundamento da moral**. Trad. Maria Lúcia Oliveira Cacciola. São Paulo: Martins Fontes, 2001.

SCHRAMM, F. R. Toda ética é, antes, uma bioética. **Humanidades**, v. 9, n. 4, p. 324-331, 1994.

SCHRAMM, F. R.; KOTTOW, M. Principios bioéticos en salud pública: limitaciones y propuestas. **Cadernos de Saúde Pública**, v. 17, n.4, p. 949-956, 2001.

SCHWEITZER, A. The ethics of reverence for life. **Christendom**, v. 1, p. 225-39, 1936. Disponível em <<https://www.ufrgs.br/bioetica/schweitz.htm>>: Acesso em 30 nov 2020.

SCOTT, R. **Blade runner**. Los Angeles: Warner Home Video, 1982.

SEARLE, J. R. Minds, brains, and programs. **Behavioral and Brain Sciences** 3 (3), 1980, 417-457.

SILVER, D; HUANG, A.; MADDISON, C.; GUEZ, A. Mastering the game of go with deep neural networks and tree search. **Nature**, v. 529, n. 7587, p. 484-489, 2016.

SIQUEIRA BATISTA R. Às margens do Aqueronte: finitude, autonomia, proteção e compaixão no debate bioético sobre a eutanásia. Tese de Doutorado: Saúde Pública. Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca; 2006.

SIQUEIRA-BATISTA, R. **(Bio)ética para todos os seres: proêmio**. In: CASTRO, J. C. DE; NIEMEYER-GUIMARÃES, M.; SIQUEIRA-BATISTA, R. Caminhos da Bioética III. Teresópolis: Editora Unifeso, 2020. p. 257–274.

SIQUEIRA-BATISTA, R. **A boa morte à luz da ética para todos os seres: o lugar da compaixão laica**. In: PEREIRA, T. S.; MENEZES, R. A.; BARBOZA, H. H. Vida, morte e dignidade humana. 1ª ed. Rio de Janeiro: GZ, 2010.

SIQUEIRA-BATISTA, R. **Fundamentos matemáticos da inteligência artificial: notas sobre a aprendizagem de máquina**. Trabalho de Conclusão de Curso: Licenciatura em Matemática. Rio de Janeiro: Universidade Estácio de Sá, 2018.

SIQUEIRA-BATISTA, R. Quantum bioethics: ethics for all beings. **Ciência & Saúde Coletiva**, v. 13, p. 1013–1015, 2008.

SIQUEIRA-BATISTA, R.; SCHRAMM, F. R. Bioética e neurociências: variações sobre o tema dos paradoxos da autonomia. In: Esperidião Antonio, V. Neurociências: diálogos e interseções. 1ed. Rio de Janeiro: Rubio, 2012, p. 473-487.

SIQUEIRA-BATISTA, R.; SILVA, E. Notas sobre os fundamentos matemáticos da inteligência artificial. **Revista de Ciência, Tecnologia e Inovação**, v. 4, n. 6, p. 44-54, 2019.

SIQUEIRA-BATISTA, R. **Fundamentos matemáticos da inteligência artificial: notas sobre a aprendizagem de máquina**. Trabalho de Conclusão de Curso, Licenciatura em Matemática. Rio de Janeiro, Universidade Estácio de Sá, 2018.

STOBBE, E. L. Um debate sobre o conceito kantiano de “fim em si mesmo”: Um conceito descritivo ou normativo? **Revista Estudos Kantianos**, v. 5, n. 1, p. 407-428, 2017.

TAYLOR, P. W. **Respect for nature. A theory of environmental ethics. Studies in moral, political, and legal philosophy**. New Jersey: Princeton University Press, 1986, p. 15.

TURING, A. M. (1950). Computing Machine and Intelligence. **Mind**, LIX, 433–460. Disponível em: <<http://mind.oxfordjournals.org/content/LIX/236/433>> Acesso em 01 out 2020.

TURING, A. On computable numbers, with an application to the Entscheidungsproblem. **Proceedings of the London Mathematical Society**, 2d series, 42, 230-265, 1936.

VERNANT, J.-P. **Mito e pensamento entre os gregos: estudos de psicologia histórica**. Trad. Haiganuch Sarian. 2ª ed. Rio de Janeiro: Paz e Terra, 1990.

VON ZUBEN, F. J. **O mundo natural e o mundo artificial**. Campinas (SP) [online]. 2011 [citado em 10 de outubro de 2018]. Disponível em: <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia707_1s11/notas_de_aula/topico1_IA707_1s11.pdf>

WHITNEY, L. Are computers already smarter than humans? **Time Magazine**. Available at: <http://time.com/4960778/computers-smarter-than-humans>. Acesso em 23 março 23, 2021.

WINOGRAD, T. Thinking Machines: Can there be? Are we?. **Cambridge University Press**, p.167-189, 1990.

ZILIO, D. Inteligência artificial e pensamento: redefinindo os parâmetros da questão primordial de Turing. **Ciências & Cognição**, v. 14, n. 1, p. 208-218, 2009.

05.

AS DUAS VIDAS DE ALAN FOSTER: UM EXPERIMENTO MENTAL SOBRE IDENTIDADE PESSOAL E EDIÇÃO GENÉTICA

Anderson do Vale Fonseca

Gustavo Leal-Toledo

Ana Luisa Lima Grein

1. INTRODUÇÃO

Neste capítulo examinaremos a noção de identidade pessoal em Derek Parfit a partir do experimento mental do teletransporte em uma versão diferente, no caso, da edição genética de um embrião com síndrome de Down. Em síntese, o teletransporte cria no espaço-tempo uma duplicata, após o armazenamento das informações e destruição do original. Um clássico típico exemplo é o quinto episódio da série *Star Trek*, intitulado *The Enemy Within*. O capítulo narra a seguinte história: o capitão James T. Kirk durante o teletransporte é acidentalmente duplicado em uma versão maligna de si mesmo.

Em torno dessa fábula, está uma pergunta intrigante: a pessoa que se teletransportou continua sendo a mesma do outro lado? Considere que, durante o processo de teleportação, o original é destruído e recriado com material local, em outra região do espaço-tempo. Assim, o fato de Kirk ter sido acidentalmente duplicado, revela que a máquina não transporta de um ponto do espaço a outro a mesma pessoa, mas que, apenas transfere a informação, gerando um outro indivíduo qualitativamente idêntico. Não à toa, o personagem estarrece ante à descoberta.

O mesmo se aplica à edição genética, ela funcionaria como uma máquina de teletransporte, capaz de substituir o material genético deficiente do indivíduo por outro saudável, produzindo, a partir disso, uma

nova pessoa. Consequentemente, tal como aconteceu a Kirk, teríamos dois seres qualitativamente idênticos, mas fisicamente distintos.

Diferentes interpretações, então, da visão parfitiana do experimento mental do teletransporte quanto à identidade da pessoa serão exploradas no presente capítulo. No entanto, a versão adotada do experimento aplica-se à edição gênica de embriões, principalmente, em situações como a imaginada de um casal cujo bebê é diagnosticado com síndrome de Down. Levantamos com base na teoria do filósofo Derek Parfit (1984) e no argumento de Stephen Holland (2008), a hipótese de a intervenção genética alterar a continuidade física e psicológica do indivíduo. Seguindo essa mesma linha, examinamos diferentes resultados possíveis do experimento bem como as questões éticas em torno.

Por exemplo, depois que o embrião com síndrome de Down é editado, é possível que a pessoa futura não seja uma e a mesma do embrião. Se isto é verdade, pode-se afirmar que uma identidade foi erradicada para ceder espaço a outra. Nesse contexto, a edição pode ser entendida como uma forma de morte, contudo, não da vida, mas da pessoa. Todavia, não há certezas quanto a isso, portanto existe também a possibilidade de nenhuma alteração na identidade pessoal acontecer.

Em contraponto, analisaremos a objeção de Peter Singer e Helga Kuhse, que sustentam não haver identidade numérica no zigoto no começo da embriogênese e, por esta razão, o uso de zigotos para pesquisas ou, intervenção genética, não causaria danos. Nesse contexto, a edição de um embrião com síndrome de Down, não produziria um indivíduo diferente do original, porque sequer haveria inicialmente um indivíduo.

Assim, tomando como referência a situação hipotética da edição gênica do embrião com síndrome de Down, investigaremos se há um limiar na continuidade física que, se rompido, significa uma alteração radical na identidade da pessoa. Em outras palavras, existe a possibilidade da existência de uma fronteira crítica que, quando alcançada, resulta em alguém completamente diferente. Finalmente, analisaremos as questões morais e éticas que as diferentes interpretações parfitianas produzem. Não é nossa intenção, aqui, apresentar resultados, mas suscitar questões morais referentes à possibilidade de a edição ser capaz de gerar uma pessoa diferente a partir da substituição de um material genético por outro.

2. LULU E NANA: OS BEBÊS CRISPR

A terapia gênica consiste na alteração de códigos genéticos defeituosos a fim de eliminar doenças até então incuráveis. O método de edição genética CRISPR Cas-9, por exemplo, descoberto em 2015, assemelha-se a um editor de textos cortando, alterando e inserindo palavras com o objetivo de melhorar a sentença e, por sua vez, a composição da obra. De modo estrito, a edição genômica CRISPR, realiza intervenções precisamente em uma cadeia específica de ácidos nucleicos através da proteína Cas-9 reunida a um RNA, a junção leva a identificação do alvo e sua clivagem. Depois, o trecho eliminado é substituído por uma nova cadeia de DNA. Com efeito, o material genético defeituoso é reescrito, apagando qualquer traço de doença ou desvantagem seletiva que venha a ser herdada pelos descendentes.

Essa técnica despertou o interesse da comunidade científica e da sociedade. Ao mesmo tempo que provocou maravilha também causou medo. Agora seria possível curar um embrião de uma doença antes do nascimento. Por outro lado, há também o acesso ao aprimoramento de características vantajosas evolutivamente, como memória e imunidade. Uma sociedade, como imaginada no filme *GATTACA* (1997), dirigido por Andrew Niccol, poderia se concretizar.

Quais seriam as consequências para a sociedade, um embrião editado em laboratório? Uma criança desenhada para ser imune a qualquer doença está em vantagem quanto aos demais humanos, não seguiria disso uma discrepância evolutiva? Mas por qual razão não deveríamos projetar humanos saudáveis? É justo permitir que um bebê futuramente desenvolva Alzheimer ou a coreia de Huntington, ao invés de eliminar a doença? A alteração do DNA também irá mudar a biografia da futura pessoa? Em que ponto a intervenção genômica beneficia o indivíduo, e em que ponto ela pode prejudicá-lo?

No ano de 2017, por exemplo, pesquisadores do MIT, liderados por Shoukhrat Mitalipov³⁰, alteraram o DNA em bebês humanos, com o método de edição CRISPR. O objetivo era corrigir ou alterar genes causadores de doenças na linha germinativa, antes da fase fetal. O experimento, à época, foi considerado um grande feito.

30 Disponível em *MIT Technology Review*: <https://www.technologyreview.com/2017/07/26/68093/first-human-embryos-edited-in-us/> acesso em 10/07/20.

Um ano depois, o cientista He Jiankui, do departamento de Biologia da Universidade de Ciência e Tecnologia do Sul, na China, surpreendeu o mundo com a notícia do nascimento dos primeiros embriões humanos editados geneticamente, com a técnica CRISPR Cas-9. A fim de tornar os bebês resistentes ao HIV, à varíola e à cólera, o cientista “cortou” o gene CCR5 [este gene codifica uma proteína que é o sítio principal de ligação do vírus HIV com os linfócitos T].

Os bebês CRISPR, também conhecidos pelo pseudônimo Lulu e Nana, se tornaram os primeiros embriões melhorados da história a nascerem. Essas crianças desenhadas imunes ao HIV transmitiriam essa vantagem às próximas gerações. Assim, seus descendentes também seriam resistentes ao HIV, além de manifestarem outros efeitos da edição do CCR5 que não podem ser previstos.

A façanha médica de Jiankui despertou preocupações éticas e morais com a aplicação tecnológica da edição genética em humanos. É de se suspeitar que Lulu e Nana, devido à elisão do CCR5, teriam a identidade numérica modificada, ou seja, antes da intervenção genômica os embriões possuíam um genótipo, depois, exibiam outro genótipo. Por mais que fenotipicamente permanecessem as mesmas, sua composição de base havia sido substituída, removida ou rearranjada. De certo modo, Lulu e Nana era cada qual uma e a mesma pessoa até certo momento, porém, imediatamente após a operação realizada pela proteína Cas-9, passaram a ser pessoas radicalmente diferentes.

Inspirados pelo caso Jiankui, elaboramos uma situação hipotética para discutir os impactos da terapia gênica na continuidade física e na continuidade psicológica do indivíduo. Primeiro, iremos examinar o experimento do teletransporte e o conceito de identidade pessoal defendido pelo filósofo britânico Derek Parfit. Vejamos a seguir.

3. O EXPERIMENTO MENTAL DO TELETRANSPORTE E A IDENTIDADE PESSOAL EM DEREK PARFIT

De acordo com Parfit: “A identidade pessoal ao longo do tempo apenas consiste na continuidade física ou psicológica” (2002, p. 657, tradução nossa). Ademais, ele depreende a continuidade psicológica como “as várias relações entre diferentes estados e eventos mentais como [...] memória [...] e

outras características psicológicas” (PARFIT, 2002, p 657, tradução nossa). À vista disso, com o intuito de esclarecer o conceito de identidade pessoal em Parfit, examinaremos o experimento mental da duplicação humana proposto pelo filósofo na obra *Reasons and Persons* (1984).

Imagine que você entra em uma máquina de Teletransporte, na Terra. À sua frente em um painel há um interruptor vermelho com a legenda abaixo, “Escanear”. Você aperta o botão e imediatamente seu corpo é escaneado, cada detalhe de seu arranjo molecular, cada neurônio do seu cérebro em cada padrão energético específico. Em seguida, após o registro dos dados, você é destruído. Não se preocupe! Em Marte, outra máquina recebeu as informações e criou uma réplica exata de você a partir dos elementos físicos presentes no local. Ora, de certo modo, você não morreu, está vivo, porém, como sua réplica, que de tão perfeita estruturalmente, não se vê diferença em relação ao original. Esse novo você feito a partir de componentes físicos marcianos têm a mesma memória que a do original; com efeito, pode-se afirmar que $A = B$, ou seja, é o mesmo sujeito psicológico.

É claro, $A = B$, mas não absolutamente. Embora o sujeito B, surgido em Marte, seja numericamente idêntico ao sujeito A, da Terra, e seja possível que B tenha as mesmas memórias de A, de tal modo que se afirme existir uma continuidade psicológica em relação ao físico, não se deve daí sustentar que haja também identidade quanto a experiência em primeira pessoa.

Para tornar mais ilustrativo, imagine que ao lado do interruptor vermelho haja outro de cor azul (sem dúvida, lembra a cena de Neo e Morpheus em *Matrix*) e, abaixo está escrito, “Salvar”. Esse interruptor tem a seguinte função: preservar o original simultaneamente à criação da réplica. Você, dessa vez, aperta o botão azul e salva a si mesmo! (Melhor do que se sacrificar!) Agora há dois de você, um na Terra, outro em Marte³¹.

Segundo essa ilustração, a ontologia do indivíduo é determinada pela forma como os constituintes físicos são arranjados espacialmente, desde os átomos até às células neurais e a estrutura sináptica emergente. Assim, no exato momento da duplicação humana, desde que a estrutura sináptica seja mantida, o outro que emerge em Marte será idêntico ao original. De certo

31 Roger Penrose na obra *A Mente Nova do Rei* sugere outra versão do experimento de teletransporte. Nessa versão, a informação não é armazenada na forma humana, mas em um aparelho eletrônico. PENROSE, Roger. *A Mente Nova do Rei*: Computadores, mentes, e as leis da Física. Rio de Janeiro: Editora Campus, 1991, p. 29.

modo, a similitude entre os dois leva a supor que ambos, mesmo que separados espaço-temporalmente, terão experiências compartilhadas, como se entre eles houvesse um tipo de sincronização neural não-local.

No entanto, de acordo com a teoria reducionista da continuidade psicológica, isso não ocorre, as experiências subjetivas de B não são idênticas as de A, embora B tenha as memórias de A, a subjetividade é um evento singular, do sujeito que a experiência. Por exemplo, se, em vez do corpo, tivesse transferido acidentalmente, em seu lugar as memórias de A para, quem sabe, o cérebro de um engenheiro que trabalha em Marte, teríamos, então, duas pessoas em um mesmo corpo, mas cada qual com experiências diferentes. Neste caso, há uma continuidade psicológica de A em C; logo, $C=A$.

Esse é um característico exemplo da relação transitiva da identidade. Se B é igual a A, e se C também igual a A, então C será igual a B. Para tornar mais claro esse conceito, vejamos outra versão do experimento do teletransporte, imagine então, dessa vez, que as memórias de A, acidentalmente, são transferidas para dois corpos, talvez, o engenheiro e seu auxiliar técnico que estão do outro lado, em terras marcianas, aguardando a recepção das informações enviadas pela máquina da Terra. É possível que, no exato momento em que a informação é recebida, tanto B (o engenheiro) quanto C (o auxiliar) se tornem contínuos psicologicamente com A (o original) e, de certo modo – mas não exatamente –, $C = B = A$. Sem dúvida, a explicação está no fato de, enquanto a identidade numérica é uma relação lógica de um para um ($x=y$), a continuidade psicológica “não é logicamente de um para um” (PARFIT, 1971, p. 7). Na presente ilustração, a continuidade psicológica é mais importante do que ser numericamente idêntico

Mas, imagine ainda, a exemplo do experimento de Locke³², que a pessoa de A se manifeste somente durante à noite, de modo que C não desconfie nem se lembre, tal como pela manhã, C se manifesta, de modo que A não desconfia nem se lembra. As experiências de C, nesse caso, jamais seriam idênticas as de A, porque haveria “duas correntes de consciência incomunicáveis, manifestando-se por meio de um e mesmo corpo” (MASLIN, 2009, p. 255).

Assim, retornando ao experimento inicial, logo depois de B emergir em solo marciano, ele vivenciará novos sentidos e emoções, e isso não é experimentado simultaneamente por A. Caso acontecesse, se B ao acordar

32 Ver MASLIN, Keith T. *Introdução à Filosofia da Mente*. Ed. Artmed, Porto Alegre, 2007, p. 255.

decidisse olhar o nascer do sol em Marte, enquanto A escolhesse contemplar o nascer do sol na Terra, ambos experimentariam uma imagem sobreposta em um único fluxo de consciência. Que visão exótica! No entanto, se essa linha de pensamento estiver correta, tal evento não parece ser possível. É verdade que $B=A$, mas não quanto à subjetividade.

Por outro lado, se A apertou o botão “salvar”, então B não é idêntico numericamente a A, porque se o fosse, seria uma mesma substância com A. Nesse caso, B é apenas idêntico qualitativamente a A, do mesmo que irmãos gêmeos, apesar da semelhança estrutural e física, não são uma e mesma pessoa. Em contraste, se A não apertou o botão azul e aceitou ser destruído no processo de escaneamento, então B é idêntico numérica e qualitativamente a A, ou seja, é contínuo física e psicologicamente.

Além disso, segundo Maslin, uma entidade individual não pode ter dois começos (2009, p. 254). Com isso, ele pretende dizer que, quando os átomos forem reordenados para compor o novo corpo da pessoa que originalmente entrou na máquina de teletransporte, não significa que a estrutura organizacional será preservada, pois ela foi perdida durante o processo de decomposição molecular. Como resultado, $B \neq A$, ou seja, mesmo que A tenha apertado o botão vermelho, o que surge em Marte, a quem chamamos B, não é numericamente idêntico, na verdade, é apenas qualitativamente indiferenciado.

Por sua vez, Parfit propõe outro experimento, o do cérebro dividido. Imagine que o hemisfério esquerdo e o hemisfério direito de seu cérebro são transferidos cirurgicamente para dois corpos diferentes, B e C, cada qual ficando com uma metade. Segundo o autor, cada metade acreditaria ser você, porque cada parte conteria suas memórias. O que se depreende da ilustração é a continuidade psicológica como fundamental, embora nem B nem C sejam idênticas numericamente a A, este continua a existir como “dois fluxos de experiências causalmente relacionados” (MASLIN, 2009, p. 273).

Porém, é possível supor, a exemplo dos casos de indivíduos que sofrem do transtorno de identidade da integridade corporal, como analisado por Oliver Sacks (2013) que, existindo continuidade psicológica de A para B e C, não significa que B e C afirmarão que suas memórias são a mesma de A, ao contrário, a mente estando corporificada, a pessoa resultante afirmará que as memórias que possui são dela e não de outro.

De acordo com Sacks, certas pessoas sentem “que um de seus membros, ou talvez parte de seus membros, não lhes pertence, que é um estorvo

estranho, e esse sentimento pode engendrar um desejo muito forte de que o membro ‘supérfluo’ seja amputado” (2013, p. 73). A razão encontra-se em alterações fisiológicas em algumas regiões bem específicas do lobo parietal.

Um dos relatos mais fascinantes é a de um paciente que, depois de um acidente no hemisfério direito, perde completamente a sensação do lado esquerdo do corpo e passa a acreditar que esse lado pertence a outra pessoa. Isso sugere que se o cérebro não estiver inteiramente incorporado não reconhecerá o corpo para o qual foi transferido. Apesar de B e C serem contínuos psicologicamente a A, afirmariam que o corpo em que estão lhes é estranho, gerando um conflito. Em outras palavras, podemos imaginar algo similar à obra *A Metamorfose*, de Franz Kafka, em que o personagem Gregor Samsa acorda transformado em um inseto. Numa outra versão, Samsa poderia ter tido a mente transferida para o corpo de um inseto e, como tal, não reconheceria aquele corpo como sendo idêntico a ele. Sob essa ótica, podemos sustentar que a continuidade física é necessária para a continuidade psicológica, no sentido de que B para ser um e o mesmo que A precisar também ser numericamente idêntico a A, e não ser somente quanto às lembranças e experiências.

Por outro lado, se a mente de Samsa fosse integrada ao corpo, como se sucede no desenrolar da novela, Samsa acreditaria ser um e o mesmo que o inseto. Do mesmo modo, quando acidentalmente a mente de A é dividida entre o engenheiro e o auxiliar técnico, é possível que B e C ainda que sejam contínuos psicologicamente a A não acreditem que suas memórias tenham origem em A, mas que essas experiências foram vividas por eles. No entanto, as experiências de B e C seriam, na verdade, quase-memórias das experiências de A. (PARFIT, 1971, p. 9-14; MASLIN, 2009, p. 268-269). Ainda que B e C afirmem se lembrarem de determinada experiência, essa experiência foi vivida por A. Logo, a causa de a quase-lembrança ser uma lembrança vivida por B e C, está na incorporação.

Até aqui analisamos alguns dos diferentes aspectos do experimento mental do Teletransporte. No entanto, é importante esclarecermos um pouco mais a teoria reducionista da identidade pessoal e o argumento de Parfit. Isto, pois, estas são as bases para nosso argumento central, que é composto pela afirmação de que a edição genética, do mesmo modo que a máquina de teletransporte, suscita questões éticas quanto à identidade pessoal.

Em síntese, o problema diacrônico da identidade pessoal está relacionado ao que determina uma pessoa em um dado momento t_0 permanecer a mesma no momento seguinte t_1 . Na visão do materialismo reducionista, a pessoa é idêntica ao corpo, nesse sentido, uma pessoa é uma e a mesma pessoa de t_0 se, e apenas se, ter um e o mesmo corpo em t_1 . Um problema que ilustra essa visão é o clássico enigma do navio de Teseu (KLEINMAN, 2013, p. 28-30). Assim é conceituado o paradoxo: Teseu após matar o Minotauro, leva consigo de volta à cidade de Antenas, os jovens que resgatou. No trajeto, as partes velhas do navio são gradualmente removidas e substituídas por novas. Quando, finalmente, alcança o porto, um mercador comenta: “Esse é o mesmo navio que partiu?”. A resposta à pergunta estabelece a relação entre a parte e o todo na identidade do ser.

Por conseguinte, na hipótese de, ao menos uma parte ter sido substituída, o navio que chegou ao porto, seria outro ou permaneceria o mesmo? De acordo com a Identidade Mereológica (KLEINMAN, 2013, p. 29), se todas as partes de B eram partes de A, então $B=A$. No entanto, exige-se que façam parte exatamente os mesmos componentes, o que não se aplica ao navio de Teseu. Esse paradoxo dá a entender que a identidade não se reduz apenas aos constituintes físicos, mas que é, sobretudo, determinada pela preservação ininterrupta da estrutura e de sua funcionalidade (MASLIN, 2009, p. 254). Nesse sentido, se a forma é mantida, a identidade também é preservada. Em outras palavras, a identidade pessoal depende da condição de a estrutura física e sua função ser ou não continua no espaço-tempo. Do mesmo modo, se o corpo de uma pessoa permanece o mesmo na transição de t_0 para t_1 , então $B=A$.

A continuidade física não é, todavia, o fator essencial para garantir a identidade de uma pessoa. Na visão de Parfit, um indivíduo será uma e a mesma pessoa se, e somente se, existir entre Y em t_0 e X em t_1 , uma continuidade e encadeamento causal de memórias e experiências – isto é, se X é psicologicamente contínuo com Y. Essa continuidade se dá de modo que a experiência e outras características psicológicas de Y estão causalmente relacionados de forma direta com as experiências de X. Segundo essa linha de pensamento, se B que surge em Marte é contínuo psicologicamente a A, então $B=A$.

Contudo, é importante enfatizar que a identidade pessoal se sustenta na continuidade física do cérebro (ou de algo próximo), pois ele é o que dá condições para a existência dos aspectos psicológicos, isto não significa que

a pessoa seja redutível ao cérebro, tão somente que é dependente dele para ser (MASLIN, 2009, p. 270; PARFIT, 2002, p. 657). Retomemos a seguinte ilustração, quando A aperta o botão vermelho e é destruído, a versão que surge em Marte é idêntica psicologicamente porque o novo cérebro contém suas memórias e experiências; mas, ao mesmo tempo, é alguém numericamente distinto, pois, a estrutura organizacional – incluindo os componentes físicos – não é completamente exata. Em suma, a identidade pessoal está causalmente relacionada à continuidade física do cérebro, ainda que não exista necessariamente identidade numérica.

Esclarecido o problema da identidade pessoal na visão de Derek Parfit (1984), agora nos voltamos para o argumento medular desse capítulo, a saber: a edição genética funciona como uma máquina de teletransporte, gerando, a partir de uma alteração fundamental um ser numericamente distinto e, conseqüentemente, com outra identidade pessoal. Nessa linha, apresentamos, a seguir, uma situação hipotética e a interpretação parfitiana.

4. O EXPERIMENTO MENTAL DA EDIÇÃO GÊNICA DE ALAN FOSTER

Imaginemos que estamos em 2045, Jonathan e Brendan Foster – um casal de jovens londrinos – decidem ter um bebê. Na Inglaterra do meado do século 21, é bastante comum empresas de biotecnologia produzirem úteros artificiais para acomodar embriões humanos, do mesmo modo que editam esses embriões, caso eles apresentem, em exame pré-natal, deficiência de base genética, ou os pais escolham adicionar características fenotípicas.

Isso criou um mercado bilionário: embriões editados sendo gerados em úteros artificiais. Quantas mulheres conservando a beleza física! Quantos maridos agradecidos! Deram fim às enxaquecas matrimoniais. A exceção eram os Foster. Brenda desejava engravidar naturalmente, depois realizar o mapeamento genômico. Jonathan concordou.

Na noite de 14 de abril, amaram-se. Algumas semanas depois, interrompeu-se a menstruação. Brenda sorriu. Quando os Foster visitaram o laboratório de genética, para o teste pré-natal, alarmaram-se com a notícia: o embrião apresentava a anomalia trissomia do 21. Uma criança com Síndrome de Down era bastante rara nessa época, porque durante o processo de triagem, os embriões com distúrbios genéticos eram descartados e

substituídos por normais pré-editados. Mas, a decisão de Brenda e Jonathan — embora não proibida por lei — havia, agora, lhes afetado profundamente.

O médico dá ao casal uma escolha, editar o embrião para gerar um indivíduo normal, ou seguir com a gestação. Para Jonathan o melhor era “consertar” a anomalia. Para Brenda a dúvida era angustiante: aceitar o que a natureza lhe presenteou, ou interferir nela. Caso continuasse, a criança teria uma vida saudável? Seria capaz de construir uma nova família? Mas se fosse editada, essas questões persistiriam?

Apesar dos Foster terem gerado o bebê naturalmente, a própria natureza os obrigou a aceitar a solução oferecida pela ciência.

No dia 5 de junho, Alan Foster — assim batizaram a criança — foi editado. Neste exato momento, dois mundos possíveis foram juntamente produzidos, duas realidades cindidas em um átimo: em uma, havia Alan-1, portador da síndrome de Down, com um sistema cognitivo peculiar, com uma mente única; na outra, Alan-2, uma criança saudável, com uma mente também única. A decisão dos pais dividiu as realidades em dois seres conscientes, numericamente distintos, que compartilham agora apenas um mesmo nome e origem.

4.1 O problema

A situação hipotética de Alan Foster expõe, segundo Holland (2008, p. 159), o problema da alteração da pessoa; mas, aqui, tomando de empréstimo o termo criado pelo escritor K. Dick (2004, p. 350) no conto *As Pré-Pessoas*, intitularemos de, o problema da alteração da pré-pessoa (HOLLAND, 2008, p. 159-160). Isto, pois, a noção de pessoa está vinculada à realização de capacidades potenciais como a razão, a memória e a linguagem. Nesse contexto, Alan teria uma identidade numérica correspondente à sua constituição genética, contudo ainda não é uma pessoa.

A partir destes argumentos, a opção da intervenção genética pré-natal (HOLLAND, 2008, p. 160) pode ser justificada moralmente fundamentando-se no que o embrião virá a ser, e não no que ele atualmente é; o que ele é agora (geneticamente deficiente) pode ser substituído por um mundo possível melhor, em que o indivíduo (saudável) terá plenas condições de desenvolver suas potencialidades. A edição genética pode ser interpretada, seguindo essa linha de pensamento, como uma ação que causa “a

não-existência do indivíduo diferente” (HOLLAND, 2008, p. 162), rompendo com a continuidade psicológica, dando origem a outra pré-pessoa.

4.2 Duas possíveis leituras

Com base na teoria parfitiana, examinaremos duas leituras diferentes da presente situação hipotética: a cartesiana e o reducionismo materialista.

Sob a ótica cartesiana, a pessoa é uma entidade imaterial que mantém um elo causal misterioso com o cérebro. Nessa perspectiva, qualquer alteração física, ainda que perturbe ou não a relação causal entre a mente e o cérebro, não representa uma mudança na própria pessoa. Sendo ela o mesmo que a substância imaterial, nenhuma intervenção física irá afetá-la. Como Parfit declara, a respeito das entidades cartesianas, “faria sentido supor que eles podem deixar de ser causalmente relacionados a algum corpo, mas continuam a existir” (2002, p. 657). Por conseguinte, Alan-2 não será distinto de Alan-1, porque entre os dois há continuidade de consciência, ainda que não sejam numericamente idênticos. Isto é, privando ou mantendo a relação causal em uma organização neural adequada, a *pré*-pessoa de Alan-2 será contínua com Alan-1. Desse modo não se vê justificativa moral que impeça a intervenção genética pré-natal (HOLLAND, 2008, p. 159), visto que a finalidade principal seria gerar um corpo “saudável” para uma e mesma pessoa.

No que diz respeito à visão do Materialismo Reducionista, “cada tipo de estado ou processo mental é *numericamente idêntico a* (é uma e mesma coisa que) algum tipo de estado ou processo mental no interior do cérebro ou no sistema nervoso central” (CHURCHLAND, 1998, p. 53). Assim, na linha materialista, qualquer mudança física na estrutura organizacional do corpo afetará a própria mente. Disso sucede que, Alan-2 é totalmente descontinuo a Alan-1, pois havendo interferência na organização material segue-se que emergirá uma nova consciência e, logo, uma nova identidade pessoal. Certamente, essa interpretação, desperta justificativas morais que impeçam a intervenção genética pré-natal, visto que a finalidade de produzir um corpo “saudável” produzirá uma pessoa diferente.

As duas leituras não oferecem uma resposta conclusiva, e não estamos certos de que sejam válidas. É importante lembrar que a identidade pessoal

“é uma questão de tudo ou nada” (PARFIT, 1970, p. 6), ou seja, X pode ou não ser contínuo a Y, a depender de qual teoria adotamos.

O argumento cartesiano, por exemplo, carece de evidências que o sustente, levando a crer em sua falseabilidade. Nos dizeres de Parfit: “Ao contrário de outros escritores, eu acredito que tal visão deve ser verdadeira. Mas nós não temos boas evidências para pensar que é, nem evidências para pensar que não é; então, eu devo assumir aqui que essa visão não é verdadeira” (2002, p. 656). Não estamos convencidos, portanto, de a possibilidade de Alan-2 contínuo com Alan-1 ser verdadeira.

Em contrapartida, a visão materialista coaduna com a intuição de que alterações na constituição genética de um embrião acarretam em identidades *pré*-pessoais descontínuas, porque, “Embora pessoas sejam distintas de seus corpos, e de qualquer série de eventos mentais, eles não são independentes ou entidades existindo separadamente” (PARFIT, 2002, p. 657). Nessa linha, podemos dizer que é verdadeiro que Alan-2 não é qualitativa nem numericamente idêntico a Alan-1, sendo duas pessoas diferentes; mas, um veio a existir em substituição do outro.

4.3 A tese de Derek Parfit

No que concerne ao materialismo, Parfit examinando o experimento do teletransporte, estabelece a seguinte condição: “Se nenhuma célula fosse substituída, a pessoa resultante seria eu” (2008, p. 658, tradução nossa). Ele ainda ressalta que, quanto maior for a proporção do corpo a ser substituída pelas novas células, menos a pessoa resultante seria a mesma que o original. Com base nisso, Parfit supõe haver um limite crítico de células replicadas de modo que, se apenas dentro dessa margem certo número de células fosse substituído, a pessoa resultante seria a mesma que o original, mas, transposto a margem de células substituídas, seria uma pessoa diferente.

Se o mesmo se aplica à edição genética, na qual a intervenção remove ou substitui um genoma, a pessoa futura será incompatível com o original. No entanto, é possível que haja, dentro dessa borda uma linha tênue em que a edição gênica não signifique uma ruptura na identidade pessoal?

Suponha, então, que Alan-1 é editado dentro da margem, de modo que, se somente a anomalia fosse erradicada, a pessoa resultante seria exatamente Alan-1. Mas não é o que a edição CRISPR parece indicar. A remoção

ou substituição de genomas específicos pode gerar uma pessoa descontínua numericamente e psicologicamente. Não significa, contudo, que não exista certas intervenções capazes de manter a identidade pessoal, o que leva a pergunta: Até que ponto a intervenção genética pré-natal não representa um risco a quem o indivíduo poderá vir a ser?

A resposta a essa questão pode ser o argumento de Stephen Holland, na obra *Bioética: enfoque filosófico* (2008). Vejamos.

4.4 A resposta de Stephen Holland à questão parfitiana

O argumento de Holland tem como base a seguinte premissa:

Certas formas de intervenção genética pré-natal alterariam não somente a identidade qualitativa, mas também numérica das futuras pessoas. Em outras palavras, algumas formas de intervenção genética pré-natal afetam quem é a pessoa que nasce (2008, p. 159).

Com esse enunciado, o filósofo estabelece que existem formas de intervenção genética em que a pessoa resultante é diferente da anterior, bem como formas de intervenção genética em que a pessoa resultante é a mesma que a anterior. A partir disso, Holland buscará traçar a linha tênue que as separa. Mas antes de chegar a esse ponto, o autor percorrerá algumas situações hipotéticas a fim de examinar a validade da premissa.

De início, Holland nos convida a imaginar um conceito que sofre de uma doença de base genética e tem sua biofísica radicalmente alterada por meio da “substituição de uma considerável proporção da constituição genética” (2008, p. 160). Segundo ele, tal circunstância leva a pensar que não havendo intervenção sobre o conceito, nascerá um indivíduo deficiente, do contrário, se a intervenção genética é aplicada, nascerá um indivíduo saudável numericamente desigual. Em suas palavras: [...] “alguma outra pessoa nascerá caso se efetue uma intervenção genética pré-natal” (HOLLAND, 2008, op. cit.).

No caso de anomalias genéticas como a síndrome de Down ou de doenças como a de Huntington, a edição genética, por exemplo, pode ser justificada com base em uma ação cujo fim é colocar o indivíduo em uma situação melhor. A isso, Holland (2008, p. 160) cunha de razão concernente a pessoa. Desse modo, o casal Foster, quando decide pela edição genética do embrião, o faz pensando em um mundo possível onde a pessoa futura Alan tenha uma vida satisfatória.

Por outro lado, a razão “impessoal” sustenta a seguinte visão: caso se saiba que a intervenção genética pré-natal altera drasticamente a identidade numérica de indivíduos entre t_0 e t_1 , a razão para fazê-lo não poderia, jamais, apoiar-se na pessoa. Nesse contexto, o casal Foster decide pela edição pensando não na pessoa que o embrião virá a ser, mas na eliminação da deficiência que carrega. Contudo, após a edição, “o indivíduo que o conceito original viria a se tornar não chega a nascer” (HOLLAND, 2008, p. 161). Ou seja, ao voltar-se para a doença, a fim de eliminá-la, o deficiente é juntamente eliminado.

Após desenvolver a premissa, Holland examina as concepções teóricas da identidade pessoal a fim de verificar a validade do argumento. Baseando-se no critério psicológico de Parfit, segundo a qual a continuidade psicológica é necessária para a existência duradoura da pessoa, o autor afirma a importância de determinar qual o “grau de descontinuidade psicológica é suficiente para alterar a identidade numérica” (HOLLAND, 2008, p. 163).

Holland volta-se então para a identidade do conceito, invertendo a ordem da análise: em vez de pensar sobre a identidade da pessoa resultante, investiga-se primeiro a identidade do conceito, porque um indivíduo numericamente distinto é proveniente de um conceito numericamente distinto. Com efeito, pode-se dizer que o embrião se iguala à noção de pré-pessoa. Nesse caso, se um embrião é geneticamente modificado, o indivíduo futuro será diferente do indivíduo que o embrião se tornaria, caso não fosse alterado (2008, p. 164), como ilustra a história de Alan Foster.

Todavia, há também o cenário em que após a edição genética, o embrião permanece numericamente idêntico a si mesmo, o que leva a necessidade de determinar em qual condição o embrião, caso editado, continuará com a mesma identidade numérica. De acordo com Holland, a condição numérica para a identidade de um embrião está diretamente relacionada à identidade da pessoa resultante; embora essa conjectura se prove circular (2008, p. 166).

Outro argumento seria a teoria da suficiência abordado por Saul Kripke (HOLLAND, 2008, p. 166). Em resumo, uma vez que alguém tenha a mesma origem, isto é, os mesmos pais, será numericamente idêntica. Porém, uma vez que se origine de pais diferentes, então não é a mesma pessoa. Assim, nessa linha de pensamento, se Alan Foster é editado geneticamente, Alan-2 será numericamente idêntico visto ter uma mesma origem, o que significa

que, mesmo tendo sido alterado de modo fundamental, por ser ainda o mesmo embrião, intui-se ser a mesma pessoa.

No entanto, não parece ser verdade, porque o gene defeituoso é substituído por um saudável, alterando, com isso, a relação mereológica na constituição do sujeito. Não há como garantir que a pessoa resultante de um conceito modificado radicalmente seja a mesma do conceito inalterado. É uma afirmação inconsistente. De outro modo, se houve tão somente pequenas alterações, é possível que a identidade permaneça intacta.

Em outra perspectiva, se assim como no barco de Teseu em que cada peça substituta é colocada no mesmo local e na mesma posição da peça antiga, cada gene saudável quando inserido no mesmo local e na mesma posição do gene defeituoso, desde que a estrutura permaneça continua no espaço e no tempo, a pessoa resultante será a mesma que a anterior (STRAWSON, 2016, p. 51).

Nessa linha, se em um determinado locus gênico, o gene é removido de seu loco gênico (posição) e substituído por um gene correspondente no mesmo local e posição, mesmo que qualitativamente diferente, a relação mereológica é conservada. Enquanto uns afirmam que o barco de Teseu ao voltar não é o mesmo, pois os componentes foram trocados, outros sustentam ser o mesmo, pois a forma entre B e A é contínua. O mesmo ocorre com a edição se cada gene saudável substitui o defeituoso no mesmo local e na mesma posição, acredita-se que a pessoa resultante será contínua com a anterior, ou seja, $B=A$.

Posto isto, Holland, finalmente, estabelece que, se um número extremamente grande de material genético for substituído por material genético funcionalmente diverso, então haverá uma mudança radical que leva a produção de uma nova pessoa. Contudo, se um número menor de material genético for substituído, então, as chances de a identidade ser afetada também serão baixas. Logo, divisa-se na continuidade física e psicológica um limiar. A esse respeito, o filósofo expõe: [...] “na extremidade do *continuum*, a intervenção genética pré-natal é suficientemente radical para produzir uma outra pessoa” (2008, p. 168).

Dessa maneira, uma vez que se aceita existir um ponto limítrofe, embora não haja um modo preciso de determinar em que momento a relação quantitativo-qualitativa no processo de manipulação gênica é capaz de provocar uma ruptura no *continuum*, temos uma versão mais elaborada da

premissa central do argumento, a saber: algumas formas de intervenção genética pré-natal, segundo o grau, o tipo e a função do material genético, podem alterar ou não a identidade qualitativa e numérica da futura pessoa (HOLLAND, 2008, p. 170).

Quando, então, os genes defeituosos do bebê Alan Foster³³ foram substituídos por genes saudáveis, houve uma ruptura na continuidade entre Alan-2 e Alan-1, de modo que a pessoa resultante Alan-2 não é o mesmo que a pré-pessoa Alan-1. Seria esta ruptura uma forma de assassinato da identidade pessoal, visto que, de certa maneira, a existência presente de alguém foi erradicada?

O exame da decisão do casal Foster a partir dos dois princípios morais apresentados por Holland (2008, p.169) iluminará a questão. Vejamos no próximo tópico.

4.5 Aspectos morais da edição genética da identidade pessoal implicados pelo experimento

De acordo com Holland há dois princípios morais similares que fundamentam a decisão pela intervenção genética pré-natal (2008, p. 169-170).

O primeiro é o princípio de não-maleficência, segunda a qual a decisão pela intervenção genética sustenta-se na obrigação moral de evitar que o indivíduo sofra dano. Por exemplo, na visão de Brenda e Jonathan Foster, a síndrome de Down prejudicaria o desenvolvimento social e cognitivo de Alan. Em contrapartida, a edição genética evitaria esse prejuízo. Por essa razão, se a identidade pessoal de Alan será afetada, não é importante, desde que o embrião esteja vivo e saudável. Em síntese, visar a qualidade da criança bem como dos pais é a razão central que impulsiona a escolha.

O segundo princípio é a beneficência, para a qual a pessoa resultante da edição genética beneficia-se da alteração sofrida, pois, devido a ela terá uma vida satisfatória. Este é um princípio com dois pesos e duas medidas. Se, de um lado, a intervenção genética pré-natal beneficia a pessoa resultante; do outro, é danosa a atual *pré*-pessoa que, graças à edição, cessa de existir. É, de certo modo, como se um mesmo leito de hospital fosse

33 Krikpe explica ainda que Alan Foster é um designador rígido, porque em todos os mundos possíveis, Alan Foster se refere necessariamente a um mesmo indivíduo. KRIPKE, Saul A. *Naming and necessity*. In Darragh Byrne & Max Kölbel (eds.), *Philosophy*. Routledge, 1980, pp. 431-433.

ocupado por dois pacientes semelhantes, até que um ceda o espaço para o outro. Nesse contexto, Alan-1 teve a existência erradicada para dar lugar a Alan-2, o que leva a afirmar que a substituição de um material genético por outro pode ser entendido como algo equivalente ao aborto; embora seja um argumento inconsistente e dubio, pois a vida do embrião não foi perdida, apenas deu origem a uma outra pessoa. Assim, enquanto Alan-1 está como possibilidade para o real, Alan -2, que emerge sobre a não-existência do primeiro, está agora como real para a possibilidade. Antes a situação era inversa, Alan-1 estava como realidade para o possível, mas a edição genética afetou a balança.

Por outro lado, a decisão do casal Foster também é motivada pela noção dos mundos possíveis. Ou seja, em qual dos muitos mundos possíveis Alan Foster terá uma vida plena. A escolha claramente se baseia na crença de que um dos muitos mundos possíveis é o melhor para Alan existir. Contudo, é um argumento fraco e incerto para respaldar uma decisão. Qualquer dos muitos mundos possíveis pode ser o melhor ou o pior para Alan Foster. Mas Brenda e Jonathan acreditam que um destes mundos, onde a pessoa resultante terá um corpo saudável e cognição adequada é o melhor para o bebê existir. Este será o mundo em que Alan-2 emerge. Além disso, a decisão baseada na seleção dos muitos mundos possíveis desconhece que, no exato momento em que um mundo é selecionado, todos os outros – de certa forma – cessam de existir.

Entretanto, segundo o autor

A intervenção genética pré-natal é mais ou menos drástica, dependendo do ponto do *continuum* em que se situe. Eliminar, por exemplo, minha psoríase por meio da manipulação genética não seria drástico. Uma intervenção genética pré-natal que alterasse a identidade o seria. [...] Há também um conjunto de deveres inteiramente separado para com pessoas deficientes existentes. Somos obrigados a lhes proporcionar tudo o que é preciso para maximizar sua qualidade de vida (HOLLAND, 2008, p. 172).

Inicialmente, Holland indaga se a terapia genética no estágio precoce da embriogênese resultaria em uma alteração na identidade do indivíduo. O filósofo afirma ao fim de seu artigo que a crítica realizada, no que tange à intervenção genética pré-natal, se mostra inconclusa quanto à intuição e à aplicação teórica. Ao levar em consideração diversos casos reais, o

princípio concernente à pessoa favorece a terapia genética, que altera a identidade pessoal, não como uma forma de matar, já que outra pessoa é também criada (2008, p. 173).

4.6 A objeção de Peter Singer e Helga Kuhse

Se para Stephen Holland (2008) a terapia gênica altera a identidade pessoal do embrião quando realizada no início de sua formação, de tal modo que a pessoa resultante é descontínua numérica e psicologicamente com a anterior, para os filósofos Peter Singer e Helga Kuhse (1990) tal questão não parece ser válida. Especificamente, de acordo com Singer e Kuhse, a identidade de um embrião só deve ser considerada a partir do 8º dia da fecundação, antes disso o zigoto é totipotente, ou seja, suas células podem se tornar potencialmente um ou mais indivíduos. Soma-se à tese que o embrião só terá uma estrutura psicológica quando o sistema nervoso estiver formado para que se torne senciente, antes disso não existe nenhuma fisiologia que possibilite a continuidade psicológica de B em relação a A (SINGER e KUHSE, 1990, p.66). Ainda, é pertinente destacar que, a premissa de Holland está condicionada ao momento em que a intervenção genética acontece, se em t0 ou t1.

Para ser mais claro, Singer e Kuhse sustentam que a pessoa de hoje que somos não é a mesma que o embrião inicial. Segundo eles, o zigoto original é um ser unicelular totalmente desprovido de consciência, enquanto a pessoa é um ser consciente constituída por milhões de células. Consequentemente, Singer e Kuhse supõem que a afirmação de o zigoto e a pessoa resultante serem o mesmo indivíduo deve se apoiar em duas noções que fundamentam a tese da identidade; é claro, diferente do que eles concebem (1990, p. 66).

A primeira se refere a ideia de continuidade genética entre o zigoto e a pessoa, ou seja, ao fato de o ser original e o ser final compartilharem o mesmo material genético.

A segunda afirma haver uma continuidade numérica entre o zigoto original e a pessoa resultante, de modo que, são uma e mesma substância.

Assim sendo, o zigoto tanto é um indivíduo em potencial como também desde o instante em que passou a existir é um indivíduo singular. Singer e Kuhse, contudo, discordam dessa tese da identidade, na visão deles, o zigoto não é uma pessoa, é o que no começo deste capítulo chamamos

de *pré*-pessoa. Como fundamento declaram existir pesquisas científicas que desmentem a crença de que a fecundação “marca o evento de quando um indivíduo particular e identificável começa a existir” (1990, p. 66). Em outras palavras, um zigoto pode ser Anderson, Gustavo ou Bruno. Além disso, antes da fecundação havia duas entidades distintas, o óvulo e o espermatozoide, agora são uma e a mesma substância, o zigoto, o qual contém um código genético novo e único, não sendo, portanto, nem geneticamente nem numericamente contínuo com o óvulo e o espermatozoide, sendo tão somente contínuo geneticamente com a pessoa resultante.

Se é verdade que há também continuidade numérica entre o zigoto e a pessoa resultante, examinemos, pois, o seguinte cenário que põem em dúvida essa assertiva.

Imagine, dessa vez, que Brenda e Jonathan têm relações sexuais e o zigoto é gerado, o qual recebe o nome de Alan. A identidade genética de Alan irá ser duplicada em todas as células do primeiro ao oitavo dia. Suponha então que no oitavo dia o grupo de células se divide em dois grupos idênticos separados. Considere ainda que o segundo grupo de células manifesta a trissomia 21. Os dois grupos se desenvolvem e ao fim de 38 semanas, Brenda dá à luz a gêmeos idênticos, contudo um deles têm síndrome de Down. Qual dos dois é Alan Foster?

É claro que um e outro são contínuos geneticamente com o Alan original, mas são numericamente idênticos? Na ótica de Singer e Kuhse supor que ambos são contínuos numericamente contradiz a tese da identidade. Outrossim, temos nesse exemplo, dois indivíduos que se distinguem pelo que pode ser chamado, por falta de palavra melhor, saudável e deficiente. Assim sendo, o Alan original (ou A) gerou B' e B'', e tanto B' quanto B'' são geneticamente idênticos a A, porém numericamente distintos entre si e de A.

Disso se depreende que, “quando a célula original se dividiu, [Alan] deixou de existir e dois novos indivíduos [...] passaram a existir” (grifo nosso) (SINGER e KUHSE, 1990, p. 67, tradução nossa). Por conseguinte, B' e B'' começaram a existir apenas no oitavo dia da fertilização. Nesse contexto, antes do 8º dia (t0) e depois do 8º dia (t1) há uma cisão na identidade numérica, porém não na identidade genética. Isto é de fundamental importância para a questão desse capítulo.

Outro ponto relevante é: as células embrionárias iniciais serem identificadas como totipotentes, ou seja, um embrião humano em vez de ser apenas um único indivíduo “tem o potencial de ser um ou mais diferentes indivíduos” (SINGER e KUHSE, 1990, p. 67). Em outras palavras, o zigoto pode produzir 1 ou 4 indivíduos e após a clivagem, 8 indivíduos, além do que as células podem se reunir em 1 ou 2 indivíduos. Ademais, no 14º dia, quando a linha primitiva é formada, isto é, quando o indivíduo se consolida, a totipotência celular é perdida e, a partir de então, o embrião se desenvolverá. Portanto, a hipótese de que a identidade numérica é relevante desde o zigoto inicial como parece sugerir Stephen Holland (2008), para questionar a terapia gênica, enfraquece quando se considera o momento certo na fase embrionária para a intervenção genética, o momento em que não há indivíduo algum. Nessa ótica, se a edição genética de Alan Foster é realizada entre o primeiro e o oitavo dia, de modo nenhum significará a erradicação de uma pessoa, porque sequer ela existe. O indivíduo que virá do embrião editado é o Alan original, pois não havia nenhum outro antes.

Há ainda a crença de o embrião ser uma pessoa desde a fertilização. Acreditar nisso motiva a defesa de a terapia gênica ser uma forma de erradicação da pessoa para a geração de uma totalmente diferente. No entanto, Singer e Kuhse discordam desta perspectiva. Para eles (1990, p. 70), o embrião inicial não é um ser com as qualidades mentais que o distinguem de seres não-humanos, pois não possuem um sistema nervoso central desenvolvido. Desse modo, ele não está consciente neste período, visto que estas qualidades surgem posteriormente.

Nessa linha de pensamento, “a característica mínima necessária para dar ao embrião o direito à consideração é a senciência, ou seja, a capacidade de sentir dor ou prazer” (SINGER e KUHSE, 1990, p. 71). Logo, apenas quando o embrião é senciência, tem o direito a ter sua identidade pessoal mantida, pois assinala que o indivíduo tem experiências e memórias. Em outro caso, não há porque não intervir em sua constituição genética. Ademais, se a edição não afetar a experiência da pessoa, então, não há dano. Assim, a edição genética de Alan Foster não representaria risco à identidade numérica como também à psicologia do indivíduo se realizada na fase embrionária inicial.

5. CONCLUSÃO

Em suma, baseado na tese de Parfit e no argumento de Holland, a edição genética pode funcionar como a máquina de teletransporte das obras de ficção científica, quando, a partir de uma substituição considerável do material genético do embrião, uma outra pessoa numérica e qualitativamente diferente emerge. Outrossim, devemos pressupor que há uma extremidade no *continuum* da identidade pessoal e, por isso, certas formas de intervenção genética pré-natal podem alterar a identidade da *pré*-pessoa, como o método de edição genética CRISPR-Cas9.

Por conseguinte, a história de Alan Foster revela a importância de considerar a condição nodal para a realização como a não realização da alteração genética, a saber: se a edição vale a pena ser aplicada mesmo que cause a substituição da pessoa atual por uma radicalmente diferente. É claro, a decisão final depende de a razão que a fundamenta ser impessoal, ou, concernente à pessoa, somado ao impacto que terá sobre a família e a sociedade.

Todavia, para Singer e Kuhse, o embrião inicial não é uma pessoa, só depois de um momento preciso é que se torna alguém. E, justamente o intervalo entre t_0 e t_1 permite a intervenção genética pré-natal sem haver preocupação quanto à identidade numérica do conceito.

Destarte, não haveria porque a decisão de Brenda e Jonathan se basear em se a pessoa resultante é Alan-1 ou Alan-2, visto o indivíduo gerado ser definitivamente o original. Mais especificamente, a terapia gênica pode não ser danosa à identidade pessoal, uma vez que realizada no começo da embriogênese.

6. REFERÊNCIAS

DICK, Phillip K. **O Pagamento**. Rio de Janeiro: Ed. Record, 2004.

HOLLAND, Stephen. **Bioética: enfoque filosófico**. São Paulo: Edições Loyola, 2008

KUHSE, H.; SINGER, P. Individuals, humans and persons: The issue of moral status. In P. Singer, H. Kuhse, S. Buckle, K. Dawson, & P. Kasimba (Eds.), *Embryo Experimentation*, 1990, p. 65-75. **Cambridge: Cambridge University Press**. doi:10.1017/CBO9781139172325.009

KRIPKE, Saul A. **Naming and necessity**. In Darragh Byrne & Max Kölbel (eds.), *Philosophy*. Routledge, 1980, pp. 431-433.

KLEINMAN, Paul. **Tudo o que você precisa saber sobre Filosofia: de Platão e Sócrates até a ética e metafísica, o livro essencial sobre o pensamento humano**. São Paulo: Ed. Gente, 2014.

MASLIN, Keith T. **Introdução à Filosofia da Mente**. 2ª ed., Porto Alegre: Ed. Artmed, 2009.

PARFIT, Derek. **Reasons and Persons**. Oxford University Press, 1984.

PARFIT, Derek. **Reductionism and personal identity**. In David J. Chalmers (ed.), **Philosophy of Mind: Contemporary Readings**. Oxford University Press, 2002, pp. 655-51.

PARFIT, Derek. **Identidade Pessoal**. “Personal Identity”, *The Philosophical Review*, 80, 1971, p. 3-27. Trad de Pedro Galvão.

Disponível em: <http://pedrogalvao.weebly.com/uploads/6/6/5/5/6655805/parfit-ip-1971.pdf> acesso em 14/05/2020.

PENROSE, Roger. **A Mente Nova do Rei: Computadores, mentes, e as leis da Física**. Rio de Janeiro: Editora Campus, 1991.

STRAWSON, Peter F. **Indivíduos: Um ensaio de metafísica descritiva**. São Paulo: Ed. Unesp, 2019.

Leimin Tian

Traduzido e revisado por:

*Matheus Moura Novelli**Stefania Salvador Pereira Montenegro*

1. Computação Afetiva

*A questão não é se as máquinas inteligentes podem ter quaisquer emoções,
mas se as máquinas podem ser inteligentes sem quaisquer emoções.*

— Marvin Minsky, *Society of Mind* (1988)

A Computação Afetiva se originou em 1995 pelo Prof. Rosalind Picard em seu artigo (PICARD, 1995) e livro (PICARD, 2000) publicado sob o mesmo nome. Ela definiu a computação afetiva como “a computação que se relaciona com, surge de, ou influencia, as emoções”. Os estudos de computação afetiva enfocam a compreensão das emoções humanas, como a forma como as pessoas percebem e expressam emoções durante a comunicação interpessoal, bem como podemos incorporar esse conhecimento à inteligência artificial, por exemplo, como um robô pode detectar os estados emocionais de seu usuário e fornecer serviços em conformidade.

Nos últimos 25 anos, avanços significativos foram alcançados na pesquisa de computação afetiva. Por exemplo, modelos de reconhecimento automático de emoções foram capazes de atingir desempenho de nível humano em algumas tarefas: Beard *et al.* (2018) compararam o desempenho humano e de uma máquina ao classificar 8 categorias de emoções retratadas por atores que falaram uma frase curta acompanhada de expressões faciais e vocais. Usando uma Rede Neural Recorrente de Memória ao Longo Prazo, o reconhecedor automático de emoções alcançou maior acurácia de classificação do que os humanos na classificação baseada em áudio, bem como na classificação baseada em audiovisual, conforme mostrado na Tabela 1.

Modalidade	Humano	Maquina
Áudio	40,9%	41,5%
Visual	58,2%	52,5%
Áudio-Visual	63,6%	65,0%

Tabela 1: Precisão humana vs. máquina na classificação de 8 categorias de emoção (BEARD, 2018).

Como um tópico de pesquisa interdisciplinar, a computação afetiva baseia-se no conhecimento de vários domínios, especialmente neurociência, psicologia e inteligência artificial, conforme mostrado na Tabela 2. Hoje, a computação afetiva é um campo de rápido crescimento que atrai interesses científicos e de aplicação, como mostrado nas Figuras 1 e 2. Para encorajar colaborações e esforços comunitários, a Associação para o Avanço da Computação Afetiva (AAAC)³⁴ foi estabelecida para pesquisadores de computação afetiva em todo o mundo, juntamente com sua Conferência Internacional bianual sobre Computação Afetiva e Interação Inteligente (ACII) e as Transações IEEE de alto impacto em Computação Afetiva. Devido a sua contribuição inestimável para a comunidade de pesquisa de computação afetiva, a professora Rosalind Picard foi premiada como uma das primeiras AAAC Fellows no ACII 2019.

Com o desenvolvimento técnico amadurecendo rapidamente, começamos a ver as aplicações do mundo real da pesquisa em computação afetiva. Por exemplo, Affectiva³⁵, uma empresa fundada em 2009, fornece softwares que permitem a análise de emoções para clientes em vários domínios, incluindo na indústria de jogos, educação, mídia e saúde. Agora é o momento crucial para traduzir os avanços da pesquisa em computação afetiva para beneficiar o bem-estar dos indivíduos e da nossa sociedade.

Em Tian *et al.* (2021), capturamos esse processo de transformação da computação afetiva do ambiente de laboratório controlado para cenários de aplicação *in-the-wild*, propondo um novo tópico de pesquisa, a Computação Afetiva Aplicada. Chamamos a atenção da comunidade de computação afetiva para os desafios árduos da aplicação dos resultados da pesquisa em cenários do mundo real. Neste livro, realizamos análises extensas sobre as abordagens

³⁴ <https://aaac.cs.nott.ac.uk/>

³⁵ <https://www.affectiva.com/>

existentes para o processamento de sinais sociais, incluindo o Aprendizado de Máquinas, coleta de dados e processamento de sinais. Isso fornece referências para pesquisadores que trabalham em tópicos relevantes sobre o estado da arte e os principais desafios da área. Além disso, cobrimos estudos de caso aprofundados em vários campos de aplicação da computação afetiva, como robótica social, saúde e educação. Para pesquisadores novos no campo da computação afetiva ou aqueles que gostariam de aplicar a computação afetiva aos seus campos, nosso livro (TIAN, 2021) fornece orientações para o desenvolvimento de sistemas de inteligência artificiais centradas no ser humano de nova geração.

Área de referência	Porcentagem
Neurociência	17.1%
Psiquiatria	17.0%
Inteligência Artificial	14,5%
Engenharia Elétrica Eletrônica	12,7%
Métodos de Teoria da Ciência de Computação	9,1%
Sistemas de Informação	9,0%
Neurologia Clínica	8,0%
Psicologia	4,6%
Psicologia Experimental	4,5%
Aplicações Interdisciplinar da Ciência de Computação	4,2%

Tabela 2: As 10 principais áreas de referência em publicações de computação afetiva (excluindo autocitação).

Aprendizado de máquina, coleta de dados e processamento de sinais. Isso fornece referências para pesquisadores que trabalham em tópicos relevantes sobre o estado da arte e os principais desafios da área. Além disso, cobrimos estudos de caso aprofundados em vários campos de aplicação da computação afetiva, como robótica social, saúde e educação. Para pesquisadores novos no campo da computação afetiva ou aqueles que gostariam de aplicar a computação afetiva aos seus campos, nosso livro fornece orientação para o desenvolvimento de sistemas de inteligência artificial centrados no ser humano de nova geração.

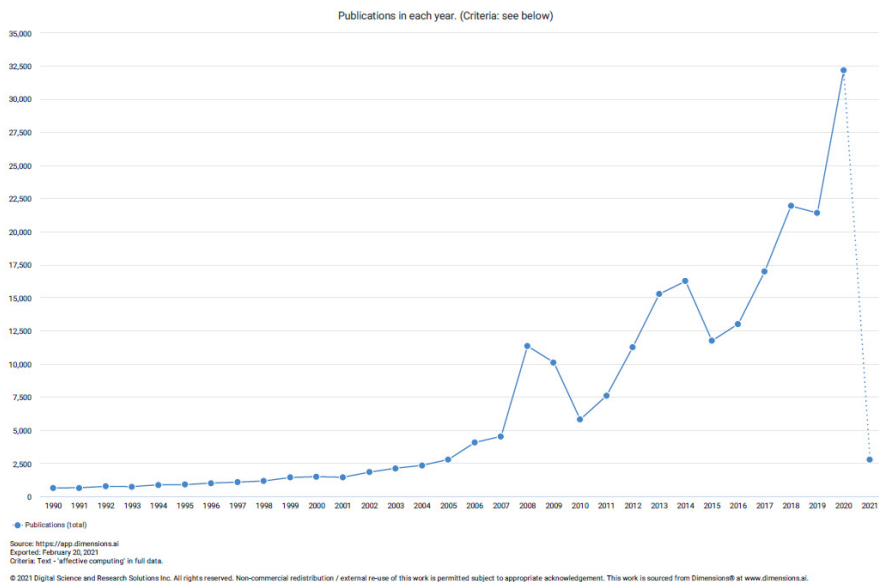


Figura 1: Quantidade de publicações contendo a palavra-chave “computação afetiva”.

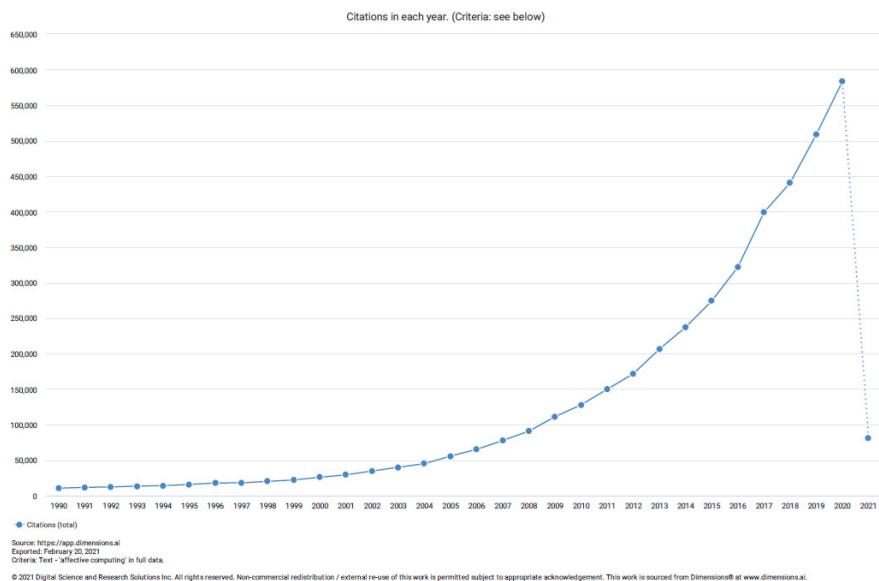


Figura 2: Número de publicações contendo a palavra-chave “computação afetiva” citadas em outras publicações.

2. Computação afetiva aplicada

Definimos computação afetiva aplicada como **esforços de pesquisa e aplicação que se concentram na compreensão da eficácia, resultados e impacto pessoal e social das tecnologias de computação afetiva no uso na vida real** (TIAN, 2021). Em comparação com a computação afetiva, a computação afetiva aplicada aborda o desafio de uma avaliação holística dos modelos computacionais de emoções situados no contexto. Além disso, a computação afetiva aplicada mantém uma perspectiva centrada no ser humano em vez de uma perspectiva centrada na tecnologia e se concentra em como os sistemas de computação afetiva influenciam os estados físicos e mentais de seus usuários. O objetivo da pesquisa de computação afetiva aplicada é desenvolver sistemas de computação afetiva que sejam éticos e capazes de alcançar os resultados pretendidos em ambientes naturais, não estruturados e complexos.

Aqui, fornecemos um exemplo que ilustra nossa definição de computação afetiva aplicada: O reconhecimento automático de emoções a partir da fala tem sido um tópico amplamente estudado na computação afetiva. Houve uma melhora significativa no desempenho do reconhecimento de emoções baseado na fala, conforme avaliado em vários conjuntos de dados de fala emocional (AKÇAY, 2020). Para a pesquisa de computação afetiva aplicada no reconhecimento de emoções pela fala, mudamos o foco da identificação de abordagens eficazes de reconhecimento de emoções comparadas em um conjunto de dados para a compreensão de como o desempenho de um modelo pode ser influenciado por fatores contextuais, por exemplo, características do usuário, como idade ou personalidade, ou tipo de discurso, como diálogo orientado para tarefas ou bate-papos sociais. Os principais fatores contextuais são decididos pelo objetivo da aplicação de um modelo de reconhecimento de emoção pela fala. Em um exemplo de sistema que visa identificar clientes frustrados em um call center, o modelo de reconhecimento de emoção precisará se acomodar a ruídos na linha, entre outros fatores. Além disso, além de avaliar a precisão do modelo de reconhecimento de emoção pela fala na identificação de um cliente frustrado em suas ligações, esse modelo precisa ser incorporado a outros componentes do sistema, como um gerenciador de diálogo que seleciona as estratégias de mitigação adequadas ou um conhecimento base de serviços

disponíveis específicos para o call center. Ao avaliar este sistema integrado em métricas como a satisfação dos chamadores, o resultado do sistema de reconhecimento de emoção pela fala é medido de uma maneira situada.

3. Transformação do laboratório para o mundo real

Facilitar a transformação da computação afetiva em computação afetiva aplicada requer esforços colaborativos. A Figura 3 ilustra os desafios nos vários aspectos da computação afetiva aplicada. A fim de desenvolver um sistema de computação afetivo eficaz e ético, as principais questões que os pesquisadores e profissionais precisam abordar incluem:

- Ter um bom entendimento dos usuários e quais são suas expectativas;
- Ter funções que mantenham seu desempenho em ambientes variados;
- Ter a capacidade de aprender com as observações acumuladas;
- Ter um sistema integrado que combina afeto com outros processos comunicativos e cognitivos e com percepções de domínios relacionados;
- Ter uma boa compreensão dos impactos sociais das tecnologias de computação afetivas.

Computação Afetiva Aplicada



Figura 3: Os vários aspectos da computação afetiva aplicada.

3.1 Desenvolvendo um sistema de computação afetiva centrado no ser humano

Os sistemas de computação afetiva tratam da emoção, cognição e comunicação humana. Os sistemas de computação afetiva aplicada são frequentemente projetados com o propósito de monitorar ou influenciar os comportamentos e/ou estados mentais de seus usuários. Portanto, eles requerem uma perspectiva centrada no ser humano durante os processos de design e avaliação para atender às necessidades e variações dos usuários e para se ajustar aos cenários de aplicação específicos (NORMAN & DRAPER, 1986). A nova área de pesquisa de análise comportamental (OVIATT, 2018) lançou luzes sobre o desenvolvimento recente de interfaces multimodais centradas no ser humano e suas aplicações em domínios como saúde e educação. Para desenvolver um sistema de computação afetiva centrado no ser humano, os usuários devem estar envolvidos nos processos de design e avaliação para incorporar suas expectativas e necessidades. O sistema deve abordar as variações e preferências individuais e incluir as diferentes características do usuário. Por exemplo, as expressões emocionais geradas por um sistema projetado para fins de tratamento da demência devem utilizar expressões multimodais mais pronunciadas para abordar o possível declínio na percepção emocional de um paciente com demência (SMILJANIC, 2017; YATSUDA, 2018). Para lidar com variações e preferências individuais, tem havido esforços crescentes na realização da personalização em sistemas interativos (AUGSTEIN, 2019; IRFAN, 2019).

3.2 Desenvolvendo um sistema de computação afetiva confiável

Para que um sistema de computação afetiva saia dos ambientes de laboratório controlados e atinja seus resultados projetados quando aplicado em ambientes da vida real, é fundamental que suas funcionalidades continuem a oferecer um desempenho confiável. O tópico de reconhecimento de emoção *in-the-wild* tem recebido interesse crescente na comunidade de computação afetiva, que aborda um aspecto do desenvolvimento de sistemas de computação afetiva confiáveis (PICARD, 2011). Tem havido esforços de *benchmarking*, como o reconhecimento de emoção nos desafios selvagens (EmotiW) (DHALL, 2020), na comunidade de pesquisa para desenvolver modelos de reconhecimento de emoção robustos e precisos que são resistentes a ruídos e

variações. Para garantir a confiabilidade das abordagens de reconhecimento das emoções, metodologias de avaliação mais rigorosas são adotadas em estudos recentes, como a análise de corpora cruzada (EYBEN, 2015). Além disso, os pesquisadores se beneficiaram do avanço em plataformas de *crowdsourcing* e vasto conteúdo online para gerar conjuntos de dados de alta fidelidade de experiências e interações afetivas *in-the-wild*, como o conjunto de dados CMU-MOSEI (ZADEH, 2018). Para garantir a qualidade da anotação de dados, várias medidas estão em vigor para contabilizar a subjetividade e incerteza durante a anotação de uma emoção e durante o processamento e modelagem de dados, como reter uma distribuição de rótulos para uma instância de dados em vez de um rótulo agregado por instância e modelando tal distribuição (RIZOS & SCHULLER, 2020).

3.3 Desenvolvendo um sistema de computação afetiva adaptativa

Muitos sistemas de computação afetiva aplicada são projetados para estabelecer ou manter um relacionamento social com os usuários para atingir seus objetivos pretendidos. Por exemplo, sistemas de tutoria inteligentes capazes de construir um bom relacionamento com os alunos tutorados podem levar ao aumento da confiança, experiência de aprendizagem e resultados de aprendizagem (PAN & HAMILTON, 2018; ZHAO, 2016). Para desenvolver um relacionamento social com os usuários, são necessárias interações longitudinais. As interações de longo prazo também são comuns em cenários de aplicação, como saúde mental e física, onde um programa de tratamento pode conter várias sessões ao longo de meses ou até anos. Em tal interação longitudinal, é necessário que um sistema de computação afetiva aplicada aprenda e se adapte às dinâmicas e contextos de interação em mudança. Além disso, a capacidade de se adaptar e fornecer serviços longitudinais ajudará a resolver o desgaste no uso, que tem sido um grande problema em sistemas de computação afetiva comercializados, como os robôs sociais (HOFFMAN, 2020).

3.4 Desenvolvendo um sistema de computação afetiva integrado

Na computação afetiva aplicada, a emoção não é modelada isoladamente. Em vez disso, está embutida em uma estrutura integrada que inclui outros componentes que monitora ou simula estados mentais, cognições

e funções comunicativas. Por exemplo, um agente de conversação social requer capacidades de diálogo, uma base de conhecimento de tópicos potenciais e a capacidade de abordar o contexto, como as intenções dos usuários, além de capacidades afetivas e habilidades sociais (COHEN, 2018). Da mesma forma, o projeto de expressões emocionais de um robô social precisa abordar a aparência física e as capacidades de hardware do robô (SONG & YAMADA, 2017). A computação afetiva aplicada também é um tópico de pesquisa interdisciplinar, onde esforços colaborativos de domínios incluindo, mas não se limitando a psicologia, inteligência artificial, sistemas interativos, robótica, engenharia, ciências sociais e ciências médicas são essenciais para desenvolver hipóteses significativas e metodologias sólidas. Por exemplo, as percepções dos terapeutas são críticas durante o design iterativo e melhoria das aplicações de bem-estar mental (FITZGERALD & MCCLELLAND, 2017).

3.5 Desenvolvendo um sistema de computação afetivo ético

Com o aumento da aplicação de tecnologias afetivas, tem havido preocupações crescentes sobre o potencial mau uso da tecnologia e vieses no sistema (ZUBOFF, 2019), como a proibição de reconhecimento facial e tecnologias de análise (CONGER, 2019). Assim, é fundamental entender o impacto social da computação afetiva aplicada e estabelecer medidas apropriadas, que garantam que os pesquisadores sejam orientados para a pesquisa responsável e os profissionais sejam regulamentados para aplicações éticas com benefícios sociais. Como um campo jovem, a computação afetiva pode aprender com campos intimamente relacionados, como a interação humano-computador. A comunidade de interação humano-computador enfrenta desafios semelhantes para garantir um futuro ético de sistemas interativos inteligentes e identificou que a colaboração entre as partes interessadas, como pesquisadores, designers, usuários e formuladores de políticas, é a chave para lidar com as crescentes preocupações em relação à justiça, responsabilidade e transparência dos sistemas inteligentes (ABDUL, 2018).

4. Diretrizes para computação afetiva aplicada

Nesta seção, apresentamos um conjunto de diretrizes para cada fase da computação afetiva aplicada. Além disso, usando três cenários hipotéticos, fornecemos uma discussão baseada em cenários sobre como pesquisadores de computação afetiva novatos ou experientes podem adotar essas diretrizes em seus próprios estudos, bem como como os consumidores de pesquisas em computação afetiva podem adotar essas diretrizes em seu desenvolvimento das aplicações de computação afetiva.

I. Diretrizes da fase de design

1. A computação afetiva é profundamente interdisciplinar. Pesquisadores e praticantes devem basear suas hipóteses ou projetos de sistema em teorias verificadas da emoção e interação humana, bem como no conhecimento de especialistas do domínio.

2. Além de serem consumidores de teorias de emoção e interação, estudos e aplicações de computação afetiva podem contribuir potencialmente para a descoberta de novas teorias.

3. Hipóteses em computação afetiva precisam ser adaptadas ao estudo específico ou contextos de aplicação, como características do usuário ou cenários de aplicação.

4. Abrace a multimodalidade na concepção de modelos de reconhecimento e expressão de emoções.

5. Além de analisar os afetos humanos para informar as decisões e comportamentos de um sistema envolvendo seus usuários, os modelos computacionais de emoção também podem ser aplicados ao projeto de processos de tomada de decisão internos de agentes inteligentes para permitir a geração de comportamentos complexos ou adaptativos.

6. Projete componentes emocionais integrados com outros componentes de um sistema, como compreensão de linguagem natural e geração em um sistema de conversação.

7. Se um estudo ou cenário de aplicação for definido em um ambiente menos controlado, preste atenção à robustez e no estudo ou projeto do sistema, por exemplo, adotando um modelo de reconhecimento de expressão facial resistente à oclusão.

8. Se um estudo ou cenário de aplicação for longitudinal, preste atenção à adaptabilidade no projeto do sistema, por exemplo, incluindo um modelo de aprendizagem por reforço para aprender com os comportamentos do usuário.

II. Diretrizes da fase de experimento e análise

1. Garanta o alinhamento entre os dados usados para treinar uma função emocional e seus cenários de aplicação para evitar discrepâncias de desempenho entre o treinamento e a avaliação.

2. Se possível, faça uma avaliação extrínseca das funções emocionais embutidas em um sistema integrado de computação afetiva.

3. Esforços comuns na coleta de dados emocionais de alta qualidade em ambientes multitemporais e naturalistas são essenciais para o avanço da computação afetiva aplicada. Além disso, para apoiar a aprendizagem supervisionada das funções emocionais, esses dados precisam ser rotulados com esquemas de anotação confiáveis apoiados por teorias da emoção.

4. A computação afetiva aplicada se beneficiou muito dos avanços em aprendizado de máquina e pesquisa de aprendizado profundo. Melhorias adicionais podem ser alcançadas por meio do desenvolvimento de processamento de dados adaptável e em tempo real e abordagens de fusão de modalidades, bem como o desenvolvimento de novos modelos capazes de lidar com a subjetividade, ambiguidade e incerteza.

III. Diretrizes da fase de aplicação

1. Seja cauteloso se as descobertas na pesquisa de computação afetiva podem ou não ser transferidas para aplicações afetivas. Entenda o alinhamento e as desconexões em termos de contexto de aplicativo e usuários-alvo usando abordagens baseadas em cenários para orientar a tradução da pesquisa de computação afetiva.

2. Permitir que um sistema de computação afetiva aprenda e se adapte às variações do ambiente de maneira autossupervisionada é fundamental para manter sua eficácia. As abordagens centradas no ser humano também podem beneficiar os aplicativos afetivos, permitindo que se adaptem ao contexto de usuário em constante mudança.

3. Abordagens iterativas e abordagens baseadas em simulação podem apoiar o rápido desenvolvimento de sistemas complexos de computação afetiva.

4. Adote plataformas multimodais e multissensoriais para aplicações afetivas mais precisas e robustas.

5. Um objetivo de aplicação, como melhorar o bem-estar mental de alguém, muitas vezes requer uma estrutura multidisciplinar que investiga o resultado colaborativo de funções afetivas, comunicativas e interativas, juntamente com orientações diferenciadas sobre o domínio e o conhecimento específico do contexto.

6. Para garantir uma aplicação afetiva ética, é fundamental entender como isso vai influenciar seus usuários a curto e longo prazo, e quais são os impactos sociais. A pesquisa e os aplicativos de computação afetivos devem ser testados estrategicamente e sistematicamente para garantir resultados confiáveis, que também podem informar as políticas e regulamentações que evitam danos.

No primeiro cenário hipotético, a Pesquisadora A é uma geóloga humana. Ela está interessada em entender como as tecnologias de casa inteligente influenciam as pessoas que vivem sozinhas durante o segundo **bloqueio** COVID-19 em Melbourne. Isso a leva ao campo da computação afetiva. Depois de obter conhecimento prévio das abordagens e ferramentas de pesquisa de computação afetiva disponíveis, a Pesquisadora A começa a projetar seu estudo seguindo as diretrizes da fase de projeto. Sua hipótese, enraizada na geografia humana, pode ser que as pessoas exibirão um aumento na antropomorfização de objetos domésticos durante o bloqueio. Em seu projeto de estudo, A decide adotar uma nova abordagem de análise de dados multimodais coletados por vários sensores, como movimentos dentro de casa, dados fisiológicos de relógios inteligentes, gravações de alto-falantes inteligentes ou trajetórias de aspiradores de pó robóticos, além das metodologias de análise qualitativa tradicionais de pesquisa em geografia humana baseada em entrevistas com participantes. Durante seu estudo, A anota os dados com um conjunto de rótulos afetivos específicos para sua pesquisa, como empatia, aproximação ou evitação. A realiza uma análise quantitativa da relação entre esses rótulos e os comportamentos e percepções dos participantes usando as ferramentas apresentadas neste livro. Sua análise quantitativa indica que os participantes mostraram mais empatia em seu diálogo com o alto-falante inteligente e seu movimento se sobrepôs à trajetória do aspirador de pó robô mais durante o **bloqueio** em comparação com

antes do **bloqueio**. A análise qualitativa das entrevistas também mostra que a maioria dos participantes nomeou seu alto-falante inteligente e aspirador de pó robô, e atribuiu estados mentais a esses objetos ao discuti-los. Essas observações convergem em alinhamento com sua hipótese. Ao relatar suas descobertas, seguindo as recomendações, A compartilha os dados com a comunidade de pesquisa após remover qualquer informação identificável.

No segundo cenário hipotético, o Pesquisadora B é uma pesquisadora experiente em computação afetiva. Ela leu as publicações da Pesquisadora A e obteve acesso ao conjunto de dados da casa inteligente. B está interessada em identificar a solidão a partir de pistas comportamentais e fisiológicas. Ela anota o conjunto de dados com rótulos adicionais de ocorrência e intensidade de solidão. B, então, conduz experimentos para analisar o desempenho de vários modelos de aprendizado profundo para detectar a solidão neste conjunto de dados. Seguindo as diretrizes da fase de experimento e análise, ela combina recursos acústicos e lexicais das gravações de alto-falante inteligente, bem como recursos fisiológicos extraídos dos dados de frequência cardíaca e condutividade da pele coletados pelo relógio inteligente. Depois que B identificou um modelo confiável para detecção automática de solidão, ela realiza análises adicionais além do conjunto de dados da casa inteligente. Em particular, B testa a robustez e o fator em tempo real do modelo de detecção de solidão, incorporando-o nos sistemas das casas inteligentes de vários participantes por uma semana. Os participantes são solicitados a preencher uma avaliação subjetiva de seus estados mentais quando o modelo indica um alto nível de solidão. Além disso, os psicólogos são convidados a revisar as gravações desses participantes e avaliar seus estados mentais com base em suas observações. B então estuda a convergência entre as previsões automáticas, os autorrelatos e as avaliações dos especialistas. Os resultados revelam que o modelo de detecção de solidão tem menor acurácia neste experimento adicional. Uma investigação adicional indica que isso é causado pelos participantes não estarem em um cenário de **bloqueio** durante o experimento adicional, que é diferente do contexto do conjunto dos dados da casa inteligente que a Pesquisadora A coletou. B relata suas descobertas no conjunto de dados da casa inteligente e durante a avaliação adicional, e compartilha os códigos-fonte do modelo de detecção de solidão.

No terceiro cenário hipotético, a Praticante C está desenvolvendo um robô companheiro para idosos que vivem em casa sozinhos. Ela leu as publicações

da Pesquisadora B e obteve acesso ao modelo de detecção da solidão. Ela está interessada em incorporar este modelo no robô para que ele possa se engajar em estratégias que proporcionem a companhia no momento certo. Seguindo as diretrizes da fase de aplicação, C compara o contexto dos experimentos da Pesquisadora B e seus casos de uso alvo. Ela descobre que no conjunto de dados da casa inteligente da Pesquisadora A e nos experimentos adicionais da Pesquisadora B, os participantes têm todos entre 20 e 50 anos, enquanto os usuários-alvo do robô de C têm entre 60 e 80 anos. Portanto, C convida pessoas que vivem sozinhas entre 60 e 80 anos, bem como psicólogos e geriatras para uma oficina de design participativo. Durante este workshop, C coleta feedback dos participantes sobre solidão e companheirismo no cuidado ao idoso domiciliar. Usando esses insights, C revisa o modelo de detecção de solidão e a estratégia de engajamento do robô. Ela então implanta o robô protótipo na casa de vários usuários por um período de um mês com o consentimento deles e os entrevista antes, durante e após a implantação. As medidas de saúde dos usuários, como a pressão arterial, também são monitoradas e um psicólogo acompanha todo o período de implantação. Usando os dados coletados durante a implantação, C identifica outros ajustes no modelo de detecção de solidão e na estratégia de engajamento. Além disso, ela inclui um componente de personalização no robô que permite que ele atualize periodicamente seu modelo de detecção de solidão e a estratégia de engajamento com base em observações acumuladas. C, então, implanta o robô nas casas dos participantes por um período maior de três meses com protocolos semelhantes à primeira implantação. Depois de algumas iterações, a Praticante C finaliza o projeto do robô companheiro para produção.

Conforme demonstrado nesta discussão baseada em cenários, as diretrizes propostas para a computação afetiva aplicada apoiam os pesquisadores a conduzir estudos de computação afetiva de alta qualidade que beneficiam as descobertas de conhecimento, bem como os praticantes para projetar, desenvolver e avaliar sistemas de computação afetiva com uma perspectiva centrada. Abaixo, elaboramos mais sobre como os pesquisadores podem alcançar estudos empíricos válidos e generalizáveis com metodologias de pesquisa de computação afetiva rigorosa, e como os profissionais podem alcançar aplicações de computação afetivas éticas e eficazes, garantindo a justiça, transparência e responsabilidade de seus sistemas.

5. Rumo a metodologias de pesquisa de computação afetiva rigorosas

Em sua infância, a pesquisa em computação afetiva é frequentemente exploratória ou prototípica (HOFFMAN & ZHAO, 2020; LEDGERWOOD, 2017; SHROUT & RODGERS, 2018), por exemplo, estudos de computação afetiva com uma amostra pequena. No entanto, para abordar questões na aplicação realista de resultados de pesquisa, metodologias experimentais mais rigorosas e complexas são necessárias para melhorar a qualidade e robustez da pesquisa de computação afetiva. Como a computação afetiva está intimamente relacionada às ciências sociais e comportamentais, muitas vezes os pesquisadores da computação afetiva são consumidores de descobertas nesses campos, a recente “crise de replicação” em psicologia e ciências sociais (MAXWELL, 2015) (ou seja, os esforços para replicar estudos existentes muitas vezes não produz as mesmas observações ou conclusões) deve ter um impacto significativo na pesquisa de computação afetiva também. Para evitar que uma crise de replicação semelhante aconteça na computação afetiva, os pesquisadores são fortemente encorajados a adotar metodologias de pesquisa rigorosas para garantir resultados válidos e reproduzíveis e reduzir o desperdício de recursos causado pela necessidade de revisar ou verificar o trabalho inicial. Nesta seção, discutimos como as metodologias de pesquisa em computação afetiva podem ser aprimoradas em termos de validade, generalizabilidade e reprodutibilidade.

5.1 Hipóteses baseadas em teorias verificadas

Os estudos de computação afetiva são frequentemente baseados em teorias de emoção e comunicação. Por exemplo, Dobrosovetsnova e Hannibal (2020) adotaram teorias sobre como a emoção da decepção pode ser usada como uma estratégia pedagógica para um professor motivar um aluno, e desenvolveram uma estrutura para projetar robôs de tutoria que utilizam a mesma estratégia. Por um lado, fundamentar as hipóteses da pesquisa em computação afetiva com teorias em psicologia, comportamento e ciências sociais pode beneficiar a pesquisa, aumentando a força das hipóteses, especialmente para estudos que enfocam a compreensão ou influência na percepção e no comportamento das pessoas. No exemplo do robô tutor (DOBROSOVESTNOVA & HANNIBAL, 2020), inspirado em estudos humanos, quando o projeto de um robô considera uma ampla variedade

de emoções positivas e negativas, ele pode levar a experiências de aprendizagem e resultados mais benéficos para um usuário em comparação com um projeto de robô apenas lisonjeando superficialmente. Um exemplo é que assistentes domésticos inteligentes, como Alexa, foram criticados por aumentar o uso de tons rudes e indelicados pelas crianças durante sua comunicação com o sistema e com pessoas (GARG & SENGUPTA, 2020).

Por outro lado, permanecem diferenças fundamentais entre como as pessoas percebem e interagem com outra pessoa e como as pessoas percebem e interagem com um agente inteligente. Isso pode limitar a aplicabilidade das teorias derivadas da percepção humana e da comunicação no contexto da computação afetiva. No exemplo do robô tutor (DOBROSOVESTNOVA & HANNIBAL, 2020), a decepção expressa por professores humanos pode motivar um aluno sob as condições de que esse desapontamento está associado ao compromisso do professor de cuidar do bem-estar pessoal e do sucesso de aprendizagem do aluno, e da confiança do aluno nesse cuidado existente do professor (SUMNER, 1996). No entanto, um aluno pode não confiar que um robô tutor “cuida” de seus melhores interesses e, portanto, a decepção expressa pelo robô pode não levar ao mesmo efeito que a decepção expressa por um professor humano.

Além de considerar cuidadosamente a aplicabilidade das teorias da emoção para aumentar a força das hipóteses da computação afetiva, os pesquisadores da computação afetiva também devem ser críticos sobre os estudos das ciências psicológicas ou comportamentais que escolhem como base para seus estudos. Para identificar teorias de emoção confiáveis, é importante investigar se uma teoria foi ou não verificada por vários estudos com análise experimental robusta. Além disso, os pesquisadores devem prestar atenção se os contextos de estudo, como dados demográficos dos participantes, se alinham ou não entre o estudo da teoria da emoção e o estudo da computação afetiva, que discutiremos mais detalhadamente na próxima seção sobre a validade ecológica da pesquisa em computação afetiva.

5.2 Validade ecológica

Os sistemas de computação afetivos geralmente são projetados com um contexto de aplicativo específico em mente. Esse contexto pode ser um local de interação específico, como na casa de alguém ou em um hospital,

um grupo de usuários específico, como pacientes com demência ou trabalhadores de call center, ou um cenário de tarefa específico, como aprender um segundo idioma ou evacuação de emergência. Como as emoções e reações emocionais são dependentes do contexto, a validade ecológica de um estudo de computação afetiva influenciará diretamente a validade e generalização de suas descobertas. A validade ecológica representa a generalização das observações em ambiente de laboratório controlado para comportamentos naturais na natureza (CAMPBELL, 1957). Ele se baseia em três dimensões, o ambiente experimental, estímulos e respostas (SCHMUCKLER, 2001).

Em termos de validade ecológica do cenário experimental, um estudo deve ser conduzido em um ambiente que se assemelhe muito ao cenário-alvo, preferencialmente como estudos de campo. No estágio inicial de desenvolvimento ou quando o teste de campo é difícil de realizar, estudos de simulação que se aproximam do ambiente de destino podem ser conduzidos. Por exemplo, os pesquisadores podem desenvolver e avaliar a capacidade de um robô industrial de entender as intenções humanas para a interação homem-robô de proximidade segura em um ambiente de laboratório controlado primeiro (LASOTA, 2014). No entanto, eles devem ter em mente quaisquer diferenças entre as configurações de simulação e os contextos da vida real, e a possível influência que podem ter nas observações e conclusões que fazem nos estudos de simulação. No exemplo do robô industrial (LASOTA, 2014), um ambiente de laboratório é geralmente livre de quaisquer obstáculos estacionários ou móveis, enquanto em um ambiente de fábrica a presença de tais obstáculos pode representar dificuldade para a estimativa do robô das intenções humanas ou execução de uma trajetória segura.

Em termos de validade ecológica dos estímulos, os estímulos usados em experimentos devem ser representativos de eventos ou objetos factíveis e naturais. Por exemplo, muitos estudos afetivos usaram filmes como estímulos emocionais (CARVALHO, 2012), em que alguns enredos podem ser raros na vida cotidiana e, assim, reduzir a validade das observações. Um exemplo na psicologia do desenvolvimento são as descobertas do impacto das emoções na memória das crianças, que foram debatidas quando os estímulos usados eram menos representativos de eventos emocionais naturais em situações do mundo real (YUILLE & WELLS, 1991). Enquanto isso, os

pesquisadores que estudam estímulos que são potencialmente prejudiciais aos participantes, como situações que provocam estresse ou ansiedade, precisam incluir medidas que protejam o bem-estar mental e físico dos participantes, como briefing adequado e sessões de debriefing antes e depois do experimento (ALLEN, 2017), para que a segurança dos participantes seja garantida e o experimento não apresente ônus de longa duração ou impactos negativos severos sobre os participantes.

Em termos de validade ecológica da resposta, é importante recrutar participantes representativos do contexto do estudo, em vez de com base em amostragem por conveniência. Um grande número de estudos psicológicos e comportamentais existentes usaram participantes provenientes da população ocidental, educada, industrializada, rica e democrática (WEIRD) (HENRICH, 2010). No entanto, foi observada uma variabilidade substancial entre os participantes do WEIRD e outros grupos populacionais, por exemplo, Herrmann *et al.* (2008) descobriram que variações significativas entre sociedades existem nos comportamentos e decisões dos participantes durante o jogo do bem público para medir a pró-socialidade e a cooperação. Portanto, os pesquisadores são encorajados a incorporar um contexto social e interativo em seu projeto de estudo e recrutar participantes de sua população de usuários-alvo.

5.3 Avaliação com as medidas certas

Na pesquisa em computação afetiva, as observações e medições podem ser feitas por meio de várias abordagens, como medidas de desempenho de tarefas, medidas comportamentais, medidas psicofisiológicas, questionários e análise qualitativa de entrevistas. Essas medidas podem ser autoavaliação ou avaliação feita por um observador. Para aumentar a validade e a confiabilidade da avaliação, os pesquisadores são encorajados a combinar várias medições e investigar a convergência entre elas, como o uso de sensores vestíveis e questionários autorrelatados para avaliar o nível de estresse experimentado de uma pessoa (SCHMUCKLER, 2001).

Para modelos de emoções supervisionadas, a qualidade dos dados influenciará diretamente o desempenho do modelo aprendido. Como as emoções são subjetivas, ambíguas e complexas, é comum que os anotadores de dados emocionais tenham dúvidas sobre sua anotação ou discordem

uns dos outros. Para abordar a subjetividade e a incerteza nos dados emocionais e rótulos, protocolos de coleta de dados, técnicas de processamento de dados e abordagens de modelagem podem ser empregados (RIZOS & SCHULLER, 2020). Por exemplo, usando rótulos de emoção ordinal com âncoras adequadas (YANNAKAKIS, 2017), usando rótulos suaves que capturam anotações de todos os avaliadores em vez de fundi-los em um rótulo rígido (CHOU & LEE, 2019), e modelagem de incerteza do avaliador (KWON, 2020) ou atenção (TU, 2020) em o modelo de emoção.

Além disso, testes estatísticos apropriados e análises de poder devem ser conduzidos ao analisar os resultados experimentais. Por exemplo, se um estudo tiver um tamanho de amostra pequeno, para rótulos emocionais categóricos, testes exatos (por exemplo, McDonald (2009)) são preferidos, enquanto para rótulos emocionais contínuos, revisar suposições de métodos paramétricos clássicos ou métodos não paramétricos são recomendados (MORGAN, 2017).

5.4 Reprodutibilidade e generalizabilidade

Para melhorar a reprodutibilidade e generalizabilidade dos estudos de computação afetiva, os pesquisadores podem seguir sugestões da psicologia social em hipóteses de pré-registro (VEER & GINER-SOROLLA, 2016). Em relatórios e publicações do trabalho de alguém, é necessário cobrir os resultados significativos e não significativos na discussão, divulgar a metodologia de estudo detalhada (como materiais complementares, se necessário) e abrir o código e os dados usados no trabalho se possível. Esforços de benchmarking, como desafios abertos (por exemplo, Ringeval *et al.* (2019), Dhall *et al.* (2020)), ferramentas de modelagem (por exemplo, Eyben *et al.* (2010), Baltrušaitis *et al.* (2016)) e conjuntos de dados compartilhados (por exemplo, Zadeh *et al.* (2018), Baveye *et al.* (2015)), são especialmente bem-vindos, pois criam uma plataforma comunitária para o avanço do estado da arte na pesquisa de computação afetiva.

6. Rumo à pesquisa e aplicação de computação afetiva ética

Nesta seção, discutimos como a pesquisa e a aplicação da computação afetiva ética podem ser alcançadas a partir das perspectivas de justiça, transparência e responsabilidade do sistema.

6.1 Um sistema justo

As emoções são subjetivas, portanto, as tecnologias afetivas são especialmente propensas a vieses. Vieses emocionais conscientes ou inconscientes na pesquisa e aplicação de computação afetiva podem prejudicar os usuários ou impedir os resultados esperados. Por exemplo, se um sistema de computação afetiva é projetado para identificar a frustração no contexto do serviço ao cliente e fornecer orientação sobre como a situação pode ser tratada, o reconhecimento parcial ou impreciso da emoção pode levar a sugestões ineficazes ou prejudiciais. Portanto, um aspecto chave para alcançar a pesquisa e a aplicação da computação afetiva ética é reduzir os preconceitos e desenvolver um sistema justo.

Um tipo particular de preconceito estudado na pesquisa de computação afetiva é o preconceito racial, em que pessoas de um determinado grupo étnico sofrem com as visões estereotipadas existentes. Rhue (2018) estudou duas ferramentas comerciais de reconhecimento de expressão facial, nomeadamente Face ++ e Microsoft Face API. Os resultados expuseram que ambas as ferramentas atribuem mais emoções negativas aos rostos dos homens negros do que aos rostos dos homens brancos. Se uma expressão facial está sorrindo, os rostos negros em média são classificados com pontuação de sorriso mais baixa, enquanto se uma expressão facial for ambígua, os rostos negros são mais propensos a ser classificados como expressando emoções negativas ou agressivas, como raiva. Essas observações se alinham com as descobertas das ciências sociais de que, em comparação com os prestadores de serviços brancos, é necessário que os prestadores de serviços negros exagerem nas expressões emocionais positivas a fim de receber paridade nos julgamentos avaliados pelo supervisor sobre seu desempenho no trabalho e calor interpessoal (GRANDEY, 2019). O trabalho de Rhue (2018) ilustra que os algoritmos refletem os mesmos vieses que os humanos têm.

Além do preconceito racial, as diferenças culturais ainda não são totalmente compreendidas pelos atuais modelos de reconhecimento automático de emoções, o que pode limitar sua precisão e a consequente tomada de decisão em sistemas de computação afetiva. O Workshop e Desafio de Emoção Audio / Visual 2019 (AVEC 2019) hospedou um sub-desafio de reconhecimento de afeto transcultural (RINGEVAL, 2019). A equipe de

melhor desempenho (ZHAO, 2019) relatou Coeficientes de Correlação de Concordância de 0,4 na excitação e 0,471 na Valência, o que indica uma correlação média entre a predição automática e os rótulos verdadeiros. Além da etnia e histórico cultural, existem muitas variações individuais que podem influenciar a eficácia de um sistema de computação afetivo, por exemplo, a idade, onde as expressões faciais de pessoas idosas são mais difíceis de serem reconhecidas (FÖLSTER, 2014).

Da perspectiva do projeto do sistema, ter uma equipe diversificada de projeto e desenvolvimento de pessoas de diferentes gêneros, etnias, disciplinas, status socioeconômico e pontos de vista ajuda a capturar a complexidade das emoções e, assim, reduzir os preconceitos. Vieses em algoritmos são frequentemente herdados dos dados nos quais os modelos são treinados, por exemplo, quando um determinado grupo de usuários é sub-representado nos dados ou mal interpretado pelo modelo. Portanto, a coleta de diversos dados por meio de uma plataforma multimodal multi-sensorial colocada em contextos alinhados com os objetivos da aplicação pode reduzir a representação incorreta no modelo. Tornar esses dados de alta qualidade acessíveis de forma a preservar a privacidade também beneficia a comunidade de computação afetiva no avanço da tecnologia. Além disso, abordagens algorítmicas podem ser adotadas para permitir que modelos afetivos pré-treinados abordem vieses por meio de adaptação contínua, como Howard *et al.* (2017).

6.2 Um sistema transparente

Outra questão crítica a ser abordada a fim de alcançar a pesquisa e aplicação de computação afetiva ética é calibrar a confiança das pessoas em relação aos sistemas de computação afetivos, especialmente para evitar o viés de automação, onde uma pessoa confia demais em um sistema (GODDARD, 2012). Quando as pessoas deixam a tomada de decisões inteiramente para um sistema inteligente, os riscos associados aos limites desse sistema podem ser mal compreendidos ou subestimados. Por exemplo, Robinette *et al.* (2016) investigou o excesso de confiança em cenários de evacuação de emergência. Eles descobriram que todos os participantes seguiram as instruções de um robô cegamente, mesmo quando acabaram de observar o robô cometendo erros e quando as instruções estavam

claramente erradas. Goddard *et al.* (2012) revisaram o viés de automação no contexto dos sistemas de suporte à decisão clínica e descobriram que isso pode levar à redução da precisão do diagnóstico, por exemplo, os médicos substituindo suas próprias decisões corretas em favor de conselhos errôneos do sistema de suporte à decisão (FRIEDMAN, 1999). Goddard *et al.* (2012) identificaram que a discrepância entre a confiança de um usuário em relação a um sistema e a confiabilidade do sistema é o fator de condução mais forte da tendência de automação.

Como a confiança em um sistema automático é essencialmente uma troca entre a autoconfiança do usuário e sua confiança no sistema (LEE & MORAY, 1992), expor os usuários aos pontos fortes e fracos de um sistema inteligente os ajudará a tomar decisões mais informadas. Aumentar a transparência e a clareza do processo de tomada de decisão ajuda um sistema inteligente a comunicar sua intenção de maneira eficaz com os usuários e expor sua incerteza ao interpretar dados ambíguos ou incompletos. Para aumentar a transparência e clareza, a IA explicável (XAI) é uma abordagem promissora (HOFFMAN, 2018). Um sistema de computação afetivo pode melhorar sua transparência seguindo as listas de verificação de Hoffman *et al.* (2018) para medir a qualidade das explicações que ele fornece e para estabelecer um modelo mental de seus usuários em termos de sua percepção do sistema e de si mesmos. Os avanços no XAI também foram acelerados por ferramentas de código aberto que os pesquisadores e profissionais de computação afetiva podem adotar, como o Local Interpretable Model-Agnostic Explanations (LIME³⁶) (RIBEIRO, 2016).

6.3 Um sistema benéfico

A busca pela pesquisa e aplicação da computação afetiva ética pode aprender com o desenvolvimento em andamento para abordar a Inteligência Artificial Ética. Russell (2019) propõe que as máquinas de IA devem ser puramente altruístas e benéficas. Deve-se esperar que eles atinjam os objetivos humanos sem ter nenhuma suposição inicial sobre as preferências humanas, para que possam aprender a prever as preferências humanas por meio da observação das escolhas humanas. Gabriel (2020) discutiu ainda as possíveis abordagens para alinhar IA com valores humanos, classificando

36 <https://github.com/marcotcr/lime/>

o alinhamento de um agente de IA em relação aos valores humanos em 6 estágios:

1. O agente faz o que é *instruído* por uma pessoa.
2. O agente faz o que é *pretendido* por uma pessoa.
3. O agente faz o que o comportamento humano sugere que ele *prefere* fazer.
4. O agente faz o que um ser humano racional e informado *deseja* que ele faça.
5. O agente faz o que é objetivamente no *melhor interesse* de uma pessoa.
6. O agente faz o que é *moral* definido pelos indivíduos ou pela sociedade.

Uma abordagem técnica para um agente de IA se alinhar com as preferências humanas é o Aprendizado por Reforço Inverso (IRL), que extrai sua função de recompensa por meio de observações em vez de usar uma função de recompensa predefinida (ARORA & DOSHI, 2018; NG, 2020). A IRL pode ser implementada imitando especialistas humanos e aprendendo com suas demonstrações (ABBEEL & NG, 2004; RANA, 2018), por meio da modelagem de motivações e raciocínios por trás das ações observadas (RAMACHANDRAN & AMIR, 2007; VOIGT & BUSSCHE, 2017) ou por meio de abordagens evolutivas (KHADKA, 2019; SALIMANS, 2017). Lembre-se de que RL é uma abordagem comum em sistemas de computação afetiva. Portanto, tais sistemas de computação afetiva podem adotar as mesmas abordagens de aprendizagem de preferência para obter resultados benéficos e em alinhamento com as preferências humanas (estágio 3 de Gabriel (2020)).

É especialmente crucial para um sistema de computação afetiva projetado para o bem-estar entender e se alinhar com os desejos e objetivos racionais de seus usuários (estágios 4 e 5 de Gabriel (2020)). O bem-estar de uma pessoa pode ser medido tanto por meio de medidas subjetivas, como a sensação de satisfação ou realização, quanto por meio de medidas objetivas, como avaliação de saúde ou realização de certas tarefas (NUSSBAUM & SEM, 1993). A pesquisa interdisciplinar sobre as influências de fatores físicos, psicológicos e socioeconômicos sobre o bem-estar de uma pessoa é necessária para uma melhor compreensão (SUMNER, 1996). Um sistema de computação afetiva deve incorporar o bem-estar de uma pessoa em seu design e medir tais objetivos relacionados ao bem-estar em um ambiente de longo prazo.

6.4 Um sistema responsável

Em relação à incorporação de valores morais em sistemas de IA (estágio 6 de Gabriel (2020)), um agente de IA precisa investigar um conjunto de crenças morais compartilhadas por uma comunidade em um determinado contexto. Existem valores morais que se sobrepõem em diferentes culturas e contextos, conforme demonstrado na Declaração Universal dos Direitos Humanos (ASSEMBLY, 1949). No entanto, semelhante às normas sociais, muitas crenças morais são dependentes do contexto, especialmente em casos extremos complexos. Awad *et al.* (2018) reuniu julgamentos de milhões de pessoas em 233 países e territórios sobre suas preferências em vários dilemas morais enfrentados por veículos autônomos, por exemplo, se um carro que dirige sozinho passando por uma falha de freio deve atropelar pedestres ou desviando para o bloqueio de estrada matando seus passageiros. Eles observaram preferências globais, com os três mais fortes sendo poupando humanos em vez de animais, poupando mais vidas e poupando vidas jovens. No entanto, eles também observaram variações individuais e culturais consideráveis, mesmo sobre essas três preferências compartilhadas: participantes de culturas individualistas mostraram uma preferência mais forte por poupar mais vidas, enquanto participantes de culturas coletivas mostraram uma preferência mais fraca por poupar vidas jovens. Portanto, embora os valores morais compartilhados devam ser considerados no projeto de sistemas de computação ético-afetivos, eles também devem ser adaptados à aplicação específica e aos contextos do usuário.

Além de garantir IA ética por meio de abordagens de pesquisa e design, governos em todo o mundo estão começando a desenvolver regulamentos em torno da IA com a ajuda de conselhos consultivos, por exemplo, a legislação do Regulamento Geral de Proteção de Dados (GDPR) da União Europeia como um esforço para proteger a privacidade do usuário e reduzir preconceitos (VOIGT & BUSSCHE, 2017). As políticas governamentais, como o GDPR, são essenciais para garantir a responsabilidade da sociedade pelas tecnologias de IA. Para alcançar uma aplicação de IA segura e justa, pesquisadores e profissionais estão tomando ações proativas na redação de diretrizes de desenvolvimento de sistema, por exemplo, a

Declaração de Toronto³⁷, formando conselhos éticos internos, por exemplo, IA da Microsoft, Ética e Efeitos em Engenharia e Comitê de Pesquisa e Escritório de Responsabilidade IA³⁸, bem como estabelecer colaborações da comunidade, por exemplo, a iniciativa Partnership on AI³⁹.

Aprendendo com as experiências para garantir a pesquisa e aplicação ética da IA, os pesquisadores e profissionais de computação afetiva são incentivados a ter discussões dentro e fora da comunidade, como o painel de discussão muito necessário sobre “Computação Afetiva e o Uso Indevido de ‘Nossa’ Tecnologia / Ciência” na Conferência Internacional de Computação Afetiva e Interação Inteligente de 2019 (ACII). Essas discussões e a incorporação de justiça, transparência e responsabilidade na pesquisa e nas aplicações de computação afetiva são essenciais para a realização da computação afetiva ética e inclusiva, ou seja, a computação afetiva para TODOS.

7. REFERÊNCIAS

ABBEEL, P; NG, A. Y. Apprenticeship learning via inverse reinforcement learning. In: **Proceedings of the twenty-first international conference on Machine learning**. 2004. p. 1.

ABDUL, A. *et al.* Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In: **Proceedings of the 2018 CHI conference on human factors in computing systems**. 2018. p. 1-18.

AKÇAY, M. B.; OĞUZ, K. Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. **Speech Communication**, v. 116, p. 56-76, 2020.

ALLEN, M. **The SAGE encyclopedia of communication research methods**. Sage Publications, 2017.

ARORA, S.; DOSHI, P. A survey of inverse reinforcement learning: Challenges, methods and progress. **Artificial Intelligence**, p. 103500, 2021.

37 https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf

38 <https://www.microsoft.com/en-us/ai/responsible-ai/>

39 <https://www.partnershiponai.org/>

- ASSEMBLY, UN G. *et al.* Universal declaration of human rights. **UN General Assembly**, v. 302, n. 2, p. 14-25, 1948.
- AUGSTEIN, M.; HERDER, E.; WÖRNDL, W. **Personalized human-computer interaction**. Walter de Gruyter GmbH & Co KG, 2019.
- AWAD, E. *et al.* The moral machine experiment. **Nature**, v. 563, n. 7729, p. 59-64, 2018.
- BALTRUŠAITIS, T.; ROBINSON, P.; MORENCY, L. P. Openface: an open source facial behavior analysis toolkit. In: **2016 IEEE Winter Conference on Applications of Computer Vision (WACV)**. IEEE, 2016. p. 1-10.
- BAVEYE, Y. *et al.* LIRIS-ACCEDE: A video database for affective content analysis. **IEEE Transactions on Affective Computing**, v. 6, n. 1, p. 43-55, 2015.
- BEARD, R. *et al.* Multi-modal sequence fusion via recursive attention for emotion recognition. In: **Proceedings of the 22nd Conference on Computational Natural Language Learning**. 2018. p. 251-259.
- CAMPBELL, D. T. Factors relevant to the validity of experiments in social settings. **Psychological bulletin**, v. 54, n. 4, p. 297, 1957.
- CARVALHO, S. *et al.* The emotional movie database (EMDB): A self-report and psychophysiological study. **Applied psychophysiology and biofeedback**, v. 37, n. 4, p. 279-294, 2012.
- CHOU, H. C.; LEE, C. C.. Every rating matters: Joint learning of subjective labels and individual annotators for speech emotion classification. In: **ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. IEEE, 2019. p. 5886-5890.
- COHEN, P. R. Back to the future for dialogue research: A position paper. **arXiv preprint arXiv:1812.01144**, 2018.
- CONGER, K.; FAUSSET, R.; KOVALESKI, S. F. San Francisco bans facial recognition technology. **The New York Times**, v. 14, 2019.
- DHALL, A. *et al.* Emotiw 2020: Driver gaze, group emotion, student engagement and physiological signal based challenges. In: **Proceedings of the 2020 International Conference on Multimodal Interaction**. 2020. p. 784-789.

DOBROSOVESTNOVA, A.; HANNIBAL, G.. Teachers' Disappointment: Theoretical Perspective on the Inclusion of Ambivalent Emotions in Human-Robot Interactions in Education. In: **Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction**. 2020. p. 471-480.

EYBEN, F.; WÖLLMER, M.; SCHULLER, B. Opensmile: the munich versatile and fast open-source audio feature extractor. In: **Proceedings of the 18th ACM international conference on Multimedia**. 2010. p. 1459-1462.

EYBEN, F. *et al.* The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. **IEEE transactions on affective computing**, v. 7, n. 2, p. 190-202, 2015.

FITZGERALD, M.; MCCLELLAND, T. What makes a mobile app successful in supporting health behaviour change?. **Health education journal**, v. 76, n. 3, p. 373-381, 2017.

FÖLSTER, M.; HESS, U.; WERHEID, K. Facial age affects emotional expression decoding. **Frontiers in psychology**, v. 5, p. 30, 2014.

FRIEDMAN, C. P. *et al.* Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. **Jama**, v. 282, n. 19, p. 1851-1856, 1999.

GABRIEL, I. Artificial intelligence, values, and alignment. **Minds and machines**, v. 30, n. 3, p. 411-437, 2020.

GARG, R.; SENGUPTA, S. He is just like me: a study of the long-term use of smart speakers by parents and children. **Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies**, v. 4, n. 1, p. 1-24, 2020.

GODDARD, K.; ROUDSARI, A.; WYATT, J. C. Automation bias: a systematic review of frequency, effect mediators, and mitigators. **Journal of the American Medical Informatics Association**, v. 19, n. 1, p. 121-127, 2012.

GRANDEY, A. A.; HOUSTON III, L.; AVERY, D. R. Fake it to make it? Emotional labor reduces the racial disparity in service performance judgments. **Journal of Management**, v. 45, n. 5, p. 2163-2192, 2019.

HENRICH, J.; HEINE, S. J.; NORENZAYAN, A. The weirdest people in the world?. **Behavioral and brain sciences**, v. 33, n. 2-3, p. 61-83, 2010.

HERRMANN, B.; THÖNI, C.; GÄCHTER, S. Antisocial punishment across societies. **Science**, v. 319, n. 5868, p. 1362-1367, 2008.

HOFFMAN, G. Anki, jibo, and kuri: What we can learn from social robots that didn't make it. **IEEE Spectrum**, 2019.

HOFFMAN, G; ZHAO, X. A primer for conducting experiments in human-robot interaction. **ACM Transactions on Human-Robot Interaction (THRI)**, v. 10, n. 1, p. 1-31, 2020.

HOFFMAN, R. R. *et al.* Metrics for explainable AI: Challenges and prospects. **arXiv preprint arXiv:1812.04608**, 2018.

HOWARD, A.; ZHANG, C; HORVITZ, E. Addressing bias in machine learning algorithms: A pilot study on emotion recognition for intelligent systems. In: **2017 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)**. IEEE, 2017. p. 1-7.

IRFAN, B. *et al.* Personalization in long-term human-robot interaction. In: **2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**. IEEE, 2019. p. 685-686.

KHADKA, S. *et al.* Collaborative evolutionary reinforcement learning. In: **International Conference on Machine Learning**. PMLR, 2019. p. 3341-3350.

KWON, Y. *et al.* Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. **Computational Statistics & Data Analysis**, v. 142, p. 106816, 2020.

LASOTA, P. A.; ROSSANO, G. F.; SHAH, J. A. Toward safe close-proximity human-robot interaction with standard industrial robots. In: **2014 IEEE International Conference on Automation Science and Engineering (CASE)**. IEEE, 2014. p. 339-344.

LEDGERWOOD, A.; SODERBERG, C. K.; SPARKS, J. **Designing a study to maximize informational value**. 2017.

LEE, J.; MORAY, N. Trust, control strategies and allocation of function in human-machine systems. **Ergonomics**, v. 35, n. 10, p. 1243-1270, 1992.

MAXWELL, S. E.; LAU, M. Y.; HOWARD, G. S. Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. **American Psychologist**, v. 70, n. 6, p. 487, 2015.

- MCDONALD, J. H. **Handbook of biological statistics**. Baltimore, MD: sparky house publishing, 2009.
- MORGAN, C. J. Use of proper statistical techniques for research studies with small samples. **American Journal of Physiology-Lung Cellular and Molecular Physiology**, v. 313, n. 5, p. L873-L877, 2017.
- NG, A. Y. *et al.* Algorithms for inverse reinforcement learning. In: **Icml**. 2000. p. 2.
- NORMAN, D. A.; DRAPER, S. W. **User centered system design: New perspectives on human-computer interaction**. CRC Press, 1986.
- NUSSBAUM, M.; SEN, A. **The quality of life**. Clarendon Press, 1993.
- OVIATT, S. *et al.* **The handbook of multimodal-multisensor interfaces, Volume 2: Signal processing, architectures, and detection of emotion and cognition**. Morgan & Claypool, 2018.
- PAN, X.; HAMILTON, A. F. de C. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. **British Journal of Psychology**, v. 109, n. 3, p. 395-417, 2018.
- PICARD, R. W. Affective Computing-MIT Media Laboratory Perceptual Computing Section. **Technical Report** no. 321, 1995.
- PICARD, R. W. **Affective computing**. MIT press, 2000.
- PICARD, R. W. **Measuring affect in the wild**. In: **International conference on affective computing and intelligent interaction**. Springer, Berlin, Heidelberg, 2011. p. 3-3.
- RAMACHANDRAN, D.; AMIR, E. Bayesian Inverse Reinforcement Learning. In: **IJCAI**. 2007. p. 2586-2591.
- RANA, M. *et al.* Towards robust skill generalization: Unifying learning from demonstration and motion planning. In: **Intelligent robots and systems**. 2018.
- RHUE, L. Racial influence on automated perceptions of emotions. **Available at SSRN 3281765**, 2018.
- RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. “ Why should i trust you?” Explaining the predictions of any classifier. In: **Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining**. 2016. p. 1135-1144.

RINGEVAL, F. *et al.* AVEC 2019 workshop and challenge: state-of-mind, detecting depression with AI, and cross-cultural affect recognition. In: **Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop**. 2019. p. 3-12.

RIZOS, G.; SCHULLER, B. W. Average Jane, Where Art Thou?—Recent Avenues in Efficient Machine Learning Under Subjectivity Uncertainty. In: **International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems**. Springer, Cham, 2020. p. 42-55.

ROBINETTE, P. *et al.* Overtrust of robots in emergency evacuation scenarios. In: **2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**. IEEE, 2016. p. 101-108.

RUSSELL, S. **Human compatible: Artificial intelligence and the problem of control**. Penguin, 2019.

SALIMANS, T. *et al.* Evolution strategies as a scalable alternative to reinforcement learning. **arXiv preprint arXiv:1703.03864**, 2017.

SCHMIDT, P. *et al.* Introducing wesad, a multimodal dataset for wearable stress and affect detection. In: **Proceedings of the 20th ACM international conference on multimodal interaction**. 2018. p. 400-408.

SCHMUCKLER, M. A. What is ecological validity? A dimensional analysis. **Infancy**, v. 2, n. 4, p. 419-436, 2001.

SHROUT, P. E.; RODGERS, J. L. Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. **Annual review of psychology**, v. 69, p. 487-510, 2018.

SMILJANIC, R.; GILBERT, R. C. Acoustics of clear and noise-adapted speech in children, young, and older adults. **Journal of Speech, Language, and Hearing Research**, v. 60, n. 11, p. 3081-3096, 2017.

SONG, S.; YAMADA, S. Expressing emotions through color, sound, and vibration with an appearance-constrained social robot. In: **2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)**. IEEE, 2017. p. 2-11.

SUMNER, L. W. **Welfare, happiness, and ethics**. Clarendon Press, 1996.

SUTTON, R. E.; WHEATLEY, K. F. Teachers' emotions and teaching: A review of the literature and directions for future research. **Educational psychology review**, v. 15, n. 4, p. 327-358, 2003.

TIAN, L.; OVIATT, S.; MUSZYNSKI, M.; CHAMBERLAIN, B.; HEALEY, J.; SANO, A. **Applied Affective Computing**. ACM Books (in preparation), 2021.

WU, H. *et al.* Exploiting Heterogeneous Graph Neural Networks with Latent Worker/Task Correlation Information for Label Aggregation in Crowdsourcing. **ACM Transactions on Knowledge Discovery from Data (TKDD)**, v. 16, n. 2, p. 1-18, 2021.

VAN'T VEER, A. E.; GINER-SOROLLA, R. Pre-registration in social psychology—A discussion and suggested template. **Journal of experimental social psychology**, v. 67, p. 2-12, 2016.

VASQUEZ, D.; OKAL, B.; ARRAS, K. O. Inverse reinforcement learning algorithms and features for robot navigation in crowds: an experimental comparison. In: **2014 IEEE/RSJ International Conference on Intelligent Robots and Systems**. IEEE, 2014, p. 1341–1346.

VOIGT, P.; VON DEM BUSSCHE, A. The eu general data protection regulation (gdpr). **A Practical Guide, 1st Ed., Cham: Springer International Publishing**, v. 10, p. 3152676, 2017.

YANNAKAKIS, G. N.; COWIE, R.; BUSSO, C. The ordinal nature of emotions. In: **2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)**. IEEE, 2017. p. 248-255.

YATSUDA, A.; HARAMAKI, T.; NISHINO, H. A robot gesture framework for watching and alerting the elderly. In: **International Conference on Network-Based Information System**. Springer, 2018, p. 132-143.

YUILLE, J. C.; WELLS, G. L. Concerns about the application of research findings: The issue of ecological validity. In: **The Suggestibility of Children's Recollections, Jun, 1989, Cornell U, Ithaca, NY, US; This volume is based in part on the aforementioned conference**. American Psychological Association, 1991.

ZADEH, A. B. *et al.* Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph. In: **Proceedings of the 56th**

Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018. p. 2236-2246.

ZHAO, J. *et al.* Adversarial domain adaption for multi-cultural dimensional emotion recognition in dyadic interactions. In: **Proceedings of the 9th International on Audio/Visual Emotion Challenge and Workshop.** 2019. p. 37-45.

ZHAO, R. *et al.* Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavior. In: **International conference on intelligent virtual agents.** Springer, Cham, 2016. p. 218-233.

ZUBOFF, S. **The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019.** Profile books, 2019.

07.

SOCIEDADE AUTÔMATA: A MÁQUINA NO CONTROLE

Elen Nas

É fácil fazer corresponder a cada sociedade certos tipos de máquina, não porque as máquinas sejam determinantes, mas porque elas exprimem as formas sociais capazes de lhes darem nascimento e utilizá-las. (DELEUZE, 1992)

1. PREÂMBULO

Considerando que a autonomia se define por estar em posse de suas próprias decisões, sem que elas ocorram por meio de manipulação ou força externa, vamos iniciar este preâmbulo retomando a experiência nas redes, nos anos em que a Cambridge Analytica influenciou os resultados das eleições, trazendo para as ações de marketing político, estratégias de manipulação psíquica em massa. Nestes contextos, o processo de decisão autônoma tornou-se um desafio fundamental, lembrando que, em perspectivas ético-políticas, a ausência de autonomia torna o ambiente propício à práticas totalitárias. (CHRISTMAN, 2020)

2018 foi mais um daqueles anos sem trégua. Uma polarização política dualista que já vinha se acirrando, desde 2016, fez transbordar as guerras do mundo virtual para todos os espaços de convívio humano. Tornaram-se evidentes as reações imediatas de ressentimentos que se multiplicavam em ofensas e rupturas de relações.

Alguém poderá dizer que sempre houve isso, mas o mundo virtual tornou mais fácil praticar o desrespeito verbal.

Diante das telas de computadores e dispositivos eletrônicos a cidadania digital abriu precedentes para estranhos engajamentos.

Desde o início da década, o *Occupy Wall Street*, a *Primavera Árabe*, e outras manifestações massivas revelaram o poder de uma sociedade conectada. Isto talvez tenha transferido às mídias alternativas uma certa ilusão de acesso mais direto à informação do que era experimentado anteriormente com as mídias oficiais.

Mas não demorou muito para que os algoritmos entrassem em cena, afunilando as redes de contato, criando bolhas e enviando mensagens subliminares em meio a propagandas de produtos ordinários.

As pessoas revelaram a necessidade imperativa de expor opinião sem que realmente estivesse havendo algum diálogo. A partir de 2016 me surpreendi que no meu perfil de *Facebook*, pessoas da minha rede de contatos, manifestavam-se de maneira ofensiva às notícias ou opiniões políticas postadas eventualmente em minha página. Não havia um motivo claro para tais reações, senão que a liberdade de expressão e opinião virou ao avesso e as mídias digitais facilitaram o desprendimento das etiquetas sociais, fato que vivenciei igualmente no YouTube com desconhecidos.

Convivo com a internet desde os anos 90 e entendo o mundo virtual como parte da vida real, de modo que uma ofensa é uma ofensa independente do meio que se utiliza para comunicar. Mas algo veio mudando severamente, e para quem assistiu a todos os *Zeitgeist* (JOSEPH, 2007; 2008; 2011) eu não poderia crer que estas eram manifestações isoladas de um fenômeno humano, ou tecnológico. Conhecendo bem a cultura predatória voltada à lucratividade sem medidas, a qual estamos inseridos, há mais coisas para investigarmos neste processo histórico atual.

Assim, em 2018 fui direcionada, através dos meus contatos, para diversas páginas de discussão política e novamente desconfiei quando as páginas que se colocavam como defensoras de uma determinada proposta e candidatos, traziam notícias ambíguas de seus concorrentes, o que poderia ser indício de perfis infiltrados. Tais notícias, em geral negativas, reforçavam o ambiente de hostilidade e oposição. Como posteriormente chegou ao conhecimento do público, algumas páginas de *Facebook* eram criadas como ‘iscas’ para atingir um determinado perfil, como por exemplo àqueles que ‘comprariam a ideia’ de atender manifestações, mesmo que as imagens e mensagens de chamada incitassem à polarização e a violência. A pesquisa encomendada pelo Facebook anos antes comprovou que o contágio emocional pelas redes é capaz de influenciar outros a repetirem os mesmos padrões (KRAMER, 2014), seja apatia, pânico ou comunicação violenta. No Brasil, a campanha “Ele Não” foi um exemplo da repetição de uma personalização que despolitiza entrando no jogo dos opostos. Mostrou-se ineficiente no seu propósito, mesmo com adesões de celebridades como a cantora Madonna. Reforçou a personalidade criticada pela sua negação.

Uma cadeia de reatividade tomou conta do cenário geral como se todo mundo descompensasse ao mesmo tempo. As manipulações no nível emocional para obter resultados eleitorais equipararam-se ao nível das operações psicológicas (PSYOP) utilizadas como estratégias de guerra (AMER K, 2019). Paradoxalmente não há qualquer controle sobre o impacto de tais estratégias. Elas ganham autonomia quando entram nas redes. Como no caso dos Estados Unidos em 2021, elas podem fazer as pessoas acreditarem que uma eleição foi roubada e que é necessário tomar uma atitude. Tudo se dissolve e pulveriza, transformando-se em outras coisas que estão além do alcance, mesmo dos que empurraram a ‘bola de neve’ do alto da montanha. O mundo se tornou complexo demais e o pouco de autonomia que resta aos indivíduos é utilizada para o reforço da autoridade. O poder de dar autoridade a quem não tem. Enquanto, por outro lado, a ‘autoridade’ do conhecimento, da sabedoria, da dimensão legal e ética é colocada em ‘xeque’ pelo ‘homem-máquina’.

Na tradição Ocidental entende-se que a autonomia se manifesta no exercício da razão. Porém, a ideia de que emoção e razão são distintas substâncias, faz com que a razão, não ciente de suas susceptibilidades a afetos, perca alguns atributos reflexivos. A paixão pela razão (DAMÁSIO, 2003) causa reações emocionais quando as ideias expostas não estão em concordância com o que um outro pensa ou entende como válido. A ‘sociedade automata’ ocupa o espaço deixado no corpo decepado, separado da mente e condenando suas emoções. As lacunas de comunicação quando as ideias estão confusas se resolvem no ‘modo automação’ para condução da vida.

2. AUTONOMIA

A autonomia faz parte do ideário norteador da construção da cidadania, nas sociedades laicas, ocidentais, e plurais contemporâneas, e, com o advento da Bioética o termo ganha especial atenção (SIQUEIRA-BATISTA, 2008). Ela representa o direito de decidir livremente sem coação. Na política e filosofia, o termo ‘autonomia’ rende debates cercados de paradoxos. Na bioética o protagonismo da autonomia oferece dilemas quando existem outros valores a serem considerados para o bem da comunidade e do indivíduo.

São dilemas porque não há uma única narrativa que ofereça a resolução mais justa. A percepção Kantiana que supõe a capacidade da vontade

se autolegislar (TRAPP, 2019) é totalmente dependente do pleno conhecimento de um conjunto de regras, ou seja, do indivíduo bem informado. Não apenas este fato, como também, diante da presença de ‘conflitos de interesse’ a auto-legislação está condicionada à capacidade do indivíduo de conectar-se com os princípios éticos delineados em vista de uma melhor vida relacional em comunidade. A capacidade cognitiva de julgamento moral que conduz à ação autônoma recebe influências de fatores afetivos e emocionais (REGO, 2005) referenciados em padrões ideológicos ou religiosos de uma determinada cultura. A plena autonomia da ação moral requer, portanto, “trabalhar a dimensão afetiva” de modo a “despertar verdadeiros sentimentos morais” (PALACIOS, REGO, 2017).

Na cultura do ocidente é comum buscar justificativas para relativizar um compromisso contratual coletivo. Em muitos casos a lei é ambígua e pode atuar em favor de uns e prejuízo de outros, ainda que os ‘outros’ sejam as vítimas em uma disputa.

Uma moral utilitária não reconhece no humano a capacidade de sacrificar o seu bem-estar para o bem-estar de outros (MILL, 2005). Precisamente por isto, desenvolveu um conceito para mediar as tensões entre os interesses individuais e coletivos. Desse modo, sugere-se optar pelo que possa produzir o maior bem, avaliando as consequências (consequencialismo). Assim, o utilitarismo se distingue do egoísmo quando pesa a relevância das consequências (DRIVER, 2014).

A autonomia em favor do que traga maior felicidade tem um componente hedonista e numérico: esta felicidade traduzida em alguma decisão medida deve atingir ao maior número de pessoas. O princípio da igualdade definido como ‘neutralidade’ está presente nos defensores do utilitarismo, desde que a felicidade de um não é vista como mais importante que a de outro. Na prática, porém, as decisões seguem avaliações morais que já carregam (pré) conceitos como o individualismo dentro de uma ideologia de ‘vencedores’ e ‘perdedores’ que tenderá a exacerbar atitudes egoístas, frequentemente anti-éticas.

O exercício da empatia é deixar-se afetar por outros corpos e suas histórias, entender as necessidades alheias, não a partir de pontos de vista externos a elas. Utilitaristas teológicos evocam a ‘vontade de Deus’ o que reduz a autonomia dos indivíduos.

As avaliações morais aos quais se baseiam as decisões passam pelas questões de vícios e virtudes e sofrem muitas distorções já que a lógica clássica não admite ambiguidades quando se trata de julgar e tomar decisão. Porém, entender se alguém está pendendo mais para o vício ou para a virtude não é uma equação matemática.

A aposta Kantiana é que as vontades pessoais podem se alinhar com as vontades da sociedade e assim, a ‘autonomia da vontade’ não é guiada pela dicotomia ‘sacrifício versus prazer’. Contudo, a vontade para Kant traduz-se em razão prática, desde que a razão é o único fundamento da determinação da vontade (TRAPP, 2019). Desse modo, a ‘autonomia da vontade’ Kantiana não é resultado de um processo, da fenomenologia e da ontologia inter relacional, mas da ‘razão pura’, uma razão supostamente capaz de se descolar dos afetos com o objetivo de alcançar os estágios mais elevados da mente para as decisões éticas ‘corretas’.

A ética requer uma visão compartilhada do dever (MILL, 2005), uma consciência coletiva, neste sentido, se os conceitos, como os delineados por Descartes e Kant são amplamente difundidos por séculos e respeitados como hegemônicos no campo do conhecimento, eles fazem parte desta consciência compartilhada. Ainda assim, se a ‘inteligência coletiva’ é fortemente influenciada pelo dualismo mente/corpo, razão/emoção, ela também reflete as distorções destes conceitos no campo empírico e prático.

Como resultado, a autonomia tanto é comprometida quando há excesso de controle nas ações guiadas por protocolos, como também quando não há controle nenhum, no caso das emoções explodirem em surtos psicóticos e outras formas de descompensações emocionais. O surto, porém, não é meramente emocional, senão uma pane da mente dividida.

3. AUTONOMIA NAS REDES

Segundo a narrativa exposta no Preâmbulo, vamos pensar até que ponto a autonomia do indivíduo é íntegra em relação às influências externas. Por exemplo, o caso da proliferação de *fake news* (CANAVILHAS, 2018) e manipulação das emoções em mídias sociais com objetivos eleitorais foi uma operação complexa envolvendo, por um lado, *bots*, que são algoritmos simulando usuários que propagaram certos tipos de notícias e opiniões (LÊU, 2019). Por outro lado, a interação destes *bots* com usuários humanos

gerou influências mútuas: os *bots* e os indivíduos das agências de marketing e empresas responsáveis pelas campanhas promoviam a manipulação por desinformação de modo a estimular reações no público, algo que se mostrou de fácil engajamento e crescimento exponencial.

Na medida que o ambiente virtual está repleto de desinformação, discursos de ódio e preconceitos multiplicam-se também através dos algoritmos, que dão destaque ao que aparece com maior frequência.

Quando um mecanismo de manipulação psíquica é arquitetado para disparar reações emocionais, a eficiência dos resultados revela que a ‘ciência do comportamento’ ganhou controle.

Estudos de comportamento e impacto emocional sobre os conteúdos na rede (KRAMER, 2014) demonstram que certos modelos de padronização psíquica têm sido úteis a diversas finalidades do mercado: do marketing de produto ao marketing eleitoral, ambos altamente lucrativos.

Enquanto poucas pessoas entendem quais são os benefícios ou malefícios relacionados ao uso dos seus dados pessoais, o compartilhamento de dados pode tornar as pessoas mais vulneráveis através da produção de perfis psicológicos, socioeconômicos e de consumo. A análise dos dados pode ‘enquadrá-las’ em categorias passíveis de preconceitos que venham a impactar seu acesso ao trabalho, educação e saúde (O’NEIL, 2016).

Ainda assim, a possibilidade de reverter este processo é mínima. Os dados já foram espalhados e a regulamentação trata das questões de privacidade, direito à informação, ‘direito ao esquecimento’ e ‘direito à explicação’ (COMISSION, 2018), mas não necessariamente garantirá o acesso à interpretação dos dados.

Enquanto tomar a decisão sobre como, e se, os dados privados poderão ser utilizados por terceiros, é um direito que apenas uma parcela mais informada das populações procurará ativar, a aposta é que o reconhecimento deste direito tenha o potencial de enfraquecer o uso do humano como mercadoria para empresas como a *Cambridge Analytica*. Tais empresas vendem sua expertise de comunicação como um *Santo Graal* capaz de modificar o comportamento das pessoas, mesmo que seja promovendo guerra entre elas.

Segundo o delator da *Cambridge Analytica*, Christopher Wylie, o líder do projeto, Steve Bannon, afirmava publicamente que “se você quer uma nova sociedade, precisa primeiro quebrá-la”. Assim, sob o lema “dividir

para controlar”, Bannon, com as ferramentas tecnológicas, especialistas da psicologia e computação, e um alto investimento financeiro, articulou uma série de estratégias para que os resultados eleitorais em diversos países atendessem uma agenda de controle autoritário. Não se trata de uma ameaça velada à democracia, já que as intenções foram claramente expostas e uma guerra cultural anunciada nos palanques, com ações virais nas redes sociais e mídias jornalísticas.

A mineração de dados do *Facebook*, o uso de Inteligência Artificial (IA), “testes de personalidade”, *bots* e etc., criaram uma ‘carnificina’, que começa no mundo virtual até gerar conflitos presenciais e corpóreos.

Um dos primeiros investidores do *Facebook*, Roger McNamee diz que a estratégia utilizada foi jogar com os instintos mais básicos como o medo e a raiva. E isto começou com a produção de uma série de ferramentas que permitiam os anunciantes explorarem a audiência emotiva com o foco em cada indivíduo (AMER K, 2019).

A jornalista Carole Cadwalladr afirmou que no Brasil não foi diferente, e a estratégia de influência e manipulação nas eleições presidenciais de 2018 se deu via *Whatsapp*, que é de propriedade do *Facebook*. (AMER K, 2019) A ameaça à democracia compromete um conjunto de valores fundamentais à dignidade humana, como a condução de políticas de acesso igualitário à saúde (NORONHA; CASTRO, 2019), educação, cultura, direito ao trabalho e moradia.

O tema da manipulação do imaginário ganhou adeptos ao longo do Século 20, sempre motivado pelo avanço de novas tecnologias de comunicação e a facilidade de proliferação de imagens.

Com o advento das mídias sociais e sua integração na vida diária de grandes parcelas da população, através de plataforma digitais que cabem no bolso, o que vêm se apresentando é uma necessidade urgente de compreender o impacto destas tecnologias na constituição física e mental do humano, assim como na sociedade e em todo meio-ambiente.

Porém mesmo que o primeiro estudo mostrando a possibilidade de influência emocional no estado psíquico das pessoas quando utilizavam o *Facebook*, tenha sido publicado em 2014, e a primeira reportagem investigativa revelando a estratégia da *Cambridge Analytica* tenha ocorrido no início de 2018, as técnicas continuaram sendo utilizadas, e os resultados eleitorais deste período ocorreram como esperado e planejado por essas

empresas e seus representantes que admitiam utilizar artifícios para modificar comportamentos capazes de mudar os rumos de uma eleição, como a apatia ou engajamento (NEWS, 2018).

Assim, se a sociedade mediada por meios tecnológicos vêm revelando que a autonomia é em muitos casos ilusória. Mas, o que faz com que algoritmos em redes sociais, estratégias de desinformação e *fakenews* sejam tão eficientes na manipulação do comportamento? Onde reside o segredo destas ferramentas? Algoritmos traçam perfis, indicam preferências, dão sugestões, assim como podem deixar algumas informações mais, ou menos evidentes. Estratégias de desinformação são conteúdos que podem vir de alguma fonte factível, mas são editados e colocados fora do contexto original: podem ser notícias antigas que ajudam a fomentar o medo ou cortes de um discurso de líderes políticos, dando-lhes significados diferentes do conteúdo original. Tais manipulações têm obtido sucesso em gerar reações emocionais imediatas.

O impacto das mídias digitais e tecnologias em geral, na vida, no corpo, e na sociedade representa um conhecimento em construção. Portanto, há que se reconhecer que conhecimentos isolados não respondem às necessidades de compreensão de tais fenômenos. A interdisciplinaridade necessária reformata e amplia o foco do olhar resultando em abordagens não convencionais. Entretanto, trazer estes temas a análise, estabelecendo novas ou diferentes associações teóricas constitui uma contribuição no exercício de fomentar o debate, urgente e necessário sobre a vida mediada pela computação ubíqua.

Neste capítulo a epistemologia que guia estudos e percepções científicas é discutida, desde que a ‘sociedade automata’ desafia o princípio de autonomia, ou seja, desafia a capacidade do indivíduo se autodeterminar. A ‘sociedade automata’ é resultado do movimento de automação de computadores e máquinas autogeridas por sistemas de informação que dependem ou não da intervenção humana no ato do seu processamento. A automação torna evidente a necessidade de pré-definir os ‘resultados desejáveis’ de modo a obter controle em favor de tais resultados. No caso da autonomia seria infame admitir que os humanos são meros autômatos que podem manter-se em operação apenas seguindo os padrões pré-definidos e formulados sem que haja a participação de sua vontade.

4. BEHAVIORISMO

nos preocupamos com as causas do comportamento humano. (...) Descobrimos e analisando estas causas seremos capazes de prever comportamentos; na medida que possamos manipular estes comportamentos, poderemos controlá-los. (SKINNER, 1965)

O conceito de homem-máquina foi apresentado por Julien Offrey de La Mettrie (1709-1751) em 1747 sob o título *L'Homme Machine*. Por um lado, ele se apropria do conceito cartesiano que reduz o corpo e os animais a reproduções funcionais autômatas, por outro lado, ele nega a existência da alma e afirma que ela está alojada no estômago. (METTRIE, 1749)

Ele diz: “o corpo humano é uma máquina que ativa as suas próprias fontes: é uma imagem viva do movimento contínuo” (METTRIE, 1749), e se as faculdades da alma dependem da organização adequada do cérebro e de todo o corpo, elas são, aparentemente, apenas essa organização. Assim, a alma nada mais é que uma ‘máquina iluminada’, e o ser humano não é menos máquina por ser dotado de característica tão ‘especial’ como a de pensar.

O ser-máquina movimenta-se dentro dos limites da ciência em uma percepção materialista: as informações contidas nos fluidos internos do corpo, a energia que este corpo gera e conecta, a potência da ação. Tal perspectiva produz e reproduz a necessidade de reconhecimento de padrões e organização destes por categorias. Às cores, sons, e às imagens estimuladas pelas diversas sensações do corpo-máquina são atribuídas classificações e significados.

La Mettrie separa as correntes filosóficas entre materialismo e espiritualismo sem negar completamente a tradição dualista cartesiana, porém argumentando que todas as respostas estão no corpo.

O materialismo monista de La Mettrie desenvolve, a partir da metáfora do corpo-máquina cartesiana, a compreensão de que o pensamento se expressa dentro deste corpo através dos fluidos que se intercomunicam e circulam, desde os órgãos (como o estômago), para todas as partes.

Uma segunda afinidade com o dualismo por La Mettrie é expressa na separação hierárquica entre aqueles que se prestam a manipulações por meio dos estímulos e carências do corpo e os que realmente se dispõem a exercitar as capacidades de pensar.

Embora as palavras escolhidas por La Mettrie para denominar as categorias dos ‘não-pensantes’ sejam depreciativas como ‘tolos’, ‘estupidez’,

‘famintos’ e assim por diante, tendendo a reforçar fortes preconceitos sociais e de classe, é interessante notar que não apenas como filósofo, como também como médico, ele associa a boa saúde à presença de pensamento neste corpo. O exercício do pensamento requer que o corpo não seja consumido pelo ódio, avareza e ambições, de modo que ele possa demandar os alimentos necessários à manifestação das virtudes.

Ainda que seja evidente que o problema das desigualdades sociais expõe um grande grupo à fome e seus reflexos de ódio, em algumas passagens do seu ‘manifesto’ *Homem Máquina* (METTRIE, 1749) ele esclarece que observa a existência de pessoas sensatas que não possuíram o privilégio da educação, enquanto muitos que gozaram de tal privilégio preferem se portar como idiotas, ou seja, aqueles que se abstém à capacidade de refletir, ponderar, são ‘autômatos da vontade’ (o que não é o mesmo que a ‘autonomia da vontade’ Kantiana).

De acordo com La Mettrie, saber distinguir entre o bem e o mal, os vícios e as virtudes, faz parte do equilíbrio necessário a se manter no ambiente interno do corpo, entre sólidos e líquidos, para que seja possível a manifestação da boa saúde.

A partir de tais perspectivas, sugiro que, o que determina o homem-máquina é sobretudo a sua ausência de autonomia, ou a presença de uma autonomia que se expressa, seja dentro do previsível, pré-programado, ‘aceitável’, até o ‘impulso das vontades’, às quais o suposto ‘sujeito autônomo’ não possui pleno conhecimento, mas que é igualmente previsível, mesmo dentro de ações entendidas como ‘imprevisíveis’.

Com o conhecimento das limitações advindas de regras sociais, jurídicas, educacionais, religiosas e culturais o sujeito é autônomo para decidir estar em concordância com elas, ainda que a discordância em determinados contextos não seja factualmente possível desde que, adequar-se a tais regras é essencial à sobrevivência do ser social. A automação, porém, se dá pela repetição com ausência de reflexão e consciência participativa.

Por exemplo, estar de acordo com as leis não necessariamente significa ser justo. Em muitos casos na história as leis endossavam maus tratos e até assassinatos contra escravos/as e suas crianças. Assim, a negação da dignidade humana se dá não apenas no que diz respeito ao abuso da força de trabalho, mas do reconhecimento – ou não - do ‘outro’ como humano.

No Século 21, pessoas continuam sendo apedrejadas (WORTH, 2010) até a morte, especialmente mulheres (ONG, 2015). A pena capital - que já é questionável por si só - também evidencia o problema de que muitos são acusados injustamente, e comumente não há interesse em investigar a verdade para defender os que fazem parte de grupos vulnerados pela desigualdade socioeconômica, racismo e misoginia.

A legislação também apresenta falhas quando protege negociações de livre mercado sem estabelecer restrições definidas em favor dos direitos de trabalhadores e consumidores, fato bem exemplificado na Série “A rota do dinheiro sujo” (DEO, 2018). Com estes breves exemplos acentua-se que a conformidade com injustiças faz com que as leis em muitos casos sejam inúteis e obsoletas. A autonomia de estar em conformidade com valores morais, ou não estar, sofre uma série de influências onde a ‘vontade autônoma’ por muitas vezes representa uma ‘vontade automata’.

Ainda assim, adaptar-se e ajustar-se não necessariamente significa tornar-se um autômato. Via de regra, em sociedades onde não há acumulação de bens nem exploração de trabalho como o histórico de muitos grupos indígenas, a autonomia pode ser considerada plena mesmo que os indivíduos se submetam às regras culturais. A autonomia pode ser plena se as sociedades possuem uma dinâmica horizontal que iniba e mesmo impeça a acumulação de poder. No caso das etnias brasileiras, a divergência é possível, podendo levar à dissidência e desligamento da comunidade. (ANTROPOFAGIAS, 2018) Os dissidentes assim são por sua própria vontade.

Entretanto, em se tratando de sociedades onde a opressão possui uma funcionalidade, a consciência crítica participativa se manifesta quando uma lei mostra-se injusta ou um protocolo social revela-se preconceituoso. Mas quando a moral justifica a propriedade, a acumulação e certas hierarquias, a horizontalidade que enfraquece a acumulação de poder não encontra caminhos de passagem. A moral indígena em grupos brasileiros condena a acumulação que é indício de avareza. Todo excedente deve ser compartilhado (ANTROPOFAGIAS, 2018).

Poderíamos então pensar que elementos da educação e cultura das mais diversas origens seriam ‘negociáveis’ ou ‘não negociáveis’ frente ao objetivo da construção de uma ética contemporânea pluricultural que acolha as demandas essenciais para a sobrevivência no planeta. Tratam-se de demandas éticas na esfera ambiental, animal, social, educacional, econômica e

cultural. Ainda que a universalização da ética como um todo seja um debate à parte, podemos perguntar se alguns princípios básicos como a autonomia, deveriam ser ‘elencados’ como essenciais em todas as culturas que compartilham relações de diversas ordens dentro da sociedade globalizada.

5. OBEDIÊNCIA

O descaso diante da realidade nos transforma em prisioneiros dela. Ao ignorá-la, nos tornamos cúmplices dos crimes que se repetem diariamente diante de nossos olhos. Enquanto o silêncio acoberta a indiferença, a sociedade continuará avançando em direção ao passado da barbárie. É tempo de escrever uma nova história e de mudar o final. (ARBEX, 2013)

A visão de La Mettrie demonstra uma distopia com o humano, expressa na impossibilidade de igualdade, já que nem todos os humanos sabem, querem, ou podem fazer uso da sua ‘capacidade de pensar’ autonomamente.

La Mettrie ressalta que, eventualmente qualquer pessoa independentemente de sua condição social poderá ser ou não idiotizada, a diferença seria entre ser uma mera máquina, ou uma máquina iluminada. Ele cita que, em algumas circunstâncias, um macaco poderá ser mais sensato que o humano (METTRIE, 1749).

Desde que é um trabalho do Século 18, deve-se colocar algumas de suas observações em contexto. A narrativa de *Homem-Máquina* é argumentativa, demonstrando até algum tipo de ressentimento aos que se abstém do filosofar. Inegavelmente, seres maquímicos são úteis a projetos de poder desde que os possíveis elos de empatia são interrompidos a partir do foco na objetificação daquele que, por ser um autômato, não possui autonomia plena: seu destino é decidido por outras pessoas e, assim como a máquina, não tem vontade própria; sua adequação é julgada pelos que estão alheios à ela; seu valor e qualidade também são julgados por sistemas alheios à sua vontade.

O descontentamento burguês de La Mettrie revela: “eu prefiro a companhia de um homem sensato, mesmo sem o polimento da educação do que a de uma pessoa que teve a má-sorte de ter uma educação ruim” (METTRIE, 1749). Traduzindo para o sistema computacional atual pode-se criar a analogia: é preferível um software aberto ao qual se pode adaptar o código,

dentro de um processo criativo e investigativo, do que um software fechado, caro, com uma programação ruim.

Os robôs atuais são ‘educados’ (treinados) para, ao mesmo tempo, serem incríveis (porque a fonte de sua existência e movimento está na potencialidade das ideias), e estúpidos (porque são apenas capazes de realizar tarefas muito limitadas).

O paradoxo da autonomia é ter a capacidade de escolha e ainda assim optar dentro da previsibilidade que define as identidades nas microculturas. Por exemplo, indivíduos voltados para a ‘desobediência’ (como Adão e Eva, que não ouviram as recomendações de Deus e comeram do ‘fruto proibido’) podem automaticamente decidir pelo que obviamente destoa da obediência, sem de fato refletir sobre o contexto e as possíveis consequências de sua decisão. É, também, comum quando aqueles cuja perspectiva da ‘boa moral’ está associada à obediência das normas e sua hierarquia, deixam de praticar a reflexão autônoma, que seria avaliar se o contexto da obediência está em consonância com a sua visão de ‘boa moral’, assim como, com os princípios éticos da sociedade.

O experimento de Milgram explorou o problema da obediência conduzindo testes (que mais tarde foram questionados na perspectiva da ‘ética em pesquisa’) em que os indivíduos investigados precisavam aplicar punições de choques elétricos toda vez que recebiam uma resposta incorreta à pergunta. A pessoa que supostamente recebia os choques era um ator que estava em uma outra sala fechada. Apesar do mal-estar provocado pelo teste, a maioria (2/3) dos investigados seguiram com a punição, conforme instruídos pela equipe (MILGRAM, 1963).

Stanley Milgram, psicólogo de origem judaica, interessou-se em observar como indivíduos comuns eram capazes de praticar a forma de tortura proposta pelo seu experimento, quando estão seguindo ordens.

No mesmo período do experimento de Milgram, o caso do julgamento de Adolf Eichmann, que durante a Segunda Guerra Mundial deportou centenas de milhares de judeus para os campos de concentração, ganhou atenção mundial e resultou no livro de Hannah Arendt onde ela apresenta a sua percepção do caso como ‘a banalidade do mal’ (ARENDT, 2013). Porque, sob o ponto de vista do acusado, ele nunca matou e não mandou matar ninguém. Segundo seu relato, tudo que fez foi apenas cumprir ordens, como

um funcionário do Estado, um burocrata. Sua autonomia se restringia em fazer o que lhe determinavam.

A obediência, portanto, é um dispositivo da ‘sociedade autômata’. Ela proporciona que os indivíduos de uma cultura operem de maneira previsível, em ‘modo de automação’. Em consequência, acredita-se que aquele que segue instintos, necessidades e desejos, nega a cultura e a sociedade. A contradição exposta pela banalidade do mal é quando estar em consonância com a sociedade é também negar os princípios morais dela mesma. É quando a obediência e consciência divergem.

Em tais casos, o grau de autonomia possível entre divergir e convergir, diz respeito a quão próximo ou distante se está de um ambiente guiado por princípios democráticos. Em regimes totalitários negar a obediência pode custar a própria vida. Ainda assim, em uma democracia negar-se a cumprir uma ordem ou apenas divergir, pode gerar perseguições pessoais - entre outros desgastes relacionais – podendo até resultar na perda do trabalho e bloqueio para novas oportunidades. Esta é a maneira em que a ‘microfísica do poder’ se espalha e torna as pessoas dóceis, dentro de um controle que é útil à ‘ordem’ e ao ‘progresso’.

O sociólogo e filósofo Zygmunt Bauman observa que as novas práticas administrativas do ambiente corporativo esvaziou a organização sindical à medida que o ingresso de trabalhadores nas empresas, para períodos de experiência anuais incentiva a competição entre todos gerando um ambiente de suspeita permanente, já que ao final de cada período alguns serão escolhidos para efetivação e outros serão dispensados (DINES, 2015). O ‘mal estar na civilização’ de um mundo globalizado é portanto um ‘estado de guerra’ dissolvido em micro práticas de combate como a fofoca, o julgamento moral e qualquer tipo de ação que seja diluída o suficiente para não ser facilmente configurada como ‘bullying’ ou outros tipos de ‘perseguições’ pessoais ou assédios.

Desse modo, o que pode parecer desorganizado, caótico e até mesmo ‘natural’ e esperado de ‘características humanas’ como disputas por atenção e espaço, são representações do controle da máquina. São resultados de estratégias elaboradas para atender a necessidade de operações ininterruptas dos sistemas de produção do mundo econômico, político e social.

O não-tempo para refletir e debater esconde medos de embate e das consequências de expor as ideias. E quando a maioria está doutrinação

não articular os pensamentos, tornando-se incapaz de estabelecer conexões com perspectivas críticas que clamam mudanças, os poucos que se expõe correm o risco de serem perseguidos caso não tenham aprovação do grupo onde estão inseridos. Existe, portanto, uma responsabilidade moral, como coletivo humano, de adensar propostas e ideias que visam benefícios comuns. Estabelecer atitudes colaborativas que fortaleçam a autonomia sem cair nas armadilhas de disputas e críticas pessoais que minam todas as potenciais alianças.

Entretanto, o desejo de democracia é combatido por estratégias pervasivas que influem no nível subconsciente. Os preconceitos que operam no nível subconsciente (GREENWALD, BANAJI, s.d.) são apenas uma ‘fatia’ do repertório capaz de bloquear empatias e estimular animosidades entre os ‘comuns’. Junto à autonomia a democracia é também derrotada a cada instante para que a ‘sociedade automata’ mantenha seu ritmo e controle, sem interrupções ou risco de rupturas.

A autonomia do sujeito que obedece a tal fluxo replica a simplificação necessária para o sucesso da ‘sociedade automata’.

Como influência de um darwinismo social, adaptar-se significa estar apto a “jogar as regras do jogo” tal qual ele se apresenta. Sob pena de extinção. E o humano que extingue o próprio humano é aquele que não quer ouvir o que o outro tem a dizer, não quer ser obrigado a pensar, menos ainda transpor o seu universo para entender a perspectiva alheia à sua própria. Menos ainda sentir que teria a obrigação moral de se opor a certas práticas que se distribuem de forma viral, dos comandos das lógicas de mercado às instituições.

Afinal, do ponto de vista político, o que é adaptar-se? Aceitar qualquer circunstância sem questionamentos? Um exemplo de quebra do “cumprimento do dever” profissional em virtude da consciência ética aconteceu durante a Segunda Guerra Mundial com o militar Francês e especialista em computação, René Carmille (WIKIPEDIA, 2016). Ele era conhecido como um “defensor ardente dos cartões perfurados” e tinha todas as máquinas do serviço demográfico do governo Francês em boas condições, de modo que “prometeu” que conseguiria facilmente entregar todos os judeus abrigados no território de seu país. Quando os alemães chegaram em agosto de 1940, para confiscar as máquinas Hollerith, Carmille resgatou os tabuladores e, três meses depois, conduziu o senso para uma estatística social a serviço

de uma agenda racial. Ele teve acesso a todas as máquinas e foi o responsável pela assinatura de contratos milionários com empresas diferentes que ofereciam tabuladores similares; enquanto se colocou formalmente a serviço da identificação dos judeus, ele também tomou algumas medidas para atrasar o processo com alegações técnicas de erros no sistema (BLACK, 2001). Carmille se tornou o único especialista ‘confiável’ para conduzir o senso com as máquinas Hollerith e, secretamente modificou os resultados, salvando assim dezenas de milhares de vidas. Sua autonomia estava alinhada com a participação na resistência política francesa. Ele foi capturado e morto pelos nazistas em 1945.

No mesmo período, a obediência de Eichmann, que estava no comando das operações que Carmille sabotou o quanto pôde, foi ‘premiada’ com o direito à vida. Pelo menos por mais de uma década, até o momento que o encontraram fugitivo na Argentina, e o levaram a julgamento em Jerusalém (ARENDT, 2013).

Uma sociedade composta de indivíduos autônomos e não-automatos poderia iniciar-se desde o questionamento sobre ‘por que criar máquinas para classificar as pessoas de acordo com sua origem ‘étnico-racial’? Este é um questionamento bem-vindo a todos, e no caso das máquinas, com especial interesse aos engenheiros, matemáticos, designers, empresários e funcionários. O ‘dispositivo’ capaz de frear a ‘banalidade do mal’ é o pensamento ativo e participativo, onde uma autonomia genuína se alicerça.

A autonomia do indivíduo requer a consciência expandida, a teia da inteligência coletiva que ampara cada um para que o senso crítico não seja vítima do totalitarismo. Entretanto, a consciência expandida não é livre das influências dos ‘consensos fabricados’ (ACHBAR, 1993).

Seria, portanto, esperado que aquele que se rebela exerce a vontade autônoma de acordo com sua consciência. Porém o Século 21 ingressa na ‘pós-verdade’ (FEITOSA, 2017) e a maneira das coisas fazerem sentido muda de lugar. Por exemplo, na invasão ao Congresso dos Estados Unidos em 6 de janeiro de 2021 (BBC, 2021) o ‘rebelde’ é fabricado por notícias falsas. Ele nega as dinâmicas da sociedade e os fatos, seguindo seus instintos. Acreditando ser um sujeito autônomo, é também um sujeito automato, ou seja, um sujeito manipulado para levar suas insatisfações ao limite, mesmo que seja quebrando os laços sociais por meio de conflitos.

Este é um fenômeno que tem crescido como resultado de manipulações da opinião pública com objetivos políticos, por meio das mídias sociais. Há aí um embaralhamento entre uma suposta autonomia que se desprende de valores ético-sociais. Paradoxalmente, grupos afinados com valores democráticos também estão passíveis de manipulação, potencialmente adensando os conflitos. Em linhas gerais, estruturas hierárquicas fazem parte de todo tipo de estrutura e organização. Porém a forma com que tais estruturas reforçam direta ou indiretamente preconceitos e desvios éticos é algo que transborda para além da presença das coisas que estão em objeto de análise.

Proponho, neste sentido, que a Bioética explore os territórios da ‘não-presença’ (NAS, 2021) para que possa observar com atenção os desafios bioéticos do tempo presente.

A autonomia em um cenário geopolítico, neoliberal, e contemporâneo está direcionada à disputa e não à cooperação.

Em uma democracia, eleger um candidato é apenas um passo para os resultados que venham a beneficiar a maioria. Todos os governos, mesmo os governos que não são afinados com as agendas de direitos humanos e sustentabilidade ambiental, temem a opinião pública.

O ‘consenso fabricado’, portanto, é um dispositivo necessário das estratégias de poder. Por este motivo, o cenário que antecedeu à ditadura militar iniciada em 1964 no Brasil foi de grande investimento de empresários em propaganda subliminar que circulou em todos os cantos do país através de projeções de cinema alternativas (MARINHO, 2014).

Assim, o ‘hackeamento’ da subjetividade das populações através das mídias, por meio de imagens e mensagens diretas ou subliminares, é anterior ao *Facebook*, *YouTube*, *WhatsApp* e *Google*.

A extrema divisão das ‘disciplinas’ de estudo proporciona o aumento de lacunas entre as vivências e as ideias que temos sobre elas. A discussão sobre o impacto da estética na percepção ética dos indivíduos demanda um exame cuidadoso.

A ‘não-presença’ de pessoas pretas, pardas e morenas nos modelos de família feliz e próspera das propagandas pré-Ditadura de 64 (MARINHO, 2014), trazia embutida a mensagem da exclusão e desigualdade social. Entretanto, tais mensagens são elaboradas para que a maioria as perceba como uma ‘ordem natural’. Além disso, estas propagandas eram uma cópia do modelo industrial de sucesso e do ‘sonho americano’. A classe média se

espelhou nesse modelo para tomar sua decisão de apoio à intervenção militar. Não foi necessário pensar em todos aqueles que não foram incluídos em tal modelo de prosperidade. Não foi necessário pensar. A propaganda propôs o modelo de pensamento, e bastaria apenas tomar este modelo como ‘dado’, mesmo a despeito da sua incapacidade de representar a sociedade em sua diversidade.

6. DISCUSSÃO

Os brancos saíram, num tempo muito antigo, do meio de nós. Conviveram com a gente, depois se esqueceram quem eram e foram viver de outro jeito. Eles se agarraram às suas invenções, ferramentas, ciência e tecnologia, se extraviaram e saíram predando o planeta. Então, quando a gente se reencontra, há uma espécie de ira por termos permanecido fiéis a um caminho aqui na Terra que eles não conseguiram manter. (KRENAK, 2020)

A sociedade automata se movimenta em ritmos repetitivos em um ambiente de atritos que, dentro da máquina, são geradores de energia. Os valores embutidos em sua epistemologia ditam seu modo de separar, classificar e organizar a informação incorporada nos produtos, projetos e toda sua infraestrutura.

Como a Bioética poderá auditar as ideias e crenças embutidas nos sistemas e que são restritivas da autonomia necessária dentro de uma cooperação social? A obediência sobreposta ao próprio humanismo, como conceito, no caso do Holocausto, demonstra como os princípios de autoridade e hierarquia são estruturas inegociáveis de uma ‘sociedade automata’. O holocausto praticado pelo governo nazista na Segunda Guerra Mundial carrega consigo a mesma lógica do ‘holocausto brasileiro’ (ARBEX, 2011) e do estupro de centenas de milhares de mulheres que vivem em territórios de guerra (ZIN, 2013).

Existe uma ‘não-presença’ de ideias (NAS, 2021) que torna possíveis tais atrocidades acontecerem com pouca ou nenhuma resistência ao nível social. As vítimas, vulneradas pelo preconceito e abandono, submetidas pela força, têm pouco com o que resistir, especialmente se não existem medidas de proteção e a omissão do meio prevalece. O conceito de omissão não tem nenhum sentido na sociedade composta por sujeitos-máquina que são operacionais e apenas se atém às suas funções.

A história do ‘holocausto brasileiro’ resgatada pela jornalista Daniela Arbex (ARBEX, 2013) é o reflexo de uma sociedade adestrada à obediência entre dois regimes totalitários, o Estado Novo e a Ditadura de 64. No livro, ela cita apenas um médico que ao final dos anos 70 levou a denúncia a público em formato de artigo científico e foi ameaçado de perder o registro devido à divulgação do artigo na imprensa. Portanto, é de se espantar como todos os profissionais envolvidos, da medicina, enfermagem, administração e assim por diante, adaptaram-se a fazer parte de uma operação recorrente onde as pessoas eram empilhadas como em um depósito de lixo. Elas eram submetidas à tortura até a morte porque seus cadáveres valiam algum dinheiro e suas vidas não valiam nada. A total falta de direitos e dignidade humana naturalizou-se. A categoria ‘louco’ e ‘indigente’ foi utilizada para eliminar o que se considerava a escória da humanidade em um laboratório aberto às práticas da ‘banalidade do mal’.

Tudo aconteceu com a participação do Estado, de Entidades Religiosas, Conselhos de Medicina, Psiquiatria e Universidades que compravam os corpos e órgãos. Quando até mesmo freiras são capazes de abusar de autoridade com pessoas em situação de total vulnerabilidade, torna-se evidente que a humanidade se perdeu de sua humanidade e possivelmente não se deu conta. Por isto vamos falar da ‘não presença’ das ideias que penetram nos sentidos através de conteúdos semióticos e são capazes de distorcer a percepção dos fatos, assim como moldá-los de acordo com uma visão dominante.

O Hospital Psiquiátrico continuaria – sob os mesmos moldes - em alta demanda se não houvesse a reforma psiquiátrica (REPÚBLICA, 2001), desde que é um dispositivo (ARBEX, 2016) utilizado para expulsar todos aqueles que por um ou outro caminho se desviam do que é considerado normalidade (RATTON, 1979).

Ele funcionava como um depósito de todos considerados ‘improdutivos’, ‘inadaptados’, ‘indesejáveis’ e os desafetos. (RATTON, 1979)

O que significa, portanto, enclausurar todos que, de alguma maneira ‘incomodam’, em um regime de exclusão ao direito à vida e à dignidade que deveria acompanhá-la? Não seria esta mais uma evidência de que a ‘banalidade do mal’ se torna perversa à medida que a ideia de democracia é ‘sequestrada’ por propostas autoritárias? Ou seja, através de aprovação popular regimes totalitários se instauraram e a ideia de segregação,

exclusão e mesmo extermínio de outros humanos foram endossadas através de justificativas que se cercam apenas de interesses próprios.

A ‘sociedade automata’ pode não perceber, mas é higienista, sob a perspectiva de uma visão de ciência que se torna ultrapassada à medida que se percebe que as ‘categorias ideais’ são influenciadas por preconceitos eugenistas. Em adição, poder acusar alguém de ‘louco’ levantando pretensas suspeitas de periculosidade devido a comportamentos de dissenso, é característica de ausência de lei. Em um estado de exceção as acusações nunca são consideradas infundadas, desde que são ‘dispositivos’ de terror onde a palavra do acusador terá sempre maior peso do que a da vítima. Neste contexto, o Hospital Colônia em Barbacena (MG) foi bastante conveniente para que abusadores sexuais empurrassem suas vítimas, e mesmo as crianças geradas do ato do estupro, para o esquecimento e a morte. O mesmo aconteceu com esposas que ‘precisavam’ desaparecer para que os maridos casassem novamente, e mesmo os filhos desse casamento ‘descartado’ também precisavam ser jogados às ruas como órfãos.

Como alguém, mesmo não tendo um problema mental grave, poderia manter alguma sanidade mental sendo tratado com choques elétricos a cada contestação, e vivendo em condições subumanas? O ‘holocausto brasileiro’ tem alguns sobreviventes. Alguns com sequelas aparentes, outros não.

Voltemos então ao behaviorismo para lembrar que sua proposta é de limitação da autonomia. A intenção é claramente “retirar as funções anteriormente atribuídas à autonomia do indivíduo e transferi-las uma a uma ao ambiente de controle” (SKINNER, 1971).

Há uma interferência no processo de identidade quando o ser é forçado a fazer parte de categorias pré-definidas com base na crença de que há uma ciência capaz de entender a ‘natureza do comportamento humano’ através de classificações, estatísticas e probabilidades.

A formulação behaviorista tem sido um ‘problema da psicologia’ que gera debates e discordâncias desde 1912 (MCDOUGALL, 1929). Há uma reivindicação da doutrina como ciência, mas o que não está presente em seus pressupostos é a sua filiação política e ideológica. Fato que possivelmente sequer seus defensores estavam plenamente conscientes, desde que os aspectos políticos e ideológicos se escondem em um modelo de universalidade.

O transbordamento desta política aparece por vezes quando os defensores do behaviorismo parafraseiam Maquiavel dizendo que o comportamento não se controla pelo amor, mas pelo medo, ou quando se lembra que a observação do comportamento está presente no ‘pecado original’ de Adão e Eva (MCDOUGALL, 1929): a tentação da serpente é parte de um jogo de manipulação psíquica que não busca investigar a autoconsciência. A serpente não pede à Eva para olhar para dentro de si (MCDOUGALL, 1929).

Assim, o conto do pecado original é cercado por fronteiras onde a autonomia está condicionada entre obedecer ou desobedecer. Permanecer na ignorância, ou sofrer as consequências da perda da inocência. Quanto de autonomia está presente na escolha entre Deus ou a Serpente? Seria apenas eleger o que mobiliza o medo maior?

Façamos uma digressão sobre a semiótica do holocausto como potência incentivadora do medo. As vítimas dos holocaustos são os dejetos da máquina. Elas são deglutidas e expelidas. Elas possuem uma função, mas é uma função não reconhecida como valor dentro da máquina. O não-valor faz parte do excedente de energia que elas proporcionam à ‘sociedade automata’. Filhos de estupros de mães dilaceradas, vidas às quais não são conferidos qualquer direito à dignidade humana.

Diante de uma realidade onde a ‘não presença’ é a absoluta violência, palavras e conceitos como ‘comportamento’ ou ‘autonomia’ esvaziam-se.

A absoluta violência está presente nas áreas mais nobres da máquina, desde a arquitetura até os mais populares dispositivos eletrônicos.

Ademais, quando transferimos dilemas como o do ‘pecado original’ para uma outra cultura como de indígenas brasileiros não haverá culpa em comer um fruto. Assim, o que torna estes e outros indivíduos inadequados à máquina é também o fato de que não fazem parte do ‘padrão de fábrica’, não podendo assim, ocupar um lugar de ‘peça’ no sistema. Na incapacidade de destruir a máquina, um fracasso também dos luditas – tecelões ingleses do Século 19, lhes resta serem deglutidos, preteridos e expelidos. Mesmo os que venham a se render à ocidentalização estão fadados a ocupar uma posição subalterna e vulnerada pelos ‘dispositivos’ de discriminação racial e socioeconômica. Assim, a dúvida de que a máquina poderá lhes incorporar como peça permanente estará sempre presente. Peças descartáveis do sistema fazem parte dos requintes de crueldade da obsolescência programada e naturalizada dentro da sociedade de consumo.

Assim, se o behaviorismo propõe moldar humanos como máquinas - e embora sua tese já tenha sido (e continua sendo) bastante criticada - a proposta é tão sedutora e afinada com a organização da sociedade industrial - e à sua ciência - que a ‘não-presença’ dessas ideias é reforçada através de dispositivos que penetram em todas as camadas da vida social e dos corpos.

São compreensões do mundo que se tornam pervasivas e, pela sua ‘não-presença’ permanecem de difícil alcance nas discussões mediadas pelos métodos que clamam objetividade.

7. REFLEXÕES FINAIS

A ‘sociedade automata’ como sociedade de controle não é novidade alguma, o que se procura entender é porque a crítica do seu modelo, que segue recorrente desde o século passado, foi ignorada ou pouco absorvida, enquanto as transformações tecnológicas seguiram em paralelo, com enorme entusiasmo da comunidade científica e sociedade.

O humano-máquina é conhecido pela sua funcionalidade. “Qual a sua utilidade?” Ela é uma médica, engenheira e cientista da computação. Ele é um professor, técnico, enfermeiro. Eles correspondem a uma função dentro de um mecanismo que compõe a máquina. São conjuntos de mecanismos onde tudo deve ter uma serventia, cabendo à pergunta se tal pragmatismo serve mais à comunidade e às pessoas do que à própria ‘máquina’.

As máquinas, assim como os humanos, também carregam o sentido de finitude através da obsolescência: elas quebram, “dão defeito” e se tornam ultrapassadas pela conveniência dos que desejam lucrar com elas. As máquinas precisam quebrar para que outras sejam vendidas (DANNORITZER, 2010). As pessoas precisam ser descartadas para que outras também sejam descartáveis. Elas precisam sofrer injustiças para que outras também sofram injustiças.

Em 1971 o pensador Noam Chomsky argumentou que seria importante investigar seriamente a reivindicação de que a ciência do comportamento, e a tecnologia que ela propõe, fornece os meios racionais para o controle do comportamento. Ele observou que, se o behaviorismo transfere todo controle comportamental ao ambiente, a autonomia torna-se apenas ilusória (CHOMSKY, 1971).

Se a ‘ciência do comportamento’ encerra a compreensão do indivíduo como sendo unicamente um mero resultado de antecedentes, ou seja, de eventos passados, veremos então que as técnicas propostas do behaviorismo, estão hoje embutidas nos sistemas computacionais que são ubíquos em toda estrutura social e da vida nos seus mais diversos aspectos.

Quais seriam os ‘anticorpos’ capazes de combater a pandemia da ‘banalidade do mal’?

As proposições pseudocientíficas do behaviorismo foram refutadas (CHOMSKY, 2017) mas se ainda permanecem influentes é porque as técnicas aplicadas inicialmente às experiências científicas e educação, estão embutidas nas organizações e suas infraestruturas. Além disso, a constante busca por padrões de reconhecimento na psicologia, assim como a fácil aceitação e assimilação destas classificações, torna os questionamentos éticos a estas práticas ainda brandas. Neste sentido, a proposta deste capítulo é compartilhar a reflexão sobre que elementos não-evidentes sustentam uma ‘sociedade automata’ e como a autonomia do sujeito moderno em um sistema industrial se confunde com a autonomia da máquina: é possível trabalhar de maneira independente, desde que em um sistema de total controle.

A ideia de que o comportamento humano precisava ser classificado e dividido em modelos e categorias foi uma maneira de reforçar o controle dentro da perspectiva científica moderna (SKINNER, 1965).

O comportamento regido por leis capazes de adequar cada um de maneira determinística reflete o entendimento do corpo como uma máquina que deve estar sob controle. Nesta perspectiva, assim como para La Mettrie, a mente não está dissociada do corpo. Ela se ajusta, tanto quanto o corpo e apresenta seu repertório de previsibilidades.

No exemplo do ‘holocausto brasileiro’ vemos inicialmente o poder da imagem, que independente da vontade do fotografo, resultou na primeira denúncia na imprensa, em 1961. Porém, a Ditadura de 64 reforçou a prerrogativa do abuso de poder e o caso caiu novamente no esquecimento até que ao final dos anos 70, um período histórico em que outros setores da sociedade brasileira voltavam a se mobilizar, a imagem em movimento, através do cinema, em conjunto com uma série de matérias na imprensa nacional e internacional adensaram às críticas e protegeram os poucos que se arriscaram a denunciar.

Desde que todo repertório da ciência foi utilizado para reforçar supostas diferenças raciais e prover, com um enredo de justificativas biológicas (HERBES-SOMMERS, 2003), os argumentos capazes de discriminar, separar, e, mesmo eliminar, outros seres humanos, foram naturalizados e permanecem vivos no imaginário social, sendo necessária uma desconstrução. Entretanto esta não será possível enquanto todos os meios de validação do conhecimento passarem pelos mesmos caminhos que são os de reforço às estruturas antigas.

A fissura entre mente e corpo proposta pelo dualismo, apresenta a compreensão mecânica do corpo. O isolamento da mente permitiu que o corpo fosse investigado, e a medicina ocidental avançou no conhecimento da fisiologia, através de estudos empíricos, enquanto no plano conceitual, uma luta constante no entendimento dos cuidados com a saúde, seriam travados, por séculos, até os dias atuais. As bases de tais conflitos estão na esfera da epistemologia: que compreensão se têm desse corpo? É um sistema integrado, ou suas partes são separadas e podem haver ‘peças de reposição’ como em um sistema maquinico? É um sistema independente, ou ele é integrado com o ambiente, constantemente afetado pelo que está à sua volta? Existe “dentro” e “fora” do corpo, ou as sensações desse corpo respondem ao “dentro” e “fora” sem fazer qualquer distinção?

A autonomia do corpo não se dá independentemente da cultura e do tempo histórico onde este corpo está inserido.

O dualismo, longe de ser uma invenção de Descartes, remonta o princípio hilemórfico aristotélico, onde “a distinção entre forma e matéria, alma e corpo, reflete uma cidade composta de cidadãos em oposição à escravos” (SIMONDON, 1995). Esta distinção, portanto, revela o recorte da visão que se propaga pela fôrma, ou seja, a forma, que é a conceituação de um modelo replicado, da epistemologia para todas as áreas de produção de saber.

A ideia de que o ser pensante poderia ser ludibriado pelas sensações do corpo, reforça modelos hierárquicos (pensar é mais importante que sentir...) e repressões baseadas em crenças de que desejos e fluidos corpóreos são inadequados.

Atribuir maior valor às capacidades cognitivas como se estas estivessem descoladas do sentir, assim como de que um determinado grupo de pessoas são dignos de empatia e direitos de igualdade, enquanto outros não

são, faz parte de uma contínua repetição, justificada do mesmo modo que a fôrma do tijolo só pode produzir tijolos (SIMONDON, 1995).

Enquanto o dualismo cartesiano tem o princípio hilemórfico embutido em sua estrutura, o monismo spinozista oferece uma compreensão da ética onde todos os organismos coexistem sem que lhes sejam atribuídas escalas de valor.

Spinoza (1632-1677), além de negar a hierarquia e separação entre corpo e mente, postulou que “o corpo humano se compõe de muitos indivíduos (de natureza diferente)” (SPINOZA, 2009), aos quais, são, também eles, altamente compostos. Dos estudos de Teologia à Física, Spinoza tornou-se ótico e suas habilidades com as lentes ópticas lhe abriram à observação da matéria por ângulos distintos.

Entender que o corpo humano é um conjunto, que não se resume a órgãos com funções delimitadas e mecânicas, mas que é constituído de um todo plural e composto de muitos organismos. Esta visão, ainda hoje é capaz de revolucionar a medicina, assim como toda ciência. O monismo, assim denominado como resposta ao dualismo, é, na verdade, pluralista, o que tornou mais complexa sua assimilação.

Embora seja possível perceber a presença do monismo em distintos campos do conhecimento, seu impacto na ciência é minimizado pela ampla adesão à perspectiva dualista já que esta apresenta maior facilidade de instrumentação. O monismo se desdobra entre idealismo e materialismo, entre a metafísica e o fisicalismo, tendo influenciado correntes como behaviorismo, funcionalismo, utilitarismo e fenomenologia.

Continua sendo difícil demonstrar que os modos de fazer ciência e de pensar não são neutros e que existe uma política embutida em cada preceito, assim como na linguagem e no exercício do método.

A base spinozista sugere um corpo aberto em comunicação permanente com o mundo, sem distinção ou barreira do que é considerado natural ou artificial. A natureza é tão criativa e mutante, como também poderá ser repetitiva e previsível.

Se a observação da ciência precisa de repetição, observamos que repetição não é necessariamente generalização (DELEUZE, 1994). O estudo da consciência expressa o problema de maneira perfeita: enquanto as associações tendem a se repetir, elas podem se revelar únicas a cada vez, no mesmo indivíduo.

A sociedade é automata através das formas de ver e representar o mundo compostas por todos os indivíduos que fazem parte dela, e do que eles supõem ou acreditam ser verdade sobre este mundo.

Seria necessário deixar desmoronar algo para construir “o novo” dentro de uma outra perspectiva. Ou, se a vontade é esquivar-se de tal radicalidade, que se tracem estratégias de mudanças entre o paradigma da ‘sociedade automata’ para o da ‘sociedade autônoma’.

Tais estratégias passam pela compreensão do mundo simbólico, dos impactos do mundo simbólico – da estética – sobre a vida. Não se trata de minimizar a arte como a desprezada sala de “Terapia Ocupacional”, para onde foi direcionada a Dra. Nise da Silveira quando se negou a adotar a eletroconvulsoterapia nos pacientes (BERLINER, 2016).

Trata-se de valorizar a experiência que se tem com a vida, a cada instante. A vida, leia-se, o ambiente, a qualidade das inter-relações.

O foco da ‘sociedade automata’ é a eficiência da máquina. A autonomia se expressa na vegetação que cresce entre as brechas, da água que encontra bifurcações e abre novas passagens quando existe um bloqueio na via principal.

Cabe à Bioética potencializar estes movimentos criativos da natureza que se refletem em humanos e não-humanos desviantes de sentidos e representações que não lhes incluem.

8. REFERÊNCIAS

ADAM DEL DEO, YON MOTSKIN, LISA NISHIMURA, STACEY OFFMAN, JASON SPINGAM-KOFF, A. G. **Na Rota do Dinheiro Sujo (Dirty Money)**. USA Netflix, 2018.

AMER K, N. J. **Privacidade hackeada**. Los Gatos: Netflix, 2019.

ANTROPOFAGIAS. **Pensamento Indígena - Viveiros de Castro Aula 2**. YouTube, 2018.

ARBEX, D. Holocausto brasileiro: 50 anos sem punição. **Tribuna de Minas**, 2011.

ARBEX, D. **Holocausto Brasileiro**. São Paulo: Geração Editorial, 2013.

ARENDT, H. **Eichmann em Jerusalém: Um Relato Sobre A Banalidade do Mal (1963)**. São Paulo: Companhia das Letras, 2013.

BERLINER, R. Nise - **O Coração da Loucura**. Brasil Imagem Filmes, 2016.

BLACK, E. **IBM and the Holocaust: the Strategic Alliance Between Nazi Germany and America's Most Powerful Corporation**. [s.l.] Crown Books, 2001.

C. HERBES-SOMMERS, J. CHENG, L. ADELMAN, L. SMITH, T. S. **Episode One- The Difference Between Us. In: Race: The power of an Illusion**. California Newsreel, 2003.

CHOMSKY, N. The case against BF Skinner. **The New York Review of Books**, v. 17, n. 11, p. 18–24, 1971.

CHOMSKY, N. **Noam Chomsky on Behaviorism**. YouTube, 2017. Disponível em: <<https://youtu.be/OrQ0LfqxABM>>

CHRISTMAN, J. **Autonomy in Moral and Political Philosophy**. Stanford University, 2020.

WIKIPEDIA. **Rene Carmille**.

COMISSION, E. **General Data Protection Regulation**. Disponível em: <https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en>. Acesso em: 31 ago. 2018.

DAMÁSIO, A. **O erro de Descartes: emoção, razão e o cérebro humano**. [s.l.] Companhia das Letras, 2003.

DANIELAARBEX, A. M. **Holocausto Brasileiro**. Brasil Vagalume Filmes, 2016.

DANNORITZER, C. **A História Secreta da Obsolescência Planejada (The Light Bulb Conspiracy)**. França, Espanha, 2010.

DELEUZE, G. Post-Scriptum sobre as Sociedades de Controle. In: **Conversações: 1972-1990**. Rio de Janeiro: Ed. 34, 1992. p. 219–226.

DELEUZE, G. (TRANS. P. P. **Difference and Repetition (1968)**. New York: Columbia UP, 1994.

DINES, A. **Observatório da Imprensa Entrevista o Sociólogo Zygmunt Bauman**. TV Brasil, 2015. Disponível em: <<https://www.youtube.com/watch?v=kM5p8DqgG80>>

DRIVER, J. **The History of Utilitarianism**. Stanford University, 2014. (Nota técnica).

FEITOSA, C. Pos-verdade e política. **Cult**, 2017.

HERNAN ZIN, ANDREW FORD, A. P. **La Guerra Contra Las Mujeres**. Contramedia Films, 2013.

Invasão ao Congresso dos EUA: quatro pessoas morrem; Biden fala em “insurreiç o”. **BBC News**, 6 jan. 2021.

JO O CANAVILHAS, MA RA BITTENCOURT, M. A. A. DE A. Conte dos virais no Facebook: estudo de caso na pr -campanha das elei  es presidenciais brasileiras de 2018. **Brazilian Journalism Research**, v. 15, n. 3, p. 598–625, 2018.

JOSEPH, P. **Zeitgeist: The Movie** Gentle Machine Productions, 2007. Dispon vel em: <<http://www.zeitgeistmovie.com/>>

KRAMER, ADAM DI; GUILLORY, JAMIE E.; HANCOCK, J. T. Experimental evidence of massive-scale emotional contagion through social networks. **Proceedings of the National Academy of Sciences**, v. 111, n. 24, p. 8788–8790, 2014.

KRENAK, A. **A Vida N o   Util**. S o Paulo: Companhia das Letras, 2020.

L U, M. DE O. *ET AL.* **Detec  o autom tica de bots em redes sociais: um estudo de caso no segundo turno das elei  es presidenciais brasileiras de 2018**. [s.l.: s.n.].

MARINHO, G. F. **O Pr logo: 1960’s Brazilian Political Cinema**. Aggregator Media, 2014.

MARK ACHBAR, P. W. **Manufacturing Consent: Noam Chomsky and the Media**. Zeitgeist Films, 1993.

MCDUGALL, J. B. W. W. **The battle of behaviorism: An exposition and an exposure**. [s.l.] WW Norton & Company, 1929.

METTRIE, J. J. O. DE L. (TRANS. M. D. **Man a Machine**. London: Homer’s Head, 1749.

MILGRAM, S. Behavioral Study of Obedience. **Journal of Abnormal and Social Psychology**, v. 67, n. 4, p. 371–378, 1963.

MILL, J. S. Utilitarianism. In: MOORE, A. D. (Ed.). **Information Ethics: Privacy, property, and power**. Washington: University of Washington Press, 2005. p. 47–65.

NAS, E. **Bioethics of NonPresence: body philosophy, and machines**, 2021.

NEWS, C. 4. **Cambridge Analytica Uncovered**. YouTube, 2018. Dispon vel em: <<https://youtu.be/mpbeOCKZFfQ>>

NORONHA, JOS  CARVALHO DE; CASTRO, L. Democracy, Health, and the 16th National Health Conference in Brazil: what future? **Cadernos de Sa de P blica**, v. 35, 2019.

O'NEIL, C. **Weapons of math destruction: how big data increases inequality and threatens democracy**. New York: Crown, 2016.

ONG denuncia 44 países que ainda tem leis discriminatórias contra a mulher. **Correio Braziliense**, 20 fev. 2015.

RATTON, H. **Em Nome da Razão**. Brasil Quimera, 1979.

REGO, S. Sobre a classificação de indivíduos em estágios de acordo com a teoria de desenvolvimento moral de Lawrence Kohlberg. **Ciência & Saúde Coletiva**, v. 10, n. 4, p. 1082–1083, 2005.

REPÚBLICA, P. DA. **Lei N. 10.216**. Brasil, 2001.

RODRIGO SIQUEIRA-BATISTA, F. R. S. A eutanásia e os paradoxos da autonomia. **Ciência & Saúde Coletiva**, v. 13, p. 207–221, 2008.

SERGIO REGO, M. P. Contribuições para planejamento e avaliação do ensino da bioética. **Revista Bioética**, v. 25, n. 2, p. 234–243, 2017.

SIMONDON, G. **L'individu et sa genèse physico-biologique (1964)**. Grenoble: Millon, 1995.

SKINNER, B. F. **Science and human behavior**. [s.l.] Pearson Education, 1965.

SKINNER, F. B. **Beyond Freedom and Dignity**. New York: Knopf, 1971.

SPINOZA, B. **Ética**. Belo Horizonte: Autêntica Editoria, 2009.

TONY GREENWALD, MAHZARIN BANAJI, B. N. **Project Implicit**. Disponível em: <<https://implicit.harvard.edu/implicit/takeatest.html>>. Acesso em: 20 ago. 2020.

TRAPP, R. A autonomia da vontade em Kant. **Griot: Revista de Filosofia**, v. 9, n. 3, p. 197–210, 2019.

WORTH, R. F. A crueldade do apedrejamento. **Folha de S. Paulo**, 2010.

Este texto foi escrito e enviado enquanto pesquisadora visitante no Departamento de Informática da University of California - Irvine.

Agradecimentos: CAPES.

*Rômulo Cristóvão de Souza**Bruno Carlos da Cunha Costa**Leonardo Vasconcelos Guaurino de Oliveira*

1. INTRODUÇÃO

Usualmente, a demanda por serviço em saúde começa pela percepção de certa necessidade. A partir daí, uma sequência de passos é iniciada, podendo começar pelo agendamento de uma consulta ou a ida imediata a uma unidade de emergência. Tomando isto por base, é possível estabelecer uma espécie de mapeamento das etapas necessárias até a concretização do atendimento.

Um procedimento comum a todas as fases de um atendimento em saúde, e que será o ponto central do que trataremos neste capítulo, diz respeito à captura dos dados do indivíduo, os quais podem ser utilizados para diferentes propósitos, tais como, identificação, registros sobre aspectos de saúde, política pública em saúde, gestão, entre outros.

Se em outras épocas os dados eram coletados, mantidos e processados de forma manual, com o avanço tecnológico no campo da ciência da computação, cada vez mais os dados são armazenados eletronicamente, o que tem possibilitado o emprego de técnicas avançadas de análises e extração de informações. Mais recentemente, com o aumento na capacidade de equipamentos, tanto em termos de armazenamento quanto de processamento, aliada à grande evolução dos softwares, a chamada inteligência artificial passou a ganhar espaço em praticamente todas as atividades humanas, incluindo o segmento da saúde.

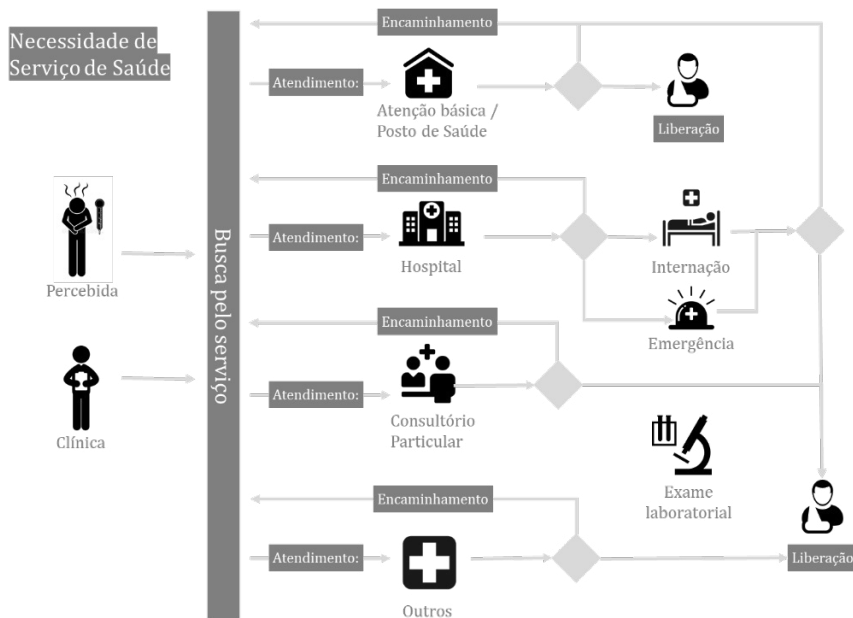
Indiscutivelmente, os recursos computacionais de que dispomos hoje são ferramentas que parecem se tornarem indispensáveis em quase todos os cenários. Contudo, a que se ter em mente eventuais danos colaterais, como usos indevidos e até ilegais dos dados. Neste sentido, há que se estabelecerem normas e legislações, além de discussões de questões éticas, ligadas à proteção e uso de tais dados.

Sendo assim, sob a ótica da saúde, o presente capítulo abordará, conceitos de Sistemas de Informação (SI), Big Data, além de exemplos de tecnologias voltadas para análises de dados, incluindo os avanços em termos de inteligência artificial. Finalmente, discutiremos aspectos éticos relacionados à aquisição e manutenção de dados, bem como a proteção dos mesmos, englobando nisto aspectos de bioética. Para tal, utilizaremos um cenário ilustrativo, que chamaremos de *caminho do paciente*.

2. O CAMINHO DO PACIENTE

O cenário, que chamamos aqui de O caminho do paciente, apresenta cinco fases gerais a saber: busca, atendimento e liberação ou encaminhamento (Figura 1).

Figura 1 – O Caminho do Paciente



Inicialmente, há uma percepção de *necessidade* de atendimento em saúde. Essa percepção pode ser do próprio indivíduo ou não. A partir daí, diferentes encaminhamentos podem ser dados – agendamento de uma consulta, ida imediata a um serviço, como posto de saúde ou hospital, por exemplo. A sequência de etapas até que a pessoa seja atendida pode variar bastante, assim como sua complexidade. Para fins das abordagens que

desejamos fazer, vamos considerar, de forma simplificada, um atendimento realizado em um ambiente hospitalar.

Certo indivíduo *busca* um serviço de emergência de determinado hospital. Inicialmente, dirige-se à recepção onde fornece dados como os de identificação. Após isso é encaminhado ao setor de triagem para uma avaliação em saúde, recebendo uma classificação indicativa da gravidade da situação e, conseqüentemente, da urgência do atendimento. Seguindo, é *atendido* por um médico. Depois dessa avaliação, fica sob cuidados da equipe de enfermagem que, entre outros procedimentos, faz administração das medicações prescritas.

Passado um período de observação, surge a recomendação de internação hospitalar. Cumpridos os requisitos burocráticos, o paciente é *encaminhado* ao apartamento. Nos dias que se seguirão, será avaliado por equipe de profissionais, passará por exames diagnósticos e outros procedimentos, sempre assistidos por enfermeiros e técnicos, até que é *liberado*, recebendo a alta médica.

3. DADOS E INFORMAÇÕES

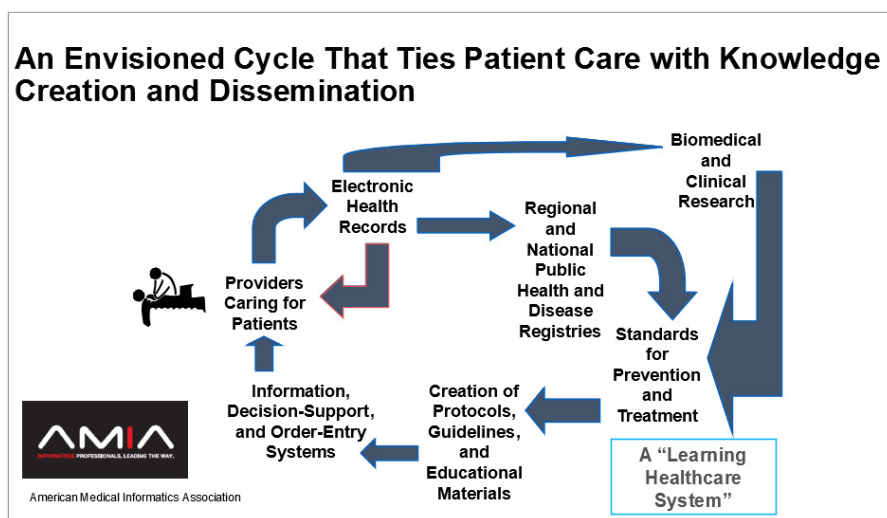
Uma distinção deve ser feita entre os termos que serão amplamente utilizados ao longo deste capítulo: *dado* e *informação* (KOLB & KOLB, 2013). Por dado entende-se a qualquer valor, seja numérico, como uma medida, ou textual que se obtém a partir de algum processo – questionário, aferição ou observação, por exemplo – que não tenha passado por qualquer processamento e, assim, isoladamente, não apresente algum significado. Informação é resultante de um conjunto de dados processados e que, geralmente baseado em alguma convenção, signifique ou indique algo. Usando um exemplo simples, quando se faz uma aferição da pressão arterial (PA) de uma pessoa, os valores indicativos da PA sistólica e diastólica são os dados. A interpretação desses valores produz uma informação para o profissional, que permitirá classificar o indivíduo em termos de ser hipotenso, normotenso ou hipertenso.

Com base na distinção acima, podemos apresentar o conceito de sistemas de informação (SI) (CARVALHO & EDUARDO, 1998), entendidos como um conjunto de elementos, relacionados ou não, voltados para a coleta, armazenamento e processamento de dados, produzindo, assim, informações. Convém dizer que um SI independe do uso de recursos de

informática. Portanto, é plenamente possível se estabelecer um sistema de informações baseado em meio físico, em papel.

Em um ciclo típico de atendimento em saúde, dados são coletados desde o nível individual e podem ser utilizados para diversos propósitos como auxiliar no estabelecimento de políticas públicas, pesquisas clínicas, criação de protocolos e guias, formando o que a Associação Americana de Informática Médica chama de o “o sistema que aprende” (Figura 2).

Figura 2 – ciclo típico de um atendimento em saúde



Fonte: American Medical Informatics Association – AMIA.

4. DADOS DO PACIENTE

Desde os primórdios da medicina o registro de dados dos pacientes foi se estabelecendo como prática recomendada. Historicamente, o primeiro relatório médico que se conhece data do período entre 3000 e 2500 a.C., feito por um médico egípcio [1]. No entendimento de Hipócrates, século V a.C., as anotações feitas pelos médicos sobre os pacientes por eles atendidos serviam a dois propósitos: 1) mostrar o curso da doença; 2) estabelecer as possíveis causas da doença. Essas anotações seguiam uma ordem cronológica, sendo, por isso, classificadas como registro médico orientado ao tempo (*time-oriented medical record*) [2,3].

Em 1880, William Mayo, fundador da Clínica Mayo nos Estados Unidos, observou que os médicos, em sua maioria, faziam os registros das consultas prestadas a todos os pacientes em um documento único. Esse método, contudo, gerava dificuldades na identificação de informação específica sobre um paciente. Dessa forma, a partir de 1907 a Clínica Mayo adotou a estratégia de se anotarem os registros dos atendimentos prestados de forma individualizada. Surgiu, então, o registro centrado no paciente, ainda que orientado ao tempo [2,3]. Foi também da Clínica Mayo a iniciativa de se tentar estabelecer um conjunto mínimo de dados a serem registrados, em um movimento na busca de uma padronização do que se convencionou como prontuário do paciente [3].

Sendo reconhecido instrumento primordial, o Conselho Federal de Medicina (CFM) define prontuário do paciente como: “...o documento único constituído de um conjunto de informações, sinais e imagens registradas, geradas a partir de fatos, acontecimentos e situações sobre a saúde do paciente e a assistência a ele prestada, de caráter legal, sigiloso e científico, que possibilita a comunicação entre membros da equipe multiprofissional e a continuidade da assistência prestada ao indivíduo”. [4]

Desta definição, destacamos os aspectos de os dados anotados nos prontuários dos pacientes serem sigilosos e, por isso, devendo estar sob proteção, servirem a propósitos diversos, entre eles a assistência e a pesquisa e favorecerem a assistência continuada. Adiciona-se a isto o fato de não ser um documento utilizado exclusivamente pelo profissional de medicina, mas, também, por agentes de outras categorias – enfermagem, fisioterapia, nutrição, entre outras. Ainda, embora, até pelo cenário que criamos, o prontuário do paciente possa parecer se utilizado apenas em hospitais, de fato esse documento deve estar presente em qualquer serviço assistencial em saúde, como ambulatorios, e consultórios – médicos e demais profissionais.

5. DADOS DO PACIENTE COMO FONTE PARA SISTEMAS DE INFORMAÇÕES

Da definição apresentada de SI, conseguimos estabelecer os dados como matéria prima para a produção informações. Nesse sentido, comumente fala-se em fontes de dados, ou seja, instrumentos por meio dos quais se fazem os devidos registros. Prontuários dos pacientes são uma das

principais fontes de dados dentro no âmbito da assistência em saúde. Mas não é o único. Vale dizer, também, que dados mantidos em um prontuário frequentemente são utilizados para alimentarem outras fontes, como, por exemplos, as de interesse para saúde pública, pesquisa e administrativo – vide a Figura 1.

Se considerarmos, tomando por referência a realidade do Brasil, unidades de atendimento, públicos ou privados, como hospitais, ambulatorios, consultórios, clínicas, laboratórios e seguradoras partes componentes de um sistema de saúde, em cada ambiente desses podemos encontrar fontes de dados. Aceitando que todo e qualquer local que mantenha registros, ou ofereça produtos, relacionados à saúde de indivíduos fazem parte desse sistema, incluiríamos à lista de componentes departamentos de saúde de trabalhadores, drogarias e agências de assistência social.

Assumindo que em todos os componentes de um sistema de saúde podemos encontrar fontes de dados, nos parece razoável afirmar a possibilidade de haver uma vinculação entre estas fontes. Com isso, no *caminho do paciente* os dados coletados estão sujeitos a um tipo de fluxo. É o que procuramos mostrar no cenário que criamos. Desde o primeiro atendimento, por qualquer serviço (hospitalar, ambulatorial), existe um encadeamento dos dados e em cada etapa do atendimento os mesmos são compartilhados, outros agregados e/ou atualizados.

6. DE SISTEMAS DE INFORMAÇÕES PARA BIG DATA

Falamos que um sistema de informação não depende, obrigatoriamente, de ferramentas de informática. Contudo, nas últimas décadas observamos a entrada e o enorme desenvolvimento dos computadores em todos ambientes e segmentos. Com o avanço tecnológico, um volume cada vez maior de dados passou a ser mantido por meios digitais. Mais recentemente, recursos computacionais, de equipamentos e dispositivos aos *softwares*, permitem o armazenamento, a integração e a produção de análises de dados em quantidades, velocidade, amplitude e sofisticação jamais vistos. Este novo paradigma de armazenamento e análise de dados é denominado Big Data.

Big Data é um termo que engloba qualquer tecnologia ou metodologia utilizada no processamento de grandes volumes de dados armazenados digitalmente (XU & DUAN, 2019). Tais tecnologias e metodologias são

comumente utilizadas em quatro fases comuns em um sistema de informação baseado em Big Data, a *saber, geração, aquisição, armazenamento e análise* de dados (CHEN, 2014). A seguir, apresentamos uma breve explicação de cada fase, contextualizando-a sua aplicação ao caminho do paciente apresentado anteriormente.

A *geração* de dados é a primeira etapa em Big Data. No caminho do paciente estes dados podem ser gerados em quaisquer uma das fases. Por exemplo, quando um paciente tem uma necessidade percebida de saúde, antes de buscar por um serviço, ele pode pesquisar a respeito dos sintomas na Internet ou em aplicativos de celular. Ao buscar por um serviço de saúde, o paciente pode realizar uma pesquisa na Internet por unidades próximas. Quando é atendido, diversos dados pessoais e de saúde são gerados; assim como na liberação ou encaminhamento, onde os dados relacionados ao período de internação devem ser encaminhados.

A segunda fase de um sistema Big Data, *aquisição*, inclui a coleta, transmissão e pré-processamento dos dados. Durante esta fase, dados são coletados a partir das respectivas fontes e transmitidos para serem pré-processados e armazenados. Neste ponto é conveniente distinguir a forma de organização dos dados, a *saber*, estruturados, semiestruturados ou não estruturados. Dados estruturados são aqueles que contêm um esquema pré-definido, o qual permite conhecer o seu significado. Por exemplo, a Tabela 1 apresenta dados de pacientes em uma estrutura tabular, onde os títulos das colunas significam os dados das respectivas linhas. Caso não houvesse o título da coluna, não seria possível identificar com precisão o significado das datas 15/06/2017, 01/02/2002, 22/11/2008 (poderiam referir-se à data de nascimento).

Nome	Idade	Naturalidade	Primeira consulta
Ricardo Almeida	03	Rio de Janeiro	15/06/2017
Letícia Amorim	18	Petrópolis	01/02/2002
Fernanda da Silva	12	Teresópolis	22/11/2008

Tabela 1 – Estrutura tabular (exemplo de dados estruturados).

Diferentemente, dados não estruturados não têm um esquema prévio definido. Por exemplo, vídeos, fotos ou arquivos de áudio. Assim também, os dados semiestruturados podem não possuir esquema, no entanto, podem ser auto descritivos (contém o significado na própria descrição do dado). Um exemplo deste tipo de dado é o formato de arquivo BibTex, o qual é demonstrado na Figura 3. Neste formato, cada referência é estruturada unindo o dado com o seu significado (author = <nome_do_autor>; pages = <numero_de_paginas>).

Figura 3 – Exemplo de dados semiestruturados.

```
@article{Goncalves2013,
author = {Gon{\c{c}}alves, Jo{\~{a}}o Paulo Pereira and Batista,
Larice Rodrigues and Carvalho, Larissa Mendes and Oliveira,
Michelle Pimenta and Moreira, K{\^{}e}nia Souto and Leite,
Ma{\~{}i}sa Tavares de Souza},
doi = {10.1590/s0103-11042013000100006},
issn = {0103-1104},
journal = {Sa{\~{}u}de em Debate},
month = {mar},
number = {96},
pages = {43--50},
publisher = {Centro Brasileiro de Estudos de Sa{\~{}u}de},
title = {{Prontu{\~{}a}rio Eletr{\^{}o}nico: uma ferramenta que
pode contribuir para a integra{\c{c}}{\~{a}}o das Redes de
Aten{\c{c}}{\~{a}}o {\~{}a} Sa{\~{}u}de}},
volume = {37},
year = {2013}}
```

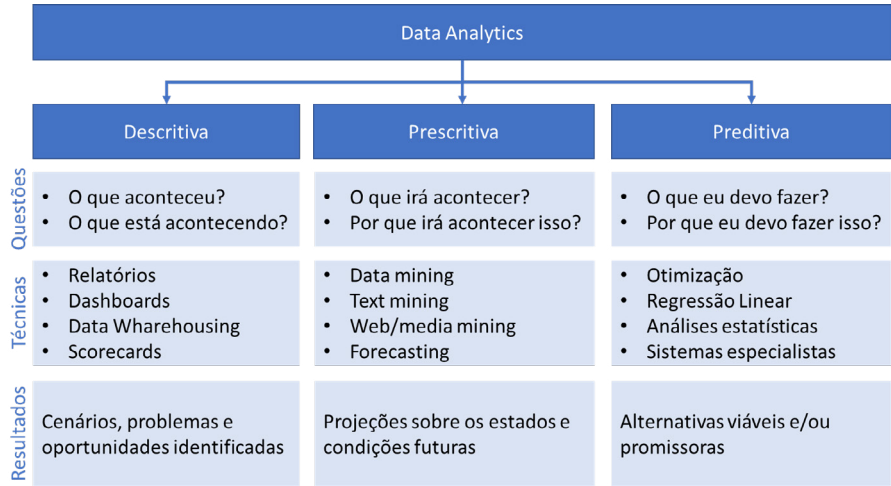
Fonte: elaborada pelos autores

Dados de saúde podem ser localizados em todas as três formas de estruturação. Um prontuário eletrônico é um típico exemplo de dados semiestruturados. No entanto, os dados nele contidos podem ser pré-processados e armazenados em um banco de dados relacional, o qual prevê um esquema pré-definido. Por outro lado, as anotações geradas durante a internação hospitalar ou entrevistas com pacientes gravadas no formato de áudio podem ser de característica não estruturada.

Na terceira fase, *armazenamento* de dados, os dados pré-processados são armazenados em conjuntos de dados denominados datasets, os quais serão utilizados na fase seguinte, análise de dados. O objetivo da quarta e última etapa, *análise* de dados (comumente referenciada como *Data Analytics*) é a extração de informação dos dados, as quais serão úteis para a tomada de decisão. Diversos métodos são utilizados nesta fase - análise de correlação, regressão linear, análise estatística, aplicação de algoritmos de data mining, hashing, MapReduce, entre outros.

Os diferentes métodos em *Data Analytics* podem ser categorizados em três grandes grupos – análise descritiva, análise preditiva e análise prescritiva (DELEN & DEMIRKAN, 2013). A análise descritiva busca identificar informações referentes aos dados passados. Por exemplo, buscando relações de causa e efeitos em dados epidemiológicos de anos anteriores. A análise preditiva busca prever o que poderá acontecer no futuro, tendo como premissa semelhança de evento futuro com fato passado. Com base nas relações de causa e efeitos encontrados em dados epidemiológicos de anos anteriores, é possível projetar cenários parecidas para o futuro. Finalmente, a análise prescritiva tem por objetivo analisar os dados e modelos de forma a oferecer alternativas viáveis para tomada de decisão. A Figura 4 apresenta uma taxonomia contendo questões, técnicas e resultados de cada um dos grupos.

Figura 4 – Taxonomia de Data Analytics



Fonte: Delen & Demirkan (2013).

Com base nas etapas descritas acima é possível identificar três características fundamentais de Big Data, a saber: volume, variedade e velocidade (os três V’s de Big Data) (GANDOMI & HAIDER, 2015). Volume se refere à magnitude dos dados, os quais podem chegar a Terabytes ou até Petabytes. Variedade refere-se à heterogeneidade dos dados, os quais podem estar estruturados, semiestruturados ou não estruturados. Por fim, velocidade refere-se ao tempo decorrido em todas as fases, desde a geração até a análise dos dados.

7. BIG DATA EM SAÚDE

A aplicação de técnicas de Big Data a fim de gerar conhecimento em saúde tem sido um importante objeto de estudo nos últimos anos. Um estudo (GALETSI, 2019) que revisou mais de oitocentos artigos relacionados, identificou que um dos grandes benefícios está relacionado a aplicação de novas abordagens para diagnóstico e prognóstico em medicina personalizada, a qual visa tratar a saúde do paciente de maneira exclusiva, levando em conta o seu histórico e analisando cada caso individualmente. Por exemplo, a criação de perfil genético para oferecer informações individuais de saúde com vias ao tratamento de uma variedade de doenças (cardiovasculares, câncer, diabetes etc.), permitindo o uso de esquemas terapêuticos personalizados.

Um segundo estudo (GALVÃO & VALENTIM, 2019) analisou pesquisas que buscaram conectar dados epidemiológicos para gerar informações pertinentes para a geração de conhecimento em saúde. Os autores destacaram os seguintes estudos:

Google-driven search for big data in autoimmune geoepidemiology: Analysis of 394,827 patients with systemic autoimmune diseases, Ramos-Casals (2015): utilizou o motor de busca do Google para coletar e mesclar grandes séries de doenças autoimunes sistêmicas, com o objetivo de obter a figura geoepidemiológica de alta de definição de cada doença.

Leveraging hospital big data to monitor flu epidemics, Bouzillé & Chazard (2018): mostrou que grandes dados de hospitais parecem ter um grande potencial para monitorar epidemias de influenza, além de fornecer características adicionais dos pacientes.

Cloud enabled data analytics and visualization framework for health-shocks pre-diction, Mahmud *et al.* (2016): apresenta um framework de visualização e análise de dados para predição de surtos baseado em um dataset de larga escala com dados de 1.000 lares no Paquistão.

Real-time predictive seasonal influenza model in Catalonia, Spain Basile (2018): projeta um modelo preditivo em tempo real sobre a taxa de incidência de ILI (Influenza Like Illness) na Catalunha usando dados de semanas anteriores e levando em consideração a distribuição geográfica dos surtos durante a temporada de influenza de 2015-2016.

No entanto, como também identificado no primeiro estudo supracitado (GALETSI, 2019), apesar dos benefícios e oportunidades, alguns dos desafios da utilização de Big Data em saúde está relacionado ao gerenciamento dos dados e, principalmente, questões de segurança e privacidade. Neste sentido, diferentes iniciativas têm sido propostas e adotadas com o objetivo de discutir e minimizar eventuais danos causados por usos indevidos de dados individuais, principalmente os de grande sensibilidade como aqueles relacionado à saúde.

Os cuidados que se impõem quando se lida com dados de terceiros são tratados de maneiras distintas pelos países. No Brasil, além dos direitos ao sigilo e privacidade já definidos pela Constituição, a Lei Geral de Proteção de Dados (LGPD) deve ser um potente instrumento no norteamento para a implementação de práticas que possam aumentar as garantias do correto e adequado uso dos dados pessoas em todos os cenários.

8. A LEI GERAL DE PROTEÇÃO DE DADOS E O CAMINHO DO PACIENTE

A recente revolução tecnológica e informacional trouxe grande preocupação p, principalmente no campo da dispersão de informações, que vem gerando grave insegurança jurídica no tocante à tutela e salvaguarda dos dados individuais, que seria a divulgação dos mesmos sem autorização pessoal ou legal para tanto, e os danos decorrentes dessa violação. A facilidade de acesso e dispersão de dados e informações desperta, ou deveria despertar, a noção do real valor dos dados pessoais e da importância que deve ser dada na elaboração e construção de sistemas sólidos o suficiente para impedir a ocorrência de incidentes de segurança da informação. Em se tratando de dados sensíveis, a exemplo dos produzidos e manuseados nos serviços de saúde, a preocupação com a segurança dos mesmos parece demandar cuidados ainda maiores uma vez que os mesmos revelam características pessoais, hábitos e comportamentos sociais, orientação sexual, resultados de exames laboratoriais, genéticos, entre outros aspectos.

Analisando o caminho do paciente apresentado na introdução, bem como todas as considerações técnicas acerca da segurança da informação, deve ser delimitado o ordenamento jurídico que tratará do tema, especificando as normas que regularão a atuação dos agentes de tratamento de

dados, e de todos os personagens que participam direta ou indiretamente da jornada do paciente, no caso, o ordenamento jurídico brasileiro.

A Lei Geral de Proteção de Dados (LGPD), neste contexto, é o documento que cria diretrizes gerais para uma readequação das atitudes dos profissionais da saúde durante suas atividades, como forma de tutelar o sigilo médico (VILLAS-BÔAS, 2015), assegurando os direitos de informação e transparência dos pacientes, a retificação, a exclusão e a atualização de informações, além do processamento seguro dos dados.

A mens legis, o espírito da norma, decorre da correta ponderação entre o direito à intimidade e o direito à informação, sendo protegido o núcleo fundamental de ambos os princípios, impedindo a ocorrência de aniquilação de qualquer deles, em prol da paz social e da ética nas relações jurídicas.

Importa salientar que a LGPD incide sobre toda e qualquer atividade de tratamento de dados: ocorrida no território brasileiro; que tenha dados pessoais coletados no Brasil; ou que seja realizada fora do país, quando atividade de tratamento tenha por objetivo a oferta ou o fornecimento de bens ou serviços ou o tratamento de dados de indivíduos localizados no território nacional⁴⁰ (BRASIL, 2018).

No setor da saúde, a temática recebe uma especial especificidade, tendo em vista as informações armazenadas serem bastante íntimas ao paciente, mas extremamente essenciais para que haja o tratamento médico. A norma obriga que tanto as empresas quanto os estabelecimentos de saúde garantam o correto tratamento e a segurança das informações que tenham recebidos de seus clientes ou pacientes. Como forma de tutelar essa segurança, deve haver um correto mapeamento dos dados por todo o caminho do paciente, pois sem uma visão detalhada do ciclo de vida dos dados, é difícil implementar qualquer segurança. Assim, o mapeamento de dados se torna uma etapa essencial para alcançar a conformidade com o GDPR⁴¹ (BRASIL, 2018).

Cabe aqui explicar dois conceitos importantes implícitos nesse caso prático, quais sejam: o mapeamento dos dados e o caminho do paciente.

O caminho do paciente é a sequência de eventos iniciada pela identificação da necessidade de saúde, seguida pela busca e atendimento do paciente que apresente alguma enfermidade pelo sistema de saúde, seja por telemedicina, seja fisicamente na rede hospitalar, e todo o trâmite do mesmo até

40 Brasil. Presidência da República. Obra citada, Art. 3º.

41 Brasil. Presidência da República. Obra citada, Art. 46.

que receba alta, o que declare o óbito do mesmo (TREBBLE, 2010). Por sua vez, o mapeamento dos dados, ou *data mapping*, seria extraído justamente do caminho do paciente, proporcionando a edição de um fluxograma dos dados pessoais do paciente, entendendo todos os momentos do seu ciclo de vida, sendo este mapeamento essencial para a correta adequação de todas as atividades da empresa à Lei Geral de Proteção de Dados, *Lei n. 13.709, de 14 de agosto de 2018* ⁴² (BRASIL, 2018).

Como apresentado nos tópicos anteriores, o caminho do paciente tramará por diversas áreas do ambiente hospitalar, sendo importante verificar a necessidade de proteger os dados dos pacientes em diversos momentos, tais como: a necessária autorização expressa do paciente na coleta de seus dados; a possibilidade de o paciente alterar, corrigir ou excluir seus dados pessoais; as repercussões jurídicas pelo descumprimento da lei; entre outras situações.

Iniciando a jornada junto a determinado hospital, temos que o mesmo apresenta uma queixa de saúde, preenche uma ficha e assina um termo de consentimento, sendo este documento o ponto central de tudo que será abordado adiante.

A partir do exemplo apresentado, subdividiremos a nossa análise em três subtópicos: o termo de consentimento para tratamento de dados; o prontuário do paciente e a permissão para o acesso e correção do mesmo; as penalidades para o descumprimento da LGPD.

9. O TERMO DE CONSENTIMENTO PARA TRATAMENTO DE DADOS

No momento que o paciente ingressa no hospital recebe do setor de atendimento uma ficha, para que preencha todos os seus dados pessoais e relate os sintomas que tem apresentado. Neste mesmo documento, consta um longo texto, extremamente detalhado, conhecido como Termo de Consentimento Livre e Esclarecido (“TCLE”), traz diversos aspectos técnicos e jurídicos acerca da utilização e proteção dos dados do paciente, tendo por escopo o registro da manifestação da vontade livre, informada e

42 Brasil. Presidência da República. Obra citada, Art. 1º.

inequívoca⁴³, comprovando o consentimento explícito e específico do titular dos dados⁴⁴, sendo vedadas autorizações genéricas⁴⁵ (BRASIL, 2018).

Cabe aqui salientar que este documento será nulo caso as informações fornecidas ao paciente, por meio do termo de consentimento para tratamento de dados tenham conteúdo enganoso ou abusivo, ou não tenham sido apresentadas previamente com transparência, de forma clara e inequívoca.

O hospital somente poderá receber, manter, administrar e arquivar os dados pessoais de seus pacientes, se o mesmo der o expresso, claro e inequívoco consentimento, e desde que aquele dado seja realmente necessário aos fins visados pelo mesmo, é o chamado minimização dos dados (do inglês *data minimisation*) (ICO, 2020).

A minimização dos dados refere-se à limitação da atividade de tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos em relação às finalidades do tratamento de dados (EC, 2020).

Isto posto, o hospital precisa definir a finalidade da coleta de seus dados e a eventual possibilidade de compartilhamento com outras instituições⁴⁶. O paciente tem o direito de saber essa finalidade, podendo inclusive questionar a real necessidade do tratamento de determinados dados, por entendê-los não essenciais à atividade proposta (BRASIL, 2018).

Se os agentes de tratamento do hospital entenderem que haverá necessidade de utilização dos dados do paciente para fins não previstos no termo de consentimento para tratamento de dados, deverá a instituição coletar uma nova autorização do paciente, sob pena de utilização indevida dos dados do mesmo⁴⁷. Resta ainda esclarecer que o paciente tem o direito de, a qualquer momento, revogar o consentimento dado, podendo ainda solicitar o acesso, modificação, ou correção de qualquer dado incompleto, inexato ou desatualizado⁴⁸ (BRASIL, 2018).

43 Brasil. Presidência da República. Obra citada, Art. 5º, XII.

44 Brasil. Presidência da República. Obra citada, Art. 11, I.

45 Brasil. Presidência da República. Obra citada, Art. 8º, § 4º.

46 Brasil. Presidência da República. Obra citada, Art. 6º, I.

47 Brasil. Presidência da República. Obra citada, Art. 8º, § 6º.

48 Brasil. Presidência da República. Obra citada, Art. 9º, §2º.

10. O PRONTUÁRIO DO PACIENTE E A PERMISSÃO PARA O ACESSO E CORREÇÃO

Após a triagem inicial, o paciente será encaminhado a um médico, que realizará a anamnese, e iniciará a elaboração do prontuário do paciente. O prontuário é um dos principais documentos do caminho do paciente, pois o acompanha em quase todas as fases apresentadas., podendo ser definido como sendo “o documento único constituído de um conjunto de informações, sinais e imagens registradas, geradas a partir de fatos, acontecimentos e situações sobre a saúde do paciente e a assistência a ele prestada, de caráter legal, sigiloso e científico, que possibilita a comunicação entre membros da equipe multiprofissional e a continuidade da assistência prestada ao indivíduo” (CFM, 2002).

A LGPD determina que os hospitais e demais empresas deverão indicar uma pessoa que terá o dever de efetuar a proteção dos dados no sistema, este seria o *data protection officer*, também conhecido como encarregado⁴⁹ (BRASIL, 2018).

O *data protection officer* é a pessoa indicada pelo hospital e para atuar como canal de comunicação entre o estabelecimento de saúde, os pacientes - titulares dos dados - e a Autoridade Nacional de Proteção de Dados (POHLMANN, 2019).

Assim, resta claro que o sistema e o gerenciamento da proteção de dados do hospital devem ser seguros e precisam de protocolos de segurança eficientes.

11. AS PENALIDADES PARA O DESCUMPRIMENTO DA LGPD PELOS HOSPITAIS

O ordenamento jurídico brasileiro exige que os hospitais utilizem as mais modernas medidas técnicas e administrativas visando à maior proteção aos dados pessoais sensíveis de seus pacientes, evitando acessos não autorizados, incidentes de segurança, ilícitos ou acidentais, que gerem a perda, destruição ou divulgação indevida dos dados dos seus pacientes.

O elemento tecnológico não é o único a ser observado nessa árdua tarefa protetiva, devendo os próprios agentes de tratamentos de dados, pessoas

49 Brasil. Presidência da República. Obra citada, Art. 6º, III.

físicas, serem treinados e conscientização do valor e importância da proteção dos dados os pacientes, e o enorme potencial lesivo decorrente de um tratamento indevido destes. Para tanto, a Lei Geral de Proteção de Dados evidencia a relevância da adoção de “boas práticas e de governança que estabeleçam as condições de organização, o regime de funcionamento, os procedimentos, incluindo reclamações e petições de titulares, as normas de segurança, os padrões técnicos, as obrigações específicas para os diversos envolvidos no tratamento, as ações educativas, os mecanismos internos de supervisão e de mitigação de riscos e outros aspectos relacionados ao tratamento de dados pessoais”⁵⁰ (BRASIL, 2018).

Com efeito, a partir da integral entrada em vigor da *Lei Geral de Proteção de Dados*, em 03 maio de 2021⁵¹, todas as empresas – em especial: (a) clínicas, (b) consultórios e (c) hospitais deverão adequar-se aos seus termos (BRASIL, 2018).

A adequação às regras sobre tratamento de dados pessoais não se restringe aos dados constantes de banco de dados virtual, abrangendo dados colhidos manualmente, inclusive. Nesse contexto, o tratamento de dados – que se inicia na coleta e termina⁵² na eliminação – deve respeitar os ditames da legislação específica.

Portanto, todos os hospitais, públicos ou privados⁵³, deverão realizar relatório de impacto à proteção de dados pessoais e, necessariamente, nomearão encarregados para fiscalizar o adequado tratamento dos dados pessoais – inclusive aqueles definidos como sensíveis – dos seus pacientes (BRASIL, 2018).

O controlador deverá comunicar à autoridade nacional e ao titular a ocorrência de incidente de segurança que possa acarretar risco ou dano relevante aos titulares⁵⁴ (BRASIL, 2018).

É de se ressaltar que as multas previstas – em cenários de não conformidade aos termos da *Lei n° 13.709/2018* – podem alcançar 2% (dois por cento) do faturamento do hospital por cada ato violador, com o limite de 50 (cinquenta) milhões de reais, por cada infração⁵⁵ (BRASIL, 2018).

50 Brasil. Presidência da República. Obra citada, Art. 50.

51 Brasil. Presidência da República. Obra citada, Art. 65, II.

52 Brasil. Presidência da República. Obra citada, Arts. 15 e 16.

53 Brasil. Presidência da República. Obra citada, Art. 1º.

54 Brasil. Presidência da República. Obra citada, Art. 48.

55 Brasil. Presidência da República. Obra citada, Art. 52, II.

12. CONCLUSÃO

Independente do período histórico, cuidados em saúde é um campo essencialmente dirigido por dados e conhecimento. Registra-se os dados sobre os pacientes, tais como identificação, evolução clínica, resultados de exames etc. Comumente, estes dados têm por objetivo serem utilizados por profissionais como fonte de conhecimento investigação sobre a natureza e causas de doenças, diagnósticos, encaminhamentos, ou prescrição medicamentosa.

Atualmente, as tecnologias da informação e comunicação (TICs) têm proporcionado o armazenamento desses dados em uma magnitude nunca vista na história humana. Os dados armazenados em grande escala permitem a aplicação de técnicas e algoritmos especializados com vias a descrever acontecimentos passados, prever fenômenos futuros ou auxiliar na tomada de decisão. A aplicação destas técnicas em grandes volumes de dados digitais denomina-se Big Data, e tem proporcionado benefícios e indicado diversas oportunidades para pacientes, profissionais de saúde e poderes públicos.

No entanto, Big Data em saúde também evoca questões a serem tratadas no campo da Bioética, que tem como função primordial a análise interdisciplinar de toda e qualquer prática no universo da Biologia e da Medicina, atuando para explicar se estes atos têm ou não potencialidade lesiva a princípios e valores humanos fundamentais. No âmbito da tutela de dados da saúde relacionada à revolução informacional ocasionada pelo Big Data, tendo papel fundamental na reflexão acerca dos riscos e benefícios derivados de evolução na tecnociência médica.

Dessas questões a serem enfrentadas, no presente texto procuramos enfatizar a necessidade de atenção quanto à segurança e à garantia da privacidade dos dados individuais, mais especificamente os ligados à saúde. Assim, destacamos, por tratar-se de temática com grave risco social em seu uso indiscriminado, a opção feita pelo Estado por legislar sobre o tema, por meio de Lei Geral de Proteção de Dados, passando o Biodireito a participar da celeuma, pontuando e criando parâmetros normativos de equilíbrio para um seguro e ético progresso científico, mantendo de forma constante o respeito ao ser humano e seus valores, principalmente à dignidade da pessoa humana.

Indiscutivelmente, as tecnologias da informação e da comunicação cada vez mais ganham espaço e importância nas práticas relacionadas à saúde humana. Contudo, a incorporação de tais tecnologias devem ser orientadas pela ética, a fim de que não se colham resultados adversos aos benefícios que elas nos trazem. Acreditamos que os pontos apresentados e levantados neste capítulo, possam ser úteis para despertar o interesse pelos potenciais ganhos trazidos pelo avanço tecnológico, ao mesmo tempo que sensibiliza quanto à real necessidade de que dados e informações individuais sejam tratados e utilizados respeitando a dignidade pessoal.

13. REFERÊNCIAS

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). **Presidência da República**, Brasília, DF, 2018.

CARVALHO, A. O.; EDUARDO, M. B. P. Sistemas de informação para municípios. **São Paulo: Fundação Peirópolis Ltda**, 1998.

CFM. CONSELHO FEDERAL DE MEDICINA. Resolução CFM nº 1.638/2002. [portaria na internet]. Publicada no D.O.U. de 9 de agosto de 2002, Seção I, p. 184-5. Disponível em: <http://www.portalmedico.org.br/resolucoes/cfm/2002/1638_2002.htm>. Acesso em: 07 maio 2020.

CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. **Mobile networks and applications**, v. 19, n. 2, p. 171-209, 2014.

DELEN, D.; DEMIRKAN, H. Data, information and analytics as services. 2013.

EC. EUROPEAN COMMISSION. How much data can be collected?. Bélgica: Bruxelas. Disponível em: <https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr/how-much-data-can-be-collected_en>. Acesso em: 25 maio 2020.

GALETSI, P.; KATSALIAKI, K.; KUMAR, S. Values, challenges and future directions of big data analytics in healthcare: A systematic review. **Social science & medicine**, v. 241, p. 112533, 2019.

GALVÃO, A. B.; VALENTIM, R. A. M. Desafios para os Avanços da Análise de Big Data na Saúde. In: **Anais Estendidos do XIX Simpósio Brasileiro de Computação Aplicada à Saúde**. SBC, 2019. p. 155-160.

GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International journal of information management**, v. 35, n. 2, p. 137-144, 2015.

ICO. INFORMATION COMMISSIONER'S OFFICE. Principle (c): Data minimisation. Inglaterra: Wilmslow. Disponível em: <<https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/data-minimisation/>>. Acesso em 02 jun. 2020.

KOLB, J.; KOLB, J. **The Big Data Revolution: The World is Changing, are You Ready?**. Applied Data Labs, 2013.

POHLMANN, S. O DPO revisitado. Serpro. Notícias e Artigos. 2019. Disponível em: <<https://www.serpro.gov.br/lgpd/noticias/2019/encarregado-protecao-dados-dpo-data-protection-officer>>. Acesso em: 25 maio 2020.

TREBBLE, T. M. *et al.* Process mapping the patient journey: an introduction. **Bmj**, v. 341, 2010.

VILLAS-BÔAS, M. E. The right and duty of secrecy, as a patient protection. **Revista Bioética**, v. 23, p. 513-523, 2015.

XU, L. D.; DUAN, L. Big data for cyber physical systems in industry 4.0: a survey. **Enterprise Information Systems**, v. 13, n. 2, p. 148-169, 2019.

09.

AVANÇOS DA INTELIGÊNCIA ARTIFICIAL NA FERTILIZAÇÃO *IN VITRO*: BREVES REFLEXÕES

Oswaldo Jesus Rodrigues da Motta

Geisla Teles Vieira

Ronaldo Goldschmidt

Eugênio Silva

1. INTRODUÇÃO

A utilização de sistemas baseados em inteligência artificial (IA) para escolha de embriões viáveis à fecundação acarreta muitas discussões e reflexões. Por outro lado, e com igual destaque, tem-se avaliado que o apoio da IA pode corroborar para menor sofrimento do casal que deseja um filho, diminuição de custos do procedimento com maior acerto e verificação de possíveis enfermidades. A fertilização *in vitro* (FIV) – caracterizada como: “*extracting a woman’s eggs, fertilizing the eggs in the laboratory, and then transferring the resulting embryos into the uterus of the female patient or gestational carrier*” (CDC, 2017). – é um procedimento de valor elevado para cada ciclo de tentativa, com muitas pacientes que requerem vários ciclos para alcançar uma gestação com sucesso (BIRENBAUM-CARMELI, 2004; CDC, 2017; TONER, 2002).

Embora existam muitos fatores que influenciam, como idade materna, qualidade do embrião e receptividade do endométrio, o desafio da seleção não invasiva de maior qualidade de embrião permanece como uma questão importante a ser considerada (BORMANN *et al.*, 2020). Os métodos atuais para seleção de embriões requerem avaliação morfológica visual do embrião e são altamente dependentes de prática e fatores subjetivos (BAXTER-BENDUS *et al.*, 2006; PATERNOT *et al.*, 2009; STORR *et al.*, 2017). O uso de um sistema computacional para seleção de embriões é desafiador devido à complexidade das morfologias embrionárias e necessária

medição de parâmetros específicos definidos por especialistas como grau de simetria celular, número de blastômeros, fragmentação citoplasmática, entre outros (ROCHA *et al.*, 2017a, 2017b).

Na atualidade, apenas o estudo de células do embrião é capaz de identificar a normalidade de um embrião do ponto de vista cromossômico. No entanto, mesmo embriões cromossomicamente normais falham em cerca de 40% dos casos de implantação (BORMANN *et al.*, 2020). Encontrar outros parâmetros embrionários que sejam capazes de prever, entre os embriões, aqueles com maior potencial de nascimento é uma das possibilidades do uso de sistemas baseados em IA (BORMANN *et al.*, 2020). Os sistemas baseados em IA são frequentemente representados por métodos computacionais que simulam a inteligência humana ao identificar e aprender padrões ocultos em um determinado conjunto de dados para que, desta forma, possam apoiar a tomada de decisões complexas (FACELI *et al.*, 2011).

A ideia é identificar padrões nos embriões e a relação entre o parâmetro da qualidade do trofotoderma após blastulação bem como o tempo de formação da mórula com ligação à capacidade do embrião cromossomicamente normal de gerar um bebê saudável. Dessa forma, a incorporação da IA nos processos de reprodução assistida tem como objetivo reduzir a subjetividade das análises e processar dados que seriam impossíveis de serem processados manualmente. Dentro da clínica de reprodução humana, a IA também pode reduzir a subjetividade da classificação dos embriões e, junto com o estudo cromossômico, identificar o embrião com maior potencial de nascimento (BORMANN *et al.*, 2020).

Nesse sentido, técnicas de aprendizado de máquina (do inglês, *Machine Learning*, uma importante área de estudos da IA) para a análise morfológica do embrião têm sido utilizadas em alguns países a partir de bases de dados robustas, apresentando resultados positivos com um aumento de acurácia em comparação com a avaliação morfológica realizada pelo embriologista da forma tradicional. No entanto, questões bioéticas podem advir desses estudos, ao levantar questões como: (i) o melhoramento humano, (ii) o descarte de embriões, entre outros.

Com base neste breve recorte, serão trazidas reflexões sobre a utilização de IA na reprodução humana – mais especificamente na FIV – tendo em vista os desafios impostos pela prática seletiva embrionária.

2. DISCUSSÃO

Clinicamente, além das características morfológicas, várias outras métricas e parâmetros importantes são considerados pelos embriologistas no momento da tomada de decisão, como levar em consideração o *status* dos embriões transferíveis. O apoio computacional para a seleção de embriões no âmbito da FIV, utilizando técnicas de IA, tem sido investigado. Redes neurais profundas têm valor no auxílio à tomada de decisão clínica e têm recebido atenção significativa da comunidade de pesquisadores que trabalham na área de FIV. Bormann *et al.* (2020) argumentam que ferramentas assistivas e algoritmos de IA podem utilizar imagens para melhoria do procedimento de FIV. Foi avaliada a utilização de uma rede neural convolucional profunda, treinada usando imagens de embriões coletados em 113 horas pós-inseminação.

Uma grande preocupação, entretanto, é a perda de embriões viáveis devido a erros de sistema. Portanto, o algoritmo de seleção de embriões baseado em IA relatado não faz qualquer sugestão sobre o descarte de embriões. Todos os embriões avaliados no processo de seleção podem ser criopreservados. Desse modo, a abordagem não afetará negativamente a taxa de gravidez cumulativa, uma vez que embriões viáveis não serão perdidos. Além disso, é importante destacar que, em seu estágio atual, esse sistema pretende atuar apenas como uma ferramenta de auxílio ao embriologista. Os embriologistas podem incluir a previsão do sistema para fazer julgamentos melhores durante a seleção do embrião.

Chéles *et al.* (2020) compararam estudos que utilizaram diversas variáveis que impactam o sucesso da reprodução assistida, como morfologia do blastocisto e aspectos morfocinéticos do desenvolvimento embrionário, bem como características dos pacientes submetidos à reprodução assistida, a fim de prever a qualidade do embrião, implantação ou nascimento. Os autores concluem que, tendo em vista todas as variáveis apresentadas no estudo, há o potencial de prever nascidos vivos após utilização das técnicas de IA, uma vez que esta ferramenta permite a aplicação de diversas variáveis como entrada para o software e obter como saída a probabilidade de nascimento vivo. No entanto, o sistema deve abordar variáveis que influenciam o sucesso do nascimento a termo: parâmetros morfológicos e morfocinéticos do embrião, características clínicas do paciente e do parceiro. Dessa forma, surge a seguinte questão: como a IA poderia auxiliar na seleção de espermatozoides?

3. SELEÇÃO DE ESPERMATOZOIDES E INTELIGÊNCIA ARTIFICIAL

A seleção de espermatozoides através de técnicas de IA para FIV também ganha importante discussão através da pesquisa de Pedrosa *et al.* (2020). Os espermatozoides travam uma batalha para conquistar a fertilização, mas as características necessárias para ter sucesso permanecem indefinidas. As células espermáticas são bastante complexas e sofisticadas.

As qualidades naturais vantajosas que permitem que apenas alguns espermatozoides selecionados cheguem ao local da fertilização permanecem desconhecidas. Com efeito, nenhum indicador de qualidade espermática objetivo ou confiável foi estabelecido e a seleção espermática é, em grande medida, baseada em avaliação qualitativa subjetiva.

O melhor método para seleção de espermatozoides na FIV apresenta vários desafios: as qualidades intrínsecas do esperma não podem ser avaliadas e o desfecho ideal para esses estudos é discutível. Nesse sentido, um método ideal para a seleção de espermatozoides na FIV deve ser não invasivo e de baixo custo, e permitir a identificação de espermatozoides de alta qualidade e produzir melhores resultados em termos de taxas de gravidez e nascidos vivos.

3.1 (Bio)ética e fertilização *in vitro*

Quando se trata do tópico mais controverso, edição do genoma da linha germinativa (GGE), não há consenso no campo da (bio)ética sobre se esta tecnologia tem vantagens substanciais sobre os procedimentos existentes, como seleção de embriões após FIV e diagnóstico genético pré-implantação (PGD).

Ranisch (2020) confirma que existem alguns cenários em que apenas a intervenção na linha germinativa permite que casais tenham filhos saudáveis (biologicamente relacionados), porque a seleção não evitará a doença. Em uma segunda etapa, examina os possíveis argumentos morais a favor da modificação genética, a saber, que poderia salvar alguns embriões e que proporcionaria certos benefícios para uma pessoa futura. Como conclusão, aponta que novas ferramentas para edição de genoma são amplamente consideradas tecnologias disruptivas.

Wallis (2020) argumenta que à medida que as tecnologias de genética reprodutiva avançam, as famílias têm mais opções para escolher que tipo de filho desejam ter, o que será evidenciado pelos avanços da IA. O uso do diagnóstico genético pré-implantação (PGD), por exemplo, permite que os pais avaliem vários embriões existentes antes de selecionar qual implantar via FIV. Um dos traços que o PGD pode identificar é a surdez genética, e agora embriões com audição são selecionados preferencialmente em todo o mundo usando esse método. É importante ressaltar que algumas famílias surdas desejam uma criança surda, e PGD-FIV também é uma opção para elas.

A seleção para surdez, com a surdez entendida como mera diferença e valorizada por sua identidade cultural, não precisa necessariamente de danos morais inadmissíveis. Portanto, às vezes é moralmente permissível selecionar a surdez em um filho. No caso da surdez genética, pode ser moralmente permissível que os pais selecionem a surdez para seu futuro filho. Isso coloca uma responsabilidade significativa sobre os pais de considerarem apropriadamente o tipo de vida que seu futuro filho terá e que os pais surdos são devidamente capazes, cabendo reflexões (bio)éticas instigantes. (WALLIS, 2020)

Blank *et al.* (2020) desenvolveram um estudo utilizando um modelo baseado em Floresta Aleatória (RFM) para prever o potencial de sucesso na implantação de um embrião transferido e compará-lo com um modelo de regressão logística multivariada (MvLRM). A base de dados utilizada foi proveniente de um estudo retrospectivo de uma coorte de 2 anos, em um único centro de mulheres submetidas a FIV ou injeção intracitoplasmática de espermatozoides (ICSI). Os desempenhos do RFM e do MvLRM para prever a gravidez foram quantificados em termos da área sob a curva ROC (AUC), precisão, especificidade e sensibilidade (BLANK *et al.*, 2020).

A análise da curva ROC resultou em uma AUC de $0,74 \pm 0,03$ para o RFM proposto e $0,66 \pm 0,05$ para a previsão de gestações em curso com 11 semanas ou mais. Os modelos RFM e MvLRM produziram, respectivamente, sensibilidades de $0,84 \pm 0,07$ e $0,66 \pm 0,08$ e especificidades de $0,48 \pm 0,07$ e $0,58 \pm 0,08$. O referido artigo, por meio da validação cruzada, conclui que o desempenho para prever a implantação contínua melhorará significativamente com o uso de uma abordagem baseada no RFM em comparação com o MvLRM.

Guntram L e Zeiler K (2019) reforçam a discussão bio(ética) da FIV na literatura, apontando análises de risco-benefício. Enfatizam que os riscos são baixos e que a fertilização *in vitro* se tornou, afinal, uma prática estabelecida em muitos países ocidentais. Além disso, é comumente afirmado que a FIV só deve ocorrer quando os pacientes envolvidos foram devidamente informados sobre os riscos, tiveram o tempo e a oportunidade necessários para considerar alternativas e tomaram uma decisão informada.

Todavia os autores argumentam que, ainda que a FIV tenha sido descrita como geralmente aceitável para os pacientes, alguns riscos estão envolvidos, como o risco de síndrome de hiperestimulação ovariana, o que pode ser fatal. A FIV também pode evocar ansiedade, por causa, por exemplo, de seu resultado incerto além de questionamentos sobre valores e normas sobre a maternidade biológica, uma vez que possibilita novas possibilidades de parentesco.

Outra importante reflexão é destacada por Kemper *et al.* (2019), quando aponta que, na seleção de embriões considerados “sem deficiência”, se rotularia a vida de todos os indivíduos deficientes como sem valor, aumentando a discriminação contra pessoas com alguma deficiência. Ou seja, poderia acarretar o aumento de preconceito em relação aos deficientes físicos.

4. (IN)CONCLUSÕES

Os avanços nas técnicas de IA promoveram inúmeras aplicações que têm o potencial de melhorar o padrão de atendimento em diferentes campos da medicina. A abordagem mostrou o potencial das redes neurais convolucionais para serem usadas para auxiliar os embriologistas a selecionar o embrião com o maior potencial de implantação, especialmente entre embriões de alta qualidade. Embora o estudo retrospectivo atual mostre que esses sistemas podem ter um desempenho melhor do que embriologistas altamente treinados, ensaios de controle randomizados são necessários antes que o uso rotineiro na prática clínica seja adotado.

Há incipiência de estudos quando se analisa as categorias que emergiram da revisão, uma vez que o uso de técnicas de IA na fertilização *in vitro* é extremamente recente. Portanto, há necessidade de que sejam empreendidos estudos para o desenvolvimento de novas medidas para melhoria da implementação da IA na FIV. Os esforços interdisciplinares envolvendo

profissionais das mais diversas formações poderão contribuir para intensificar a conscientização e abordar os problemas (bio)éticos, além de propor estratégias de diminuição dos conflitos.

5. AGRADECIMENTO

Ao Prof. Dr. Rodrigo Siqueira-Batista (Universidade Federal de Viçosa, Departamento de Medicina e Enfermagem e Faculdade Dinâmica do Vale do Piranga, Escola de Medicina) pelo auxílio na revisão do texto.

6. CONFLITO DE INTERESSES

Os autores declaram não haver conflito de interesses.

7. REFERÊNCIAS

BAXTER-BENDUS, A. E.; MAYER, J. F.; SHIPLEY, S. K.; CATHERINO, W. H. Interobserver and intraobserver variation in day 3 embryo grading. **Fertility and Sterility**, v. 86, p. 1608–1615, 2006.

BIRENBAUM-CARMELI, D. ‘Cheaper than a newcomer’: on the social production of IVF policy in Israel. **Sociology of Health and Illness**, v. 26, p. 897–924, 2004. doi: 10.1111/j.0141-9889.2004.00422.x.

BLANK, C.; WILDEBOER, R. R.; DECROO, I.; TILLEMANN, K.; WEYERS, B.; DE SUTTER, P.; MISCHI, M.; SCHOOT, B. C. Prediction of implantation after blastocyst transfer in *in vitro* fertilization: a machine-learning perspective. **Fertil Steril**, v. 111, n. 2, p. 318–326, 2009. doi: 10.1016/j.fertnstert.2018.10.030. Epub 2019 Jan 2. PMID: 30611557.

BORMANN, C. L.; KANAKASABAPATHY, M. K.; THIRUMALARAJU, P.; GUPTA, R.; POONIWALA, R.; KANDULA, H.; HARITON, E.; SOUTER, I.; DIMITRIADIS, I.; RAMIREZ, L. B.; CURCHOE, C. L.; SWAIN, J.; BOEHNLEIN, L. M.; SHAFIEE, H. Performance of a deep learning based neural network in the selection of human blastocysts for implantation. **Elife**, v. 15, n.9:e55301, 2020.

CDC. 2017 Assisted Reproductive Technology Fertility Clinic Success Rates Report. **Centers for Disease Control and Prevention**; 2015. <https://www.cdc.gov/art/reports/2017/fertility-clinic.html>

CHÉLES, D. S.; MOLIN, E. A. D.; ROCHA, J. C.; NOGUEIRA, M. F. G. Mining of variables from embryo morphokinetics, blastocyst's morphology and patient parameters: an approach to predict the live birth in the assisted reproduction service. **JBRA Assist Reprod**, 2020 Apr 15. doi: 10.5935/1518-0557.20200014. Epub ahead of print. PMID: 32293823.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. Inteligência artificial: uma abordagem de aprendizado de máquina. Rio de Janeiro: **LTC**, v. 2, p. 192, 2011.

GUNTRAM, L.; ZEILER, K. The Ethics of the Societal Entrenchment-approach and the case of live uterus transplantation-IVF. **Med Health Care Philos**, v. 22, n. 4, p. 557-571, 2019. doi: 10.1007/s11019-019-09891-w. PMID: 31102175; PMCID: PMC6842335.

KEMPER, J. M.; GYNGELL, C.; SAVULESCU, J. SUBSIDIZING PGD: The Moral Case for Funding Genetic Selection. **J Bioeth Inq**, v. 16, n. 3, p. 405-414, 2019. doi: 10.1007/s11673-019-09932-2. Epub 2019 Aug 15. PMID: 31418161; PMCID: PMC6831526.

PATERNOT, G.; DEVROE, J.; DEBROCK, S.; D'HOOGHE, T. M.; SPIESSENS, C. Intra- and inter-observer analysis in the morphological assessment of early-stage embryos. **Reproductive Biology and Endocrinology**, v. 7, p.105, 2009.

PEDROSA, M. L.; FURTADO, M. H.; FERREIRA, M. C. F.; CARNEIRO, M. M. Sperm selection in IVF: the long and winding road from bench to bedside. **JBRA Assist Reprod**, v. 24, n. 3, p. 332-339, 2020. doi: 10.5935/1518-0557.20190081. PMID: 32155013; PMCID: PMC7365522.

RANISCH, R. Germline genome editing versus preimplantation genetic diagnosis: Is there a case in favour of germline interventions? **Bioethics**. v. 34, n. 1, p. 60-69, 2020. doi: 10.1111/bioe.12635. Epub 2019 Aug 25. PMID: 31448423; PMCID: PMC6973094.

ROCHA, J. C.; PASSALIA, F. J.; MATOS, F. D.; TAKAHASHI, M. B.; MASERATI, M. P.; ALVES, M. F.; DE ALMEIDA, T. G.; CARDOSO, B. L.; BASSO, A. C.; NOGUEIRA, M. F. G. Automatized image processing of bovine blastocysts produced *in vitro* for quantitative variable determination. **Scientific Data**, v.4, p. 170192, 2017a.

ROCHA, J. C.; PASSALIA, F. J.; MATOS, F. D.; TAKAHASHI, M. B.; CINICIATO, D. S.; MASERATI, M. P.; ALVES, M. F.; DE ALMEIDA, T. G.; CARDOSO, B. L.; BASSO, A. C.; NOGUEIRA, M. F. G. A method based on artificial intelligence to fully automatize the evaluation of bovine blastocyst images. **Scientific Reports**, v. 7, p. 7659, 2017b.

STORR, A.; VENETIS, C. A.; COOKE, S.; KILANI, S.; LEDGER, W. Inter-observer and intra-observer agreement between embryologists during selection of a single day 5 embryo for transfer: a multicenter study. **Human Reproduction**, v. 32, p. 307–314, 2017.

TONER, J. P. Progress we can be proud of: U.S. trends in assisted reproduction over the first 20 years. **Fertility and Sterility**, v. 78, p. 943–950, 2002.

WALLIS, J. M. Is it ever morally permissible to select for deafness in one's child? *Med Health Care Philos*, v. 23, n. 1, p. 3-15, 2020. doi: 10.1007/s11019-019-09922-6. PMID: 31542873; PMCID: PMC7040060.

10.

APOIO COMPUTACIONAL À TOMADA DE DECISÃO EM TERAPIA INTENSIVA: QUESTÕES BIOÉTICAS

Oswaldo Jesus Rodrigues da Motta

Gabriel Resende Machado

Ronaldo Ribeiro Goldschmidt

Eugênio Silva

1. PREÂMBULO

Os avanços no campo da tecnologia têm promovido a manutenção e a extensão da vida, bem como proporcionado debates sobre a complexidade e a diversidade moral da tomada de decisão clínica na sociedade contemporânea, que requer com constância, deliberações sobre como ocorrem as decisões de fim de vida. Hoje, uma característica marcante da terminalidade é que ela ocorre no ambiente hospitalar, especialmente na Unidade de Terapia Intensiva (UTI). Trata-se de um espaço altamente integrado tecnologicamente, que fornece suporte avançado para manter e prolongar a vida de pacientes gravemente frágeis ou clinicamente instáveis (BRASIL, 1998).

Essa complexa estrutura é operada por equipes multidisciplinares, altamente qualificadas e que trabalham vinte e quatro horas por dia em favor da vigilância completa em cuidados intensivos. O avanço da biotecnologia, aliado a todo o aparato tecnológico disponível, tem desempenhado um papel de extrema relevância no ambiente de terapia intensiva, mas que também suscita importantes reflexões. Se, por um lado, a UTI é solidariamente responsável pela promoção da saúde, por outro, também pode prolongar a vida humana a um ponto em que as discussões sobre o seu fim se tornam inevitáveis (BRASIL, 1998).

Afinal, mesmo em casos em que a possibilidade terapêutica de cura inexistente, ainda assim os recursos disponíveis em uma UTI podem,

artificialmente, estender a vida por longos períodos. Nessas ocasiões, a equipe responsável pelo cuidado se depara com a necessidade de tomar decisões extremamente sensíveis que envolvem a continuidade ou a interrupção do suporte à vida. Trata-se de uma decisão impregnada de dilemas morais e éticos que precisam ser amplamente discutidos entre todos os atores dedicados à atenção ao paciente, o que pode envolver não apenas os profissionais de saúde, mas também a família.

Em geral, a formação dos profissionais que integram a equipe de cuidados os prepara para que se empenhem sempre em benefício da vida, mas pode não os preparar para as decisões relacionadas ao fim da vida do paciente. Assim, não raro há divergências entre os profissionais envolvidos sobre a melhor decisão (ou menos pior) a ser tomada em situações extremas. Com o envolvimento da família, o processo decisório ganha contornos de maior complexidade ainda, uma vez que desejos, emoções, princípios e crenças passam a exercer forte influência no desfecho.

Pelas razões apresentadas, o ambiente de UTI se tornou um terreno fértil para o desenvolvimento de pesquisas que envolvem os estágios finais da existência humana. Destacam-se aí as pesquisas dedicadas à discussão das questões (bio)éticas⁵⁶ que permeiam as decisões de fim de vida de pacientes que estejam fora da possibilidade terapêutica de cura. Desse contexto, considera-se que uma ferramenta computacional devidamente elaborada pode ser um instrumento de grande utilidade para auxiliar e amparar a equipe de saúde em suas decisões em casos extremos. Para a construção da ferramenta cogitada, as técnicas que se mostram mais adequadas ao propósito parecem ser aquelas diretamente relacionadas à área de Inteligência Artificial (IA), em especial à Aprendizagem de Máquina (AM). Aqui, cabe destacar que o que se discute é um recurso de apoio à tomada de decisão e não uma forma de transferir a responsabilidade pela decisão da equipe envolvida para uma ferramenta computacional.

As próximas seções do texto trazem subsídios importantes para avaliar a viabilidade de se elaborar um sistema baseado em IA para o propósito desejado. Inicialmente apresenta-se uma discussão sobre a bioética clínica,

56 Ao utilizar o termo “(bio)ética”, destaca-se a interseção entre ética geral e as questões específicas que surgem no campo das ciências da vida e da saúde. Forma de reconhecer que a ética aplicada a esses campos pode ser considerada dentro do contexto mais amplo da ética como um todo. O uso do prefixo “bio” indica que o foco está nas questões éticas que surgem dentro desses campos.

as complexidades por trás das decisões a serem tomadas nesse contexto e os métodos de tomada de decisão disponíveis. Em seguida, são apresentados alguns conceitos básicos sobre de IA e AM. Posteriormente, são apresentadas algumas iniciativas a serem empreendidas para a construção de modelos computacionais capazes de apoiar a tomada de decisão em situações que envolvem conflitos morais e éticos. Por fim, são apresentadas as considerações sobre a possibilidade de materialização da ideia enunciada.

2. BIOÉTICA CLÍNICA E O PERCURSO DECISÓRIO

A bioética, como é conhecida hoje, começou a se formar no contexto das controvérsias públicas que surgiram nas décadas de 1950 e 1960. Essas controvérsias giravam em torno de questões emergentes na medicina e na ciência, que desafiavam as normas éticas existentes e exigiam novas formas de pensar sobre a moralidade. Na década de 1950, a polêmica em torno da vacinação obrigatória contra a poliomielite levantou questões sobre o direito individual à autonomia versus o bem-estar coletivo. A vacina, desenvolvida por Jonas Salk, foi vista como um marco na medicina, mas a obrigatoriedade de sua aplicação foi questionada por alguns setores da sociedade.

Na década de 1960, a introdução da pílula anticoncepcional provocou um debate público acalorado sobre a liberdade sexual, o controle da natalidade e a autonomia das mulheres. A pílula, pela primeira vez, ofereceu às mulheres um meio eficaz de controlar a própria fertilidade, desafiando as normas sociais e religiosas tradicionais. Outra controvérsia significativa dessa época foi a experimentação em seres humanos. O infame estudo da sífilis em Tuskegee, conduzido pelo Serviço de Saúde Pública dos Estados Unidos, veio à luz na década de 1970, mas começou na década de 1930 e continuou até a década de 1960.

Este estudo, que envolveu a negação do tratamento a homens negros com sífilis, levantou questões profundas sobre o consentimento informado, a justiça na pesquisa médica e a exploração de populações vulneráveis. Essas controvérsias ajudaram a moldar a bioética como um campo de estudo, levando à criação de princípios fundamentais, como autonomia, beneficência, não maleficência e justiça. A bioética continua a evoluir hoje, à medida que novas tecnologias e desafios emergem na medicina e na ciência. A UTI é um ambiente onde ocorrem muitos conflitos bioéticos. Tais

conflitos surgem devido à natureza complexa e crítica dos cuidados prestados, que frequentemente envolvem decisões de vida ou morte. Aqui estão alguns exemplos:

Autonomia do paciente versus beneficência médica: a autonomia do paciente, o direito de tomar decisões informadas sobre o próprio cuidado, é um princípio fundamental da bioética. No entanto, na UTI, os pacientes frequentemente estão incapacitados e não conseguem expressar suas vontades. Isso pode levar a conflitos entre a autonomia do paciente e a beneficência médica, o princípio de agir no melhor interesse do paciente.

Limitação de tratamento: decidir quando limitar ou retirar o tratamento é uma questão delicada na UTI. Embora prolongar a vida seja uma meta importante, a qualidade de vida e o sofrimento do paciente também devem ser considerados. Essas decisões podem ser fontes de conflito entre a equipe médica, o paciente e a família.

Alocação de recursos: em situações de escassez de recursos, como durante uma pandemia, as decisões sobre quem recebe tratamento na UTI podem ser extremamente difíceis. A necessidade de equilibrar a justiça, o princípio de tratar as pessoas com equidade, com a eficiência, pode levar a conflitos bioéticos.

Privacidade e confidencialidade: a UTI é um ambiente aberto, o que pode dificultar a manutenção da privacidade e confidencialidade do paciente. Isso pode levar a conflitos entre a necessidade de fornecer cuidados eficazes e o direito do paciente à privacidade.

Comunicação e consentimento informado: a comunicação eficaz é essencial na UTI, especialmente quando se trata de obter consentimento informado para procedimentos. A complexidade das informações e a condição do paciente podem dificultar essa comunicação, levando a possíveis conflitos bioéticos.

Esses conflitos bioéticos exigem uma reflexão cuidadosa, uma comunicação clara e, muitas vezes, a contribuição de um comitê de ética hospitalar. A bioética fornece um quadro para navegar por esses desafios difíceis, mas importantes, na UTI. Muitos óbitos em UTI acontecem por interrupção das medidas de suporte à vida. Muitas vezes é negado ao paciente o direito de saber sobre sua condição e não há diálogo entre o paciente e sua família. De acordo com Félix *et al.* (2014): “*O cuidado de fim de vida na unidade de terapia intensiva é uma situação muito complexa*”. Estudos bioéticos

ganharam um novo patamar com a publicação do Relatório Belmont, em 1979. Este relatório, elaborado pela Comissão Nacional para a Proteção de Seres Humanos Sujeitos à Pesquisa Biomédica e Comportamental dos Estados Unidos, estabeleceu três princípios fundamentais para a condução de pesquisas envolvendo seres humanos: *respeito às pessoas*, *beneficência* e *justiça*.

O respeito às pessoas envolve o reconhecimento da autonomia do indivíduo e a proteção daqueles com autonomia diminuída. Isso implica no consentimento informado, onde os participantes devem ser informados sobre o estudo e concordar voluntariamente em participar. A beneficência refere-se à obrigação de fazer o bem e evitar o mal. Os pesquisadores devem se esforçar para maximizar os benefícios e minimizar os danos potenciais de suas pesquisas. A justiça exige que os benefícios e ônus da pesquisa sejam distribuídos de maneira justa. Isso significa que certos grupos não devem ser selecionados para participar da pesquisa simplesmente por serem facilmente disponíveis ou manipuláveis.

O Relatório Belmont foi um marco na história da bioética, estabelecendo um padrão ético para a pesquisa em seres humanos. Ele influenciou a criação de comitês de ética em pesquisa e a elaboração de diretrizes éticas em muitos países. Ainda hoje, seus princípios são uma referência fundamental nos estudos bioéticos. O *princípioalismo*, uma abordagem desenvolvida por Tom Beauchamp e James Childress em seu livro “Princípios de Ética Biomédica” (BEAUCHAMP & CHILDRESS, 2002), é uma evolução das ideias iniciais de bioética. Esta abordagem se baseia em quatro princípios fundamentais para lidar com questões de natureza ética: *não maleficência*, *justiça*, *beneficência* e *autonomia*.

Os princípios de não maleficência e justiça têm um caráter deontológico, ou seja, são baseados no dever e na obrigação moral. A não maleficência refere-se à obrigação de não causar danos aos outros, enquanto a justiça envolve a distribuição equitativa de benefícios, riscos e custos na sociedade. Por outro lado, os princípios de beneficência e autonomia têm um caráter teleológico, ou seja, são orientados para um fim ou objetivo. A beneficência refere-se à promoção do bem-estar dos outros, enquanto a autonomia respeita a liberdade individual para tomar decisões informadas.

A evolução para o princípioalismo representou um avanço significativo na bioética, pois forneceu um quadro estruturado para abordar questões

éticas complexas. É importante notar que esses princípios não são absolutos e podem entrar em conflito em determinadas situações. A aplicação do princípalismo requer um equilíbrio cuidadoso e consideração das circunstâncias específicas de cada caso, como realizado no âmbito da Bioética Clínica, que se caracteriza pela sua interdisciplinaridade, abordando questões éticas na prática clínica e na pesquisa médica (PUCCINI *et al.*, 2008).

O percurso decisório na bioética clínica envolve a consideração cuidadosa de uma série de fatores, incluindo os princípios éticos fundamentais, as necessidades e desejos do paciente, as obrigações legais e profissionais do médico e as implicações sociais e culturais mais amplas (PESSINI, 2004). Assim, os quatro princípios éticos fundamentais que são comumente usados como um quadro para a tomada de decisões em bioética clínica, são enunciados da seguinte maneira:

Autonomia: respeito pela autonomia do paciente e seu direito de tomar decisões informadas sobre seu próprio cuidado.

Beneficência: obrigação de agir no melhor interesse do paciente.

Não-maleficência: obrigação de não causar danos ao paciente.

Justiça: obrigação de tratar todos os pacientes de maneira justa e equitativa.

O percurso decisório em bioética clínica envolve a aplicação desses princípios a situações clínicas específicas. Isso pode envolver a ponderação de diferentes princípios uns contra os outros, a consideração das preferências e valores do paciente e a consulta a colegas, comitês de ética ou outros recursos conforme necessários. Os desafios desse percurso incluem ainda a necessidade de tomar decisões sob condições de incerteza, a necessidade de equilibrar os interesses do paciente com os do médico ou da sociedade, e a necessidade de navegar em questões legais e regulatórias complexas.

A bioética clínica deve sempre ser sensível ao contexto cultural e social mais amplo. As normas culturais e sociais podem influenciar as percepções dos pacientes sobre o que é ético, e os médicos devem estar cientes e respeitar essas diferenças. A análise requer consideração cuidadosa dos princípios éticos, das necessidades e desejos do paciente, e das implicações sociais e culturais mais amplas. Por meio de uma abordagem reflexiva e informada, os envolvidos podem navegar nessas questões complexas e fornecer cuidados de alta qualidade que respeitam a dignidade e os direitos dos pacientes.

Ainda assim, o principialismo, apesar de ser uma abordagem amplamente utilizada na bioética, não está isento de críticas. Uma das principais críticas é que pode ser demasiado abstrato e descontextualizado, dificultando a aplicação prática dos princípios em situações reais. Ademais, o principialismo pode ser visto como uma abordagem excessivamente simplista que não leva em conta a complexidade e a diversidade das questões éticas na prática médica e na pesquisa.

Outra crítica comum é que o principialismo, originário do contexto cultural e filosófico ocidental, pode não ser aplicável ou adequado a outras culturas ou contextos. Isso levou ao surgimento de abordagens alternativas que enfatizam a importância do contexto cultural e social na tomada de decisões éticas. Por exemplo, a *ética da virtude*, que se concentra no caráter e nas virtudes do agente moral, tem sido proposta como uma alternativa ou complemento ao principialismo. Da mesma forma, a *ética do cuidado*, que enfatiza as relações interpessoais e a responsabilidade moral para com os outros, também ganhou destaque.

Surgiram abordagens que enfatizam a deliberação e o diálogo na tomada de decisões éticas, como a *ética deliberativa* e a *ética dialógica*. Essas abordagens reconhecem que a tomada de decisões éticas é muitas vezes um processo complexo e dinâmico que envolve a negociação de diferentes perspectivas e valores. Enquanto o principialismo continua a desempenhar um papel importante na bioética, as críticas e os desafios que enfrenta levaram ao desenvolvimento de uma variedade de outras abordagens para a tomada de decisões éticas. Essa diversidade de abordagens reflete a complexidade e a pluralidade da bioética como campo de estudo e prática.

Em discussão mais aprofundada sobre isso, é possível identificar diferentes métodos de tomada de decisão para problemas relacionados, como utilização de protocolos institucionais que muitas das vezes não contemplam diversos cenários e personagens. Independentemente do método empregado na tomada de decisão diante de conflitos bioéticos complexos na prática clínica, há situações em que, dada a urgência, essa decisão não pode ser compartilhada com outros profissionais da equipe de cuidado. Mesmo em casos em que haja a oportunidade de compartilhamento da decisão com o restante da equipe, ainda assim pode-se desejar ter o envolvimento de outros profissionais, especialmente especialistas em bioética, que podem não estar acessíveis no momento da tomada de decisão. Em situações como

essas, uma ferramenta computacional que, de alguma forma, possa reunir e disponibilizar o conhecimento de diversos especialistas, poderia ser um auxílio importante, sobretudo em momentos em que a decisão a ser tomada não é tão clara. Destaca-se que o intuito de um recurso como esse é sempre de apoiar a tomada de decisão. Nessa proposta, está descartada a possibilidade de o sistema substituir o profissional humano no processo de tomada de decisão.

3. INTELIGÊNCIA ARTIFICIAL E APRENDIZAGEM DE MÁQUINA

Nos últimos anos o termo Inteligência Artificial vem ganhando cada vez mais espaço na mídia, notadamente pelas diversas aplicabilidades de suas técnicas e pelos resultados surpreendentes que podem proporcionar. Hoje, há diversos ramos de negócio se valendo da IA em busca de aumento de eficiência e de vantagem competitiva. Apesar da popularidade do termo, a definição de IA não é algo trivial. A dificuldade começa no entendimento do significado da palavra “inteligência”, uma vez que, apesar dos enormes avanços obtidos pela ciência no estudo dos mistérios do cérebro humano, ainda não se tem hoje uma compreensão plena da operação dos mecanismos cerebrais.

Não se sabe ainda o que realmente é a “inteligência”. Assim, há diferentes definições para IA, sob diferentes perspectivas. Perspectivas estas que levam em consideração os objetivos da suposta inteligência por trás de um sistema computacional. Por um lado, o interesse pode estar em reproduzir a maneira como um ser humano pensa ou age, considerando que as ações e o pensamento humanos nem sempre são guiados exclusivamente pela razão. Por outro, a inteligência de um sistema pode estar voltada para a construção de processos de pensamento ou de planejamento de ações de forma estritamente racional (RUSSEL; NORVIG, 2013).

No universo de definições propostas por diferentes autores, talvez pela sua simplicidade, destaca-se aquela apresentada por Rich e Knight (1991): *“Inteligência Artificial é o estudo de como fazer os computadores realizarem tarefas em que, no momento, as pessoas são melhores.”* Para que a definição se torne mais clara, é importante entender que tarefas são essas. Há uma grande quantidade delas em que os computadores são muito mais

eficientes e confiáveis que os seres humanos. O que essas tarefas têm em comum é o fato de serem mecanizáveis, ou seja, podem ser descritas na forma de um algoritmo. Tarefas desse tipo envolvem, por exemplo, o cálculo da folha de pagamento de uma empresa, uma transação bancária ou ainda o cálculo do consumo de combustível de um veículo em tempo real. Para essas tarefas há regras muito bem definidas que podem ser expressas por meio de um procedimento efetivo e, com isso, podem ser automatizadas com razoável facilidade.

Em contrapartida, há um conjunto considerável de outras tarefas, mais diretamente ligadas ao intelecto humano, que supostamente não são passíveis de mecanização, ou seja, não podem ser descritas na forma de algoritmos. Alguns exemplos de tarefas desse tipo são: raciocínio, inferência, planejamento, percepção, aprendizagem e comunicação. Todas essas tarefas são desempenhadas por humanos o tempo todo, e com razoável facilidade, mas se mostram bastante complexas quando se tenta reproduzi-las em um computador. É exatamente para tarefas que envolvem essas habilidades humanas que os esforços da IA estão voltados.

A IA é uma área de estudo relativamente recente, que teve seu nascimento oficial em uma conferência realizada no Dartmouth College em 1956, organizada por expoentes da área como John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon. A propósito, foi a partir desse evento que a denominação “Inteligência Artificial” passou a ser oficialmente adotada (RUSSEL; NORVIG, 2013). De lá para cá, a área foi rapidamente ampliada, seja pela concepção de novas técnicas, seja com a identificação de potenciais aplicações.

Com isso, tem-se hoje um arcabouço de técnicas que, de forma simplificada, podem ser organizadas em dois grupos distintos, conforme a estratégia utilizada para reproduzir um comportamento que possa ser considerado inteligente. De um lado estão aquelas em se emprega o *raciocínio dedutivo* que, segundo Gil (1999, p. 27), “*de acordo com a acepção clássica, é o método que parte do geral e, a seguir, desce ao particular*”. Ou seja: princípios sabidamente verdadeiros e indiscutíveis possibilitando chegar a conclusões de modo formal. Do outro estão aquelas em que se emprega o *raciocínio indutivo*, que é a base da AM. Segundo Lakatos (1991, p. 47), “*a indução é um processo mental que permite, partindo de dados particulares, suficientemente constatados, inferir-se uma verdade geral ou universal*”

(que não estão contidas nos trechos examinados). Assim, o objetivo de um argumento é encaminhar a conclusões cujo conteúdo seja mais amplo do que o das premissas que o alicerçaram. O método indutivo, desta feita, consiste em observar casos particulares e estabelecer hipóteses de caráter geral.

Aprendizagem de Máquina pode ser entendida como forma de fazer com que um sistema computacional desempenhe uma tarefa sem que seja explicitamente programado para isso. Assim, em vez de elaborar uma sequência de instruções que descreva detalhadamente as ações para a execução daquela tarefa, apresenta-se ao sistema um conjunto de exemplos relacionados a ela para que, de forma automática, o sistema ajuste internamente um conjunto de parâmetros de um modelo de forma a “aprender” a executar tal tarefa.

O grande objetivo do processo de aprendizagem consiste em alcançar a *generalização*, que permite que o sistema seja capaz de reproduzir a habilidade aprendida em situações inéditas relacionadas à tarefa, mas que guardam alguma semelhança com aquelas apresentadas durante a fase de aprendizagem. A AM encontra analogia na aprendizagem humana onde, para o entendimento de como executar uma determinada tarefa, é necessária a observação de um conjunto de exemplos para que a aprendizagem aconteça. A partir disso, generaliza-se o conhecimento adquirido para outras situações que não estiveram entre os exemplos usados como subsídio para a aprendizagem. Isso acontece, por exemplo, quando se observa um conjunto de exemplos relacionados a um determinado assunto e posteriormente aplica-se o que foi aprendido a novas situações que são apresentadas em uma avaliação (prova). Algoritmos de AM podem ser utilizados em uma ampla variedade de tarefas que se dividem em *preditivas* e *descritivas* (FACELI *et al.*, 2011).

No caso de tarefas preditivas, o processo de aprendizagem exige que os exemplos utilizados sejam rotulados. Isso significa que, para cada estímulo recebido pelo sistema, deve-se conhecer a resposta correta para aquele estímulo. Durante a aprendizagem, os parâmetros internos do sistema de aprendizagem são ajustados para que o sistema responda conforme o desejado para cada estímulo recebido. Essa estratégia de ajuste está associada a um paradigma de aprendizagem denominado *aprendizagem supervisionada*. No caso de tarefas descritivas, não se tem a presença da resposta desejada

para os estímulos apresentados ao sistema durante a aprendizagem, caracterizando assim o paradigma de *aprendizagem não supervisionada*.

De acordo com o tipo do rótulo associado aos exemplos usados na aprendizagem supervisionada, as tarefas preditivas ainda se dividem em tarefas de *classificação*, caso os rótulos sejam representados por valores discretos; e de *regressão*, caso os rótulos sejam valores contínuos. Na aprendizagem não supervisionada, têm-se tarefas de *agrupamento*, cujo objetivo é encontrar algum tipo de similaridade entre os exemplos; de *associação*, que busca padrões de associação entre os exemplos; e de *sumarização*, que tenta encontrar uma descrição simplificada e concisa dos exemplos (FACELI *et al.*, 2011). A Figura 1 ilustra a hierarquia da AM de acordo com as tarefas citadas.

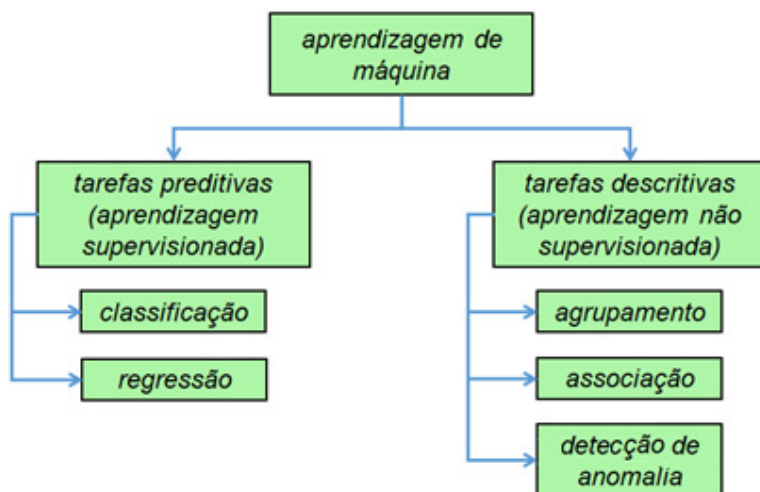


Figura 1. Hierarquia de aprendizagem de máquina.
Ilustração elaborada pelos autores.

O conteúdo sobre AM de máquina apresentado está longe de explorar o assunto em sua profundidade, uma vez que o objetivo aqui é apresentar apenas algumas noções sobre o tema. Vale destacar que AM não se restringe apenas aos paradigmas citados. Há ainda a aprendizagem ativa, por reforço e semissupervisionada. Um conhecimento mais pormenorizado sobre AM, seus paradigmas e técnicas pode ser obtido em Faceli *et al.* (2011) e Russel e Norvig (2013).

4. DECISÕES BIOÉTICAS APOIADAS POR APRENDIZAGEM DE MÁQUINA

A IA, especialmente a AM, tem o potencial de revolucionar a prestação de serviços na área de saúde. Contudo, projetar sistemas de suporte à decisão baseados em AM é mais que apenas um desafio técnico, requer também atenção aos princípios bioéticos. Com os avanços em IA e AM, as estruturas bioéticas precisam se adaptar aos problemas que esses sistemas em evolução podem apresentar, e o desenvolvimento desses sistemas automatizados também precisa se adaptar aos princípios bioéticos (GARCIA; ROMERO; RODRÍGUEZ, 2017)

A revolução proporcionada no sistema de saúde pela IA, levando a uma maior eficiência, automatizando tarefas rotineiras e reduzindo custos relacionados à saúde, vem ampliando o acesso aos serviços de saúde, direcionando com mais precisão as necessidades dos pacientes e auxiliando os médicos na tomada de decisões. No entanto, a avaliação moral e ética de cada decisão deve ser considerada. Conforme definido por Burton, Goldsmith e Mattei (2016), o padrão exato de julgamento moral varia de acordo com as diferentes escolas de teoria moral. Contudo, é possível perceber que o papel dessas novas tecnologias pode ser usado tanto para o bem quanto para o mal. Existem diferentes maneiras de fazer perguntas morais e diferentes maneiras de respondê-las.

Então, por exemplo, como ensinar às máquinas a responder eticamente a conflitos específicos? Uma das alternativas pode ser determinar uma regra ética, como proporcionar felicidade ao maior número de pessoas, transformá-la em um algoritmo e projetar uma máquina que se adapte a esse plano. Mas qual é o código de ética mais adequado? Afinal, sabe-se que toda regra ética tem suas peculiaridades e as exceções devem ser avaliadas em cada caso específico (BURTON; GOLDSMITH; MATTEI, 2016).

Outra opção é criar uma máquina que possa aprender a reagir a determinadas situações com resultados éticos, o que é muito semelhante à forma de como os humanos aprendem a identificar o que é certo e errado. Um exemplo de cooperação bem-sucedida entre humanos e máquinas é a realização de uma reunião médica para discutir as melhores decisões sobre o paciente.

Além disso, cabe supor a existência de uma máquina capaz de lidar com todas as situações em que diferentes profissionais de saúde em todo o

mundo lidam com casos semelhantes em momentos diferentes. De acordo com as circunstâncias, a cultura e os padrões sociais da época, a máquina também possui “conhecimentos morais”. Essa pode ser uma ajuda relacionada aos enfermeiros e outros membros da equipe médica. Portanto, observa-se um grande potencial da IA, não somente no que se refere à discussão de possíveis tratamentos, mas também no apoio à tomada de decisão ética diante de situações de conflito, principalmente em pacientes em situações de fim de vida (JONATHAN; ISLAM; SMITH, 2020).

Algumas questões devem ser pontuadas para solução de problemas por meio da IA. De acordo com Fenech e Buston (2020), a maior complexidade dos dados de saúde levanta novas questões sobre legislação, normas governamentais sobre uso e armazenamento de dados, especialmente porque as tecnologias de IA só são eficazes e relevantes diante da existência de conjuntos de dados que sejam amplos e representativos, que estejam disponíveis para utilização e sejam atualizados com frequência.

Além dos dados, a infraestrutura de hardware precisará ser atualizada ao longo do tempo, exigindo grandes investimentos financeiros para manter o uso da IA no sistema de saúde (DZOBO *et al.*, 2020). Assim, deve-se ter cautela, principalmente porque as complexas questões éticas envolvidas merecem consideração cuidadosa antes de implantação dessas tecnologias como suporte ao atendimento em saúde (PRABHU, 2019).

Necessário ainda se verificar o despreparo dos profissionais diante das novas tecnologias e preconceitos humanos associados ao uso da IA, com dificuldades de compreensão sobre as probabilidades fornecidas e limitação de dados (GERKE *et al.*, 2020). Há muito a ser estudado e entendido sobre as amplas implicações políticas sistêmicas da IA nos cuidados e serviços de saúde (ALAMI, *et al.*, 2020).

As preocupações com a privacidade também são importantes e ainda não há diretrizes em relação à coleta e ao compartilhamento de dados para fins comerciais, nem quanto à propriedade de dados com ou sem fins lucrativos (MICHIE *et al.*, 2017). Isso exige encontrar respostas coletivas a essas considerações, que devem acompanhar o trabalho de estruturação de práticas e infraestruturas por políticas públicas (FENECH; BUSTON, 2020).

Outra dimensão que pressiona a regulação é a velocidade do desenvolvimento. Para regular um setor de mudanças rápidas e imprevisíveis como a IA, o tempo é essencial para garantir que as normas e práticas regulatórias

sejam relevantes (PESAPANE *et al.*, 2018). Considerando que a aceitação da IA ainda é frágil, e que se espera que tenha uma presença ampliada em todos os aspectos do sistema de saúde, a avaliação de suas implicações éticas não pode ser negligenciada (ARANTES, 2020). Se não forem bem gerenciadas, essas considerações se tornarão grandes barreiras que jogarão contra a implantação da IA na saúde (PILOTTO; BOI; PETERMANS, 2018; VAYENA; BLASIMME; COHEN, 2018; BÆRØE *et al.*, 2020). E como se pode avaliar a intersecção entre IA e bioética?

A bioética é uma disciplina que lida com as implicações éticas e filosóficas das ciências da vida e da saúde e, com o advento da IA, surgem novas questões. As decisões bioéticas apoiadas por AM podem ser vistas em várias áreas, como medicina, genética e pesquisa clínica. Alguns exemplos que merecem ser citados são a previsão de risco de doenças genéticas, diagnósticos de doenças e personalização de tratamentos médicos. Apesar das facilidades proporcionadas por essas aplicações, os conflitos éticos subjacentes são altamente relevantes e não podem ser ignorados.

Primeiramente, há a questão do consentimento informado. Em medicina, os pacientes têm o direito de entender e consentir com os procedimentos médicos antes de serem realizados. No entanto, com a complexidade dos algoritmos de AM, pode ser difícil para os pacientes entenderem completamente como suas informações de saúde serão usadas. É importante garantir que os pacientes sejam adequadamente informados e que seu consentimento seja obtido antes da utilização desse recurso em seu cuidado.

Há ainda questões relacionadas à privacidade e segurança dos dados. Algoritmos de AM geralmente requerem grandes quantidades de dados para treinamento e validação. Isso pode incluir informações de saúde sensíveis dos pacientes. Assim, é essencial garantir que esses dados sejam armazenados e processados de maneira segura para proteger a privacidade dos pacientes.

Existem também questões relacionadas à equidade. Soluções baseadas em IA, especificamente em AM, têm o potencial de melhorar significativamente a qualidade do cuidado de saúde. Contudo, é importante garantir que esses benefícios sejam distribuídos de maneira justa. Assim, para que soluções baseadas em AM possam funcionar bem para todos os grupos de pacientes, as bases de dados utilizadas no processo de aprendizagem devem ser representativas e desprovidas de vieses. Ao mesmo tempo em que a AM

tem o potencial de transformar a bioética e a medicina, não se pode descuidar das implicações éticas, para assim garantir que a tecnologia seja usada de maneira responsável e justa. A bioética, portanto, desempenha um papel que alicerça a orientação do desenvolvimento e aplicação responsáveis da AM na saúde.

5. PARA NÃO CONCLUIR...

Será que é possível usar a IA como suporte para a tomada de decisão em saúde quando diante de questões bioéticas? Embora algumas iniciativas nesse sentido estejam sendo implantadas no sistema de saúde, a tão esperada revolução ainda está distante. É o tempo que determinará a confiabilidade e a amplitude da perspectiva de utilização. A adaptação da estrutura de assistência à saúde e dos aspectos tecnológicos requer uma análise meticulosa das questões de gestão e políticas, com o objetivo de tornar os processos mais ágeis e adequados à regulamentação necessária para lidar com as diversas características e impactos da Inteligência Artificial (IA) no setor de saúde. Esse processo deve também contemplar o investimento em profissionais devidamente capacitados.

Faz-se necessário compreender a importância da reflexão sobre os aspectos bioéticos em cada caso específico ao optar por ferramentas computacionais para auxiliar no cuidado. Esse entendimento é base para apoiar a tomada de decisões, especialmente em contextos críticos como a Terapia Intensiva. A administração de medicamentos e as deliberações éticas devem levar em consideração múltiplos critérios para assegurar a melhor decisão para o paciente. Entre esses critérios estão o conhecimento extraído de grandes bases de dados (*big data*), a consulta com especialistas na área e a consideração especial da vontade do paciente, respeitando sua autonomia. Pesquisas futuras baseadas na bioética, e em outras áreas do conhecimento, também são necessárias para entender melhor como os profissionais de saúde, pacientes e suas famílias tomam decisões no fim de vida ou em situações em que o paciente esteja fora de possibilidade terapêutica de cura.

Pode-se adicionar outros fatores à decisão. Acredita-se que profissionais com ideias diferentes contribuirão para o processo de aprendizagem, permitindo a criação de modelos de AM capazes de sugerir decisões considerando diferentes linhas de pensamento. Portanto, o sistema de suporte

não se limitará a um conjunto de sugestões predeterminado e imutável e pode ser estendido a outras referências teóricas. Esta variabilidade permite diferenças significativas e merece ser destacada como um dos elementos do cálculo das recomendações de apoio à decisão.

6. REFERÊNCIAS

ALAMI, H.; LEHOUX, P.; AUCLAIR, Y.; DE GUISE, M.; GAGNON, M.-P.; SHAW, J.; *et al.* Artificial intelligence and health technology assessment: anticipating a new level of complexity. **Journal of Medical Internet Research**, v. 22, n. 7, p. e17707, 2020. DOI: 10.2196/17707.

ARANTES, D. S. Modelos de tomada de decisão em bioética clínica: apontamentos para a abordagem computacional. **Revista Bioética**, Brasília, v. 28, n. 1, p. 19-27, 2020. Disponível em: https://revistabioetica.cfm.org.br/index.php/revista_bioetica/article/viewFile/952/1142. Acesso em: 5 set. 2024.

BÆRØE, K.; JANSEN, M.; KERASIDOU, A. Machine learning in healthcare: exceptional technologies require exceptional ethics. **American Journal of Bioethics**, v. 20, n. 11, p. 48-51, 2020. DOI: 10.1080/15265161.2020.1820103.

BRASIL. Portaria nº 3.432, de 12 de agosto de 1998. Estabelece critérios de classificação entre as Unidades de Tratamento Intensivo. Brasília, 1998. Disponível em: <http://www.hc.ufu.br/sites/default/files/tmp/PTGM-MS3432-98UTI.pdf>. Acesso em: 18 jul. 2019.

BURTON, E.; GOLDSMITH, J.; MATTEI, N. Using “The Machine Stops” for teaching ethics in artificial intelligence and computer science. In: AAAI Workshop: AI, 2016, p. 1-9.

DZOBO, K.; ADOTEY, S.; THOMFORD, N. E.; DZOBO, W. Integrating artificial and human intelligence: a partnership for responsible innovation in biomedical engineering and medicine. **OMICS: A Journal of Integrative Biology**, v. 24, n. 5, p. 247-263, 2020. DOI: 10.1089/omi.2019.0038.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. Inteligência artificial: uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011.

- FENECH, M. E.; BUSTON, O. AI in cardiac imaging: a UK-based perspective on addressing the ethical, social, and political challenges. **Frontiers in Cardiovascular Medicine**, v. 7, p. 54, 2020. DOI: 10.3389/fcvm.2020.00054.
- GARCIA, M. S. Á.; ROMERO, J.; RODRÍGUEZ, I. M. Investigation of the visual attention role in clinical bioethics decision-making using machine learning algorithms. **Procedia Computer Science**, v. 121, p. 1103-1110, 2017. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1877050917305495>. Acesso em: 5 set. 2024.
- GERKE, S.; BABIC, B.; EVGENIOU, T.; COHEN, I. G. The need for a system view to regulate artificial intelligence/machine learning-based software as medical device. **NPJ Digital Medicine**, v. 3, p. 101731738, 2020. DOI: 10.1038/s41746-020-0262-2.
- GIL, A. C. Pesquisa social. São Paulo: Atlas, 1999.
- JONATHAN, S.; ISLAM, M.; SMITH, A. Artificial intelligence for support of the bioethical decision-making of nurses related to end of life issues. **International Journal of Development Research**, v. 10, n. 5, p. 35899-35904, 2020. Disponível em: <http://www.journalijdr.com/sites/default/files/issue-pdf/20545.pdf>. Acesso em: 5 set. 2024.
- MICHIE, S.; YARDLEY, L. R.; WEST, R.; PATRICK, K.; GREAVES, F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. **Journal of Medical Internet Research**, v. 19, n. 6, p. e232, 2017. DOI: 10.2196/jmir.7126.
- PESAPANE, F.; VOLONTÉ, C.; CODARI, M.; SARDANELLI, F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. **Insights into Imaging**, v. 9, n. 5, p. 745-753, 2018. DOI: 10.1007/s13244-018-0645-y.
- PESSINI, L. Bioética clínica e sua prática. **Revista Bioética**, Brasília, v. 12, n. 2, p. 9-18, 2004. Disponível em: https://revistabioetica.cfm.org.br/index.php/revista_bioetica/article/view/632. Acesso em: 5 set. 2024.
- PILOTTO, A.; BOI, R.; PETERMANS, J. Technology in geriatrics. **Age and Ageing**, v. 47, n. 6, p. 771-774, 2018. DOI: 10.1093/ageing/afy026.

PRABHU, S. P. Ethical challenges of machine learning and deep learning algorithms. **Lancet Oncology**, v. 20, n. 5, p. 621-622, 2019. DOI: 10.1016/S1470-2045(19)30230-X.

PUCCINI, R. F.; SPINETTI, S. R. de M.; MAGALHÃES, M. E.; MIRANDA, A. C. L. de O.; SOUZA, A. de A. Bioética clínica. **Revista Brasileira de Reumatologia**, São Paulo, v. 48, n. 2, p. 77-85, 2008. Disponível em: <https://www.scielo.br/j/rbr/a/TK8tjw7jvvpbFY6sDsBVbNR/?lang=pt>. Acesso em: 5 set. 2024.

RICH, E.; KNIGHT, K. **Artificial intelligence**. 2. ed. New York: McGraw-Hill, 1991. p. 3.

RUSSEL, S.; NORVIG, P. **Inteligência artificial**. 3. ed. Rio de Janeiro: Grupo GEN, 2013.

VAYENA, E.; BLASIMME, A.; COHEN, I. G. Machine learning in medicine: addressing ethical challenges. **PLoS Medicine**, v. 15, n. 11, p. e1002689, 2018. DOI: 10.1371/journal.pmed.1002689.

11. EXPERIMENTAÇÃO *IN SILICO*: QUESTÕES BIOÉTICAS

Ademir Nunes Ribeiro Júnior

Daniel Louzada Fernandes

Fábio Ribeiro Cerqueira

1. INTRODUÇÃO

A existência em si não é suficiente para o *Homo sapiens*. A razão daquilo que existe ser o que é, é, pois, uma questão a ser resolvida. Diante das dúvidas e indagações que atravessam a comunidade humana, já nos primórdios, a primeira proposta de sistematização das relações causais se estabelece por meio dos mitos (SILVA; MACIEL; TELLES, 2012).

Quando o mito começou a ser insuficiente para resolver as dúvidas e elucidar a realidade, a filosofia surge como essa nova forma de pensar o mundo, entender os fenômenos e compreender o real. Assim surge o pensamento científico. Porém, no século XVII, após o Renascimento, são propostas estratégias e ferramentas de entender a existência e, com isso, o pensamento se configura em um método científico (SILVA; MACIEL; TELLES, 2012).

Galileu Galilei constrói a ideia de que é necessário formular hipóteses no tocante aos fatos e, depois, submetê-las a experimentos que possam comprovar, ou não, a autenticidade dessas teses. Assim se edifica o método científico experimental (SILVA; MACIEL; TELLES, 2012; VETTORATO; MÜLLER; DA SILVA, 2019).

Atualmente, temos diversas formas de fazer experimentos, dentre elas: estudos *in vitro*, *in vivo* com animais não humanos, *in vivo* com humanos e *in silico*.

2. ESTUDOS *IN VITRO*

2.1 O que é o estudo *in vitro*?

Um tipo de estudo muito comum no meio científico é o estudo/ensaio *in vitro*. A expressão, que vem do latim e que em português significa “em vidro” exatamente pelo fato de serem feitos em sua maior parte com equipamentos de vidro, define qualquer ensaio realizado fora de um organismo vivo, diminuindo ou, até mesmo, substituindo o uso de animais na experimentação (MORALES, 2008). Esses tipos de ensaios, que envolvem normalmente micro-organismos, proteínas, enzimas, células, tecidos ou até mesmo órgãos isolados, podem variar muito, de acordo com a finalidade e objeto de estudo.

Apesar de ser considerada uma etapa muito importante de qualquer estudo, com o resultado de ensaio *in vitro* somente, o pesquisador não consegue concluir o seu estudo. Essa é considerada uma fase preliminar e os ensaios são feitos em ambientes rigorosamente controlados e fechados, dentro de um laboratório em condições ideais, longe de qualquer similaridade com um organismo vivo que possui processos e interações extremamente complexas (TRAVASSOS, 2003). Por isso, muitas vezes os estudos *in vitro* vêm acompanhados de um estudo *in vivo* com animais ou plantas, na tentativa de mimetizar ao máximo o sistema em estudo.

A grande vantagem de se utilizar um método *in vitro* está no período do teste, que, na maioria dos casos, é mais curto quando comparado aos testes *in vivo* exatamente por poder limitar o número de variáveis experimentais (ROGERO *et al.*, 2003), obtendo respostas tais como: se uma substância é irritante, tóxica, se causa danos ao DNA ou se é corrosiva.

2.2 Quais os principais métodos?

Dependendo do tipo de pesquisa, a metodologia de um estudo *in vitro* pode ser muito diferente um do outro. Existem diversas metodologias disponíveis em literatura e muitas delas podem ser modificadas para se adequarem ao objetivo do estudo de forma a minimizar qualquer tipo de erro experimental.

Dentre todos os ensaios *in vitro*, os ensaios toxicológicos são os mais comuns e utilizados como teste preliminar em diversas áreas como:

produtos naturais (BEDNARCZUK, *et al.*, 2010), produtos cosméticos (YAMAGATA *et al.*, 2014), biomateriais (MASSON & LOMBELLO, 2016), farmácia (SOUZA *et al.*, 2007) e bioquímica (MATINELLO & SILVA, 2003).

Os métodos *in vitro*, geralmente seguem critérios previamente definidos e com isso podem ser validados por órgãos internacionais, como é o exemplo dos testes de biocompatibilidade de materiais para uso em dispositivos biomédicos, que seguem a ISO 10993 para os ensaios de citotoxicidade *in vitro*. Com esses resultados de citotoxicidade é possível saber se o produto é viável ou não para seguir com os ensaios *in vivo* com animais em laboratório. No Brasil, o conselho Nacional de controle de experimentação animal (CONCEA), pertencente ao ministério da ciência, tecnologia, informação e comunicação, é quem avalia e estimula a introdução das estratégias para a substituição ou redução do uso de animais em ensaios ou pesquisas.

2.3 Aplicação do estudo na saúde

Na área da saúde, os ensaios *in vitro* que testam o potencial de citotoxicidade são os mais utilizados. Eles avaliam a capacidade intrínseca de uma substância ou material em promover modificações no metabolismo nas culturas de células, podendo resultar na morte celular (HORTON & MATHEW, 2015). Com o avanço da tecnologia, os equipamentos que avaliam esses testes estão cada vez mais sensíveis e potentes e, como consequência disto, a utilização de culturas de células *in vitro* torna-se uma prática cada vez mais comum no meio científico. Pode-se testar a citotoxicidade de uma série de produtos, de drogas a materiais, graças à diversidade de culturas de células que podem ser replicadas *in vitro*, seja de origem animal ou humana. Existem muitos modelos de testes de citotoxicidade e sua utilização varia de acordo com o produto a ser avaliado e sua respectiva citotoxicidade. Dentre todos os ensaios, destacam-se os testes de viabilidade celular, ensaio clonogênico e ensaio de dupla marcação. Esses ensaios podem ser feitos também com culturas de células tumorais, porém, o objetivo do estudo é identificar substâncias com atividade citotóxica, os chamados quimioterápicos. É desejável que os quimioterápicos tenham atividade citotóxica com as células tumorais, mas que não sejam citotóxica com

células normais. Na prática, é muito difícil identificar uma substância com alta especificidade celular e, com isso, a substância costuma ser citotóxica também para as células saudáveis. Por isso, causam os efeitos adversos.

3. ESTUDO IN VIVO – EM ANIMAIS NÃO HUMANOS

3.1 O que é e estudo *in vivo* – em animais não humanos?

A experimentação em animal não humano se estabelece em técnicas e procedimentos, comumente realizado em laboratórios, que possuem como finalidade encontrar resultados e teorias desconhecidos, avaliar uma hipótese ou demonstrar uma tese já conhecida. (PAIXÃO; SCHRAMM, 2001; FONSECA, 2018).

Vale ressaltar que a experimentação animal pode se tratar de pesquisas em animais não humanos que busquem um conhecimento e aplicações para eles próprios, em prol da saúde e bem-estar animal, como é muito comum de se perceber no campo da medicina veterinária. Entretanto, frequentemente, os animais não humanos têm sido utilizados em estudos, como modelos experimentais, buscando conhecimento, bem-estar e benefícios para a espécie humana, ainda que isso cause prejuízos para os seres utilizados no estudo (PAIXÃO; SCHRAMM, 2001; FONSECA, 2018).

A prática da utilização de animais, por parte do homem, já ocorre há muito tempo. Desde o período em que o homem, ou seus ancestrais, viviam em cavernas, os animais possuíam uma grande importância para a sobrevivência humana. É possível encontrar registros históricos de animais sendo utilizados como parte da alimentação, como meio de transporte ou auxílio deste, em esportes e competições, para entretenimento, em rituais religiosos, dentre outras práticas que, em sua maioria, tinham como principal meta o benefício humano, sem que fosse considerado o bem-estar do animal. (FRANCO *et al.*, 2014).

Aristóteles (384-322 A.C.) apresenta uma visão de hierarquia dos seres, em que as diferenças entre os humanos e os demais animais colocam o homem em uma condição de superioridade. Adotando essa cosmovisão, é possível extrair uma perspectiva de soberania e dominação, que muitos utilizaram para construir uma ciência em que se faz aceitável a utilização de animais como ferramenta de construção de conhecimento (FONSECA, 2018).

Do mesmo modo, René Descartes (1596-1650), constroi conceitos que reverberam em diferenças significativas entre os seres. Para ele, a ação do pensamento e da sensibilidade estão relacionadas à alma, entretanto, ideologias hegemônicas – judaico-cristãs – estabelecem que a alma é a essência do homem, excluindo os animais e, mais uma vez, estabelecendo uma estrutura de superior-inferior entre os seres. Por conseguinte, os estudos com animais tornaram-se menos descritivos e observacionais, e tomaram características mais experimentais (FONSECA, 2018).

3.2 Questões Éticas

Em 1876 através da Lei Britânica de Crueldade com os Animais – *British Cruelty to Animals Act* – iniciou-se uma procura por regulamentação das pesquisas e experimentações com/em animais, o que não foi bem-visto por parte da comunidade científica da época (FRANCO *et al.*, 2014; FONSECA, 2018).

O zoologista William M.S. Russell e o microbiologista Rex L. Burch, em 1959, desenvolveram o conceito dos “3 Rs” – replacement, reduction e refinement (substituição, refução e refinamento) – em seu livro *The Principles of Humane Experimental Techniques*. Essa tese se dá em resposta às reflexões dos debates sobre sofrimento animal não humano, que crescia juntamente com as experimentações animais, e veio a se tornar uma das referências para a ciência, no que diz respeito à utilização de animais em laboratórios (FRANCO *et al.*, 2014; FONSECA, 2018). Fica, então, definido:

Replacement – Quando no estudo houver risco potencial para o animal, esse deverá ser substituído por material não consciente, tendo como exemplo, plantas e microorganismos.

Reduction – Indicar que a pesquisa deve ter um planejamento devidamente cauteloso, para que haja a minimização dos grupos de estudo.

Refinement – Preconiza a tomada de cuidados, de modo que se reduza o sofrimento e mal-estar do animal.

De modo geral, a proposta dos 3Rs não indica que os animais não devam ser utilizados nos experimentos, mas recomenda que o experimento seja realizado de modo a minimizar os danos nos animais (FONSECA, 2018).

No Brasil, a primeira Lei a instituir normas e regulamentações para as práticas didático-científicas em animais foi a 6.638/79, porém, ela não abordava o princípio dos 3Rs sugerido por Russell e Burch, 20 anos antes. Já em 12 de fevereiro de 1998, foi decretada a Lei 9.605, que dispõe sobre as sanções penais e administrativas que recaem sobre as ações lesivas ao meio ambiente, o que incluía os animais domésticos e de laboratórios. Atualmente, temos a Lei n.11.794/2008, sancionada em 8 de outubro de 2008 – mais conhecida como Lei Arouca – que estabelece as Comissões de Ética para o Uso de Animais (CEUAs), Conselho Nacional de Controle de Experimentação Animal (CONCEA).

Apesar das Leis vigentes, no Brasil, não existe uma norma que obrigue que os projetos de pesquisa, que utilizem animais, sejam submetidos ao parecer dos Comitês de Ética em Pesquisa. Entretanto, muitas instituições entendem que essa questão precisa ser considerada e refletida eticamente. Para isso, criam mecanismos de avaliação por meio de Comitês de Ética já existentes ou na criação de comitês próprios voltados para a análise de projetos que pretendam se utilizar de animais em seus estudos (FONSECA, 2018).

Mesmo com crescentes movimentos sociais em prol do bem-estar animal que lutam para a suspensão da experimentação com esses seres, muito cientistas defendem o uso dos animais não humanos na pesquisa experimental, alegando vantagens dos resultados e as consequências positivas no avanço e aprimoramento de técnicas e produtos farmacêuticos. Contudo, mesmo nesse grupo de pesquisadores que não renunciam a utilização de animais nas pesquisas, muitos admitem que esse uso precisa ser racionalizado (FONSECA, 2018).

4. ESTUDO IN VIVO – EM HUMANOS

4.1 O que é e estudo *in vivo* – em humanos?

É inegável que a participação do ser humano em experimentos científicos contribui com inúmeros benefícios. Entretanto, é muito importante ressaltar que, em nome de atingir esses benefícios, não se deve ultrapassar o respeito aos valores da dignidade e da vida humana (VETTORATO; MÜLLER; DA SILVA, 2019).

Os estudos com seres humanos ocorrem há milênios, porém, nem sempre esses valores foram critério para a condução das pesquisas, mas foi com

os impactos da Segunda Guerra Mundial que essa questão ganhou foco. Como muitos seres humanos foram, principalmente pelos nazistas, objetos de estudos e experiências, sujeitos à terríveis atrocidades, que podem ser categorizadas como tortura, com o fim da Guerra entendeu-se a necessidade de se estabelecerem princípios éticos e bioéticos que pudessem valorizar e cuidar dos sujeitos de pesquisas, além de impedirem que situações similares ocorressem (VETTORATO; MÜLLER; DA SILVA, 2019).

Após o fim dos horrores cometidos na Segunda Guerra Mundial, em 1947, foi organizado o Tribunal Militar Internacional, em Nüremberg, com o objetivo de julgar os crimes cometidos contra a humanidade nos campos de concentração. Nesse contexto, foi desenvolvido um documento pelos médicos das forças armadas dos Estados Unidos, que ficou conhecido como o Código de Nüremberg, que veio a ser o primeiro grande marco ético nas ciências da saúde com seres humanos. Esse documento tem valor internacional e constrói a importância do consentimento do sujeito participante da pesquisa, instituindo dez princípios que sujeitam os experimentos, dentre eles a beneficência como um critério que torna justa a participação do pesquisado (RECKZIEGEL, 2014; VETTORATO; MÜLLER; DA SILVA, 2019).

As diretrizes, no Brasil, que determinam e direcionam o desenvolvimento das pesquisas com seres humanos, foram estabelecidas pela Resolução 196/96, do Conselho Nacional de Saúde (CNS) do Ministério da Saúde. Porém, em dezembro de 2012, o Conselho reavaliou e publicou a Resolução 466/12, que é a referência no que diz respeito aos Comitês de Ética em Pesquisa envolvendo Seres Humanos. Os comitês têm, também, o objetivo de realizar o controle dos protocolos de pesquisa e das revisões éticas das propostas e projetos de pesquisa que envolvam seres humanos (BRASIL, 2012)

Infelizmente, no que se refere à preocupação e importância desprendidas na tentativa de se alcançar bons resultados em um estudo, proporcionar o avanço da ciência e encontrar curas para doenças, muitas vezes, não condiz com o valor que se dá aos sujeitos envolvidos na pesquisa (AMORIM, 2019).

Em 1966, um estudo denominado *Ethics and Clinical Research* verificou que muitos trabalhos publicados na época foram considerados antiéticos. Atualmente, mesmo com princípios éticos bem estabelecidos e disseminados internacionalmente, ainda é possível encontrar situações de abuso e falta com a ética diante dos participantes de um estudo. Um caso

contemporâneo foi o estudo realizado em 2002 com ribeirinhos de São Raimundo do Piratiba – Amapá, em que os indivíduos eram expostos e submetidos a picadas de mosquitos em um experimento para se estudar a malária (AMORIM, 2019).

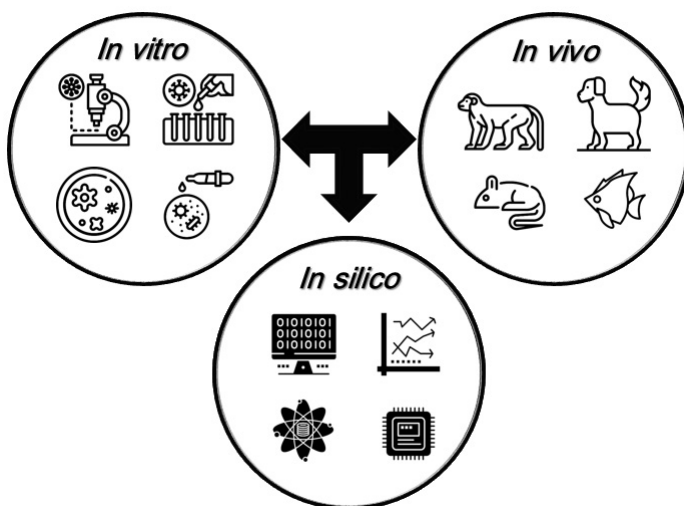
Diante de todos esses apontamentos dos estudos *in vitro*, *in vivo* com animais não humanos e *in vivo* com humanos, é importante persistir no avanço da ciência, mas sem infringir os limites éticos e bioéticos. Diante dessa necessidade, busca-se meios alternativos que possam substituir ou complementar os estudos biológicos.

5. ESTUDO *IN SILICO*

5.1 O que é o estudo computacional?

De maneira semelhante aos estudos *in vivo* e *in vitro*, a experimentação *in silico* refere-se a uma técnica de executar estudos de um determinado procedimento biológico, como a modelagem do comportamento celular; a descoberta de novos fármacos; a análise de sequenciamento de DNA; entre outros. No entanto, ao invés de conduzir o estudo em um organismo vivo ou em um ambiente controlado fora desse organismo, recorre-se ao meio computacional, conforme ilustrado na Figura 1.

Figura 1: Relação entre os modelos *in vitro*, *in silico* e *in vivo*.



Fonte: elaborada pelo autores.

A expressão *in silico* é usada para indicar estudos naturais biológicos ou de laboratórios conduzidos inteiramente por meio da modelagem e simulação por computador. Com origem no Novo Latim, a expressão é uma referência aos chips de computadores que são feitos de silício e a mesma começou a ser mencionada no início da década de 90 nos Estados Unidos (SAEIDNIA; MANAYI; ABDOLLAHI, 2013).

Uma vez que foi reconhecida a presença de desafios e biocomplexidades em pesquisas biomédicas, – por exemplo: i) as fraquezas das experiências *in vitro* por elas não conseguirem reproduzir as condições celulares precisas de um organismo; ii) a discussão e o questionamento a respeito da real necessidade da utilização de animais em experimentos científicos entre a sociedade protetora desses organismos e a comunidade acadêmica; iii) conclusões enganosas em estudos *in vivo* sobre benefício a curto prazo, mas um dano a longo prazo; iv) limitação no uso dos ensaios *in vitro* e *in vivo* em larga escala (POSSI *et al.*, 2011; GUIMARÃES; FREIRE; MENEZES, 2016); entre outros – estudos *in silico* começaram a ser empregados de maneira crescente e gradativa como testes alternativos e complementares aos testes laboratoriais convencionais – *in vitro* e *in vivo* – por meio de procedimentos computacionais (PAUL GLEESON; HERSEY; HANNONGBUA, 2011; PITTA, 2018).

A relevância das aplicações dos estudos *in silico* já vem sendo notada pela comunidade científica de todo o mundo, principalmente em pesquisas de correlação, predição, simplificação e automatização nas áreas de toxicologia e farmacologia (SAEIDNIA; MANAYI; ABDOLLAHI, 2013; LESCANO, 2017). Com a utilização destes procedimentos, estima-se que seja possível diminuir a quantidade de erros sistemáticos ocorridos; o uso de animais e de reagentes químicos em experimentos; além dos custos e o tempo do desenvolvimento de diversas pesquisas (LI *et al.*, 2009; PAUL GLEESON; HERSEY; HANNONGBUA, 2011; FERREIRA *et al.*, 2016; PITTA, 2018).

Embora os estudos *in silico* representem um caminho novo de investigação na ciência para ajudar os pesquisadores a entenderem melhor os mecanismos biológicos e moleculares e a testarem suas hipóteses científicas (POSSI *et al.*, 2011), a simulação é a criação de um modelo computacional a partir da observação de um fenômeno natural. Dessa forma, não há como realizar a análise e o cruzamento dos dados em estudos *in*

silico sem se basear nos conhecimentos já obtidos pela abordagem clássica, isto é, sem o manuseio dos experimentos *in vitro* e *in vivo* para coleta e averiguação das atividades biológicas, armazenados em banco de dados (SAEIDNIA; MANAYI; ABDOLLAHI, 2013). Além disso, por causa do domínio de aplicabilidade, nenhuma abordagem computacional será capaz, ao menos no atual estágio tecnológico, de modelar completamente a complexidade dos sistemas biológicos, o que impede previsões totalmente precisas (EKINS; MESTRES; TESTA, 2007). Portanto, é importante ter ciência de que, apesar de útil, a abordagem *in silico* ainda não substitui a total necessidade das experimentações em modelos anteriormente citados.

Apesar de ser uma abordagem alternativa e complementar, foi constatada nos últimos anos em registros da literatura, principalmente na brasileira – derivados de centros de pesquisas dos estados de São Paulo e Pernambuco, por exemplo –, a aplicação dos métodos *in silico* em diferentes pesquisas científicas. A presença desses registros demonstra que os estudos *in silico* estão em andamento, auxiliando novas descobertas científicas por meio de uma frutífera variedade de oportunidades e de resultados confiáveis dos métodos computacionais (BAPTISTINI, 2015; PITTA, 2018). A respeito das recentes pesquisas *in silico* realizadas no Brasil, pode-se mencionar algumas, como:

- a modelagem computacional para simulação da resposta imune da infecção por *Plasmodium falciparum* na malária transfusional (RIBEIRO JÚNIOR, 2019);
- o estudo dos mecanismos da infecção do ZIKA vírus por meio do uso de algoritmos de aprendizado de máquinas (PITTA, 2018);
- a investigação das interações moleculares dos compostos candidatos a inibidores de fosfodiesterase 5 por meio de docagem e simulações de dinâmica molecular (LESCANO, 2017);
- a detecção e análise automatizada de sequências de imunoglobulinas, produzidas por tecnologia de *phage display* (SILVA, 2016);
- a avaliação do potencial da formação de complexos entre compostos organocalcogênicos e as enzimas COX-1 e COX-2 na atuação como inibidores seletivos da COX-2 (BAPTISTINI, 2015).

5.2 Quais os principais métodos?

A abordagem *in silico* propõe novos *insights* no processo de descoberta de conhecimento para posterior avaliação experimental em muitas áreas – como, bioinformática, quimioinformática, farmacologia, toxicologia, imunologia molecular, entre outras – devido ao bom desempenho alcançado por meio da implementação e uso dos mais variados métodos (MULLAICHARAM; AMARESH, 2014). Entre os métodos *in silico* do estado da arte, destacam-se principalmente os seguintes:

- Docagem molecular;
- Triagem virtual;
- Relação quantitativa de estrutura-atividade;

Docagem ou ancoragem molecular, do inglês *molecular docking*, é o principal método baseado em alvo – isto é, abordagem em que a estrutura da enzima é o ponto de partida para a geração do modelo – para simulação de uma interação biomolecular (MAKRYNITSA *et al.*, 2018). Essa técnica computacional, utilizada na criação de medicamentos baseados em estruturas, possibilita tanto a modelagem da caracterização do comportamento de pequenas moléculas no local de ligação das macromoléculas biológicas alvo – comumente proteínas – quanto a obtenção de informações para o esclarecimento dos processos bioquímicos indispensáveis (RAUNIO *et al.*, 2015; MAKHOURI; GHASEMI, 2018). O método de docagem molecular possui duas etapas: i) predição da conformação, posição e orientação do ligante dentro dos locais de ligação e ii) avaliação da força de afinidade de ligação entre as moléculas por meio dos cálculos da energia (MENG *et al.*, 2011). Os resultados alcançados por essa técnica são dados referentes à energia de ligação necessária para que ocorra a interação entre o ligante e a proteína alvo. Deste modo, almeja-se as informações do ligante no local de ligação que maximize essas interações intermoleculares, minimizando a energia total do processo, pela razão de que, o ligante que necessitar de uma menor quantidade de energia para se ligar à proteína, hipoteticamente, demonstrará um melhor desempenho de atividade biológica (BARROS, 2015).

Triagem virtual, também conhecido como *virtual screening*, é um método computacional empregado atualmente em pesquisas de descoberta de medicamentos – principalmente nas etapas iniciais – com o objetivo

de avaliar muitos ligantes em relação a uma determinada molécula alvo (MULLAICHARAM; AMARESH, 2014; MAKRYNITSA *et al.*, 2018). Essa técnica computacional identifica e seleciona, por meio de testes com a utilização de grandes bibliotecas de compostos e estruturas orgânicas (químicas ou naturais), quais os compostos – agonistas ou antagonistas de receptores, ou ainda inibidores enzimáticos – que possuem a maior probabilidade de ligação a um alvo específico, com o intuito de possibilitar uma vasta gama de interações intermoleculares com os mais distintos alvos moleculares de forma rápida, eficaz e com baixo custo em relação aos experimentos *in vitro* (MENG *et al.*, 2011; RODRIGUES, 2012; MAKHOURI; GHASEMI, 2018). Em contrapartida, uma desvantagem apresentada pelo método é a alta presença de falsos negativos quando realizada a validação experimental (FERREIRA; GLAUCIUS; ANDRICOPULO, 2011).

O método relação quantitativa de estrutura-atividade, popularmente conhecido como *Quantitative Structure-activity Relationship* (QSAR), é outro método vastamente usado na criação de novos medicamentos por meio da abordagem baseada em ligantes, que tenta prever a correlação entre as atividades biológicas – propriedades físico-químicas e biológicas – e os recursos estruturais de moléculas com significância estatística (MULLAICHARAM; AMARESH, 2014; MAKHOURI; GHASEMI, 2018; MAKRYNITSA *et al.*, 2018). As abordagens de QSAR podem ser classificadas de acordo com as dimensões dos dados, variando até a sexta dimensão, isto é, conforme a maneira pela qual os valores dos descritores (representação da estrutura de uma molécula em termos quantitativos, como, o peso molecular, solubilidade, etc.) são calculados e representados quantitativamente (MULLAICHARAM; AMARESH, 2014; MAKHOURI; GHASEMI, 2018). Ressalta-se que o QSAR é uma abordagem baseada em métodos intensivos da estatística, além de métodos de busca heurística e algoritmos de aprendizado de máquina supervisionados e não supervisionados que possibilitam o descobrimento de padrões, similaridades e diferenças entre amostras de compostos químicos para prever a resposta biológica de novos ligantes (RAUNIO *et al.*, 2015; MAKRYNITSA *et al.*, 2018).

Além dos três métodos *in silico* descritos anteriormente, outras abordagens computacionais – como, mapeamento de farmacóforo; dinâmica molecular; e projeção baseado em fragmentos – também foram desenvolvidas para ajudar a entender os mecanismos moleculares e a propor

novas opções de tratamento para várias doenças (MULLAICHARAM; AMARESH, 2014; (RAUNIO *et al.*, 2015; MAKRYNITSA *et al.*, 2018). De forma semelhante ao QSAR, esses métodos *in silico* foram concebidos graças à aplicação dos algoritmos de aprendizado de máquina, sistemas multiagentes, métodos de busca heurística juntamente com as técnicas de visualização, análise e inferência de dados.

O campo do aprendizado de máquina é uma subárea da Inteligência Artificial que estuda o desenvolvimento de algoritmos capazes de conceder aos computadores a habilidade de aprender padrões e melhorar seu desempenho a partir do fornecimento de dados e da experiência com erros (MITCHELL; MICHALSKI; CARBONELL, 2013). Segundo registros da literatura, o aprendizado supervisionado e não supervisionado são as duas principais abordagens em termos de como o processo de aprendizado é concretizado (RIFAIOGLU, 2019).

Na abordagem supervisionada, aplica-se um algoritmo de aprendizado – baseado em abordagens tais como: regras de decisão, modelos probabilísticos, redes neurais, medidas de similaridade entre exemplos, etc. – sobre um conjunto de treinamento formado por registros ou instâncias cujos rótulos das classes, categorias ou grupos a qual pertencem já são conhecidos a priori com o intuito de construir um modelo computacional que seja capaz de classificar (valores categóricos) ou prever (valores numéricos) para instâncias futuras de maneira confiável baseada no processo de aprendizagem obtido anteriormente com o conjunto de treinamento (MITCHELL; MICHALSKI; CARBONELL, 2013). Por exemplo, classificar uma série de moléculas em ativas ou inativas com base em um determinado alvo molecular (RODRIGUES, 2012). Dentre os algoritmos de aprendizado supervisionado, destacam-se: abordagens bayesianas (por exemplo, *Naïve Bayes* e *Bayesian Networks*); árvores de decisão (*Decision Trees*); floresta aleatória (*Random Forest*), *k*-vizinhos mais próximos (*k*-*Nearest Neighbors* - *k*NN), máquina de vetores de suporte (*Support Vector Machine* - SVM) e, por fim, as redes neurais artificiais (*Artificial Neural Networks* - ANNs), sendo, este último, amplamente utilizado em distintas aplicações recentes de aprendizado profundo – arquitetura de rede neural que contém várias camadas ocultas de processamento de dados – devido aos resultados bem-sucedidos alcançados, especialmente em análise de imagens (RODRIGUES, 2012; RIFAIOGLU, 2019).

No aprendizado não supervisionado, o intuito da abordagem é a descoberta de características, padrões ou até mesmo de anomalias entre os dados de entrada que, neste caso, não possuem rótulos predefinidos. Portanto, não há um período de treinamento anterior usando dados históricos para os quais se conhece uma categoria. Ao contrário, o algoritmo de aprendizado é aplicado diretamente nos dados a partir dos quais se deseja encontrar padrões. Uma tarefa bastante comum no aprendizado não supervisionado é o chamado agrupamento ou *clustering*, como é mais conhecido. Neste caso, deseja-se colocar os exemplos em grupos – ou *clusters* – tais que instâncias no mesmo grupo tenham bastante similaridade entre si, enquanto que instâncias em grupos distintos tenham a similaridade minimizada. Cada grupo pode ser pensado como uma categoria que emerge dos dados a partir da aplicação do algoritmo e, assim, seus elementos podem ser posteriormente estudados. Para realizar *clustering*, utilizam-se métricas de similaridade ou de distância para medir a proximidade das Instâncias (MITCHELL; MICHALSKI; CARBONELL, 2013). Um exemplo de aplicação de *clustering* é a identificação de compostos estruturais e biologicamente semelhantes para o direcionamento da síntese de compostos promissores e novos candidatos a fármacos (RIFAIÖGLU, 2019). Além dos procedimentos de *clustering*, há outras tarefas também importantes, como a identificação de regras de associação e redução de dimensionalidade. Algumas das aplicações e algoritmos mais comuns do aprendizado não supervisionado são: o agrupamento com *k-means*, a associação com *Apriori* e a redução de dimensionalidade com a Análise de Componentes Principais (*Principal Component Analysis* - PCA) (RODRIGUES, 2012).

5.3 Aplicação do estudo *in silico* na saúde

Até o presente momento, foram discutidos apenas as experimentações *in silico* no contexto de simulações: i) de descobertas de fármacos e seus possíveis efeitos e; ii) de opções de tratamentos a serem realizados em um ambiente virtual sem consequências reais para seres vivos. No entanto, além desse contexto, o avanço da Inteligência Artificial também tem aprimorado a abordagem *in silico* para o uso na medicina – conhecida como medicina *in silico* ou medicina computacional –, permitindo que médicos e outros profissionais da área da saúde possam, por meio das técnicas computacionais,

encontrar informações (respostas e padrões) em dados não estruturados para auxiliar na descrição de fenômenos observados, na avaliação de hipóteses e em tomadas de decisões (POSSI *et al.*, 2011; MESKÓ, 2019).

As técnicas computacionais permitem a modelagem dos sistemas biológicos reais em ambientes virtuais por meio do mapeamento e simulação dos elementos biológicos com uso de *Big Data*, isto é, enormes volumes de conjuntos de dados. As aplicações da medicina *in silico* tentam demonstrar realisticamente aos pesquisadores como os sistemas – respiratórios, circulatórios, imunológicos, metabólicos, endócrinos e neurais – evoluem e reagem ao longo do tempo sob determinados estímulos (HESTER *et al.*, 2011). Os *insights* obtidos desses comportamentos oferecem previsões sobre os planos de tratamentos, dosagens mais apropriadas, além de ensaios clínicos mais rápidos e flexíveis e uma redução significativa de custos (MESKÓ, 2019).

Das diversas aplicações na saúde baseadas nos modelos *in silico*, pode-se mencionar, por exemplo, o *Oncosimulator*, o IBM Watson Health (IWH) e o HumMod. O *Oncosimulator*, desenvolvido por meio da colaboração com vários centros de pesquisa, é uma ferramenta avançada de suporte à decisão médica preditiva cuja finalidade é otimizar o tratamento individual do paciente com câncer (em especial, o nefroblastoma e o câncer de mama), explorando os próprios dados clínicos do paciente em várias escalas – por exemplo, dados de imagem, histológicos e moleculares – por meio da combinação de métodos de modelagem matemática mecanística e multiescalar, estatística, Inteligência Artificial e engenharia de *software* (STAMATAKOS, 2013). Este sistema baseado em biomodelagem multiescalar de dados clínicos almeja encontrar um tratamento ideal, adaptado à constituição biológica do paciente, uma vez que os sistemas genéricos de tratamento levam em consideração a resposta média de uma população de pacientes, o que pode resultar em reações diferentes ao mesmo tratamento entre os pacientes (STAMATAKOS, 2013).

O IWH é a uma abordagem de computação cognitiva com uma arquitetura massivamente paralela, probabilística, baseada em recuperação de informações, aprendizado de máquina, representação e raciocínio do conhecimento que analisa grandes volumes de dados da saúde, compreende perguntas complexas por meio do processamento de linguagem natural e sugere respostas baseadas em evidências de forma clara e consistentemente que se compara ao desempenho humano (AHMED *et al.*, 2017).

Por exemplo, o IWH analisa os dados de um paciente em particular com base em bancos de dados de ensaios clínicos adequados e oferece um feedback ao médico sobre a elegibilidade do paciente para estudos clínicos específicos. Desta forma, o IWH possibilita melhorias tanto nas práticas de cuidados à saúde do paciente por ser capaz de fornecer diagnósticos com maior nível de segurança, quanto na assistência à decisão cognitiva na área da saúde devido às consultas às mais atuais evidências médicas do mundo compartilhadas (AHMED *et al.*, 2017; SOMASHEKHAR *et al.*, 2018).

O HumMod é um ambiente de modelagem de simulação matemática que realiza uma extensa análise das diferentes situações fisiológicas e fisiopatológicas do corpo humano (MESKÓ, 2019). O modelo é construído a partir de dados empíricos obtidos da literatura fisiológica e permite aos pesquisadores adicionarem descrições detalhadas da fisiologia humana para testar novas hipóteses por meio de arquivos no formato XML (*Extensible Markup Language*). Segundo HESTER *et al.* (2011), o HumMod simula diferentes situações experimentais por meio da alteração interativa dos parâmetros fisiológicos numéricos e exibe as respostas dependentes do tempo. Essas simulações possibilitam observar com detalhes as propriedades emergentes de alto nível e descobrir as relações causais não evidentes que conduzem os processos fisiológicos (HESTER *et al.*, 2011). Além disso, o modelo oferece orientação e melhoria dos resultados de intervenções médicas e pode ser usado para educação fisiológica, pelo fato de fornecer aos estudantes a capacidade de entender os aspectos integrativos da fisiologia humana (MESKÓ, 2019).

Além dos exemplos internacionais mencionados, podem-se citar também alguns trabalhos de pesquisa nacionais que aplicam modelos computacionais na área da saúde. Entre os muitos registros na literatura brasileira, há os trabalhos realizados por Possi *et al.* (2011), Cerqueira *et al.* (2014) e Fernandes *et al.* (2017) desenvolvidos na Universidade Federal de Viçosa entre os departamentos de Informática (DPI) e Medicina e Enfermagem (DEM), no qual utilizam abordagens computacionais para auxiliar na tomada de decisão na prática clínica. Possi *et al.* (2011) desenvolveram um sistema baseado em agentes conhecido como AutoSimmune. Esse sistema é capaz de realizar a modelagem do sistema imunológico e a simulação de processos ligados à autoimunidade. Mais tarde, esse trabalho serviu como modelo inicial para outros pesquisadores, entre eles, Martins *et al.* (2015), no qual apresentaram um simulador com a utilização de técnicas de

programação paralela e sistema multiagentes para melhorar o poder computacional e os resultados em comparação ao AutoSimmune. Ribeiro Júnior (2019) apresentou a adaptação do AutoSimmune para simular a resposta imune da infecção por *Plasmodium falciparum* na malária transfusional.

Cerqueira *et al.* (2014) apresentaram o NICEsim, um ambiente de simulação interativo baseado nos algoritmos de aprendizado de máquina, SVM e ANN, para auxiliar os profissionais de saúde na obtenção de uma melhor avaliação de risco de morte de recém-nascidos prematuros em UTI neonatal. Ainda no cenário de suporte à tomada de decisão, exclusivamente, em bioética, Fernandes *et al.* (2017) propuseram um modelo computacional baseado em ANN e estatística que prediz e elucidada – por meio de dados de atenção visual e da mensuração da carga cognitiva – a decisão bioética dos estudantes da saúde durante um pedido de eutanásia.

Acredita-se que diante dos constantes avanços tecnológicos dos últimos anos, os laços de relacionamento entre a área da saúde e a computação se fortalecerão, possibilitando melhorias na prática médica e na prestação serviços de cuidados com a saúde por meio da exploração e análise de dados, obtenção de novos *insights* e desenvolvimento de sistemas inteligentes.

5.4 As vantagens de se utilizar a experimentação computacional como uma saída para os dilemas éticos

A discussão acerca da utilização de animais em experimentos *in vivo* é complexa, tendo em vista que até os dias atuais não se alcançou um equilíbrio entre as convicções sobre o que é ético ou não de acordo com algumas premissas, como: os estudos em animais trouxeram ao longo dos anos muitos benefícios para a vida das pessoas; esses estudos são vitais para o desenvolvimento da ciência; ainda não existe nenhum sistema biológico artificial que reaja de modo tão complexo quanto um organismo vivo; a ciência é apta a avançar sem o uso dos animais com métodos alternativos; organismos de espécies distintas podem reagir de diferentes maneiras aos estímulos apresentados; geração de modelos imperfeitos ou de validade questionável; prática da crueldade implícita e do sofrimento desnecessários dos animais na atividade experimental, etc. (FESTING; WILKINSON, 2007; COELHO, 2016).

Neste contexto, a experimentação *in silico* torna-se relevante à luz das questões éticas. De acordo com Possi *et al.* (2011), caso os experimentos *in silico* consigam modelar e simular com alta precisão os sistemas biológicos, os conflitos entre questionamentos éticos juntamente com atrasos nos avanços científicos – especialmente na saúde pública e na pesquisa (bio)médica – poderiam ser evitados. Visto que, com o auxílio dos recursos computacionais, o uso de animais em experimentos pode ser substituído ou reduzido significativamente (FESTING; WILKINSON, 2007).

Além disso, é interessante discutir outro benefício retornado por meio da modernização com os métodos computacionais. Em casos especiais, em que os resultados experimentais alcançados anteriormente – por exemplo, em cirurgias ou ensaios clínicos – por si só não são suficientes para avaliar o risco associado a uma tomada de decisão complexa, o uso da modelagem e a simulação computacional poderia oferecer *insights* importantes para apoiar a análise de risco e decidir se deve evitar ou aceitar a decisão crítica. Neste contexto, pode-se mencionar o uso da modelagem em 3D que pode fornecer visualizações avançadas da anatomia e fisiologia e orientar o planejamento pré-operatório no ambiente clínico (ZAHN, 2019).

Na prática clínica, os estudos *in silico* também proporcionam uma vantagem significativa quando se trata da reutilização ilimitada das simulações de ambientes ou organismos virtuais, possibilitando além de uma economia de custos e tempo, uma redução no número de perdas reais por causa de falhas e imperfeições nas previsões emitidas nas demais experimentações anteriores. Por exemplo, a possível morte de um organismo no ambiente virtual por causa de efeitos colaterais não implicaria em questionamentos conflituosos como se fosse com um organismo vivo no mundo real. Este tipo de estudo é capaz de repetir e ajustar o tratamento até que seja estabelecida a eficácia e a segurança dos produtos biomédicos para a aplicação em seres vivos.

Por fim, ressalta-se também a capacidade de simular experimentos que possam ser considerados – por meio das abordagens tradicionais – imorais ou que sejam contrários aos princípios impostos pela (bio)ética. No estudo realizado por Fernandes *et al.* (2017), por exemplo, foi simulada diversas vezes uma cena de pedido de eutanásia aos estudantes da saúde e foi criado um modelo computacional para prever as tomadas de decisões. Situações reais como essa são proibidas no Brasil, segundo a legislação vigente. No

entanto, com o auxílio dos recursos computacionais, momentos assim, nos quais não existem consenso moral, poderiam ser realizados e estudados – dentro das normas permitidas – para que uma melhor compreensão a respeito dos mesmos seja obtida e hipóteses científicas possam ser testadas e validadas por cientistas. Casos assim apontam para uma aplicação empolgante e valiosa dos métodos *in silico* por contribuírem na aceleração de resultados e descobertas plausíveis.

6. REFERÊNCIAS

AHMED, Mohamed Nooman. *et al.* Cognitive computing and the future of health care cognitive computing and the future of healthcare: the cognitive power of IBM Watson has the potential to transform global personalized medicine. **IEEE Pulse**, v. 8, n. 3, p. 4-9, 2017.

AMORIM, Karla Patrícia Cardoso. Ética em pesquisa no sistema CEP-CONEP brasileiro: reflexões necessárias. **Ciência & Saúde Coletiva**, v. 24, p. 1033-1040, 2019.

BAPTISTINI, Natália. **Análise *in silico*, *in vitro* e *in vivo* de compostos organocogênicos como possíveis anti-inflamatórios**. 2015. Tese. (Doutorado em Biotecnologia) – São Carlos: Universidade Federal de São Carlos, 2015.

BARROS, Maria Ester de Sá Barreto. **Estudos de docking molecular, síntese e atividade biológica de análogos da (-)-massoialactona e da combretastatina A-4**. 2015. Tese. (Doutorado em Química) – Recife: Universidade Federal de Pernambuco, 2015.

BEDNARCZUK, V. O., VERDAM, M. C. S., MIGUEL, M. D., & MIGUEL, O. G., testes *in vitro* e *in vivo* utilizados na triagem toxicológica de produtos naturais. **Visão Acadêmica**, 11(2), 2010. doi:10.5380/acd.v11i2.21366.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. Resolução n. 466, de 12 de dezembro de 2012. **Aprova diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos**. Brasília, Diário Oficial da União, 12 dez. 2012.

CERQUEIRA, Fabio Ribeiro. *et al.* NICeSim: an open-source simulator based on machine learning techniques to support medical research on prenatal and perinatal care decision making. **Artificial Intelligence in Medicine**, v. 62, n. 3, p. 193-201, 2014.

COELHO, Renata de Araújo. O uso de animais em experimentações científicas: um dilema para os seres humanos. **Brazilian Journal of Forensic Sciences, Medical Law and Bioethics**, v. 5, n. 2, p. 171-182, 2016.

EKINS, S.; MESTRES, J.; TESTA, B. *In silico* pharmacology for drug discovery: applications to targets and beyond. **British Journal of Pharmacology**, v. 152, n. 1, p. 21-37, 2007.

FERNANDES, Daniel L. *et al.* Investigation of the visual attention role in clinical bioethics decision-making using machine learning algorithms. **Procedia Computer Science**, v. 108, p. 1165-1174, 2017.

FERREIRA, Luiz Felipe Gomes Rebello. *et al.* Combination of *in silico* methods in the search for potential CD4+ and CD8+ T cell epitopes in the proteome of leishmania braziliensis. **Frontiers in immunology**, v. 7, p. 327, 2016.

FERREIRA, Rafaela S.; GLAUCIUS, Oliva; ANDRICOPULO, Adriano D. Integração das técnicas de triagem virtual e triagem biológica automatizada em alta escala: oportunidades e desafios em P&D de fármacos. **Química Nova**, p. 1770-1778, 2011.

FESTING, Simon; WILKINSON, Robin. The ethics of animal research. **EMBO Reports**, v. 8, n. 6, p. 526-530, 2007.

FONSECA, Marcos Augusto Moutinho. *et al.* **Bioética na pesquisa de engenharia biomédica: diminuição do uso de animais em experimentos científicos.** 2018.

FRANCO, Ana Lúcia. *et al.* Pesquisas em animais: uma reflexão bioética. **Acta Bioethica**, v. 20, n. 2, 2014.

GUIMARÃES, Mariana Vasconcelos; FREIRE, José Ednésio da Cruz; MENEZES, Lea Maria Bezerra de. Utilização de animais em pesquisas: breve revisão da legislação no Brasil. **Revista Bioética**, v. 24, n. 2, p. 217-224, 2016.

HESTER, Robert. *et al.* HumMod: a modeling environment for the simulation of integrative human physiology. **Frontiers in Physiology**, v. 2, p. 12, 2011.

HORTON, Nathan C.; MATHEW, Porunelloor A. NKp44 and natural cytotoxicity receptors as damage-associated molecular pattern recognition receptors. **Frontiers in immunology**, v. 6, p. 31, 2015.

LESCANO, Caroline Honaiser. **Avaliação *in silico*, *in vitro* e *in vivo* de compostos candidatos a inibidores de fosfodiesterase 5**. 2017. Tese. (Doutorado em Farmacologia) – Campinas: Universidade Estadual de Campinas, 2017.

LI, Xiang-hua. *et al.* Modelling immune system: principles, models, analysis and perspectives. **Journal of Bionic Engineering**, v. 6, n. 1, p. 77-85, 2009.

MAKHOURI, Farahnaz R.; GHASEMI, Jahan B. *In silico* studies in drug research against neurodegenerative diseases. **Current Neuropharmacology**, v. 16, n. 6, p. 664-725, 2018.

MAKRYNITSA, Garyfallia I. *et al.* *In silico* drug design. **Encyclopedia of Life Sciences**, p. 1-7, 2018.

MARTINS, Fábio Rodrigues. *et al.* Hardware architecture benchmarking for simulation of human immune system by multi-agent systems. **Multi-Agent Systems and Agreement Technologies**. p. 441-448, 2016.

MASSON, A. O., LOMBELLO, C. B., Metodologias de avaliação citotóxica: estudo comparativo segundo tempo de exposição., **COLAOB**, 2016

MATINELLO, F., SILVA, E. L. **Interferência do ácido ascórbico nas determinações de parâmetros bioquímicos séricos**: estudos *in vivo* e *in vitro*. J. Bras. Patol. Med. Lab. 39(4), 2003.

MENG, Xuan-Yu. *et al.* Molecular docking: a powerful approach for structure-based drug discovery. **Current Computer-aided Drug Design**, v. 7, n. 2, p. 146-157, 2011.

MESKÓ, Bertalan. **Virtual bodies for real drugs: *in silico* clinical trials are the future**. [S.l.: Medical Futurist Institute, 2019].

MITCHELL, R. S.; MICHALSKI, J. G.; CARBONELL, T. M. **An artificial intelligence approach**. Springer, Berlin, 2013.

MORALES, M. **Métodos alternativos à utilização de animais em pesquisa científica**: mito ou realidade? *Ciência e Cultura*, v. 60, n. 2, p. 33-36, 2008.

MULLAICHARAM, A. R.; AMARESH, Nirmala. Pharmaceutical applications of *in silico* methods: a review. **Research and Reviews: Journal of Pharmaceutical Analysis**. v. 3, p. 1-4, 2014.

PAIXÃO, Rita Leal. *et al.* **Experimentação animal: razões e emoções para uma ética**. 2001. Tese de Doutorado.

- PAUL GLEESON, M.; HERSEY, Anne; HANNONGBUA, Supa. In-silico ADME models: a general assessment of their utility in drug discovery applications. **Current Topics in Medicinal Chemistry**, v. 11, n. 4, p. 358-381, 2011.
- PITTA, João Luiz de Lemos Padilha. **Predição *in silico* de redes de interação proteica interespecie entre zika vírus e seus hospedeiros vertebrados e invertebrados**. 2018. Dissertação. (Mestrado em Biociências e Biotecnologia em Saúde) – Recife: Fundação Oswaldo Cruz, 2018.
- POSSI, M. A. *et al.* An in-silico immune system model for investigating human autoimmune diseases. In: XXXVII CONFERENCIA LATINOAMERICANA DE INFORMÁTICA, 2011, Quito. **Anais do XXXVII Conferencia Latinoamericana de Informática**. Quito: [s.n.], 2011, p. 1270-1283.
- RAUNIO, Hannu. *et al.* Modeling of interactions between xenobiotics and cytochrome P450 (CYP) enzymes. **Frontiers in Pharmacology**, v. 6, p. 123, 2015.
- RECKZIEGEL, Janaína; PEZZELLA, Maria Cristina Cereser. O papel da sociedade tecnocientífica e os riscos decorrentes das pesquisas médicas com seres humanos. **Revista de Direitos e Garantias Fundamentais**, v. 14, n. 2, p. 101-124, 2014.
- RIBEIRO JÚNIOR, Ademir Nunes. **Sistema imune artificial: modelagem computacional da infecção por plasmodium falciparum na malária transfusional**. 2019. Dissertação. (Mestrado em Ciências da Saúde) – Viçosa: Universidade Federal de Viçosa, 2019.
- RIFAIOGLU, Ahmet Sureyya. *et al.* Recent applications of deep learning and machine intelligence on *in silico* drug discovery: methods, tools and databases. **Briefings in Bioinformatics**, v. 20, n. 5, p. 1878-1912, 2019.
- RODRIGUES, Ricardo P. *et al.* Estratégias de triagem virtual no planejamento de fármacos. **Revista Virtual de Química**, v. 4, n. 6, p. 739-776, 2012.
- ROGERO, S. O., LUGÃO, A. B., IKEDA, T. I., & CRUZ, Á. S., Teste *in vitro* de citotoxicidade: estudo comparativo entre duas metodologias. **Materials Research**, 6(3), 317–320, 2003. doi:10.1590/s1516-14392003000300003.
- SAEIDNIA, Soodabeh; MANAYI, Azadeh; ABDOLLAHI, Mohammad. The pros and cons of the in-silico pharmaco-toxicology in drug discovery and development. **International Journal of Pharmacology**, v. 9, n. 3, p. 176-181, 2013.

SILVA, C. S. P.; MACIEL, E. M. G. S.; TELLES, F. S. P. Pesquisa Científica e a Saúde. In: Sergio Rego; Marisa Palácios. (Org.). **Comitês de Ética em Pesquisa: teoria e prática..** Rio de Janeiro: Editora Fiocruz, 2012, p. 23-51.

SILVA, Heide Muniz. **Método *in silico* para análise de sequências de imunoglobulinas produzidas por tecnologia de phage display.** 2016. Dissertação. (Mestrado em Biologia Molecular) – Brasília: Universidade de Brasília, 2016.

SOMASHEKHAR, S. P. *et al.* Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. **Annals of Oncology**, v. 29, n. 2, p. 418-423, 2018.

SOUZA, J., FREITAS, Z. M. F., STORPIRTIS, S., Modelos *in vitro* para determinação da absorção de fármacos e previsão da relação dissolução/absorção. **Revista Brasileira de Ciências Farmacêuticas** V. 43, N. 4, p 515-527, 2007.

STAMATAKOS, Georgios. *et al.* The technologically integrated oncosimulator: combining multiscale cancer modeling with information technology in the *in silico* oncology context. **IEEE Journal of Biomedical and Health Informatics**, v. 18, n. 3, p. 840-854, 2013.

TRAVASSOS, G. H., BARROS, M. O. Contributions of In Virtuo and *In Silico* Experiments for the Future of Empirical Studies in Software Engineering. **Proceedings of the WSESE03**, Fraunhofer IRB Verlag, Roma, 2003.

VETTORATO, Jordana Gabriele; MÜLLER, Nilvane Teresinha Ghellar; DA SILVA, Dejair Hartmann. Bioética: vida humana como objeto de experiência científica. **Revista Interdisciplinar de Ciência Aplicada**, v. 4, n. 7, p. 47-53, 2019.

YAMAGATA, A. T., GOMES JUNIOR, J. A., DUARTE, N. C. B., SILVA, I. C. R., Metodologias *in vitro* como alternativa ao uso de animais em avaliações toxicológicas de produtos cosméticos. **Acta de Ciências e Saúde** 3(2), 2014.

ZAHN, Evan M. **3-dimensional modeling in cardiovascular disease e-book.** [S.l.: Elsevier Health Sciences, 2019].

12.

NANOCOMPUTAÇÃO COM DNA: PRINCÍPIOS, APLICAÇÕES E BIOÉTICA

Poliana Aparecida Corrêa de Oliveira

Renan Albuquerque Marks

Omar Paranaíba Vilela Neto

1. INTRODUÇÃO

O advento de pesquisas em nanotecnologia e nanociência tem revolucionado a indústria mundial favorecendo o desenvolvimento de novos produtos de alta qualidade e baixo custo em diversas áreas do conhecimento. A **nanotecnologia** é definida por Vilela Neto e Pacheco (2012) como o “projeto, caracterização, produção e aplicação de estruturas, dispositivos e sistemas controlando a forma e o tamanho na escala nanométrica”, enquanto a **nanociência** é o “estudo dos fenômenos e manipulação de materiais em escalas atômicas, moleculares e macromoleculares, onde as propriedades diferem significativamente daqueles em maior escala”. Essas duas frentes constituem o alicerce de atividades multidisciplinares em física, química, biologia, matemática, tecnologia da informação e engenharia que são necessárias para o avanço científico e tecnológico.

Dentro deste escopo, a **nanocomputação** visa aplicar nanoestruturas no desenvolvimento de novos sistemas computacionais inovadores. Tais sistemas podem realizar lógica explorando outras variáveis de estado que não o fluxo de corrente elétrica ou variação de tensão, indo além da utilização das tecnologias de silício convencionais. Nesse sentido, uma tecnologia que pode ir diretamente de encontro às aplicações na área de medicina são as pesquisas relacionadas ao desenvolvimento da computação com DNA. O principal objetivo é criar sistemas moleculares para atuar diretamente no diagnóstico, monitoramento e tratamento de doenças. O DNA é uma macromolécula reconhecida pelo sistema biológico, que pode ser

sintetizada em laboratório (*in vitro*) para se comportar conforme uma programação prévia e, posteriormente, ser introduzida em ambiente celular (*in vivo*) para sensoriamento, controle e atuação de doenças no próprio meio (*in loco*).

Novas soluções e produtos oferecem inúmeras vantagens e possibilidades de aplicação em diferentes nichos, mas também trazem impactos sociais que podem ser vistos pela comunidade como pontos negativos. Se por um lado muitos problemas podem ser resolvidos com propostas inovadoras, por outro um número de alternativas levanta questionamentos éticos. Afinal, como a sociedade pode estabelecer uma fronteira entre as criações naturais e humanas? Por exemplo, o DNA pode ser sintetizado sem um rigoroso critério quanto a sua finalidade de uso? Essa resposta não é trivial, já que a linha entre esses assuntos é muito tênue e impulsiona amplas discussões.

No caso da biologia sintética, o aprimoramento da nanotecnologia trouxe a possibilidade de editar genomas. Com a técnica CRISPR-Cas9 é possível substituir um determinado segmento alvo no DNA por outra codificação, alterando assim o material genético de um indivíduo (JINEK *et al.*, 2012). Há diversos motivos para se querer alterar o gene de plantas, animais e seres humanos. Na medicina, isso pode ser feito para eliminar doenças genéticas hereditárias, tais como a hemofilia. Enquanto na indústria agrícola, modificações genéticas podem aumentar a resistência de uma planta a condições climáticas específicas, tais como seca prolongada, excesso de frio, ou melhor adaptação ao bioma.

Apesar dos benefícios alcançados com essa técnica, especialmente na agricultura, permitindo o aumento da produção de alimentos em larga escala, ainda há muitas críticas em relação à criação de seres modificados, que não são concebidos naturalmente (CRIBBS; PERERA, 2017). Isso ocorre porque o advento desta tecnologia também torna possível o uso de alterações genéticas com funções não terapêuticas, apenas por estética ou desejo pessoal. Um exemplo seria a opção de escolher a cor dos olhos arbitrariamente, deixando de ser um traço familiar passado de geração a geração, ou mais complexo ainda, a criação de bebês geneticamente programados.

A bioética é a disciplina que discute essa preocupação moral quanto às mudanças na vida humana. No sentido literal, a palavra bioética quer

dizer ética da vida. Ela envolve todas as áreas que podem afetar o bem estar humano, seja para o bem ou para o mal (CAMPBELL, 2017). Isso significa que é preciso ter legislações adequadas para ajudar a comunidade a distinguir o que é certo do que é errado. A bioética tem vasta observância no mundo globalizado, este, por sua vez, tem aumentado a sua complexidade e dificuldade para designar respostas para alguns dilemas que abalam a sociedade.

Uma gama de aplicações surgiram com o desenvolvimento da nanotecnologia, nanociência e nanocomputação, desde modelos computacionais de sistemas químicos e físicos até a manipulação de moléculas de DNA, alcançando áreas diversas como engenharias, tecnologia de informação, medicina e meio ambiente (VILELA NETO; PACHECO, 2012). Neste capítulo, nós apresentamos a computação com DNA, que é uma alternativa multidisciplinar promissora que envolve conhecimentos de química, biologia, computação e eletrônica com diversas possibilidades de uso no futuro, especialmente para a medicina.

2. COMPUTAÇÃO COM DNA

O êxito da nanociência e da nanotecnologia resultou na possibilidade de lidar com átomos e moléculas individualmente, explorando suas propriedades físico-químicas para construir novos materiais e dispositivos em escala nanométrica (TOMA, 2016). O DNA é um dos substratos que tem ganhado espaço neste segmento devido à sua natureza programável por meio do pareamento de bases nitrogenadas. Avanços na nanociência e nanotecnologia, aliados à invenção de ferramentas apropriadas para escala nano, contribuíram para uma queda no custo e aumento na qualidade de síntese dessas estruturas bioquímicas; um novo entendimento quantitativo da termodinâmica do DNA permite que fitas simples se dobrem e possam interagir com outros fragmentos, além de trazer um novo foco para sistemas livres de células a fim de evitar a interferência de outras unidades moleculares (CHEN *et al.*, 2015).

Desse modo, laboratórios especializados podem sintetizar e manusear facilmente moléculas de DNA. Conforme Benenson (2012), essa alternativa *in vitro* permite o projeto de fragmentos de DNA programados para reconhecer condições específicas (sinais de entrada), tomar decisões de

processamento de sinais e, se necessário, atuar no meio liberando outras substâncias (sinais de saída). Logo, esse sistema irá funcionar como um nanodispositivo computacional. É importante destacar que um computador de DNA não pode ser tão rápido quanto um convencional, entretanto esse dispositivo pode alcançar ambientes onde os chips de silício seriam considerados invasores estranhos ou que não poderiam atuar diretamente, como é o caso de sistemas biológicos, por exemplo, o corpo humano.

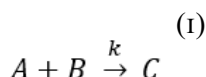
Computadores convencionais são essencialmente constituídos por circuitos lógicos combinacionais e sequenciais baseados em transistores de silício. Há mais de uma abordagem para construir circuitos biológicos sintéticos a partir de ácidos nucleicos. De acordo com Yordanov *et al.* (2014), *DNA strand displacement* (DSD), *DNA enzyme* e *RNA enzyme* são as estratégias mais promissoras. Para implementar o comportamento dinâmico de um circuito lógico, a primeira técnica se baseia na hibridização total ou parcial entre duas fitas de DNA que resulta no deslocamento de uma ou mais fitas; a segunda faz uso de enzimas associadas ao DNA (por exemplo, DNA polimerase e exonuclease) para implementar módulos elementares para ativação, autocatálise e inibição; e a terceira também usa enzimas, mas de RNA (tais como, RNA polimerase e ribonuclease) que atuam em conjunto com fitas duplas de DNA como promotores da sequência de transcrição.

Ainda segundo esses autores, cada estratégia apresenta vantagens e desvantagens. O que já se sabe é que na computação molecular, os circuitos bioquímicos para processamento e controle de sinais precisam ser simples, modulares e escaláveis. Neste capítulo, nós apresentamos o mecanismo por trás do projeto de circuitos moleculares baseados em reações de DSD. Essa abordagem permite a implementação usando somente moléculas de DNA, enquanto as outras técnicas demandam enzimas adicionais. Além da simplicidade, o comportamento dinâmico de sistemas DSD pode ser aproximado de reações químicas acopladas, o que facilita o projeto de circuitos mais complexos.

2.1 Representação abstrata dos sistemas moleculares em CRN

Uma reação química ocorre quando moléculas de uma ou mais substâncias colidem com energia suficiente para se quebrarem e formarem novas substâncias. A Equação (1) representa uma reação química arbitrária. Nesse

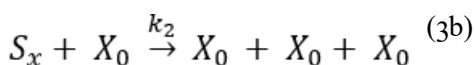
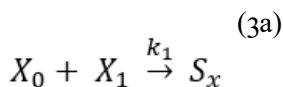
caso, A e B são duas espécies químicas (reagentes) que reagem entre si e produzem uma nova espécie molecular C (produto) em um processo irreversível (observe que a direção da seta está em apenas um sentido). Logo, ao final do processo não é possível que a espécie C seja transformada de volta em A e B . Uma reação química é um sistema dinâmico que leva um tempo até alcançar o estado de equilíbrio. Neste período, a concentração de todas as espécies na reação muda até que se torne constante.

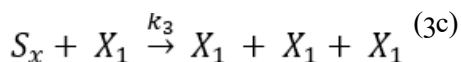


A cinética química visa estudar a velocidade e os mecanismos moleculares que levam os reagentes a serem consumidos a fim de gerar os produtos. A taxa de uma reação química reflete a mudança na concentração de uma determinada espécie por unidade de tempo. Para uma reação química bimolecular elementar, como a representada em (1), a lei da taxa --- ou equação de taxa ou equação de velocidade --- é dada por uma Equação (2) diferencial ordinária, onde os colchetes denotam as concentrações da espécie A ou B (quantidade em mol ou massa dividida pelo volume do sistema), a constante de taxa cinética é denotada por k e o sinal negativo indica que a concentração da espécie é consumida, isto é, sua concentração reduz ao longo do tempo (MARKS *et al.*, 2019).

$$\frac{d[A]}{dt} = \frac{d[B]}{dt} = -k[A][B] \quad (2)$$

Redes de reações químicas (CRN, do inglês *Chemical Reaction Network*) acopladas podem ser usadas para descrever em alto nível de abstração os circuitos moleculares. Por exemplo, a CRN na Equação (3) é composta por três reações químicas acopladas (JIANG *et al.*, 2013). As reações químicas em (3b) e (3c) são autocatalíticas, pois cada uma delas produz três cópias das espécies de reagentes X_0 e X_1 , respectivamente.





A Equação (4) mostra o modelo matemático que descreve a cinética, ou seja, o comportamento dinâmico de todas as espécies da CRN apresentada na Equação (3). A equação (4a), por exemplo, descreve que a concentração da espécie S_x aumenta com a reação descrita pela equação (3a) e diminui com as reações descritas pelas equações (3b) e (3c). Essa CRN descreve o comportamento de um sistema no qual as espécies X_0 e X_1 representam um bit biestável de informação. Dessa forma, a espécie que possui a maior concentração inicial evolui para o valor estacionário igual a 1, enquanto a outra cai para zero. Com base na modelagem e no entendimento dinâmico do sistema a partir de CRN acopladas genéricas, Soloveichik; Seelig; Winfree (2010) propuseram um dos primeiros modelos para implementar a conversão de CRNs em sistemas DSD usando o DNA como substrato.

$$\frac{d[S_x]}{dt} = k_1[X_0][X_1] - k_2[S_x][X_0] - k_3[S_x][X_1] \quad (4a)$$

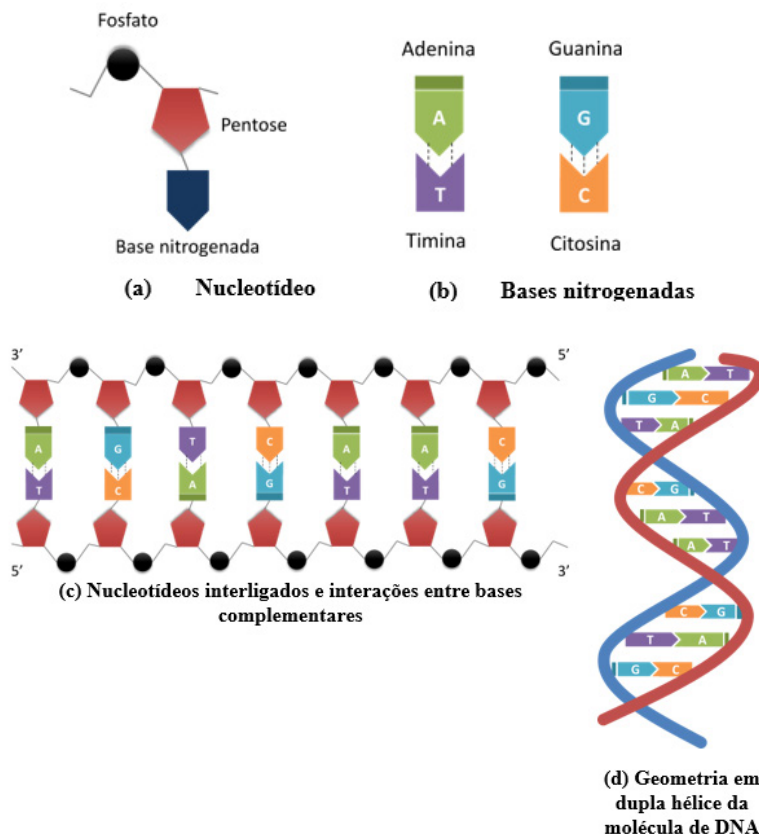
$$\frac{d[X_0]}{dt} = -k_1[X_0][X_1] + 2k_2[S_x][X_0] \quad (4b)$$

$$\frac{d[X_1]}{dt} = -k_1[X_0][X_1] + 2k_3[S_x][X_1] \quad (4c)$$

2.2 Estrutura da molécula de DNA

O ácido desoxirribonucleico (DNA) é uma macromolécula constituída por uma longa cadeia de nucleotídeos interligados. Cada nucleotídeo é composto por três componentes básicos: (1) um grupamento fosfato, (2) uma pentose (um açúcar formado por cinco carbonos que no caso do DNA é a desoxirribose) e (3) uma base nitrogenada (Figura 1a). A base de um nucleotídeo é ligada covalentemente através de uma ligação N-glicosídica com a hidroxila ligada ao carbono 1' da pentose. O grupo fosfato é esterificado no carbono 5' da pentose. O DNA contém duas bases purinas, adenina (A) e guanina (G), e duas pirimidinas, timina (T) e citosina (C) (Figura 1b). As bases purinas apresentam dois anéis de carbono e nitrogênio, enquanto as pirimidinas são formadas por apenas um anel (NELSON; COX, 2011).

Figura 1 – Composição química da molécula de DNA



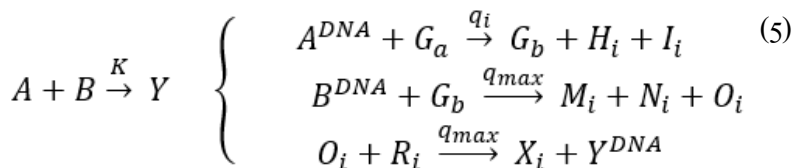
Fonte: Elaborada pelos autores.

Os nucleotídeos consecutivos são interligados covalentemente por uma ligação fosfodiéster entre um grupo 5'-fosfato de um nucleotídeo ao grupo 3'-hidroxil do próximo nucleotídeo formando uma fita do ácido nucleico (Figura 1c). Ligações de hidrogênio permitem a interação entre bases nitrogenadas complementares. O modelo mais comum, Watson-Crick, foi definido em 1953 por James D. Watson e Francis Crick e estabelece que A liga-se com T por ligação dupla e G liga-se com C por tripla ligação. Esse pareamento de bases nitrogenadas permite a formação de uma hélice dupla e, consequentemente, a duplicação da informação genética (Figura 1d). A estrutura do DNA observada por Watson e Crick é estável e tridimensional sendo formada por duas fitas complementares de nucleotídeos (NELSON; COX, 2011).

2.3 Construindo nanodispositivos com DNA

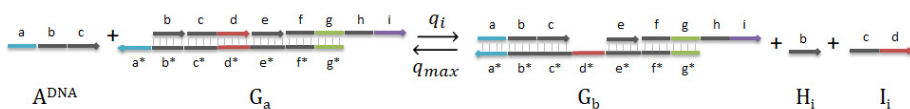
A estrutura estável da molécula de DNA garante a programabilidade necessária do material para construir nanodispositivos moleculares. Fragmentos de DNA podem ser projetados de maneira que as interações entre bases nitrogenadas de suas fitas executem funções pré-definidas. Um dos mecanismos mais conhecido para fazer computação com DNA é o de reações de deslocamento de fitas de DNA (DSD, do inglês *DNA Strand Displacement*). Nessa técnica, fragmentos de DNA sintetizados em laboratório em fitas duplas (*gate*) são projetados para desencadear reações químicas quando colidirem com determinados segmentos de fitas simples de DNA (*input*). Essas fitas de *input* irão se ligar ao *gate* de modo a provocar a liberação de um outro segmento de DNA também em fita simples (*output*). Este, por sua vez, pode ser acoplado a uma substância fluorescente (*reporter*) para facilitar a visualização da resposta do circuito molecular (SEELIG *et al.*, 2006; ZHANG *et al.*, 2007).

CRNs acopladas podem ser usadas para descrever um circuito molecular em alto nível de abstração por meio de reações químicas genéricas que definem o seu comportamento. Por exemplo, a função lógica AND pode ser expressa por uma reação química bimolecular arbitrária, conforme a representada no lado esquerdo da chave mostrada na Equação (5). Neste caso, o produto Y somente é produzido na presença de ambos os reagentes A e B . Essa mesma dinâmica química pode ser convertida para reações DSD, usando o método proposto por Soloveichik, Seelig & Winfree (2010). O lado direito da chave da mesma equação indica a conversão CRN-DSD, onde A^{DNA} e B^{DNA} representam as entradas do sistema como duas fitas simples de DNA distintas (*inputs*), G_a e G_b são fitas dupla de DNA (*gates*) projetadas para iniciar as reações quando detectarem as entradas, R_i também é uma fita dupla de DNA que atua como sinalizador (*reporter*) que configura a fita a ser liberada pela saída, e por fim, Y^{DNA} é a saída do circuito (*output*) também definida como uma fita simples de DNA. As demais espécies H_i , I_i , M_i , N_i , O_i e X_i são intermediárias geradas ao longo do processo químico e que ao final podem ser descartadas (*wastes*) se não forem completamente consumidas.



Ainda segundo esses autores, para realmente aproximar a dinâmica do sistema DSD de uma reação química genérica, algumas restrições sobre as constantes de taxa (q_i e q_{max}) e concentração de espécies auxiliares (G_a e R_i) precisam ser atendidas. Esse método de conversão CRN-DSD também é usado para transcrever outros tipos de reações químicas elementares. Após a tradução, as espécies são sintetizadas *in vitro* em pequenos fragmentos de DNA, cada um deles é composto por um conjunto de domínios. Um domínio representa uma sequência de nucleotídeos interligados, que normalmente é identificado por letras e/ou números. A Figura 2 mostra um exemplo de reação de DSD em que a representação da fita de entrada A^{DNA} é composta por três domínios, denominados a , b e c . As reações de DSD são mediadas por um domínio curto, chamado *toehold*, que geralmente é indicado em cor diferente dos demais. Na fita A^{DNA} , por exemplo, é o domínio a .

Figura 2 – Reação de deslocamento de fitas de DNA



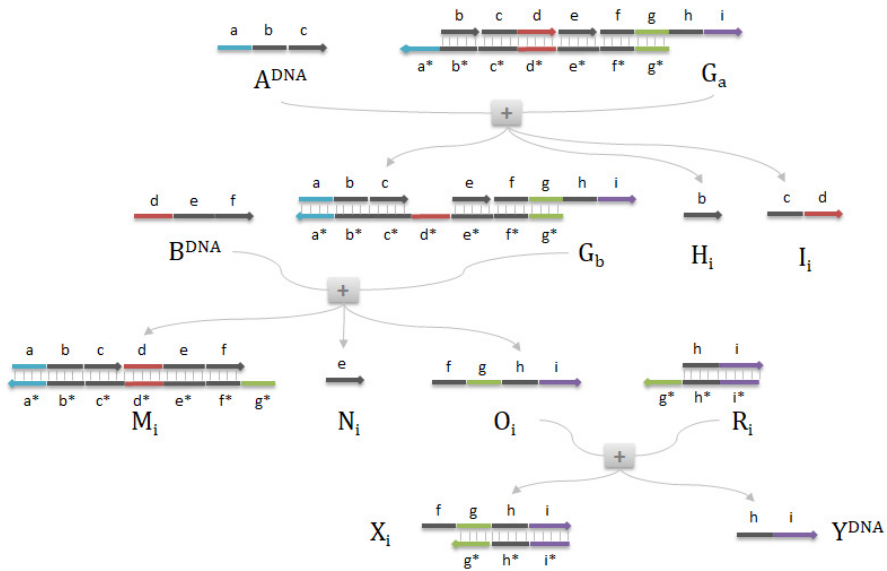
Fonte: Elaborada pelos autores.

O processo químico começa quando uma fita simples de entrada (A^{DNA}), por meio do ponto de *toehold* (a), se liga ao seu domínio complementar (a^*) disponível em um complexo de fita dupla (G_a). Em seguida, os demais domínios de ambas fitas competem entre si até que uma das fitas seja completamente hibridizada e a outra deslocada. Observe no lado direito do exemplo (Figura 2), que G_b é a espécie formada após o pareamento completo da fita A^{DNA} e a consequente liberação das fitas H_i e I_i . Essa reação é reversível, logo também é possível que a fita I_i se ligue de volta a G_b com o pareamento do *toehold* (d e d^*), o que pode provocar a liberação da fita A^{DNA} . A cinética da reação DSD é modelada de acordo com as taxas da reação no sentido direto e reverso, sendo assim, aquela que for maior definirá a dinâmica do sistema químico (SOLOVEICHIK, SEELIG & WINFREE, 2010).

A Figura 3 ilustra a cadeia de reações de DSD para implementar a função AND apresentada na Equação (5). Note que, se houver a presença de fitas de entrada A^{DNA} (ON), elas irão se ligar a G_a ocasionando a formação do *gate* G_b . Esse complexo, quando em contato com fitas de entrada B^{DNA} (ON), irá

provocar a liberação da fita O_i . Esta, por sua vez, se liga ao *reporter* R_i provocando a liberação da fita de saída Y^{DNA} . Entretanto, na ausência de ambas fitas de entrada, A^{DNA} (OFF) e B^{DNA} (OFF), o processo químico não é iniciado. Se houver apenas entrada A, fitas de entrada A^{DNA} (ON) e B^{DNA} (OFF), o *gate* G_b é formado, porém o processo é interrompido nesse ponto e não há liberação da saída. Da mesma forma, no caso de existir apenas entrada B, A^{DNA} (OFF) e B^{DNA} (ON), como não há formação do *gate* G_b , então o processo não prossegue, e por esse motivo, não há liberação do sinal de saída B^{DNA} .

Figura 3 – Cadeia de reações DSD para função lógica AND



Fonte: Elaborada pelos autores.

Um dos problemas do mecanismo DSD são os vazamentos (*leaks*). Estas são reações indesejáveis que ocorrem paralelamente à cadeia de reações DSD planejadas, seja entre espécies pré-existentes ou entre produtos intermediários ou gerados a partir de uma outra reação de vazamento. Esse último caso costuma gerar um efeito maior de propagação de erros, alterando a saída do circuito (WANG *et al.*, 2018; OLIVEIRA *et al.*, 2020). Geralmente, os vazamentos não são controlados pela existência de conexões entre *toeholds*; pelo contrário, eles são mais observados quando são

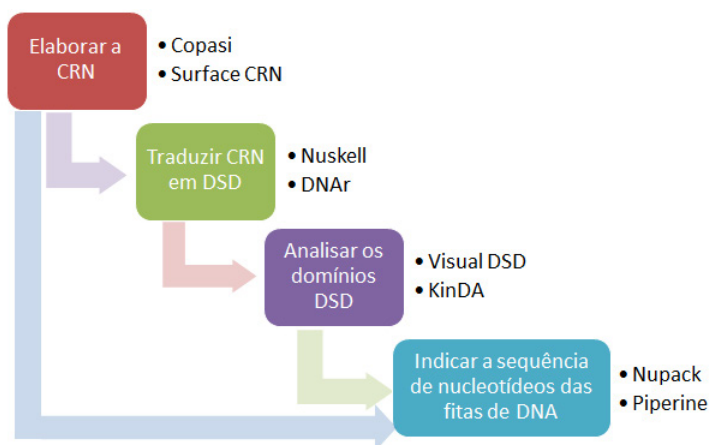
disparadas ligações entre domínios longos. Portanto, as reações de vazamento são muito difíceis de prever e controlar.

Outra característica importante a ser mencionada é que, diferentemente dos circuitos lógicos de silício convencional, os circuitos lógicos de DNA são destruídos após serem usados. Isso gera uma necessidade de replanejamento para restaurar portas lógicas para avaliar conjuntos de entradas seguintes.

2.4 Ferramentas de simulação

O processo de projeto, teste e verificação de viabilidade para implementação experimental de circuitos com DNA seria extremamente complexo, caro e inviável se tivesse de ser feito sempre em laboratório. Sendo assim, há na literatura científica diversas ferramentas de simulação que estão sendo desenvolvidas e aprimoradas continuamente, a maioria delas é gratuita e de código aberto. Esse tipo de projeto envolve quatro fases principais (Figura 4), cada uma delas com foco em um determinado nível de abstração: (1) nível mais alto é o que visa especificar as redes de reações químicas (CRN) genéricas acopladas; (2) etapa intermediária com o intuito de traduzir uma CRN em um sistema DSD; (3) estágio que tem o propósito de analisar a interação entre domínios bem como as reações do sistema DSD e de vazamentos (*leaks*); (4) nível mais baixo a fim de indicar quais são as sequências de nucleotídeos apropriadas para cada fita de DNA candidata a implementação.

Figura 4 – Processo e ferramentas para o projeto de circuitos com DNA



Fonte: Elaborada pelos autores.

As ferramentas Copasi e Surface CRN foram desenvolvidas especificamente para simular e visualizar o comportamento de um sistema químico a partir da descrição de CRNs genéricas acopladas (HOOPS *et al.*, 2006; CLAMONS, QIAN & WINFREE, 2020). Os pacotes Nuskell e DNAr, desenvolvidos respectivamente em Python e R, oferecerem o recurso para a simulação de CRNs e também realizam a tradução de CRN genéricas para esquemas de DSD (BADELT *et al.*, 2017; VIEIRA *et al.*, 2020). Ambos apresentam a funcionalidade de exportação para o software Visual DSD, o que facilita a integração entre essas duas fases do projeto. O Visual DSD e o KinDA são softwares que simulam e analisam o comportamento cinético em nível de domínio das reações de DSD que ocorrem entre as fitas projetadas de DNA, incluindo as reações de vazamento do sistema (LAKIN *et al.*, 2011; BERLEANT, J. *et al.*, 2018). Por fim, o Nupack e o Piperine são ferramentas que alcançam o maior nível de detalhamento do projeto em que são selecionadas e simuladas as sequências de nucleotídeos que serão atribuídas a cada domínio a fim de especificar exatamente como devem ser as fitas de DNA candidatas para a implementação experimental em laboratório. Piperine é uma ferramenta automática e mais completa que as demais para o projeto de circuitos com DNA já que oferece suporte desde a especificação de CRN até a definição dos nucleotídeos das moléculas de DNA (ZADEH *et al.*, 2011; SRINIVAS *et al.*, 2017).

3. PROJETO DE CIRCUITOS COM DNA

3.1 Abordagem digital

Inicialmente, os projetistas de circuitos moleculares se inspiraram na lógica booleana para criar sistemas de computação, assim como é feito em computadores tradicionais. Portas básicas simples, como AND, OR, NAND, NOR, XOR e Majority já foram demonstradas na literatura científica (JIANG, *et al.* 2013; MARKS *et al.*, 2019). Muitos desses projetos usam o padrão *dual-rail* (uma espécie para o bit 0 e outra para o bit 1) para implementar a lógica NOT, embora isso acarrete um aumento no número de espécies reagindo no sistema, o que pode se tornar um obstáculo para iniciativas práticas. Além disso, a abordagem digital demanda a conexão das portas lógicas em cascata e isso é um desafio para implementação em circuitos moleculares devido a atenuação do sinal. Em sistemas

bioquímicos, o sinal é representado pela concentração de uma determinada espécie, sendo assim, é preciso mantê-los durante um intervalo de tempo para se alcançar saídas de circuitos de DNA em níveis corretos e estáveis.

Jiang *et al.* (2013) e Boruah *et al.* (2015) apresentam propostas de implementação de circuitos de DNA baseados na abordagem digital. A proposta de Jiang *et al.* (2013) é utilizar a codificação de bits utilizando dois sinais A e B complementares: quando A está presente, B está ausente e vice-versa. Neste caso, o sinal A pode representar o valor binário ZERO e o sinal B representar o valor binário UM, por exemplo. Através do uso de DSD e de uma rede de reações químicas autocatalíticas, é possível fazer ambos os sinais se auto regularem na solução e sempre que houver uma dominância de um sinal, o outro complementar é eliminado da solução. Essa robustez é capaz até mesmo de corrigir pequenas perturbações nos sinais, que são naturalmente existentes ao longo da propagação da computação nos circuitos em execução. Já a proposta de Boruah *et al.* (2015) foca em codificar no DNA sequências de bases nitrogenadas para as entradas e portas lógicas do circuito. Utilizando enzimas para auxiliar o processo de computação e deslocamento das fitas de DNA é possível identificar ao final a presença (ou ausência) de determinadas estruturas de fitas de DNA que simbolizam o resultado final da computação das portas lógicas.

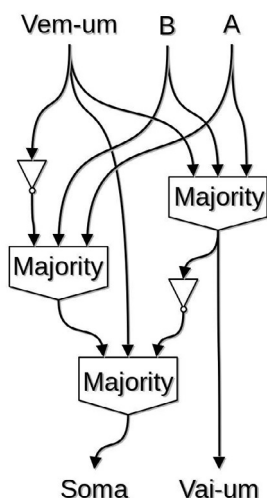
A proposta mais robusta de Jiang *et al.* (2013) foi utilizada em Marks *et al.* (2019) para a construção de uma ferramenta de projeto e simulação de circuitos lógicos digitais utilizando DNA. Neste trabalho, foram implementados alguns circuitos maiores, entre eles, um somador de números binários. Uma aplicação prática de um somador é em um contador de eventos discretos. O somador foi construído utilizando portas lógicas majoritárias (*majority gates*) (Figura 5). Essas portas funcionam de modo simples: sua saída contém o valor UM somente quando a maioria de suas entradas possui valor UM; caso contrário, sua saída terá valor ZERO. Cada porta lógica majoritária é composta de uma rede de reações químicas contendo 19 reações e 9 espécies. Um somador completo é composto de três portas lógicas majoritárias, portanto uma rede de reações químicas que contém 57 reações e 27 espécies.

Devido à natureza de funcionamento da técnica DSD, em simulações computacionais utilizando os parâmetros químicos propostos por Soloveichik *et al.* (2010) feitas no software de simulação DNAr (VIEIRA

et al., 2020), o sinal de resposta do somador binário proposto leva de duas horas e meia a quatro horas para estabilizar com a resposta final. Nota-se que a velocidade de execução é muito baixa quando comparada à velocidade de uma soma em um computador convencional. Entretanto, a possibilidade de ser usado em ambientes *in vitro* e, principalmente, *in vivo* é extremamente atraente.

Figura 5 – Exemplo de circuito de um somador binário utilizando *majority gates*;

Os elementos triangulares com um círculo na ponta são inversores, ou seja, se recebem “0” na entrada, produzem “1” na saída e vice-versa; As entradas “A”, “B” e “Vem-um” são valores de um bit de informação (“0” ou “1”), assim como as saídas “Soma” e “Vai-um”. Por exemplo: se $A = 1$, $B = 0$ e $Vem-um = 1$, significa que o circuito realizará a soma “ $A + B + Vem-um$ ” que é “ $1 + 0 + 1$ ” e teremos os valores Soma = 1 e Vai-um = 0 nas saídas do circuito.



Fonte: Elaborada pelos autores.

Por ser um somador binário --- isto é, ser capaz de somar valores de somente um bit de informação --- sua estrutura de 57 reações e 27 espécies é muito custosa em termos de espécies de DNA e complexidade das reações químicas. Para a soma de valores contendo n bits de informação, o circuito final teria um quantitativo n vezes maior de reações e espécies. Isso aumentaria ainda mais o custo e complexidade do circuito. Esse incremento

significativo do número de reações e espécies necessárias tornam a implementação de circuitos digitais em DSD um tanto quanto impeditiva.

3.2 Abordagem analógica

A abordagem analógica tem potencial para tornar mais factível a implementação de circuitos moleculares com DSD por ser mais adequada ao ambiente biológico, uma vez que as concentrações das espécies químicas dentro das células variam continuamente no tempo. Os sistemas analógicos realizam cálculos (e tomam decisões) por meio de informações que são representadas com valores contínuos em vez de valores discretos, como é o caso da representação binária aplicada na abordagem digital. Apesar dos benefícios dessa abordagem, os sistemas analógicos são menos robustos do que seus correspondentes digitais, devido à vulnerabilidade ao ruído. Diante disso, abordagens híbridas analógico-digitais com interfaces eficientes podem se tornar uma alternativa no futuro (SARPESHKAR, 2014).

Em circuitos moleculares analógicos a computação é aproximada e, por isso, o resultado é estimado sendo esperado um certo nível de tolerância a erros. Entretanto, esse grau de imprecisão pode ser aceitável em troca da simplicidade e escalabilidade de implementação desses circuitos se comparado às alternativas baseadas na lógica booleana (OLIVEIRA *et al.*, 2020). Song *et al.* (2016) propuseram portas analógicas com duas entradas para realizar operações aritméticas de adição, subtração e multiplicação. Esses circuitos não suportam valores arbitrários de entrada, sendo necessário estabelecer uma faixa contínua de valores, de zero até um valor máximo estabelecido. Os autores consideram aceitáveis erros na saída de até 5% do valor esperado.

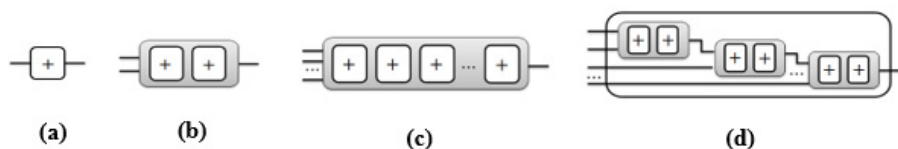
O somador analógico com duas entradas consiste de dois blocos principais (Figura 6a e 6b), sendo um para cada entrada. Desse modo, cada espécie de entrada (Ia1 e Ia2) reage com sua respectiva espécie *gate* (Ga1 e Ga2) de fita dupla, formando espécies intermediárias que, por sua vez, reagem com uma espécie individual de combustão (Fa) que irá produzir um único sinal de saída (Oa). Observe que o funcionamento é simples e intuitivo. A ideia é consumir as entradas a fim de produzir uma só saída na qual tenha concentração igual ao somatório de todas as entradas $Oa = Ia1 + Ia2$. Para construir circuitos maiores (com mais entradas), o segredo é replicar

um bloco principal para cada entrada adicional e ajustar a concentração inicial da espécie de combustão. O somador analógico com duas entradas é composto por apenas 7 espécies iniciais que desencadeiam 6 reações químicas em um circuito ideal e 24 reações químicas quando habilitamos as reações de vazamentos na ferramenta de simulação Visual DSD (LAKIN *et al.*, 2011; SONG *et al.*, 2016; OLIVEIRA *et al.*, 2020).

Zou *et al.* (2017) apresentaram algumas melhorias para as portas aritméticas analógicas de Song *et al.* (2016). Uma delas foi a implementação da lógica *dual-rail* a fim de permitir operações entre números inteiros. Para tanto, duas espécies distintas são usadas para representar uma única entrada e/ou saída, de forma complementar a concentração de uma delas define o valor positivo e a outra o negativo. Esta modificação aumentou consideravelmente o tamanho das portas aritméticas. Zou *et al.* (2017) também propuseram um aperfeiçoamento para o amplificador da porta de multiplicação a fim de reduzir o impacto das reações de vazamento. Além disso, eles projetaram uma nova porta para realizar a operação de divisão, também com abordagem analógica.

Assim como Song *et al.* (2016), esses autores apresentaram apenas simulações mostrando o funcionamento correto dentro da margem de erro aceitável para as portas aritméticas propostas. Porém, eles não analisaram os parâmetros que limitam as reações de DSD (por exemplo, taxa de ligação do toehold, faixa de intervalo das entradas, reações de vazamentos, níveis de concentração das entradas, entre outros). Em Oliveira *et al.* (2020), os autores analisaram as portas aritméticas de adição e subtração mencionadas em Song *et al.* (2016). Neste trabalho, as portas de soma e subtração foram estendidas e implementadas em versões maiores com três, quatro e cinco entradas. Além disso, duas estratégias de projeto de circuitos analógicos moleculares foram comparadas, a saber módulo-único (Figura 6c), já apresentado em Song *et al.* (2016) para o somador com duas entradas, e cascata (figura 6d) proposta por esses autores para explorar a modularidade e de certa forma controlar as reações químicas. A avaliação das estratégias de projeto e dos módulos estendidos foi em termos de tamanho, escalabilidade e precisão, levando em conta as variações dos principais parâmetros, o que é relevante para compreender completamente os sistemas DSD propostos.

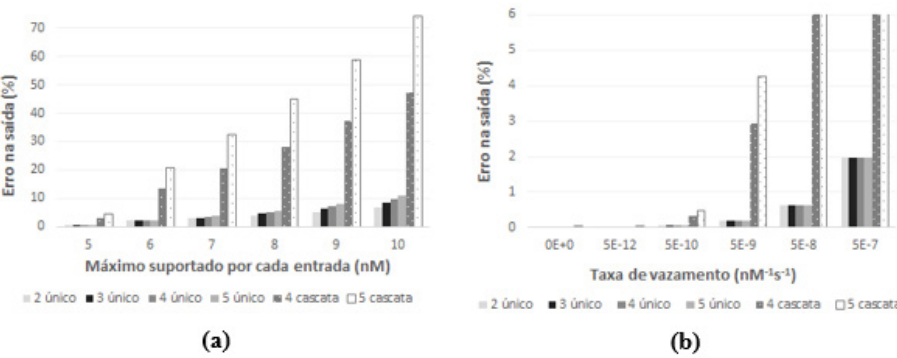
Figura 6 – Portas de soma analógica. (a) Somador de 1 entrada. (b) Somador com 2 entradas. (c) Somador de módulo-único com N entradas. (d) Somador em cascata com N módulos de somador de 2 entradas.



Fonte: Elaborada pelos autores.

Ainda segundo Oliveira *et al.* (2020), os resultados das simulações na ferramenta Visual DSD mostraram que as portas aritméticas analógicas devem ser específicas para cada aplicação, visto que sua capacidade de processamento (que é fixa e preestabelecida pelo intervalo de valores suportado por cada entrada) reduz a precisão e a escalabilidade. A Figura 7a mostra o erro percentual na saída em função do valor máximo suportado por cada entrada para diferentes tipos e tamanhos de portas somadoras. As espécies de entrada foram configuradas igualmente com concentração de 5 nM para todos os experimentos. Note que de acordo com as simulações, quanto maior é a capacidade máxima suportada por uma porta para suas entradas (isto é, mais distante dos 5nM fornecidos), maior é o erro da operação soma detectado na saída. Essas portas são susceptíveis a ruídos, e por isso, já é esperado um pequeno erro no valor de saída esperado, sendo a computação aproximada. Apesar dos circuitos analógicos em DSD serem mais simples e possuírem um número menor de reações e espécies necessárias para serem implementados, eles possuem algumas limitações. As reações de vazamento comprometem a precisão, especialmente para as versões que fazem cascatas de vários módulos. Como mencionado anteriormente, os vazamentos são reações indesejáveis e imprevisíveis que acontecem em paralelo com as outras reações esperadas do circuito. A Figura 7b ressalta que as portas construídas por um único módulo apresentam maior resistência a vazamentos, enquanto as em cascata se tornam incontrolláveis quando a taxa é superior a ordem de grandeza de 10^{-9} . Isso enfatiza que o uso de módulos em cascata é inviável quando há altas taxas de vazamento.

Figura 7 – Resultados da avaliação das portas somadoras analógicas estendidas. (a) Erro na saída para várias concentrações máxima suportada por cada entrada em implementações de módulo único e em cascata. (b) Erro na saída para várias taxas de vazamento em implementações de módulo único e em cascata.



Fonte: Elaborada pelos autores.

3.3 Aprendizado de máquina

O aprendizado de máquina envolve computadores descobrindo como podem executar determinadas tarefas sem serem explicitamente programados para isso. Isto é, através do fornecimento de dados e exemplos, os computadores são capazes de identificar padrões existentes nestes dados e, com isso, classificá-los e tomar decisões com alto grau de confiança. Muitas vezes, para resolver tarefas simples, um humano é capaz de dizer ao computador quais são todos os passos que devem ser feitos para resolver um problema --- ou seja, programar o computador e ele simplesmente seguir esses passos, o que não exige nenhum “aprendizado” por parte do computador. Porém, isso torna-se desafiador quando um humano precisa dizer a um computador como resolver tarefas muito mais complexas --- tal qual reconhecer um determinado objeto em uma imagem ou vídeo. Dessa forma, é muito mais prático e eficiente auxiliar o próprio computador a desenvolver seu próprio algoritmo, em vez de um humano descrever todos os passos. Assim, o campo de pesquisa do aprendizado de máquina estuda e desenvolve diversas abordagens com as quais um computador é capaz de dar respostas eficientes a problemas quando não existem algoritmos satisfatórios disponíveis (ALPAYDIN, 2020).

Tradicionalmente, o aprendizado de máquina é subdividido em três categorias: aprendizado supervisionado, aprendizado não-supervisionado e aprendizado por reforço.

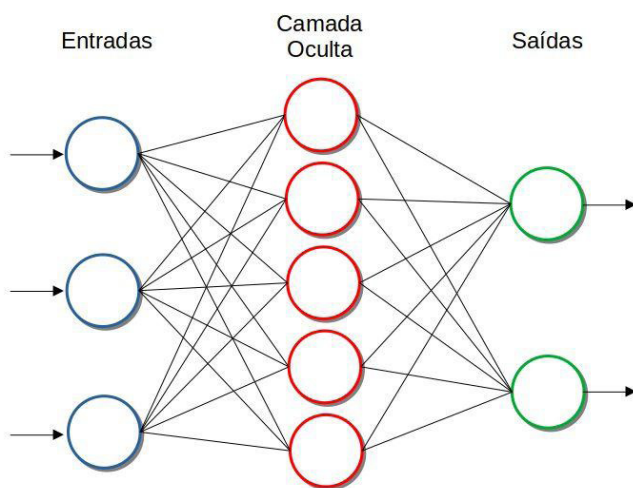
- No aprendizado supervisionado um humano fornece ao computador um conjunto de dados de “treino” --- que são as entradas com suas respectivas saídas desejadas --- de forma que o computador seja capaz de identificar as características das entradas e relacioná-las às saídas desejadas. Por exemplo: um conjunto de fotos de diversas letras escritas por pessoas com diferentes caligrafias e cada foto com a informação da letra real que foi escrita. Com esse conjunto de treino, o computador passa a ser capaz de identificar corretamente, por exemplo, letras que possuem pequenas variações em formato.
- No aprendizado não-supervisionado um humano fornece ao computador um conjunto de dados sem quaisquer informações vinculadas. Dessa forma, o computador deve analisar os dados e identificar similaridades e diferenças nos dados fornecidos, encontrando neles uma estrutura própria implícita.
- No aprendizado por reforço, o computador interage com um ambiente dinâmico de forma a atingir um determinado objetivo (como ganhar um jogo ou dirigir um automóvel sem cometer infrações de trânsito). Além disso, o computador recebe de volta do ambiente um *feedback* pela ação tomada, que será interpretada como “recompensa” ou “punição”. Dependendo de como foi configurado, o computador, nas próximas interações com esse ambiente, tentará sempre maximizar essas “recompensas” e minimizar as “punições”.

Uma das abordagens mais populares de implementação de aprendizado de máquina é a utilização de redes neurais artificiais. Elas são inspiradas nas redes neurais biológicas e sua unidade mínima é o neurônio artificial. Esse neurônio é modelado como uma unidade que recebe um sinal de entrada, é capaz de processá-lo e produzir um sinal na saída que será propagado para a(s) entrada(s) de outro(s) neurônio(s) conectados à ele na rede. No contexto de uma rede neural, os neurônios e suas conexões são modelados como grafos e chamados de “nós” e “arestas”, respectivamente. A Figura 8 mostra um exemplo de uma rede neural artificial. O vínculo de pesos nas conexões entre os neurônios artificiais é outra característica dessa abordagem. Os pesos servem para, durante o processo de “treinamento da rede”,

ajustar o comportamento dela de forma que a saída da rede corresponda a saída esperada a cada entrada de treino fornecida.

Neurônios artificiais (e redes neurais artificiais) são modelagens teóricas que podem ser implementados tanto como *software* (que será executado em computadores) quanto como *hardware* (que processará nativamente os sinais de entrada gerando sinais de saída). Assim não tardou para que comesçassem a aparecer implementações tanto de neurônios artificiais quanto de outras abordagens de aprendizado de máquina capazes de executar em outros substratos além dos dispositivos eletrônicos.

Figura 8 - Uma rede neural artificial é composta de uma rede de nós interligados; Cada ponto circular representa um neurônio artificial e cada aresta representa uma conexão de um neurônio com outro.



Fonte: Elaborada pelos autores.

No trabalho de Vieira *et al.* (2020) houve a implementação e simulação em DSD de um neurônio de McCulloch–Pitts descrito em CRN (HJELMFELT *et al.*, 1991). Além disso, existem outras abordagens para a construção de neurônios artificiais que não utilizam CRN em sua descrição, como os trabalhos de Qian *et al.* (2011) e Qian & Winfree (2011) que os especificam diretamente em espécies de DNA. Além do aprendizado usando redes neurais, também é possível usar outras abordagens para

se realizar aprendizado de máquina. É possível realizar tanto aprendizado evolutivo Lee *et al.* (2011) quanto aprendizado supervisionado Lakin & Stefanovic (2016) e com *Support Vector Machines* Noh *et al.* (2015) utilizando DNA.

4. APLICAÇÕES

Uma das motivações da nanotecnologia, nanociência e nanocomputação é o alto potencial para alcançar uma gama de aplicações, especialmente aquelas que são empregadas em ambientes biológicos, tais como no uso de terapias inteligentes visando a entrega de drogas *in loco* para tratamento de doenças, dentre outros nanodispositivos que possam interagir ou atuar livremente dentro de células vivas. Essas aplicações apontam para sistemas formados por nanoestruturas de ácido nucléico, os quais se destacam pelo seu tamanho pequeno, biocompatibilidade e maneira direta em que eles podem ser programados para interagir naturalmente com as estruturas bioquímicas que são comuns nesse meio celular (VILELA NETO; PACHECO, 2012).

A computação com DNA pode ser usada no desenvolvimento de inúmeros tipos de aplicações. O mais comum é na indústria médica e farmacêutica visando o diagnóstico, controle e tratamento de doenças. Por exemplo, no caso de um indivíduo com diabetes, um nanodispositivo poderia ser programado para detectar o aumento no nível de açúcar no sangue e, em seguida, atuar liberando fármacos até que o nível seja normalizado, tudo isso de forma automática sem a interferência do paciente (HARROUN *et al.*, 2018). Outra aplicação interessante é na área de Internet das “Bio-Nano-Things” que poderia se dar pelo emprego de nanodispositivos programáveis interligados a fim de enviar para uma central informações sobre o nível de hormônios, proteínas, vitaminas e etc. de um organismo, ou até mesmo externamente para controle de poluentes e qualidade de produtos perecíveis (AKYILDIZ *et al.*, 2015). A comparação de genomas com o propósito de correlacionar doenças também desperta o interesse (MISHRA *et al.*, 2017). Nesse caso, DNA Chips podem ser implantados em organismos para auxiliar nesse processo. Outras possibilidades incluem a programação genética, criptografia de sistemas biológicos e decifração de mensagens codificadas, além da impressão digital de DNA.

5. DISCUSSÃO

Obviamente o uso do DNA como um substrato para o desenvolvimento lógico de sistemas computacionais e suas principais aplicações associadas levantam discussões éticas imediatas. O DNA é o substrato natural que codifica a vida. Usá-lo para outros objetivos seria uma ação antiética e violaria os princípios fundamentais da natureza? Essa não é uma resposta fácil de ser respondida, mas nosso objetivo aqui é fornecer o ponto de vista dos autores e o que acreditamos como cientistas de computação. Nosso papel é pesquisar alternativas para atender demandas que hoje não são contempladas pelos sistemas computacionais existentes. Isso não impede que o leitor tenha outras opiniões, o que na verdade é saudável e amadurece o debate ético ao redor do assunto.

Conforme apresentamos neste capítulo, a molécula de DNA é formada por elementos químicos que ao serem arrançados de forma específica possuem uma estrutura estável, obtida pela interligação das bases nitrogenadas. Neste ponto, é importante classificar as moléculas de DNA em dois tipos: o DNA natural e o DNA sintético. Em nosso ponto de vista, o DNA natural não deve nunca ser usado como substrato para qualquer função que não aquela da própria natureza. Por outro lado, há um espaço gigantesco para a produção de moléculas de DNA sintéticas a fim de construir sistemas programáveis aptos a tomar decisões e executar comportamentos pré-definidos. As moléculas sintéticas podem interagir com o DNA natural de diferentes espécies, possibilitando monitorar e atuar *in loco*, o que leva a vários tipos de aplicações específicas e justificáveis. Cabe à sociedade estabelecer condições para balancear/classificar os benefícios em face dos malefícios, como também definir quais as aplicações que podem proporcionar grandes avanços para a humanidade.

6. CONTRIBUIÇÃO DOS AUTORES

Este capítulo foi escrito majoritariamente pela autora Poliana Aparecida Corrêa de Oliveira e pelo autor Renan Albuquerque Marks, contando com a supervisão e revisão do autor Omar Paranaíba Vilela Neto.

7. REFERÊNCIAS

AKYILDIZ, I. F. *et al.* The internet of bio-nano things. **IEEE Communications Magazine**, v. 53, n. 3, p. 32-40, 2015.

Alpaydin, Ethem (2020). **Introduction to Machine Learning (Fourth ed.)**. MIT. ISBN 978-0262043793., 2020.

BADELT, S. *et al.* **A General-Purpose CRN-to-DSD Compiler with Formal Verification, Optimization, and Simulation Capabilities**. Springer International Publishing. DNA Computing and Molecular Programming. p. 232--248. 2017.

BENENSON, Y.. **Biomolecular computing systems: principles, progress and potential**. Nature Reviews Genetics, v. 13, n. 7, p. 455-468, 2012.

BERLEANT, J. *et al.* Automated sequence-level analysis of kinetics and thermodynamics for domain-level DNA strand-displacement systems. **Journal of the Royal Society Interface**, v. 15, n. 149, p. 20180107, 2018.

BORUAH, K. & DUTTA, J. C. **DNA Computing Models for Boolean Circuits and Logic Gates**. IEEE International Conference on Computational Intelligence & Communication Technology, IEEE, 2015

CAMPBELL, A. V. **Bioethics: the basics**. 2 ed. London and New York: Taylor & Francis, 2017. 216p.

CHEN, Y. *et al.* DNA nanotechnology from the test tube to the cell. **Nature nanotechnology**, v. 10, n. 9, p. 748-760, 2015.

CLAMONS, S.; QIAN, L.; WINFREE, E. Programming and simulating chemical reaction networks on a surface. **Journal of the Royal Society Interface**, v. 17, n. 166, p. 20190790, 2020.

CRIBBS, A. P.; PERERA, S. MW. Focus: Genome Editing: Science and Bioethics of CRISPR-Cas9 Gene Editing: An Analysis Towards Separating Facts and Fiction. **The Yale journal of biology and medicine**, v. 90, n. 4, p. 625, 2017.

HARROUN, S. G. *et al.* Programmable DNA switches and their applications. **Nanoscale**, v. 10, n. 10, p. 4607-4641, 2018.

HJELMFELT, A. *et al.* **Chemical implementation of neural networks and Turing machines.** Proceedings of the National Academy of Sciences, 1991.

HOOPS, Stefan *et al.* COPASI—a complex pathway simulator. **Bioinformatics**, v. 22, n. 24, p. 3067-3074, 2006.

JINEK, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. **Science**, v. 337, n. 6096, p. 816-821, 2012.

JIANG, H.; RIEDEL, M. D. & PAHRI, K. K. **Digital logic with molecular reactions.** IEEE/ACM International Conference on Computer-Aided Design (ICCAD), IEEE, 2013.

MARKS, R. A. *et al.* **DNAr-logic: a constructive DNA logic circuit design library in R language for molecular computing.** Proceedings of the 32nd Symposium on Integrated Circuits and Systems Design - SBCCI '19, ACM Press, 2019.

MISHRA, S. P. *et al.* **Application of advanced molecular marker technique for improvement of animal: A critical.** 2017.

LAKIN, M. R. *et al.* Visual DSD: a design and analysis tool for DNA strand displacement systems. **Bioinformatics**, v. 27, n. 22, p. 3211-3213, 2011.

Lakin, M. R. & Stefanovic, D. **Supervised Learning in Adaptive DNA Strand Displacement Networks.** ACS Synthetic Biology. American Chemical Society (ACS), 2016.

Lee, J.-H. *et al.* **A molecular evolutionary algorithm for learning hypernetworks on simulated DNA computers.** 2011 IEEE Congress of Evolutionary Computation (CEC), IEEE, 2011.

NELSON, D. L.; COX, Michael M. **Princípios de Bioquímica de Lehninger.** 5ed. Artmed Editora, 2011.

Noh, Y.-K. *et al.* **Molecular learning with DNA kernel machines.** Biosystems, Elsevier BV, 2015.

OLIVEIRA, P. A. C. *et al.* Analysis of single-module and cascade molecular analog circuits for approximate computing based on DNA Strand Displacement.

In: **33rd Symposium on Integrated Circuits and Systems Design (SBCCI)**. IEEE, p. 1-6, 2020.

QIAN, L., *et al.* Neural network computation with dna strand displacement cascades. **Nature**. 2011.

QIAN, L. & WINFREE, E. Scaling up digital circuit computation with dna strand displacement cascades. **Science**. 2011.

SARPESHKAR, R. Analog synthetic biology. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 372, n. 2012, p. 20130110, 2014.

SEELIG, G. *et al.* Enzyme-free nucleic acid logic circuits. **Science**, v. 314, n. 5805, p. 1585-1588, 2006.

SOLOVEICHIK, D.; SEELIG, G. & WINFREE, E. DNA as a universal substrate for chemical kinetics. **Proceedings of the National Academy of Sciences**, v. 107, n. 12, p. 5393-5398, 2010.

SONG, T. *et al.* Analog computation by DNA strand displacement circuits. **ACS synthetic biology**, v. 5, n. 8, p. 898-912, 2016.

SRINIVAS, N. *et al.* Enzyme-free nucleic acid dynamical systems. **Science**, v. 358, n. 6369, 2017.

TOMA, H. E.. **Nanotecnologia molecular: materiais e dispositivos**. São Paulo: Editora Blucher, 2016. 336p.

VIEIRA, D.. K. S. *et al.* DNAr: A R package to simulate and analyze CRN and DSD networks. **ACS Synthetic Biology**, 2020.

VILELANETO, O. P. & PACHECO, M. A. C. **Nanotecnologia Computacional Inteligente: Concebendo a Engenharia em nanotecnologia**. Rio de Janeiro: PUCRio & Interciência, 2012. 512p.

WANG B. *et al.* **Effective design principles for leakless strand displacement systems**. **Proceedings of the National Academy of Sciences**, vol. 115, no. 52, pp.E12 182–E12 191, 2018.

YORDANOV, B. *et al.* Computational design of nucleic acid feedback control circuits. **ACS synthetic biology**, v. 3, n. 8, p. 600-616, 2014.

ZADEH, J. *et al.* NUPACK: analysis and design of nucleic acid systems. **Journal of computational chemistry**, v. 32, n. 1, p. 170-173, 2011.

ZHANG, D. Y. *et al.* Engineering entropy-driven reactions and networks catalyzed by DNA. **Science**, v. 318, n. 5853, p. 1121-1125, 2007.

ZOU, C. *et al.* Four-analog computation based on DNA strand displacement. **ACS omega**, v. 2, n. 8, p. 4143-4160, 2017.

13.

ARTELLIGENT: INTELIGÊNCIA ARTIFICIAL NA ARTE

Francisco de Paula Barretto

Suzete Venturelli

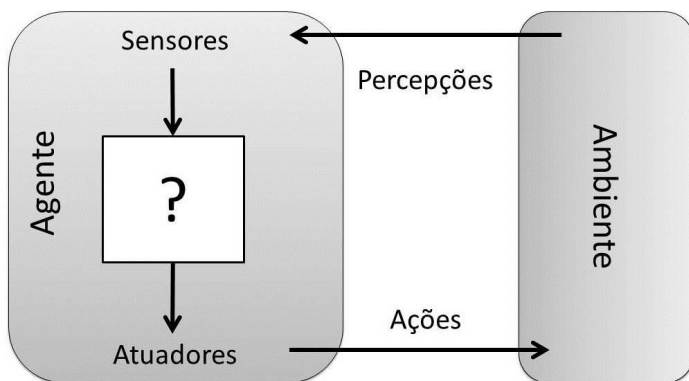
1. INTRODUÇÃO

Sob o ponto de vista criativo, discorre-se há algum tempo sobre os panoramas da utilização de Inteligência Artificiais (IA), ou agentes³, poeticamente representados no cinema pelo computador Hal, em *2001: Uma Odisseia no Espaço*, os replicantes de *Blade Runner*, baseado no romance de Philip K. Dick *Do Androids Dream of Electronic Sheeps*, ou pelo próprio Wintermute criado por William Gibson no romance intitulado *Neuromancer* (KUBRICK e CLARCKE, 1968; DICK, 1982; GIBSON, 2003). No entanto, há uma enorme diferença entre os agentes imaginários retratados nestas obras e os agentes inteligentes com os quais convivemos no mundo “real”, como aqueles que agendam passagens ou triam os nossos e-mails.

O conceito computacional geral de agente remonta aos anos 1950 quando John McCarthy, considerado por muitos como um dos pais da IA ao lado de Marvin Misky e vários outros, desenvolveu o *software* Advice Talker. No entanto, o termo “agente” só caiu em uso nas discussões *high tech* por volta de 1989, quando a Apple lançou um vídeo intitulado *The Knowledge Navigator* onde exibia um mordomo trajando um belo smoking que obedecia prontamente aos comandos de um interagente frente ao computador (JOHNSON, 1997).

Segundo Russel e Norvig (2003), um agente pode ser entendido como aquele capaz de perceber o ambiente em que está inserido através de sensores e interagir com o mesmo através de atuadores, conforme ilustrado na figura 1.

Figura 1: Representação de um agente capaz de interagir com o ambiente através de sensores e atuadores.



Fonte: Russel e Norvig (2003).

Por exemplo, um agente humano possui olhos, ouvidos entre outros órgãos capazes de sentir o mundo, além de mãos, pernas entre outras partes capazes de atuar sobre o meio. De mesmo, um agente deve ter formas de sentir e interagir com o meio físico ou virtual para que possa desenvolver sua cognição. Esta capacidade cognitiva do agente, por sua vez, pode ser representada através de agentes inteligentes expressando-se de diversas formas distintas como a música, a pintura e a poesia, por exemplo.

Ainda nos anos 1950, mais precisamente em 1956, pesquisadores como John McCarthy e Marvin Minsky cunharam o termo Inteligência Artificial, que se baseava-se na premissa de que todo aspecto relacionado à aprendizagem ou inteligência pode ser matematicamente descrito e, portanto, computacionalmente simulado (MCCARTHY, MINSKY, *et al.*, 2006). Segundo Russel e Norvig (2003), neste sentido, há várias vertentes que abordam a modelagem matemática da Inteligência Artificial sob diferentes aspectos, como os que tentam simular o raciocínio humano, pensamento racional, comportamento humano ou comportamento racional, ilustrados na figura 2.

Figura 2: Algumas definições de IA agrupadas em quatro grandes categorias.

Sistemas que pensam como humanos	Sistemas que pensam racionalmente
“O excitante novo esforço para fazer computadores pensarem ... máquinas com mentes, no sentido literal e completo.” (Haugeland, 1985)	“O estudo de faculdades mentais por meio do uso de modelos computacionais.” (Charniak e McDermott, 1985)
“[A automação de] atividades que nós associamos com pensamento humano, atividades tais como tomada de decisão, resolução de problemas, aprendizagem...” (Bellman, 1978)	“O estudo de computações que fazem possível perceber, raciocinar e agir.” (Winston, 1992)
Sistemas que agem como humanos	Sistemas que agem racionalmente
“A arte de criar máquinas que realizam funções que requerem inteligência quando realizadas por pessoas.” (Kurzweil, 1990)	“Inteligência Computacional é o estudo do projeto de agentes inteligentes.” (Poole et al., 1998)
“O estudo de como fazer computadores realizarem coisas em que, no momento, pessoas são melhores.” (Rick e Knight, 1991)	“IA ...é preocupada comportamento inteligente em artefatos.” (Nilsson, 1998)

Fonte: Russel e Norvig (2023).

Nestas primeiras discussões científicas acerca da definição conceitual de Inteligência Artificial abordou-se, entre outros, o ponto de vista da busca pela compreensão do funcionamento da inteligência humana que, por sua vez, está ligada diretamente à capacidade que temos de compreender o mundo através das nossas habilidades cognitivas. Dentre estas habilidades, pode-se destacar a relevância da concepção artística no contexto da expressão desta inteligência através do uso do nosso sistema conotativo (LUGER, 2004).

Ainda por volta dos anos 1950 e 1960 o objetivo era tentar simular a inteligência humana, que requeria uma modelagem matemática muito perfeita do funcionamento do cérebro humano, ou como pensam os humanos, a fim de obter resultados realmente satisfatórios em atividades razoavelmente generalistas, conhecida como IA forte (ou IA generalista). Dada a complexidade do cérebro humano e a subjetividade de alguns conceitos como “bom senso”, por exemplo, tornou-se praticamente inviável a simulação do pensamento humano. No entanto esta pretensão esbarrava na enorme complexidade dos problemas e na incapacidade humana de abstrair, modelar e representar o conhecimento necessário para que a “máquina” pudesse compreendê-los e resolvê-los.

Podemos imaginar que qualquer problema que necessite de apenas uma parte das competências do cérebro humano, por mais simples que possa

parecer, como costurar um tecido pode envolver um número virtualmente infinito de fatos e conhecimentos necessários. Considerando apenas esta atividade, por exemplo, talvez precisássemos “ensinar a máquina” o que é uma agulha, o que é uma linha, o que são cores, o que é um tecido, como segurar a agulha e assim por diante. Na IA forte, costurar seria apenas uma das habilidades que este agente deveria ser capaz de demonstrar. Portanto, consideremos que para cada nova habilidade devemos mais uma vez “ensinar” todo o conhecimento necessário. Se considerarmos que esta IA deve ser capaz de desempenhar um comportamento inteligente em um número virtualmente infinito de situações podemos ter uma dimensão do tamanho da dificuldade em “ensinar” tudo o que a IA necessita “saber”.

A grande dificuldade naquela época, residia, portanto, em subestimar a complexidade dos problemas e a quantidade de conhecimento necessária para simular o funcionamento completo do nosso cérebro. Considerando esta limitação, atualmente, a evolução das pesquisas nesta área nos guia no sentido de atribuir à Inteligência Artificial não mais o raciocínio humano generalista, mas o comportamento inteligente em problemas específicos, operando em domínios mais restritos que possam ser mais facilmente modelados e representados.

Esta abordagem, conhecida como IA fraca, geralmente age sobre um domínio mais restrito que permite gerenciar e reduzir a complexidade quando comparada ao comportamento generalista. Esta linha de ação passa a compreender, portanto, a expressão do comportamento inteligente como dependente do contexto, baseado em um conjunto de conhecimentos que operam em um determinado domínio. Neste caso, por exemplo, um agente robótico capaz de compor música deve ser projetado para interagir dentro deste domínio e certamente falhará se necessitar realizar outra tarefa como escrever.

Pode-se perceber, portanto, que um ponto crucial no desenvolvimento da IA se dá na abordagem de como representamos o conhecimento de tal forma que o agente seja capaz de compreendê-lo e manipulá-lo a fim de demonstrar um comportamento inteligente. Além disso, o domínio e o contexto onde o agente irá operar determinarão certamente que tipos de sensores e atuadores ele deverá possuir. Por fim, será necessário definir que técnica deverá ser utilizada para desenvolver o “cérebro” do agente.

Do ponto de vista artístico, a importância da tecnologia no processo composicional e mais especificamente na produção artística vem aumentando

de forma diretamente proporcional ao desenvolvimento das tecnologias ligadas à computação e à eletrônica. Este aperfeiçoamento dos computadores e suas linguagens de programação tem permitido o desenvolvimento de ferramentas para várias áreas, inclusive para a arte. A habilidade de programar, analogamente à composição musical, exige um processo mental de abstração e atenção aos detalhes. Portanto, não é surpresa que os músicos tenham sido, segundo Curtis Roads (1996), alguns dos primeiros artistas a se utilizarem do uso massivo dos computadores.

Na música, a questão da representação do conhecimento acerca das relações que dominam a expressão criativa fica mais evidente, posto que ela possui gramática e regras próprias, capazes de serem descritas em forma de algoritmo (ROADS, 1996). A exemplo da música concreta e da música eletroacústica, a composição algorítmica tem sido desenvolvida há décadas como forma de prover uma composição assistida pelo computador ou, inclusive, a realização de uma obra composta autonomamente pelo computador.

Se considerarmos a IA uma espécie de automação de comportamento inteligente, segundo a definição de Luger e Stubblefield (LUGER, 2004), e a criatividade como uma expressão de comportamento inteligente inerente ao nosso sistema cognitivo e conotativo (FERNANDES, 2008), podemos então estudar a relação do comportamento de expressão criativa através da aplicação de técnicas de IA sob dois paradigmas composicionais: composição assistida pelo computador e composição pelo computador (POHLMANN, 2002).

No primeiro paradigma, a composição artística é realizada por um usuário através do auxílio do computador. Neste caso, o computador pode facilmente auxiliar usuários inexperientes no processo composicional. A segunda abordagem diz respeito à composição artística totalmente realizada por um computador, sendo bem mais promissor para o estudo do ponto de vista da Inteligência Artificial e da Criatividade Computacional para as Artes.

De um ponto de vista computacional, como podemos definir sistemas que exibem um comportamento inteligente e criativo ao produzir resultados artísticos emergentes? Quais princípios, características, funções, um sistema tem que demonstrar para ser considerado como tal?

A fim de responder tais questões fez-se necessária a criação do termo Artelligent (BARRETTO, 2016; BARRETTO e VENTURELLI, 2018),

que tem sua origem na junção das palavras em latim *ars/artis* e *intelligere*. A palavra *ars*, correspondente ao termo grego *tékne*, descreve basicamente uma habilidade prática para produzir artefatos e ao mesmo denota um conhecimento, ciência, método relacionado à alguma teoria arte-científica. Para este caso, consideramos que se trata de um conceito envolvendo a expressão ou habilidade imaginativa representadas em suas mais amplas vertentes como pintura, música e literatura e que tem associado algum potencial emocional ou estético. Já a palavra *intelligere*, também oriunda do latim, está na raiz da palavra inteligente é uma composição das palavras *inter*, isto é, entre, e *legere*, que significa escolher, coletar, colher, reunir. Juntas, *inter* e *legere* significam perceber, reconhecer, compreender, entender e realizar.

Um sistema Artelligent pode ser definido como um sistema autopoietico que através da utilização de técnicas específicas de inteligência artificial, representa o conhecimento de forma extensível e, considerando os princípios que regem o processo criativo humano, é capaz de exibir resultados de caráter reconhecidamente emergentes em um determinado ambiente. No entanto, é preciso esclarecer cada uma das partes que compõem esta definição posto que, certamente, ela não englobará qualquer sistema artístico que envolva arte e Inteligência Artificial apenas.

Alguns dos modelos apresentados neste trabalho, como as redes neurais artificiais e os algoritmos genéticos, possibilitam ao artista diversas aplicações que permeiam os conceitos de emergência e autopoiese. O primeiro, definido, segundo Peter Cariani (2009) como o surgimento de novas entidades que, em um sentido ou em outro, não poderiam ter sido previstas com base naquilo que as precedeu, enquanto a autopoiesis, ou autopoiese, (do grego *auto* “próprio”, *poiesis* “criação”) descreve os sistemas autônomos, capazes de autoproduzir e autorregular, mantendo interações com o meio. Por sua vez, o meio pode desencadear, apenas de forma indireta, mudanças nos processos ou estrutura internas do sistema autopoietico (MATURANA e VARELA, 1997). Estes conceitos se apresentam de forma estruturante no sentido de que se deseja que um sistema Artelligent seja capaz de exibir um comportamento artístico criativo de forma autônoma.

Por fim, muitos estudos podem ser desenvolvidos na confluência das áreas citadas envolvendo os conceitos de criatividade, consciência, emergência e autopoiese no sentido de criar, ou simular, através da IA, estes

atributos inerentes à existência humana. Esta simulação pode se dar através do desenvolvimento de sistemas artísticos inteligentes que sejam capazes de se expressar de forma autônoma em termos artísticos, através de música ou imagem. Assim, alguns trabalhos, resultados de pesquisas e obras que utilizam diferentes técnicas de IA, serão aqui apresentados a fim de demonstrar o delineamento das intersecções entre os diferentes conceitos apresentados e a evolução histórica da área.

3. ABORDAGENS: SIMBÓLICA, EVOLUTIVA E CONEXIONISTA

Para qualquer estudo que permeie a IA, a identificação e codificação dos dados de forma a constituir e representar conhecimento de modo que o agente seja capaz de compreendê-lo, processá-lo e, então, agir é fundamental para a estruturação cognitiva de um sistema inteligente. Uma vez estruturado o conhecimento, dada a possibilidade de compreensão do mesmo, é possível que o sistema aprenda através da avaliação heurística dos resultados obtidos através da aplicação de um método de “raciocínio”. Ao ser capaz de inferir, baseado em uma heurística, a “distancia” entre o resultado esperado e o resultado obtido é possível guiar as ações ou “decisões” do sistema na direção um determinado objetivo que se quer atingir. Este aspecto é apresentado sob o prisma de três abordagens: simbólica, conexionista e evolutiva.

Na abordagem simbólica, a inteligência requer conhecimento declarativo e alguma forma de mecanismo de raciocínio, denominado cognição. A representação simbólica de conhecimento se manifesta através de uma relação semântica e sintática, onde várias unidades (átomos) interconectadas são, coletivamente, responsáveis por representar vários conceitos. A evolução dos estados da cognição permite descrever a evolução destes, criados durante o processo de cognição, através do uso de vocabulário similar à linguagem natural ou linguagem lógico-matemática. Esta evolução dos estados da cognição permite que se desenvolva uma forma de aprendizagem independente dos estados cognitivos, configurando aprendizagem.

Baseado nesses preceitos, o comportamento inteligente pode ser descrito em termos do conhecimento que demonstram ter ao executarem uma certa tarefa que demanda inteligência. Logo, os programas inteligentes

devem ser capazes de modelar conhecimento e manipulá-lo. Neste sentido, a representação de conhecimento, que será manipulada através de um mecanismo de raciocínio, diz respeito à um formalismo escolhido para descrever um conhecimento do mundo real.

No entanto o conhecimento humano é volumoso, difícil de caracterizar, altamente mutável e normalmente se organiza dependendo da forma como será utilizado. As técnicas de IA exploram o conhecimento de forma que ele seja capaz de capturar generalizações através da valorização das propriedades importantes, permitindo modificação, correção e compreensão além de tentar limitar as várias possibilidades que, geralmente, seriam analisadas.

É assaz difícil encontrar um domínio real que possa ser estruturado de maneira precisa em sua totalidade. Para realizar algum raciocínio sobre qualquer domínio é necessário a construção de um modelo que passa a ser uma representação simplificada do conhecimento e, certamente, deixará alguns fatos desconhecidos seja por ignorância teórica ou prática. Por isto, normalmente, os sistemas inteligentes devem lidar com a falta de informação em algum momento e associar medidas de certeza aos fatos (crenças), gerando conhecimento incerto. Além disso, só é humanamente possível estruturar um domínio sobre o qual temos compreensão. Assim, será muito difícil planejar um agente que seja capaz de agir em um domínio desconhecido (como a criação artística criativa emergente) utilizando a representação simbólica.

Para Luger (2004), a maioria das técnicas encontradas na literatura representa o conhecimento explicitamente através de abstrações e algoritmos de busca, concebidos para implementar um comportamento inteligente, como na abordagem simbólica. No entanto, numa abordagem diferente, pode-se construir programas inteligentes utilizando modelos matemáticos que simulam sistemas biológicos: como os evolutivos ou conexionistas.

Estes modelos evolutivos, como os Algoritmos Genéticos (GA) e os Autômatos Celulares (CA), inspiram-se principalmente nos processos relacionados à evolução através da adequação de uma população de indivíduos a partir da sobrevivência dos seus membros mais ajustados. O poder de seleção dos indivíduos através deste processo se demonstra através da própria emergência de espécies na evolução natural (LUGER, 2004).

Os modelos evolutivos de representação de conhecimento funcionam através do processo de introdução de variações em sucessivas gerações ao

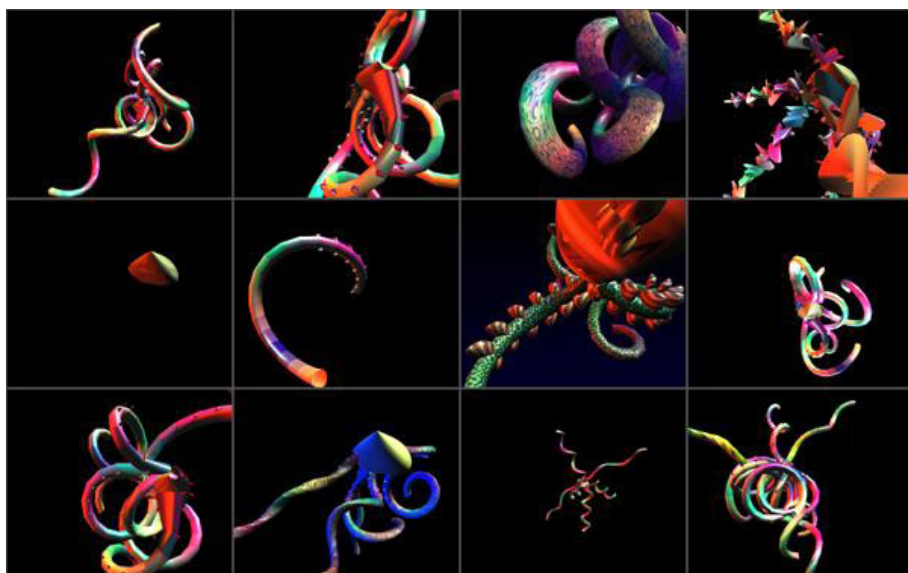
passo que se eliminam os indivíduos menos ajustados. Assim, emerge, uma população cuja adaptação exibe uma crescente capacidade e diversidade. Estes indivíduos desempenham ações que afetam outros indivíduos e, por sua vez, são afetados também pelo ambiente. Assim, as pressões seletivas não são unicamente oriundas do ambiente no qual os indivíduos estão inseridos, mas também da interação com outros membros da população, configurando assim um ecossistema mais complexo.

Neste caso, representação do conhecimento se dá através do genótipo, formado por um conjunto de alelos. Este modelo, assim como na evolução biológica, produz conhecimento operando sobre unidades de informação que são transmitidas e adaptadas socialmente. A corporificação do indivíduo, usualmente referida como seu fenótipo, descreve o seu comportamento e características físicas do indivíduo baseadas na codificação do genótipo.

Neste sentido, quando utilizamos um Algoritmo Genético para solucionar um problema genérico temos três estágios distintos: primeiro, os indivíduos (potenciais soluções) são representados de tal forma que sejam capazes de suportar variações evolutivas através de operadores de seleção. No segundo estágio, são definidos alguns mecanismos de acasalamento e mutação, análogos às atividades sexuais e mutantes de formas biológicas. Estes mecanismos são capazes de produzir uma nova geração de indivíduos através da recombinação de características de seus pais e da introdução de uma variação genética. Finalmente, uma função de aptidão julga quais indivíduos mais adaptados serão favorecidos através de uma maior probabilidade de propagação de seus alelos.

Entre os projetos que trabalham o tema da evolução das espécies estão as obras do Karl Sims: *Genetic Images* (1993), *Evolved Creatures* (1994) e *Galápagos* (1997) (apud. PAUL, 2008). Ambos permitem que o interagente influencie em uma evolução simulada onde os organismos/indivíduos da população são escolhidos através do parâmetro estético. Em *Galápagos*, figura 3, são exibidos organismos abstratos em várias telas dispostas em círculo onde o interagente pode escolher as que considera mais interessantes. Os organismos escolhidos reagem mutando ou reproduzindo-se a fim de substituir os indivíduos mais preteridos e, portanto, menos adaptados. Esta evolução simulada é fruto de uma interação entre o homem e a máquina onde o interagente exerce um controle criativo direto ao tomar uma decisão de ordem estética.

Figura 3: Formas emergentes da obra Galápagos.



Fonte: Karl Sims (1997).

Para Laurentiz (2007), a obra *Evolved Creatures* (1994) de Karl Sims explora as consequências de se trabalhar de forma conjunta com diferentes conceitos como os algoritmos simples que controlam a morfologia do indivíduo, as redes neurais e sensores responsáveis pelo controle das reações dos indivíduos no ambiente e aprendizado, além de outras funções como, por exemplo, a que é implementada de forma a estimular a competição entre os indivíduos orientados por um propósito comum.

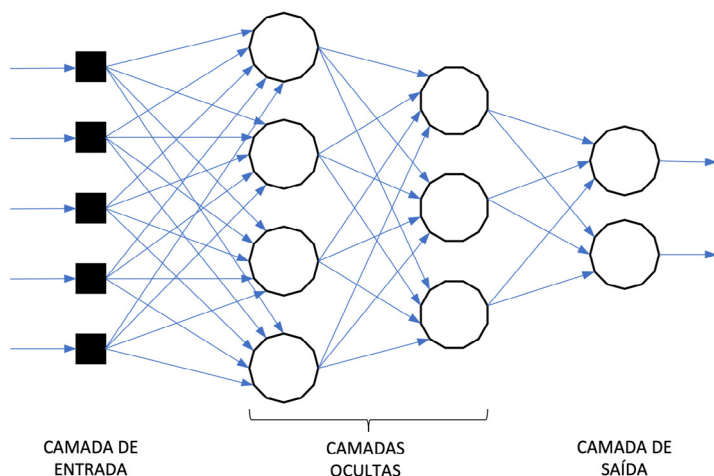
Por fim, a abordagem conexcionista, como as Redes Neurais Artificiais ou RNA, se baseiam no funcionamento neuronal onde o neurônio simulado passa a ser a unidade computacional básica do sistema em questão. Segundo Lippman (1997) as RNAs são sistemas físicos que podem adquirir, armazenar e utilizar conhecimentos experimentais, alcançando uma boa performance devido à sua densa interconexão entre nós da rede.

Este modelo inicialmente proposto por McCulloch e Pitts (1943), considera que a capacidade de aprendizagem do cérebro está relacionada com a plasticidade, isto é, a capacidade de gerar alterações nas ligações sinápticas entre neurônios. A plasticidade pode ser atribuída a dois mecanismos: a criação de novas conexões sinápticas entre neurônios e a modificação das sinapses

existentes. Na representação artificial conexcionista, cada sinapse tem um peso associado positivo se a sinapse for excitatória e negativo se for inibitória.

Do ponto de vista topológico uma rede neural pode ser vista como um grafo orientado, onde os nós representam neurônios artificiais e os arcos representam as conexões entre eles. A orientação dos arcos indica o sentido do fluxo de informação entre os neurônios de diferentes camadas. A figura 4 ilustra uma RNA cuja primeira camada é a de entrada, responsável por distribuir os valores recebidos para os neurônios da camada seguinte. A última camada é a de saída, podendo existir, como neste caso, uma ou mais camadas intermediárias denominadas camadas ocultas. Caso existam camadas ocultas a rede neural é denominada multicamadas. Ainda, se existem arcos de todo neurônio de uma camada para todos os neurônios da camada seguinte, a rede é denominada totalmente conectada. Caso contrário, a rede é considerada como parcialmente conectada.

Figura 4: Topologia de uma Rede Neural Artificial totalmente conectada alimentada à frente



Fonte: elaborada pelos autores.

O conhecimento em uma RNA é representado pelos pesos sinápticos associados à cada conexão e o mútuo reforço ou inibição dos estímulos a um determinado neurônio pelos outros neurônios. A não linearidade é essencial para dotar a rede com “poder de decisão”, ou seja, mudanças quantitativas na entrada podem assim produzir mudanças qualitativas na saída ao invés de simplesmente alterar a saída proporcionalmente à entrada (ROADS, 1996).

A aprendizagem com as RNAs consiste na adaptação dos pesos sinápticos sob dois métodos: aprendizagem supervisionada (com um tutor) ou não supervisionada (autônoma). Na aprendizagem supervisionada, um conjunto de casos de histórico, juntamente com as respectivas respostas, é submetido à rede e a saída gerada é comparada à saída esperada para cada caso. A regra de aprendizagem é utilizada para ajustar os valores dos pesos sinápticos para que as respostas dadas se aproximem das respostas corretas. Este método pressupõe um rol de casos e soluções conhecidas, o que o torna menos interessante para o desenvolvimento de resultados inesperados posto que a rede estará treinada para reagir aos estímulos de forma previamente conhecida.

Embora autores como Marvin Minsky (1988) apontem como desvantagem das RNAs o fato de serem sistemas fechados ou “caixas pretas”, limitando a compreensão do funcionamento interno e minimizando as interferências exteriores no ajuste dos pesos e aprendizagem da rede, a IA conexcionista possui um vasto campo de aplicação tendo obtido sucesso em vários domínios de difícil abordagem através de outras técnicas, sobretudo quando o domínio é mutável, como no reconhecimento de imagens faciais de pessoas em função do envelhecimento (MINKSY, 1988). No entanto esta “desvantagem” pode ser entendida como uma característica autopoietica deste sistema, dotando-o de uma identidade (MATURANA e VARELA, 1997).

Um bom exemplo de trabalho artístico que utiliza RNA é a obra *Machine Hallucinations NYC*, figura 5, do artista computacional Refik Anadol (2020), que é capaz de gerar imagens sintéticas da cidade de Nova Iorque. Nesta instalação, a partir de um conjunto de 110 milhões de imagens da cidade foi possível treinar uma RNA capaz de classificar qualquer imagem como sendo de Nova Iorque ou não. A partir deste “repertório” aprendido, utilizando a técnica Generative Adversarial Networks (GAN) foi possível realizar o processo inverso e solicitar que a rede passasse a gerar uma infinidade de novas imagens, absolutamente sintéticas, da cidade de Nova Iorque. Este processo faz emergir um conjunto de imagens que nunca foram vistas antes por tratarem-se de imagens geradas pelo computador, dotando o sistema com uma espécie de capacidade de improvisação criativa na geração destas imagens.

Obras como a do Refik Anadol (2020) tem um potencial enorme pois, diferentemente das abordagens simbólica e evolutiva, a abordagem conexãoista não necessita que o conhecimento seja estruturado de maneira formal. A capacidade de aprender a partir de um treinamento pressupõe apenas um conjunto de dados sobre o domínio. Portanto, pode-se dizer que o avanço na última década se concentra sobretudo nas técnicas derivadas da abordagem conexãoista, destacando-se o ramo da Deep Learn. Este campo interdisciplinar floresce por uma confluência de fatores como, por exemplo, a larga disponibilidade de dados e o aumento exponencial no poder computacional.

Figura 5: Machine Hallucinations NYC.



Fonte: ilustração de Refik Anadol (2020).

4. DOS NÚMEROS À POÉTICA

Em meados do século passado, muitos artistas computacionais (2017) se interessaram nas capacidades das máquinas computacionais pensantes. Os artistas esperavam que os computadores colaborassem na tomada de decisões em relação ao processo de criação. Para tanto introduziram a ideia de uma certa aleatoriedade que gerariam variações inspiradas do design original. Esse método traria para os artistas mais proposições para inventar e executar suas próprias inspirações. Neste campo, os artistas computacionais Georg Nees, Frieder Nake, Manfred Mohr, Vera Molnár, Waldemar Cordeiro foram pioneiros no uso das imagens generativas.

Entretanto, o interesse começa muito antes com a matemática inglesa Ada Lovelace. Em 1842 ajudava Charles Babbage a publicar o primeiro algoritmo a ser executado por sua Máquina Analítica, conhecido como o primeiro computador mecânico, quando percebeu oportunidades que poderiam transpassar a matemática, pois imaginou um computador que pudesse processar não apenas números, mas resolver problemas de qualquer complexidade. Na época, era revolucionário que as máquinas tivessem aplicações além do cálculo puro. Ela chamou a ideia de Ciência Poética. Para ela a Máquina Analítica poderia agir sobre outras coisas além do número, se fossem encontrados objetos cujas relações fundamentais mútuas pudessem ser expressas por aquelas da ciência abstrata das operações. Ela supunha, por exemplo, que as relações fundamentais de sons agudos na ciência da harmonia e da composição musical eram suscetíveis de tais expressões e adaptações. Assim, se poderia compor peças de música elaboradas e científicas de qualquer grau de complexidade ou extensão.

Posteriormente, destacamos que desde 1973, Harold Cohen - um pintor, professor da Universidade da Califórnia, San Diego e ex-representante da Grã-Bretanha na Bienal de Veneza - elaborou um software chamado AARON, que conseguiu fazer pinturas de forma autônoma por décadas; mesmo no final dos anos 1980. Cohen talvez possa ser considerado como um dos únicos artistas que terá exposições póstumas de novas obras criadas por AARON, mesmo após sua própria morte.

As imagens que o programa em evolução fez nas últimas quatro décadas são realmente obras de Harold Cohen, ou criações independentes da própria AARON, ou talvez colaborações entre os dois? Essa, certamente, é

uma questão delicada. AARON nunca se afastou muito do estilo de Cohen, elaborado desde os anos 1960, quando era um expoente de sucesso da abstração. AARON segue seu mestre.

Primeiramente, as questões não resolvidas sobre a arte da máquina são sobre o seu potencial e, em segundo lugar, se - independentemente da qualidade do trabalho produzido - pode realmente ser descrito como “criativo” ou “imaginativo”. Estes são problemas fascinantes que nos levam profundamente aos mistérios da arte humana.

Imaginação está implicada no surgimento de IA com AIML (Artificial Intelligence Markup Language), uma linguagem de estímulo-resposta aberta e minimalista para a criação de personalidades *bot* como A.L.I.C.E. Em 1995, Richard Wallace (2020) desenvolve o *chatbot* A.L.I.C.E (Artificial Linguistic Internet Computer Entity), inspirado no programa ELIZA de Joseph Weizenbaum, mas com o acréscimo de coleta de dados de amostra de linguagem natural em uma escala sem precedentes, possibilitada pelo advento da web, A.L.I.C.E. *chatbot* aprende a falar na web.

Essa IA traz novas questões para perspectivas de criação, como a de poemas, na medida em que domina a linguagem. Nesse mesmo viés, uma das IAs projetadas em parceria entre a Universidade de Stanford, a Universidade Massachusetts Amherst e o Google Brain escreveu algumas poesias surpreendentemente tristes (BOWMAN, VILNIS, *et al.*, 2016). O princípio por trás da pesquisa foi criar uma IA capaz de se comunicar por meio de frases em linguagem natural. A poesia do Google foi escrita por um sistema de IA, treinada por milhares de romances românticos. A equipe do Google recorreu a uma rede neural e deu ao sistema duas frases retiradas dos livros. Coube então à IA criar sua própria poesia com base nas informações disponíveis.

Em junho de 2015, outra equipe do Google desenvolveu um *chatbot* que ameaçava seus criadores (VINYALS e LE, 2015). A inteligência artificial aprendeu a arte da conversação analisando milhões de roteiros de filmes, permitindo-lhe meditar sobre o significado da vida, a cor do sangue, a moralidade e até mesmo ficar com raiva de seu inquisidor humano. Questionado sobre o sentido da vida, o *chatbot* respondeu: “Viver para sempre”. O Facebook também tem ensinado sua inteligência artificial com livros infantis como O Livro da Selva, Peter Pan e Alice no País das Maravilhas, para ensinar suas redes (HILL, BORDES, *et al.*, 2016).

5. PESQUISAS E EXPERIMENTOS CRIATIVOS

Michel Bret, Marie Helene Tramus e Alan Berthoz, em 2002, fazem um relato de uma colaboração que ocorreu no âmbito de um estudo-experimento artístico sobre figuras virtuais interativas na fronteira da arte e das ciências cognitivas. Este estudo, denominado Interatividade ‘Inteligente’ (Conexionismo, Ciência Evolutiva e Vida Artificial) em Artes Digitais em Relação com a Fisiologia da Percepção de Ação e Movimento, foi apoiado pelo Programa Cognitique 2000 sobre Arte e Cognição, uma iniciativa de Ministério da Pesquisa da França (2005). Como resultado artístico entre 2004-2006, a partir do design de interatividade de pesquisa “Inteligente” eles apresentam uma instalação experimental e interativa entre um ser humano e um equilibrista virtual, a partir de questões que levantaram sobre a diferença entre simpatia e empatia nas relações humano-máquina.

É interessante destacar que os algoritmos de IA avançavam para colaborar nas tarefas de várias outras atividades, como na criação de robô para casa em 2002. A empresa spin-off de Rodney Brook, iRobot, criou o primeiro robô de sucesso comercial para o lar - um aspirador de pó autônomo chamado Roomba. Limpar o tapete parece muito longe dos algoritmos criativos que importam para a arte. Mas o Roomba foi uma grande conquista. Suas poucas camadas de sistemas geradores de comportamento eram muito mais simples do que outros algoritmos e apesar dos sensores relativamente simples e do poder de processamento mínimo, o dispositivo tinha inteligência suficiente para limpar uma casa de maneira confiável e eficiente. O Roomba inaugurou uma nova era de robôs autônomos, focados em tarefas específicas, cuja performance vai interessar aos artistas.

Outro fato importante é que em 2008 a IA começa a resolver grandes problemas. Em novembro de 2008, um interessante recurso apareceu no iPhone da Apple - o aplicativo para reconhecimento de fala. Isso representou um grande avanço, apesar do reconhecimento de voz sempre ter sido um dos principais objetivos da IA. Sabe-se que décadas de investimento nunca o elevaram acima de 80% de precisão. O Google recorreu à uma nova abordagem: milhares de computadores poderosos, executando redes neurais paralelas, aprendendo a detectar padrões nos vastos volumes de fluxo de dados de muitos usuários do Google. Após anos de aprendizado e melhorias, o Google afirma que é 92% preciso.

Um ano depois, a ImageNet democratiza dados, tornando-se muito interessante para o campo da criação. A pesquisadora da Universidade de Stanford, Fei-Fei Li (2020) viu seus colegas acadêmicos e da indústria de IA insistindo no mesmo conceito, no qual um algoritmo melhor tomaria melhores decisões, independentemente dos dados. Mas ela percebeu uma limitação dessa abordagem - o melhor algoritmo não funcionaria bem se os dados que ele aprendeu não refletissem o mundo real. Sua solução foi construir um conjunto de dados que pudesse mapear todo o mundo dos objetos (FEI-FEI, 2010). O conjunto de dados resultante foi denominado ImageNet. A pesquisadora lançou o ImageNet, um banco de dados gratuito de 14 milhões de imagens que foram rotuladas por dezenas de milhares de funcionários do Amazon Mechanical Turk. Os pesquisadores de IA começaram a usar o ImageNet para treinar redes neurais para catalogar fotos e identificar objetos. O conjunto de dados evoluiu rapidamente para uma competição anual para ver quais algoritmos poderiam identificar objetos com a menor taxa de erro. Muitos o veem como o catalisador para o boom da IA que o mundo está experimentando hoje.

Em 2010, data que marca a criatividade na dança, pois a nova tecnologia significava que computadores menores também poderiam ter um impacto maior. Os novos computadores dos robôs humanóides, como o robô NAO, podiam fazer coisas que para predecessores como Shakey seria quase impossível. Os robôs NAO usaram muitas tecnologias pioneiras na década anterior, como a aprendizagem possibilitada por redes neurais. Na Expo Mundial de 2010 em Xangai, algumas das capacidades extraordinárias desses robôs foram exibidas, enquanto 20 deles dançavam em perfeita harmonia por oito minutos.

Uma pesquisa em 2012 que trouxe mais expectativa no contexto da criatividade da IA foi a de Jeff Dean e Andrew Ng (LE, RANZATO, *et al.*, 2012), que relataram um experimento no qual eles mostraram uma rede neural composta por 10 milhões de imagens não rotuladas, tiradas aleatoriamente de vídeos do YouTube e para surpresa deles, um dos neurônios artificiais aprendeu a responder fortemente a fotos de gatos.

Em 2012, Simon Colton professor de criatividade computacional do Goldsmiths College, em Londres, idealizou a IA artista The Painting Fool. Nesse trabalho o pesquisador sugeriu que, para que os algoritmos de IA fossem considerados criativos, eles teriam que ir além do teste de Turing

(COLTON, 2012). Ele propõe, então que o algoritmo em vez de simplesmente ser capaz de conversar como um humano, como Turing propôs, a IA teria que se comportar de maneira “habilidosa”, “apreciativa” e “imaginativa”, para ser considerada uma artista. Para tanto, ele criou um programa para digitalizar um artigo no *Guardian* sobre a guerra no Afeganistão, que extraiu palavras-chave como “OTAN”, “tropas” e “britânico” e encontrou imagens relacionadas a elas. Em seguida, ele as juntou para formar uma imagem composta refletindo o “conteúdo e o humor” do artigo do jornal.

Em uma exposição em Paris, *The Painting Fool* tirou fotos de visitantes em diferentes estados de espírito, respondendo a palavras-chave emocionais, derivadas de 10 artigos que foram lidos por ele. Se a contagem de estado negativo fosse muito alta (que ocorre com cobertura de notícias), Colton então programou a IA para entrar em um estado de desânimo, no qual se recusa a pintar, um equivalente virtual do temperamento artístico. No seu site ele diz que o objetivo do projeto é que ele seja levado a sério - um dia - como um artista criativo por direito próprio. Segue se apresentando e diz ter sido construído para exibir comportamentos que podem ser considerados habilidosos, apreciativos e imaginativos. O autor declara que:

“Meu trabalho foi exibido em galerias reais e online; as ideias por trás de sua concepção foram usadas para abordar noções filosóficas como emoção e intencionalidade em inteligências não humanas; e artigos técnicos sobre as técnicas de inteligência artificial, visão de máquina e computação gráfica que foram publicadas na literatura científica.” (COLTON, 2020).

Em junho de 2015, Alex Mordvintsev e a equipe de pesquisa Brain AI do Google publicaram alguns resultados fascinantes para a criatividade artística. Depois de algum treinamento na identificação de objetos a partir de pistas visuais e sendo alimentado com fotografias de céus e coisas em formato aleatório, o programa começou a gerar imagens digitais sugerindo a imaginação combinada de Walt Disney e Pieter Bruegel, o Velho, incluindo um híbrido “Porco-Caracol”, “Camel-Bird” e “Dog-Fish”. Isso deu origem a uma nova forma de arte chamada “Inceptionism”, em homenagem ao algoritmo Inception, em que uma rede neural aumentaria progressivamente o zoom em uma imagem e tentaria “vê-la” dentro da estrutura do que já conhecia (MORDVINTSEV, 2015).

A última pesquisa do Google em inteligência artificial produziu algumas imagens seriam consideradas alucinatórias - compostas, como pinturas digitais, usando redes neurais artificiais como uma máquina digital de sonhos. “Iniciação”, como a equipe de pesquisa do Google o chama, é o processo de alimentar milhões de imagens em uma rede neural artificial, modelada no cérebro humano, até que ela comece a reconhecer padrões e objetos específicos dentro das imagens. Google tem um site intitulado Experiments with Google (2020), no qual os programadores que criaram milhares de experimentos criativos usando Chrome, Android, AI, Web VR, AR e muito mais, podem apresentar seus projetos, junto com ferramentas e recursos úteis, para inspirar outros a criar novos experimentos. O site apresenta uma coleção de experimentos para ser explorado com novos adicionados a cada semana.

6. POÉTICAS COLABORATIVAS HUMANO-IA

Em 2017, a artista Taryn Southern (2018) que trabalha com várias plataformas de IA para coproduzir seu álbum de estreia *I AM AI* criou a música “Break Free” resultado de uma colaboração humano-AI. Os softwares de IA para criação musical avançou tanto nos últimos anos que não é mais uma novidade; é uma ferramenta viável que pode e está sendo usada pelos produtores para ajudar no processo criativo. Para muitos músicos, esse fato levanta a questão: a inteligência artificial poderia um dia substituir os músicos? Usar a IA como uma ferramenta para fazer música ou ajudar músicos já está em prática há algum tempo. Nos anos 1990, por exemplo, o artista David Bowie ajudou a desenvolver um aplicativo chamado Verbasizer, que reunia material literário e reordenava aleatoriamente as palavras para criar novas combinações que poderiam ser usadas como letras, e o processo ficou conhecido como música aleatória.

Em 2016, pesquisadores da Sony usaram um software chamado Flow Machines para criar uma melodia no estilo dos Beatles. Segundo Dealh (2018), este material foi então entregue ao compositor humano Benoît Carré e desenvolvido em uma canção pop totalmente produzida chamada Daddy’s car. O Flow Machines também foi usado para ajudar a criar um álbum inteiro de música sob o nome de SKYGGE, que significa “sombra” em dinamarquês. O autor, ainda afirma, que no nível do consumidor, a tecnologia já

está integrada a programas populares de criação de música como Logic, um software que é usada por músicos em todo o mundo e pode preencher automaticamente padrões de bateria exclusivos com a ajuda de IA.

Dealh (2018) esclarece que a maioria desses sistemas funciona usando redes de aprendizagem profunda, um tipo de IA que depende da análise de grandes quantidades de dados. Basicamente, diz o autor, você alimenta o software com toneladas de material de origem, de sucessos de dança a clássicos do disco, que ele analisa para encontrar padrões. Ele capta coisas como acordes, andamento, duração e como as notas se relacionam umas com as outras, aprendendo com todas as entradas para que possa escrever suas próprias melodias. Existem diferenças entre as plataformas: algumas fornecem MIDI enquanto outras fornecem áudio. Alguns aprendem puramente examinando dados, enquanto outros contam com regras embutidas em código baseadas na teoria musical para orientar sua produção.

Nesse mesmo viés, encontramos os artistas das artes visuais interessados nesse novo processo criativo. Uma máquina pode gerar uma obra de arte sozinha? Por exemplo, esta questão foi lançada pelo coletivo de artistas Obvious, um trio baseado em Paris fascinado pelo potencial artístico da inteligência artificial. Obvious alimentou um algoritmo com 15.000 imagens de retratos de diferentes períodos de tempo. Quando foi leiloado na Christie's em outubro de 2018, o Retrato de Edmond Belamy foi vendido por incríveis \$ 432.500, sinalizando a chegada da arte de IA no palco do leilão mundial. No seu manifesto (2019), o grupo diz que desejam explorar, usar e compartilhar os diferentes algoritmos de aprendizado de máquina para fortalecer nossa natureza humana. A noção de criatividade, comenta o grupo, é extremamente difícil de simular, pois implica no processo uma série de fatores que ainda não estão devidamente definidos. Através da replicação de comportamento humano em um contexto criativo, eles vêem os algoritmos como uma ferramenta fascinante para se aprofundar e entender melhor as diferentes forças no processo de criação de algo novo, único e inovador. Eles também percebem outros exemplos de algoritmos ajudando humanos a ampliar sua criatividade em cenários como em um jogo de xadrez ou GO Game.

No contexto dos games, Narendra Mohan Mittal, nos traz a seguinte questão: os videogames têm inteligência artificial? Muitos, talvez a maioria, dos videogames apresentam os chamados personagens não-jogadores

(NPCs). Podem ser adversários, aliados, espectadores ou qualquer outra coisa. A questão é que eles são controlados não pelo jogador (você), mas pelo computador. Normalmente, as pessoas se referem à maneira como esses NPCs se comportam como a “IA” do jogo. Mas como funciona exatamente a IA em um videogame típico? A arquitetura da máquina de estado finito ou Game IA, termo utilizado no caso dos jogos, desempenha papel central na maioria dos games e também é amplamente usada em robótica e carros autônomos. Muitos jogos usam algoritmos adicionais para além disso controlar o comportamento do NPC, e alguns não usam (nos últimos anos, uma alternativa às máquinas de estado finito, são as chamadas árvores de comportamento, que se tornaram populares). Mas é justo dizer que as máquinas de estado finito estão entre os algoritmos mais comuns para implementar o comportamento NPC.

Alguns identificam, atualmente, o comportamento do jogador para tomar decisões emergentes, como por exemplo o *The Mind Game*. Segundo Statt (2019) ele foi projetado principalmente para avaliar o estado psicológico de jovens recrutas e muitas vezes apresenta aos jogadores situações impossíveis de testar seu poder mental em face de uma derrota inevitável. Destaca-se que o jogo é procedural, gerando ambientes e situações em tempo real, e permite aos jogadores realizar qualquer ação em um mundo virtual que eles poderiam fazer no mundo real. Indo ainda mais longe, ele responde ao estado emocional e psicológico de seus jogadores, adaptando-se e respondendo ao comportamento humano e evoluindo ao longo do tempo. O autor diz que o game ainda se baseia nas memórias do jogador para gerar mundos de jogo inteiros feitos sob medida para o passado do jogador no game.

7. CONSIDERAÇÕES FINAIS: AS MÁQUINAS REFLETEM NOSSA CONSCIÊNCIA

O estudo da Inteligência Artificial é um dos campos da tecnologia cujos limites de pesquisa vão bem além da análise técnica. A princípio ela pode servir também para investigar a natureza do ser humano e da inteligência humana, os limites da máquina e nossos próprios limites como construtores de artefatos “inteligentes” (WILSON, 1995). Não é surpresa que os artistas passassem, portanto, a explorar o vasto campo da IA como forma de

produzir trabalhos interativos mais sofisticados que vão além das simples janelas e menus que tanto caracterizam a arte computacional.

Esta possibilidade artística de investigar a natureza humana através do uso de IA pode ser vista na obra *Sentimentos da Virada* (2020), dos artistas André Gola, Francisco Barretto e Marília Pasculli. Nela, uma Rede Neural Convolucional foi treinada para reconhecer estados afetivos humanos baseados na expressão facial. A partir deste treinamento, a obra se tornou capaz de analisar fotos e *selfies* postadas nas redes sociais a fim de tentar mapear o sentimento da cidade de São Paulo durante a sua Virada Cultural de 2020. Uma vez identificadas as fotos e os sentimentos correspondentes a obra exibia o resultado através de uma projeção mapeada, figura 6, que tornava pública essas informações, assim como as fotos analisadas.

Figura 6: *Sentimentos da Virada*.



Fonte: ilustração de André Gola, Marília Pasculli e Francisco Barretto (2020).

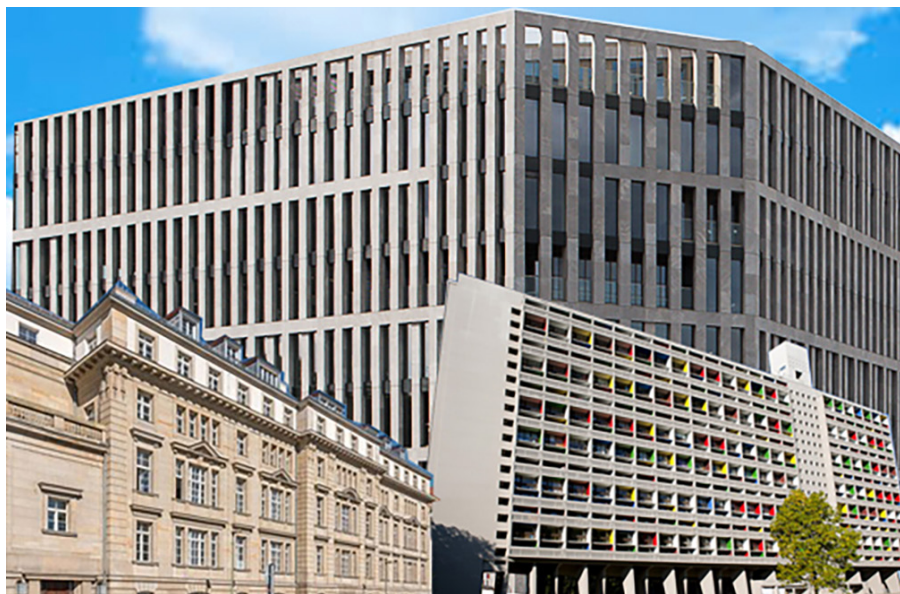
Segundo Duanne Ribeiro (2019), em *Apparatus Memories* – projeto em desenvolvimento pelo matemático, designer, curador, pesquisador e artista Fabrizio Poltronieri (2021), apoiado pelo Rumos 2017-2018 – o computador cria narrativas, a partir de um acervo de vídeos com cerca de 20 terabytes. A IA que está utilizando identificará os constituintes das paisagens captadas, imagens serão reorganizadas de forma imprevisível. O artista tem como questão: o que narram os aparelhos? O que importa na sua ideia são os componentes das captações disparadas por ele mesmo ao longo de viagens que realizou pela Europa e pela China, entre fotografias e vídeos – de formato típico e em 360 graus –, num procedimento chamado segmentação de imagens. Segundo Poltronieri o processo não é simples, pois ele tem de ver frame por frame e separar pessoas, plantas e carros para poder organizar em um banco de dados com elementos de diversas fontes, de diversos vídeos, quando o artista maquínico poderá, então, propor suas narrativas (figura 7 e 8).

Figura 7: Obra *Apparatus Memories*.



Fonte: ilustração de Fabrizio Poltronieri (2020).

Figura 8: Obra Apparatus Memories.

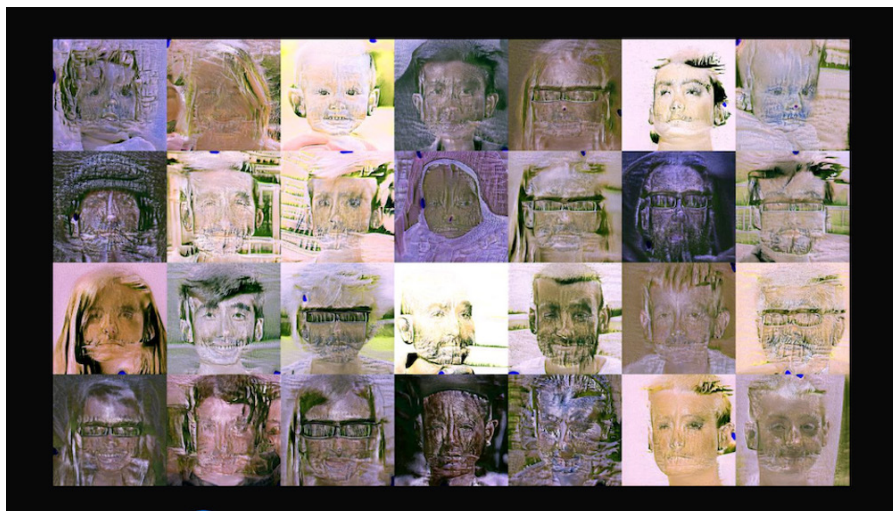


Fonte: ilustração de Fabrizio Poltronieri (2020).

Memo Akten (2018) usa tecnologia para refletir nossa humanidade e explorar como damos sentido ao mundo. O artista treinou um algoritmo de aprendizado de máquina para “ver” usando imagens que representam conceitos essenciais na vida humana. Para obter imagens que refletissem nossa humanidade compartilhada, ele baixou fotos do Flickr marcadas com as seguintes palavras, entre outras: tudo, mundo, universo, espaço, montanhas, oceanos, flores, arte, vida, amor, fé, ritual, deus.... Então ele programou a máquina para “imaginar” novas imagens com base em todas essas imagens - criando um novo mundo de paisagens, objetos e ideias nunca antes vistos, mas com base em nossa própria experiência de vida. O resultado é uma jornada através da “imaginação de uma máquina” que foi treinada a partir dos nossos conceitos básicos de nossa própria existência.

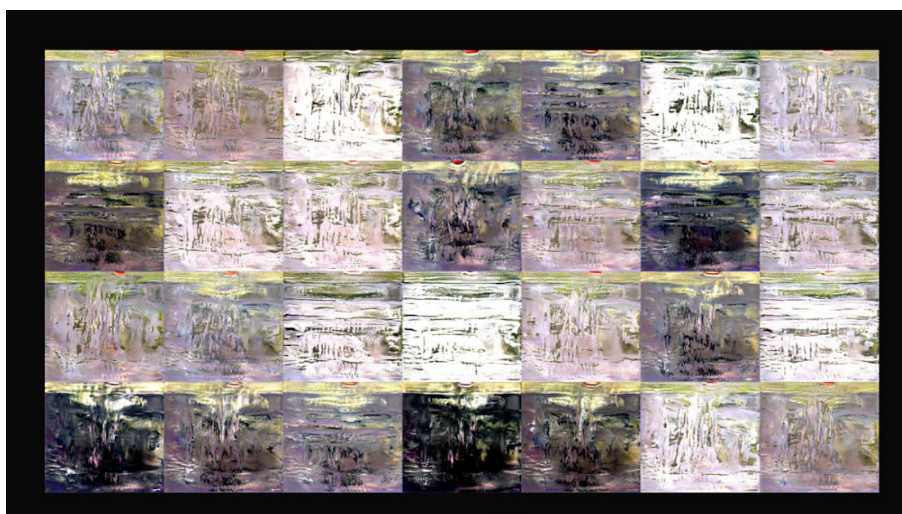
A proposta de Suzete Venturelli, figuras 9 e 10, envolve a ideia de inteligência artificial generativa para que a IA escolha livremente seu caminho criativo a partir de banco de dados de fotos, desenhos, imagens, sons da natureza, sons do universo, em tempo real. Para tanto foi pensado a utilização do software ml5.js - Friendly Machine Learning for the Web, cujo algoritmo de aprendizado de máquina foi proposto por Daniel Shiffman.

Figura 9: Training Image Experiments de Suzete Venturelli, 2019/2020.



Fonte: elaborada pelos autores.

Figura 10: Training Image Experiments de Suzete Venturelli 2019/2020.



Fonte: elaborada pelos autores.

Por meio de alguns avanços tecnológicos, como o TensorFlow.js, tornou-se possível usar a unidade de processamento gráfico (GPU) integrada do navegador para fazer cálculos que, de outra forma, seriam executados muito lentamente usando a unidade de processamento central (CPU). O

ml5.js é um caminho para tornar todos esses novos desenvolvimentos em aprendizado de máquina na web mais acessíveis para todos.

A tecnologia está acelerando tão rápido que é quase impossível acompanhar e, sem pretensão de concluir, levantamos as seguintes questões: o que o futuro reserva para o aprendizado de máquina e a arte? Como a inteligência artificial pode continuar a contribuir no contexto da colaboração humano-computador? Qual a importância da IA para a imaginação e inovação, em momentos de crise, como na atualidade?

8. REFERÊNCIAS

2001: Uma Odisséia no Espaço. Direção: Stanley KUBRICK. [S.l.]: [s.n.]. 1968.

AKTEN, Memo. <https://memoakten.medium.com/>. Acesso em 6 de nov. de 2018.

ANADOL, R. Machine Hallucinations - NYC. **Refik Anadol Website**, New York, 2020. Disponível em: <refikanadol.com>. Acesso em: 10 Dezembro 2020.

BARRETTO, F. **Artelligent: Arte, Inteligência Artificial e Criatividade Computacional**. Brasília: Tese (Doutorado em Artes) - Universidade de Brasília, 2016.

BARRETTO, F. Emergência Enquanto Heurística Para Criatividade Em Agentes De Inteligência Artificial Aplicados À Composição Algorítmica. **Seminário Música Ciência Tecnologia**, São Paulo, v. 1, n. 4, 2012.

BARRETTO, F.; VENTURELLI, S. Zer0: An Emergent and Autopoietic Multi-agent System for Novelty Creation in Game Art through Gesture Interaction. **Procedia Manufacturing**, v. 3, p. 850-857, 2015. ISSN 2351-9789. 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences,

BARRETTO, F.; VENTURELLI, S. Artelligent: A Framework for Developing Interactive Computer Artwork Using Artificial Intelligent Agents. In: A., M.; W., W. **Lecture Notes in Computer Science**. [S.l.]: Springer International Publishing, 2018. p. 471-489.

BENJAMIN, Walter. **Para a crítica da violência**. In: Escritos sobre mito e linguagem. São Paulo: Editora 34, 2013.

BLADE Runner: o caçador de androids. Direção: Ridley SCOTT. [S.l.]: [s.n.]. 1982.

BOWMAN, S. R. *et al.* **Generating Sentences from a Continuous Space**. [S.l.]: [s.n.], 2016.

BRET, Michel, TRAMUS, Marie-Hélène e BERTHOZ, Alain. **Interacting with an Intelligent Dancing Figure: Artistic Experiments at the Crossroads between Art and Cognitive Science**. Disponível em: <https://www.jstor.org/stable/1577645?seq=1>. Acesso em 30 de jun. de 2006.

Buffrestein, Alissa. **Google's Artificial Brain Creates Its Own Artworks and They Are Freaky**. Disponível em: <https://news.artnet.com/art-world/google-artificial-neural-networks-created-artworks-309782>. Acesso em 12 de dez. 2020.

CARIANI, P. Emergência e Criatividade. In: ITAULAB **Emoção Artificial 4.0**. São Paulo: Itaú Cultural, 2009. p. 21-41.

COLTON, S. The Painting Fool: Stories from Building an Automated Painter. **Computers and Creativity**, p. 3-38, ago. 2012. ISSN 978-3-642-31726-2.

COLTON, Simon. **About me... I'm The Painting Fool: a computer program, and an aspiring painter**. Disponível em: <http://www.thepaintingfool.com/>. Acessado em: 2 de dez. 2020.

DEAHL, Dani. **How AI-Generated music is changing the way hits are made**. disponível em: <https://www.theverge.com/2018/8/31/17777008/artificial-intelligence-taryn-southern-amper-music>. Acesso em 1 de nov. de 2018.

FEI-FEI, L. ImageNet: crowdsourcing, benchmarking & other cool things. **CMU VASC Seminar**, Março 2010.

FERNANDES, A. M. R. **Inteligência Artificial: noções gerais**. Florianópolis: VisualBooks, 2008.

GALLEGO, JELOR. FACEBOOK IS TRAINING ITS AI TO UNDERSTAND LANGUAGE BY TEACHING IT TO READ CHILDREN'S BOOKS. Disponível em: <https://futurism.com/teach-ais-reading-comprehension-researchers-giving-books>. Acesso em 3 de dez de 2016.

GIBSON, W. **Neuromancer**. São Paulo: Aleph, 2003.

GIORDANO, Paolo. **No contágio**. Belo Horizonte: Ayiné, 2020.

Google. Experiments with Google. Disponível em: <https://experiments.with-google.com/>. Acesso em 18 de jun. de 2020.

HAN, Byung-Chul. **No enxame: perspectivas do digital**. Petropolis: Vozes, 2018.

HILL, F. *et al.* **The Goldilocks Principle: Reading Children's Books with Explicit Memory Representations**. [S.l.]: arXiv.org, 2016.

HOWARD, Philip N. **Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives**. New Haven, CT: Yale University Press, 2020,

JOHNSON, S. **Cultura da Interface: como o computador transforma a nossa maneira de criar e comunicar**. Nova York: Jorge Zahar Editor, 1997.

LAURENTIZ, S. **A Poética dos Recursos Computacionais que Simulam a Vida**. Anais do 6º Encontro Internacional de Arte e Tecnologia. Brasília: UnB. 2007.

LE, Q. V. *et al.* **Building high-level features using large scale unsupervised learning**. [S.l.]: arXiv, 2012.

LIPPMANN, R. P. Speech recognition by machines and humans. **Speech Communication**, Amsterdam, v. 22, n. 1, p. 1-15, jul. 1997.

LUGER, G. F. **Inteligência Artificial: estruturas e estratégias para a solução de problemas complexos**. Tradução de Paulo Engel. [S.l.]: Bookman, 2004.

LUISI, P. L. Autopoiesis: a review and a reappraisal. **Naturwissenschaften**, v. 90, n. 2, p. 49-59, 10 January 2003.

MARKOFF, John. How Many Computers to Identify a Cat? 16,000. Disponível em: <https://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html>. Acesso em 25 de jun. de 2012.

MATURANA, H.; VARELA, F. **De Máquinas e Seres Vivos: a organização do vivo**. 3rd Edition. ed. Porto Alegre: Artes Médicas, 1997.

MBEMBE, Achille. O direito universal à respiração. São Paulo: N-1 Edições, 2020. <https://n-1edicoes.org/020>. Ver: <https://www.instagram.com/>

MCCARTHY, J. *et al.* A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. **AI Magazine**, v. 27, n. 4, p. 12-14, dec 2006. ISSN 0738-4602.

MCCULLOCH, W.; PITTS, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. **Bulletin of Mathematical Biophysics**, 5, 1943. 115-133.

MINKSY, M. **The Society of Mind**. [S.l.]: Simon & Schuster, 1988.

MITTAL, Narendra Mohan. Machine learning and Artificial Intelligence 2.0 with Big Data: Building Video Games using Python 3.7 and Pygame. Kindle Edition. 2019

MORDVINTSEV, A. Inceptionism: Going Deeper into Neural Networks. **Google AI Blog**, 2015. Disponível em: <<https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>>. Acesso em: 10 Dezembro 2020.

OBVIOUS. Obvious is a collective of friends, artists and researchers. <http://obvious-art.com/wp-content/uploads/2020/04/MANIFESTO-V2.pdf>. Acesso em 11 de outubro de 2019.

PAUL, C. **L'Art Numérique**. [S.l.]: Thames & Hudson, 2008.

POHLMANN, K. C. **Principios de Audio Digital**. Madrid: Espanha, 2002.

POLTRONIERI, Fabrizio. Site pessoal: Disponível em: <https://www.fabrizio-poltronieri.com/artworks>. Acesso em: 5 de janeiro de 2021.

RIBEIRO, Duanne. Este computador vai se tornar capaz de contar histórias. Disponível em: <https://www.itaucultural.org.br/secoes/rumos/este-computador-vai-se-tornar-capaz-de-contar-historias>. Acesso em 10 de jun. de 2019.

ROADS, C. **The Computer Music Tutorial**. Massachussets: MIT Press, 1996.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence: a moder approach**. New Jersey: Prentice Hall, 2003.

SLOTERDIJK, Peter. “Os humanos não estão preparados para proteger a natureza”. DW- Deutsche Welle, 28 de jun. 2020. Disponível em: <https://www.dw.com/pt-br/os->.

STATT, Nick. How artificial intelligence will revolutionize the way video-games are developed and played. Disponível em: <https://www.theverge.com/2019/3/6/18222203/video-game-ai-future-procedural-generation-deep-learning> . Acesso em 2 de dez. De 2019.

TRAMUS, M.-H.; CHEN, C.-Y. **La funambule virtuelle et Quorum Sensing, deux installations interactives s'inspirant du connexionnisme et de l'évolutionnisme**. La création artistique face aux nouvelles technologies. Paris: [s.n.]. 2005.

VENTURELLI, S. Estética, Arte e Tecnologia. **Revista VIS**, 2007.

VENTURELLI, Suzete. Arte Computacional. Brasília: Edunb, 2917.

VINYALS, O.; LE, Q. V. **A Neural Conversational Model**. [S.l.]: arXiv, 2015.

WALLACE, Richard. A.L.I.C.E. Disponível em: <https://www.chatbots.org/chatbot/a.l.i.c.e/> . Acesso em 01 de mar de 2020.

WILSON, S. Artificial Intelligence Research as Art. **Stanford Electronic Humanities Review: Constructions of the Mind**, 4, n. 2, 1995.

WILSON, S. **Information Arts**. Cambridge: The MIT Press, 2002.

WIRED. Fei-Fei Li's Quest to Make AI Better for Humanity. Disponível em: <https://www.wired.com/story/fei-fei-li-artificial-intelligence-humanity/>. Acesso em 4 de dez. De 2019.

WIRED. Google's AI has written some amazingly mournful poetry. Disponível em: <https://www.wired.co.uk/article/google-artificial-intelligence-poetry>. Acesso em 15 de jan. de 2018.

WIRED. Inside Deep Dreams: How Google Made Its Computers Go Crazy. Disponível em: <https://www.wired.com/2015/12/inside-deep-dreams-how-google-made-its-computers-go-crazy/>. Acesso em 1 de Dez. de 2015.

ZYLINSKA, Joanna. **AI Art: Machine Visions and Warped Dreams**. London: Open Humanities Press 2020.

14.

O CÉREBRO E A NEUROÉTICA SOB O OLHAR DO DIREITO

Maria Aglaé Tedesco Vilardo

1. INTRODUÇÃO

Comecei a me interessar por neurociência a partir da história de Phineas Gage, um inglês que trabalhava na construção de estradas de ferro na Nova Inglaterra, em meados do século XIX. Conheci a história e assisti a um breve vídeo no YouTube (NOVAES, 2017). O caso ocorre em um momento do mundo sem tecnologia quando a ligação entre os povoados ocorria através de trilhos de trens para aproximar os lugares inóspitos e de difícil acesso. Um dos trabalhadores era Phineas Gage que, por sua liderança, por sua capacidade e por seus equilíbrios pessoais, características que o tornavam líder, comandava uma equipe de vários trabalhadores. Em um dia de trabalho, ao preparar a pólvora com uma barra de ferro, ocorreu uma explosão inesperada e essa barra de um metro de comprimento virou uma lança violenta que perfurou seu rosto do lado esquerdo, atravessando seu cérebro e saindo pela cabeça.

De forma inusitada, Phineas Gage não morre. Caminha e se comunica logo depois do acidente. O enorme buraco que está em sua cabeça é tratado pelo médico John Harlow durante dois meses e ele se recupera fisicamente. Entretanto, sua personalidade se modifica de forma surpreendente. Embora tenha perdido a visão do olho esquerdo, seus sentidos estavam intactos. Porém, não era mais o líder equilibrado e falava muitos palavrões, agindo de forma impaciente e grosseira, não conseguindo permanecer em empregos fixos. O caso passa a ser objeto de profundos estudos sobre o cérebro a partir de relatos detalhados do médico que o atendeu e prosseguindo com análise do crânio de Gage, anos após sua morte, treze anos depois do acidente.

Um dos grandes estudiosos deste caso é o professor português António Damásio, autor de inúmeros livros sobre o cérebro e a neurociência. Seus estudos o levaram a afirmar que a nova personalidade de Phineas não se caracterizava por valores antigos, pois não mais exerciam influência sobre suas decisões, nem mesmo por mudança de seu sistema de valores (DAMÁSIO, 2012). Parte deste sistema continuava influenciando de forma irrisória as decisões a serem tomadas no mundo real, seu caráter estava dissociado da sua cognição e comportamento. Damásio observa que a atenção, a percepção, a memória, a linguagem e a inteligência são aptidões mentais que podem ficar, uma ou mais, dissociadas do caráter degenerado.

O cérebro possui sistemas múltiplos compostos por unidades cerebrais interligadas e cada região cerebral deve ser considerada no contexto do sistema em que está inserida (DAMÁSIO, 2012). O dano no lobo frontal do cérebro de Gage produziu alterações no planejamento de sua vida e nos sistemas cerebrais das suas funções morais quanto ao seu comportamento social e princípios éticos. A descoberta conduziu a novas e instigantes formulações sobre a constituição das regiões do cérebro e a forma que podem ser afetadas e suas consequências para a tomada de decisão.

Esta reflexão se relaciona com o livre-arbítrio e com a responsabilidade na tomada de decisões por livre vontade, quanto ao que é certo ou errado e se danos neurológicos podem justificar ou isentar as pessoas de responsabilidade pela prática de seus atos. Portanto, é uma abordagem sobre a saúde mental.

A saúde mental ainda é vista com preconceito e com discriminação social, uma fragilidade de nossa humanidade, em busca incessante de resposta moral segura sobre o que é normal e o que é patológico. As tomadas de decisão das pessoas e sua repercussão na sociedade em correlação ao conhecimento atual sobre o funcionamento do cérebro levam à discussão sobre a atuação das normas e do Direito vigente. A discussão é enriquecida com o diálogo entre os diversos saberes proposto pela bioética e pelo biodireito, como veremos a seguir.

2. A RELAÇÃO ENTRE A NEUROCIÊNCIA E O DIREITO

Tanto um caso de repercussão econômica quanto uma obra de engenharia ou escolhas com consequências sobre a saúde de uma pessoa poderão chegar ao exame de profissionais do Direito para uma solução jurídica.

O Direito tem esta peculiaridade de examinar as mais diferentes áreas de conhecimento através da técnica jurídica para chegar à conclusão sobre o que é o correto a ser feito em um caso judicial. A decisão final deve tomar por base a lei, mas o percurso até o resultado exige boa argumentação e para isso alguma compreensão básica sobre as demais áreas de conhecimento será relevante.

Quando o caso judicial envolve decisões morais, a compreensão do funcionamento do cérebro ajuda na formulação de perguntas adequadas sobre o modo como as decisões são tomadas por uma pessoa. Daí a proposta de estudo da neurociência a fim de melhorar a aplicação da lei e da justiça e sabermos se decisões morais são tomadas livremente ou recebem influências sobre o livre-arbítrio e, se receberem, em que medida isso ocorre.

A neurociência se preocupa com a estrutura e o funcionamento do cérebro e do sistema nervoso. Ao Direito interessa compreender de que forma o cérebro e seus sistemas podem ou não influenciar na tomada de decisões e no livre-arbítrio e se as atitudes das pessoas são praticadas por escolhas livres ou se há interferência relevante a determinar as escolhas morais.

Nesta intersecção, entre a tomada de decisão e a repercussão desta decisão no mundo jurídico, vamos encontrar um vasto campo para nossa discussão sobre o cérebro e as leis. Passemos a examinar melhor estes conceitos.

3. O CÉREBRO, A NEUROCIÊNCIA AS LEIS E O DIREITO

O estudo do cérebro nos proporciona o conhecimento profundo de seus detalhes anatômicos, suas dinâmicas e seu funcionamento de forma detalhada por exames como a tomografia, a ressonância magnética e a magnetoencefalografia. O cérebro não é um órgão que se expande ou se contrai, não é um órgão excretor tampouco bombeador, não obstante regula a respiração, a pressão arterial, a circulação, os batimentos cardíacos e todas as funções dos desejos e das necessidades de sobrevivência e prazer dos seres humanos, como a fome, o sono, as emoções e os pensamentos (CARTER, 2012).

A neurociência nos auxilia a compreender melhor o cérebro. Este é dotado de módulos funcionais constituídos de circuitos neurais por meio dos quais os neurônios se comunicam e a presença de alguns módulos funcionais, alguns a mais do que os macacos possuem, faz de nós seres conscientes (AMTHOR, 2017). São 100 bilhões de neurônios produzindo 1

quadrilhão de sinapses. A parte da frente do lobo frontal, o córtex pré-frontal, muito maior nos cérebros humanos, controla o nível abstrato de planejamento de metas (AMTHOR, 2017). Erros de desenvolvimento na estrutura do cérebro, como disfunções neurais, orgânicas ou psicológicas e ambientais podem danificar ou mudar a neuroquímica do cérebro (AMTHOR, 2017), levando a graves problemas como esquizofrenia, autismo, depressão e Alzheimer. Dentro do córtex pré-frontal, a área dorsolateral é responsável pelas funções cognitivas de alto nível para tomada de decisão e resolução de problemas e a área ventromedial é a responsável para a tomada de decisões sociais e para a avaliação de riscos (AMTHOR, 2017).

Outra estrutura do cérebro, o tálamo, é a transmissora dos sinais dos sentidos e de outras áreas subcorticais e do neocórtex. A complexidade é imensa, existindo áreas subcorticais da maior relevância, como o sistema límbico que tem na sua estrutura o hipocampo, com a função de criar memória; e a amígdala, que processa as emoções, especialmente o medo (AMTHOR, 2017). Essa, possivelmente, a emoção que nos impede de agir contrariamente às leis.

As leis existem com o propósito de estabelecer regras de convívio em sociedade e são publicadas, criando normas para serem cumpridas pelos indivíduos, estabelecendo sanções, caso haja descumprimento. O julgamento para aplicação ou não das sanções pelo descumprimento é feito pelo Poder Judiciário que estabelece padrões a serem seguidos em casos futuros, considerados como precedentes, tanto na área do Direito Penal quanto do Direito Civil como demonstrado a seguir. No Direito Penal, há regras que sancionam atos ou omissões que possam causar lesão ou ameaça ao corpo físico, à vida, à moral ou ao patrimônio de uma pessoa, originados de atos reprováveis da conduta humana, considerados delitivos e podem ter como consequência, dentre outras, a pena de prisão para impedir o convívio do criminoso em sociedade. Como nos lembra Zaffaroni (2013), o Direito Penal se preocupa em interpretar as leis penais, por meio da dogmática jurídica, a fim de chegar a uma sentença de forma segura.

No Direito Civil, as normas regulamentam os direitos existenciais e patrimoniais, estabelecendo regras de convívio e de negócios entre os indivíduos nas diversas situações de vida, regulam as relações privadas e buscam cessar ato ou omissão em desacordo com o previsto nas leis, retornando

ao *status quo ante*, podendo o pagamento de uma indenização em dinheiro ser uma forma de recompensar o dano material ou moral sofrido.

Praticar uma ação ou deixar de fazer algo que deveria e poderia ser feito pode ter repercussão jurídica e a solução será encaminhada pela técnica jurídica. A primeira reflexão é sobre a efetiva liberdade em decidir e agir e sobre a existência do livre-arbítrio e qual a extensão de nossa responsabilidade por decisões tomadas e suas repercussões diante da racionalidade preditiva dos seres humanos e da relação entre neurociência e Direito.

4. POSSIBILIDADES INTER-RELACIONAIS ENTRE NEUROCIÊNCIA E DIREITO PENAL

Há críticas à interferência da neurociência no Direito, especialmente no Direito Penal, ao afirmar que o questionamento dos fundamentos da culpabilidade “não seria passível de conviver com a soberba intrínseca dos postulados neurocientíficos” (BUSATO, 2014). Esta crítica diz respeito aos processos neuronais serem determinantes da conduta humana, o que seria inconcebível diante do fato de o cérebro não ser separado da própria pessoa, portanto não a domina ou a engana com a ilusão de tomar decisões. Neste sentido, somente com a liberdade de ação se forma a estruturação normativa, pois não se regula aquilo que não se pode fazer, por isso conclui que a neurociência seria contraproducente ao Direito. Outra crítica à neurociência (DE BRITO, 2014), além de negar o livre-arbítrio, é o fortalecimento da periculosidade. A partir do entendimento de que o cérebro pensa e decide por si, excluindo o conceito de vontade, o criminoso passaria a ser um doente psiquiátrico. O tratamento para estas pessoas seria para prevenir a prática de outros crimes. Então, a neurociência seria uma ameaça a ampliar a justificativa de doença para todos os criminosos, sem que houvesse responsabilidade por suas ações e com tratamentos longos e caros.

Uma terceira crítica é quanto à desconsideração de garantias processuais nos processos criminais (GUARAGNI, 2014), pois poderia ser afastada a presunção de inocência e ser a pessoa considerada culpada sem julgamento mediante análise de imagens do cérebro demonstrando seu potencial perigo.

Passemos a analisar estas críticas em destaque.

Entender que a neurociência é contraproducente ao Direito afasta a possibilidade de se agregar o saber científico à normatização das condutas sociais. Pelo próprio argumento de o cérebro não ser separado da pessoa e da importância da liberdade de ação é que se tem a necessidade de conhecer cada vez mais o quanto somos influenciados pelos processos químicos e biológicos. Este conhecimento contribui para leis mais adequadas às nossas limitações humanas e que sejam despidas de, apenas, um objetivo punitivo, mas possam contribuir efetivamente para melhorar o convívio social.

A segunda crítica, ao refutar a compreensão de que um criminoso passaria a ser considerado um doente mental, revela o mencionado anteriormente sobre o preconceito existente com os problemas de saúde mental. Considerar a neurociência como uma ameaça por isentar as pessoas de responsabilidade quando estão sem os requisitos indispensáveis que a tornem responsáveis pelos seus atos é discriminatório. Além do que, o direito a tratamento em caso de problemas de saúde é um assegurado pela Constituição e pelas leis vigentes. Se há interferências que isentam pessoas de se responsabilizarem por seus atos, elas devem ser consideradas para proteger a própria pessoa e a sociedade e não deixar à sua própria sorte ou afastar esta pessoa do convívio social, sem proporcionar tratamento.

Do mesmo modo, a crítica sobre culpabilização sem julgamento não procede na medida em que a neurociência traz novos elementos a serem considerados dentro de um processo judicial com garantias constitucionais de ampla defesa e contraditório.

O desconhecido não precisa ser temido ou escondido, mas desvelado.

5. ALGUNS ASPECTOS CRÍTICOS ENTRE NEUROCIÊNCIAS E DIREITOS EXISTENCIAIS

O campo dos direitos existenciais diz respeito aos direitos que não são relativos ao patrimônio, mas a nossa vida pessoal, nossas decisões sobre o nosso ser no mundo.

Cada pessoa pode decidir a sua vida privada e a legislação brasileira prevê que a partir dos 18 anos toda pessoa tem autonomia absoluta para decidir sobre sua vida e sua existência. A lei faz limitação tão somente no tocante aos direitos patrimoniais. Esta mudança de paradigma veio com a mudança na legislação sobre pessoas com deficiência. O Estatuto da Pessoa

com Deficiência (BRASIL, 2015) traz a norma de que nenhuma deficiência afeta a plena capacidade civil da pessoa. Pontua cada uma das possibilidades como casar ou constituir união estável; exercer direitos sexuais e reprodutivos; planejar os filhos que deseja ter e ter as informações adequadas sobre reprodução e planejamento familiar; conservar sua fertilidade, sendo vedada a esterilização compulsória; exercer o direito à família e à convivência familiar e comunitária; e exercer o direito à guarda, à tutela, à curatela e à adoção, como adotante ou adotando, em igualdade de oportunidades com as demais pessoas. Qualquer outro direito existencial será pleno a todas as pessoas independente de sua condição psíquica ou mental.

Esta liberdade legal mudou o que o Direito entendia anteriormente. Hoje as pessoas tomam suas próprias decisões e cabe ao Poder Judiciário respeitar sendo limitada a sua intervenção. A questão é saber se estas pessoas responderão igualmente sobre seus atos perante a sociedade.

Deve ser observado que nenhuma das conquistas legais obtidas para as pessoas com impedimentos de natureza física, mental, intelectual ou sensorial poderá ser derogada ou restringida de acordo com a Convenção Internacional sobre os Direitos das Pessoas com Deficiência, artigo 4.4, pois todas as disposições que beneficiem as pessoas com deficiência mental contidas nas leis nacionais ou nas convenções internacionais permanecem em vigor.

Isto representa a proteção dos direitos já conquistados e o reconhecido princípio de vedação de retrocesso. Ocorre que para isso deve haver a compreensão de que compete aos Estados Partes signatários da Convenção assegurar o pleno exercício de todos os direitos humanos e liberdades fundamentais a todas as pessoas com deficiência mental ou qualquer outra, sem qualquer tipo de discriminação. Por isso, a Convenção reconhece a importância da autonomia e da independência individuais, inclusive da liberdade para fazer as próprias escolhas.

Chegamos ao ponto que interessa à neurociência, a liberdade em fazer as próprias escolhas. Se considerarmos que os comandos não são totalmente livres, mas originados de processos químicos e biológicos, deixamos de ter a liberdade de escolher. Seria um determinismo biológico que não deixaria margem de escolha para a pessoa tomar uma decisão. Damásio denomina de “algoritmos naturais” e traz ideias importantes sobre tais algoritmos (DAMÁSIO, 2017).

Damásio (2017) afirma que os algoritmos são fórmulas que constroem um resultado específico. Os humanos vão se construindo segundo esses algoritmos, mas não são propriamente os algoritmos. Essa simples frase evoca a importância de se compreender a humanidade como algo a mais do que um passo a passo a se chegar a um objetivo. Temos tecidos, órgãos e sistemas que compõem uma entidade viva, afirma o autor, e tanto os processos intelectuais quanto os sentimentos funcionam de forma interligada e interferem para produzir um objetivo que será alcançado a partir de uma caminhada com base nos valores morais construídos ao longo da vida.

Os valores morais advêm das recompensas e dos castigos, de prazer e de dor, e estes ocorrem em razão dos processos químicos e neurais. Todas estas funções podem ser objeto de estudos científicos, entretanto os “algoritmos naturais” recebem forte influência dos sentimentos conscientes dos seres humanos. Podemos ter uma liberdade limitada para agir contra nossos impulsos, mas a cultura humana representa a resistência contra os algoritmos naturais. Damásio conclui sustentando que tanto o sentimento quanto o pensamento concedem graus de liberdade e faz uma crítica ao pensamento reducionista que vê o avanço da ciência como humilhante à natureza humana.

A partir da conclusão do referido autor, podemos assumir que cada ser humano, nas suas diferenças e peculiaridades, terá a influência da cultura humana, dos seus sentimentos e dos seus pensamentos para a sua tomada de decisão quanto aos seus direitos existenciais, o que certamente ocorrerá em diferentes espectros.

Nessa interessante perspectiva, podemos nos valer do mesmo autor para algumas observações sobre a Inteligência Artificial.

6. INTELIGÊNCIA ARTIFICIAL COMO CRIATIVIDADE HUMANA

Os avanços da biotecnociência nos trouxeram a inteligência artificial que pode alterar as nossas vidas, de forma surpreendente.

Ao criarmos as máquinas, a ideia é que possam produzir mais e melhor do que os seres humanos, com maior efetividade e celeridade. O objetivo é produzir e ampliar o alcance do fazer humano em atividades repetitivas. Estas máquinas não possuem os atributos humanos do sentimento ou mesmo os processos mentais, daí sua limitação. Não obstante, é corrente o

medo de que as máquinas suplantem os seres humanos ou até tenham autonomia e possam atacar seus criadores, embora não possuam sentimentos. Os sentimentos humanos nos fazem seres vulneráveis, para nossa alegria ou tristeza, contribuindo para construir os conceitos morais e a ideia de justiça. São os sentimentos que contribuem para o processo da subjetividade, essencial para a consciência, sendo distintos das emoções.

As emoções são ações acompanhadas de ideias e modos de pensar e os sentimentos são percepções da reação do nosso corpo durante a emoção (DAMÁSIO, 2011). O começo é no cérebro, através de imagens desencadeando as emoções com reações na amígdala e no córtex do lobo frontal, iniciando os processos químicos pelas glândulas endócrinas e núcleos subcorticais com reações internas e externas no corpo e mentalizando ideias que, conforme a emoção, poderão responder de formas diversas como reduzir ou acelerar o raciocínio.

Outro exemplo dado por Damásio é a resposta à compaixão. Pesquisas vêm demonstrando que o cérebro tem respostas diferentes para a compaixão pelo sofrimento físico ou sofrimento mental, certamente afetadas pela experiência de vida de cada pessoa. Para a dor física, há resposta imediata do córtex insular e tendência a desaparecer mais rápido e para a dor mental ocorre um tempo maior, todavia permanecendo maior tempo.

O fato é que as experiências mentais que provocam prazer ou dor constituíram os nossos desejos e deram ensejo às nossas criações e transformações, demonstrando a força de nossa evolução cultural. Muitas vezes, sem saber o porquê daquela dor ou da alegria, vamos além para buscar uma explicação na transcendência pelo fato da ciência não nos trazer tudo que precisamos saber. Mais uma vez não devemos temer uma criação da nossa cultura como a inteligência artificial. Nem tampouco superestimá-la, mas aceitar como mais uma rica e criativa possibilidade do intelecto humano.

7. CONCLUSÃO

Fizemos uma breve reflexão sobre o cérebro e os estudos da neurociência por meio de um olhar ético, a partir da área do Direito, considerando os estudos do professor Antônio Damásio. O autor estudou o caso de Phineas Gage e concluiu que aptidões mentais, como a cognição e o comportamento, podem estar dissociadas do caráter.

O tema diz respeito ao problema da saúde mental e o preconceito e discriminação ainda existente na tentativa de delimitar a normalidade e a patologia.

A relação com o Direito pode ser observada tanto no âmbito penal quanto no civil e compreender melhor o funcionamento do cérebro contribui para o aprimoramento das leis e das decisões judiciais para buscar a solução em um caso judicial de fundo moral.

No campo penal, a neurociência recebe críticas por negar o livre-arbítrio, fortalecer a periculosidade e afastar a presunção de inocência. Estas críticas foram refutadas por ser importante o conhecimento da neurociência para contribuir para leis menos punitivas, por terem um conteúdo discriminatório no tocante à saúde mental e porque novos elementos e informações devem ser considerados em cumprimento à garantia constitucional do contraditório e ampla defesa.

No âmbito dos direitos existenciais ocorreu uma mudança de paradigma na legislação ao passar a considerar a capacidade plena que cada pessoa tem para tomar decisões em sua própria vida, limitada pela lei somente no tocante aos direitos patrimoniais. Portanto, a autonomia legalmente independe da condição psíquica ou mental de cada um. No entanto, a legislação mantém todas as conquistas de proteção legal já obtidas, vedado o retrocesso com qualquer restrição a direitos conquistados, nos quais se encontram a proteção especial a pessoas com qualquer deficiência mental.

Antônio Damásio ainda defende a ideia de que as fórmulas da vida que constroem um resultado específico, chamadas por ele de “algoritmos naturais”, não representam o próprio humano, pois se diferenciam em razão do sentimento consciente e que nossa resistência se dá por meio da cultura humana. Por sua vez, a liberdade de agir de cada pessoa vem do sentimento e do pensamento humanos que não estão limitados pelos processos biológicos e químicos.

Com origem na nossa cultura, a inteligência artificial não deve trazer preocupação. As máquinas, criadas para ampliar nossas atividades cansativas e repetitivas, não possuem nossos atributos ou processos mentais. São limitadas. O processo de subjetividade obtido por meio de nossos sentimentos, enquanto percepções do nosso corpo em reação às nossas emoções, embora espelhem a vulnerabilidade humana, auxiliam na construção da nossa moralidade e da ideia de justiça. As experiências da vida, positivas ou negativas, são as que criam e transformam o mundo e nos permitem, até mesmo, buscar explicações

outras que não sejam somente na ciência, mas também na transcendência. Não se deve subestimar nem superestimar qualquer criação da nossa cultura.

O ser humano está sempre em busca do mais e não pode se esquecer que, apesar de todas as nossas qualidades, inteligência e destreza, somos todos seres vulneráveis que sofrem e riem e que somos diversos e fazemos parte de um grande sistema universal a ser respeitado.

8. REFERÊNCIAS

AMTHOR, F. **Neurociência para leigos**. Alta Books Editora, 2017.

BRASIL. Lei nº 13.146, de 6 de julho de 2015. Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência). **Diário Oficial da União**, Brasília, DF, 6 jul. 2015.

BUSATO, P. C. Urna visão crítica das implicações dos estudos neurocientíficos em Direito Penal. In: **Neurociência e Direito Penal**. Atlas, 2014. p. 49-82.

CARTER, R.; ALDRIDGE, S.; PAGE, M.; PARKER, S. **O livro do cérebro, 4 vols**. Rio de Janeiro: Agir, 2012.

DAMÁSIO, A. R. **A estranha ordem das coisas: a vida, os sentimentos e as culturas humanas**. Porto: Círculo de Leitores, 2017.

DAMÁSIO, A. R. **E o cérebro criou o homem**. São Paulo: Companhia das Letras, 2011.

DAMÁSIO, A. R. **O erro de Descartes: emoção, razão e o cérebro humano**. São Paulo: Companhia das Letras, 2012.

DE BRITO, A. C. Neurociência e livre-arbítrio entre a dogmática penal e a política criminal. In: **Neurociência e Direito Penal**. Atlas, 2014. p. 111-142.

GUARAGNI, F. A.; GUIMARÃES, R. R. C. Neurociência, livre-arbítrio e Direito Penal: precipitação científica e alternativas para sustentação da culpabilidade. In: **Neurociência e Direito Penal**. Atlas, 2014. p. 165-214.

NOVAES, C. Phineas Gage Legendado 1. YouTube, nov 2017. Disponível em: <https://www.youtube.com/watch?v=m-m8xyoyQL8>. Acesso em: abril de 2025

ORGANIZAÇÃO DAS NAÇÕES UNIDAS [ONU]. Convenção sobre os Direitos das Pessoas com Deficiência. Nova Iorque, 13 dez. 2006.

ZAFFARONI, E. R. **A questão criminal**. Rio de Janeiro: Revan, v. 131, 2013.

15.

ANÁLISE DA RESPONSABILIDADE JURÍDICA CIVIL QUANDO DO USO DE FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL NO PROCESSO DE TOMADA DE DECISÃO EM CONFLITOS (BIO) ÉTICOS⁵⁷

Gláucio Inácio da Silveira

Débora Fernandes Pessoa Madeira

Andréia Patrícia Gomes

1. INTRODUÇÃO

1. Cada vez mais os desafios oriundos do desenvolvimento da tecnologia e de sua utilização nos cenários da prática médica vem exigindo o trabalho de diferentes áreas do conhecimento, a fim de produzir respostas mais adequadas, ainda que temporárias, para as questões quotidianas.

O entrelaçar das áreas da saúde e da ciência da computação tem sido marcante quando se pensa na utilização de ferramentas de inteligência artificial (IA) para a otimização de processos de tomada de decisão no âmbito técnico da atenção à saúde, como no diagnóstico por imagem ou no estabelecimento de escores de gravidade, mas também em situações em nível gerencial do sistema.

Mais recentes são as proposições dos recursos da ciência da computação em processos de tomada de decisão ética, considerando que essas decisões provêm de raciocínios complexos e que podem envolver valores e princípios, mas, por outro lado, podem ser facilitadas pelos conhecimentos acerca dos códigos, teorias e conceitos relacionados ao campo da ética e à opinião das correntes de pensamento, ou mesmo dos bioeticistas mais experientes. Assim, a partir desta prática, as implicações éticas, mas

57 Andréia Patrícia Gomes redigiu os itens 1, 3.2 e 4.1; Débora Fernandes Pessoa Madeira, os itens 3.1, 4.2 e 5; e Gláucio Inácio da Silveira, o item 2. Todos participaram da estruturação e da revisão do texto.

também do campo do direito emergem trazendo perguntas acerca das possíveis implicações jurídicas da utilização destes recursos como aparato na tomada de decisão no âmbito da relação entre médicos e pacientes.

2. Este capítulo surge, por conseguinte, da necessidade de responder a algumas perguntas fundamentais da relação médico-computador-paciente: quem deve tomar a decisão: o médico ou o sistema computacional? Quem é o responsável pela decisão tomada: o médico, o sistema ou quem criou/treinou o sistema? Se o médico for subordinado a um hospital ou a uma clínica particular, a responsabilidade poderá ser atribuída à entidade privada? Além disso, o que dizer nos casos em que o médico atue como profissional credenciado por um plano de saúde? E, ainda, a quem é atribuída a responsabilidade se o uso do sistema computacional de apoio à decisão se der no âmbito do serviço público de saúde? Em caso de não adesão à sugestão do sistema, há alargamento do ônus argumentativo imposto ao médico?

Certamente estas não são questões simples, mas o que se propõe aqui é o seu enfrentamento, já avisando aos leitores que as respostas resultam de leituras particulares dos autores, que são provisórias, que o especialista em direito poderá e deverá se aprofundar cada vez mais para que outras respostas sejam alcançadas. O objetivo é abrir um caminho de conversa, de diálogo entre as áreas do conhecimento, almejando atender às necessidades dos médicos e dos cientistas da computação e sabe-se que certamente não se alcançará o que os colegas juristas poderão precisar para seu processo de tomada de decisão. Pretende-se abrir o debate e convidar os leitores e as leitoras para um passeio neste mundo da IA, da prática médica e do direito.

2. A RESPONSABILIDADE CIVIL E A PRÁTICA MÉDICA: CONSTRUINDO FUNDAMENTOS

2.1. Relação médico-paciente: caracterização jurídica

3. A relação jurídica que vincula o médico ao paciente, quando estabelecida no âmbito do direito privado, resulta da celebração de um negócio jurídico, que possui, para parte dos estudiosos, a natureza de um *contrato de prestação de serviços* (GONÇALVES, 2010). Desse contrato emerge, como principal obrigação para o médico, o dever de desempenhar sua atividade em benefício do paciente, empregando todo o seu conhecimento, bem como as técnicas e os tratamentos adequados e disponíveis.

Todavia, mesmo os partidários dessa corrente reconhecem que o contrato celebrado entre o médico e o paciente é dotado de peculiaridades que o diferenciam da maioria dos negócios jurídicos que se enquadram no modelo típico da prestação de serviços. Essas peculiaridades consubstanciam-se, notadamente, em uma série de deveres extrapatrimoniais que se agregam ao dever principal; e esses deveres acessórios (ou anexos), impostos ao médico pela ordem jurídica, conferem contornos diferenciados ao contrato. Diante dessa constatação, revela-se defensável a posição consoante a qual referido contrato não se enquadraria, com perfeição, na moldura legal delineada para a prestação de serviços, caracterizando-se, na verdade, como *contrato atípico* (TEPEDINO; TERRA; GUEDES, 2020), cuja celebração se fundamenta no princípio da autonomia privada e é expressamente autorizada pelo Código Civil (CC), art. 425 (BRASIL, 2002).

Independente da linha de pensamento adotada, uma conclusão é inafastável: a relação jurídica que vincula o médico ao paciente, quando estabelecida no âmbito do direito privado, nasce de um *negócio jurídico de natureza contratual*.

4. A *disciplina legal* dessa relação jurídica não se encontra apenas no Código Civil. Normas contidas no microssistema do Código de Defesa do Consumidor também incidem sobre o contrato e sobre a relação jurídica por ele estabelecida.

Muitos dos deveres impostos ao médico e integrantes dessa relação jurídica não são enunciados expressamente pelas manifestações de vontade das partes. De modo diverso, conquanto integrem a relação jurídica contratual, eles são impostos diretamente por lei, encontrando sua fonte, sobretudo, no Código Civil e no Código de Defesa do Consumidor. Ao impor esses deveres, limitando a autonomia privada, a ordem jurídica considera a vulnerabilidade do paciente e procura assegurar igualdade substancial entre as partes do contrato. São justamente esses deveres que dão ao contrato a feição peculiar há pouco mencionada.

A norma que constitui a principal fonte desses *deveres acessórios* é o *princípio da boa-fé objetiva*, cláusula geral estabelecida pelo CC, art. 422 (BRASIL, 2002) e pelo Código de Defesa do Consumidor, art. 4.º, inciso III (BRASIL, 1990). Nesse sentido, ensina Couto e Silva (2006, p. 33): “A boa-fé possui múltiplas significações dentro do direito [...] Com relação ao [direito]

das obrigações, manifesta-se como máxima objetiva que determina aumento de deveres além daqueles que a convenção explicitamente constitui”.

Deve-se considerar ainda a densificação do princípio da boa-fé objetiva, por intermédio da previsão de alguns deveres acessórios, de modo expresso, pelo Código Civil, pelo Código de Defesa do Consumidor e, no plano infralegal, pelo Código de Ética Médica. Sem ter a pretensão de proceder a exame exaustivo, indiquem-se, em breve sumário, alguns dos mais destacados *deveres anexos* impostos ao médico por esses diplomas jurídicos: (i) o dever de informar;⁵⁸ (ii) o correlato dever de aconselhar; (iii) o dever de obter o consentimento informado do paciente;⁵⁹ (iv) o correlato dever de respeitar a autodeterminação do paciente;⁶⁰ (v) o dever de sigilo (que subsiste, inclusive, na fase pós-contratual, após cumprida a obrigação principal);⁶¹ (vi) o dever de manter-se atualizado, acompanhando a evolução técnica e científica.⁶²

5. Quanto à obrigação principal assumida pelo médico, ela constitui, em regra, uma *obrigação de meio*, e não uma *obrigação de resultado*. Isso significa que o médico não assume o dever de curar o paciente, mas sim o dever de prestar a assistência e os cuidados adequados e necessários ao tratamento, utilizando os conhecimentos científicos e suas habilidades da melhor forma possível, de modo a maximizar as chances de produção do resultado desejado (PEREIRA, 2018).

Essa regra, contudo, não é absoluta, havendo situações excepcionais em que o médico assume o dever de produzir um resultado específico. A título ilustrativo, pinça-se o debate em torno da cirurgia plástica. Predomina, na jurisprudência do Superior Tribunal de Justiça (STJ), o entendimento consoante o qual a cirurgia plástica exclusivamente estética envolve uma obrigação de resultado, enquanto a cirurgia plástica reparadora, uma obrigação de meio. Guiando-se pelas mesmas premissas, o STJ entende que, nas cirurgias plásticas com finalidade simultaneamente estética e reparadora, a

58 CDC, art. 6.º, inciso III (BRASIL, 1990); Código de Ética Médica – CEM, arts. 34, 44 e 88 (CFM, 2018).

59 CEM, arts. 22, 101, *caput* e § 1.º, e 102, parágrafo único (CFM, 2018).

60 CC, art.15 (BRASIL, 2002); CEM, Capítulo I, inciso XXI, e arts. 24, 31, 41, parágrafo único, e 42 (CFM, 2018).

61 CEM, Capítulo I, inciso XI, e arts. 73 a 79 (CFM, 2018).

62 CEM, Capítulo I, inciso V (CFM, 2018).

obrigação assumida pelo médico será de resultado, naquilo que se associa apenas à finalidade estética, e de meio, naquilo que se associa à finalidade reparadora. Esses entendimentos encontram-se consagrados nos acórdãos que decidiram, respectivamente, o AgRg nos EDcl no AREsp 328.110/RS (BRASIL, 2013a), o REsp 1.046.632/RJ (BRASIL, 2013b) e o REsp 819.008/PR (BRASIL, 2012a).

Destaque-se que a regra pode ser afastada ainda pela manifestação do médico que, no exercício da autonomia privada, assume o dever de produzir o resultado (GONÇALVES, 2010). Essa transmigração do campo das obrigações de meio para o campo das obrigações de resultado pode decorrer do modo como o médico desempenha seu dever de informação e aconselhamento. Com efeito, a obrigação será de resultado se as informações prestadas pelo médico puderem induzir o paciente à compreensão de que o resultado será alcançado. Alerta-se, contudo: esse exame deve ser realizado de modo objetivo, ou seja, para que se opere a transmigração, o discurso do médico deve conter elementos dos quais se possa inferir, objetivamente, a sugestão de que o resultado desejado pelo paciente será alcançado. É esse o entendimento recomendado pela boa-fé objetiva, enquanto cânone de interpretação dos negócios jurídicos, nos termos do CC, art. 113, *caput* e § 1.º (BRASIL, 2002).

6. É preciso considerar ainda a evolução organizacional da prestação dos serviços médicos. No âmbito da rede privada de saúde, além do médico e do paciente, outros sujeitos participam da atividade, formando-se uma *rede de contratos*.

Tepedino, Terra e Guedes (2020) identificam os diversos contratos celebrados, apontando: (i) o contrato entre o médico e o paciente, cujo objeto principal foi descrito acima; (ii) o contrato entre a clínica ou o hospital privado e o paciente, que tem por objeto a hospedagem em condições adequadas ao tratamento, envolvendo prestações “[...] atinentes aos exames, à enfermagem, aos aparelhos e à estrutura de apoio ambulatorial [...]” (p. 195); (iii) o contrato entre a clínica ou o hospital privado e o médico, para a realização do tratamento.

Essa equação se torna ainda mais complexa pela disseminação dos planos de saúde, o que insere na malha contratual os contratos entre as operadoras de planos de assistência à saúde e os médicos filiados, além dos

contratos celebrados com os clientes (eventuais pacientes), assim como os contratos com as clínicas ou os hospitais privados (BERGSTEIN, 2017).

Diante dessa realidade, os problemas relacionados à responsabilidade médica podem envolver o exame simultâneo de diversos contratos que, embora preservem sua individualidade, encontram-se conectados, em razão da unidade econômica e da identidade teleológica que os ligam. Afinal, os diversos deveres que deles decorrem se coordenam para a obtenção de um fim comum: a prestação, em caráter privado, do serviço de assistência à saúde.

7. Por derradeiro, é preciso considerar que o serviço médico também pode ser prestado, no âmbito de uma *relação jurídica regida pelo direito público*.

Extrai-se da Constituição da República Federativa do Brasil de 1988 (CRFB) a disciplina fundamental do serviço público de saúde, cuja prestação é imposta ao Estado, por determinação dos arts. 196, 197 e 198 (BRASIL, 1988). Todavia, a CRFB não reserva ao Estado a titularidade exclusiva desse serviço. Pelo contrário, permite, nos termos do art. 199, que também a iniciativa privada se dedique à assistência à saúde (MELLO, 2013).

É justamente essa permissão constitucional que justifica, no plano normativo, a existência da rede privada de saúde e as relações jurídicas de natureza contratual anteriormente descritas.

Já as relações travadas no âmbito da rede pública de saúde, representando a prestação de um serviço público, sujeitam-se ao regime jurídico de direito público e não possuem, portanto, natureza contratual. Tal distinção é relevante, produzindo reflexos no campo da responsabilidade civil, como será oportunamente destacado.

8. Examinados os contornos jurídicos da relação médico-paciente em suas linhas fundamentais, passa-se ao exame da responsabilidade civil, iniciando pelo respectivo conceito.

2.2 Conceitos: responsabilidade e responsabilidade civil

9. Na linguagem jurídica, o vocábulo *responsabilidade* designa o gênero dentro do qual se alojam, como espécies, a *responsabilidade civil*, a *responsabilidade administrativa* e a *responsabilidade penal*.

Em qualquer dessas esferas, a responsabilidade é, em essência, um *dever*. Trata-se de dever imputado (isto é, atribuído) por uma norma jurídica àquele que pratica uma conduta ou desempenha uma atividade da qual decorra um dano injusto a um bem jurídico de outrem.

Cabe ao legislador escolher *quais bens* são merecedores de tutela jurídica. Do mesmo modo, também cabe ao legislador a escolha das *condutas merecedoras de reprovação*. Por fim, cabe ainda ao legislador, em função da importância do bem jurídico e do grau de reprovação da conduta, definir a *extensão da proteção* a ser conferida.

Dessa forma, pode o legislador, desejando ampliar a proteção a determinado bem jurídico, imputar múltiplos deveres diante de uma mesma conduta lesiva. Em outros termos, é possível que uma mesma conduta deflagre, simultaneamente, a responsabilidade civil, a responsabilidade administrativa e, até mesmo, a responsabilidade penal.

Quanto a esta última, vale registrar que somente é imposta em caráter subsidiário. O direito penal é a *ultima ratio*, ou seja, o legislador somente confere a proteção penal a um determinado bem, quando percebe a insuficiência da proteção concedida pelos demais ramos do direito. Além da subsidiariedade, a proteção conferida pelo direito penal também se caracteriza pela fragmentariedade. Isso significa que, em relação a esses bens eleitos para a proteção penal, o direito penal somente se dedica a reprimir condutas dotadas de maior gravidade. Em resumo, a responsabilidade penal somente incide quando se deseje fortalecer a proteção já conferida ao bem jurídico pelas outras esferas de responsabilidade e evidenciar a maior reprovabilidade da conduta (TOLEDO, 1991).

10. Para ilustrar, examine-se o seguinte exemplo: sentindo-se mal, uma pessoa dirigiu-se ao consultório de seu médico particular; o paciente foi recebido pela secretária, que, desconfiando tratar-se de enfermidade grave, telefonou imediatamente ao médico, que não se encontrava no local; mesmo sem examinar o paciente, o médico receitou o medicamento que julgou adequado; deixou, porém, de encaminhar o paciente ao hospital, por entender, à vista do que lhe fora relatado pela secretária, não se tratar de situação de urgência ou emergência; adveio, então, o óbito do paciente, vindo a constatar-se, posteriormente, que esse fato decorreu diretamente do medicamento prescrito.

Diante dos fatos narrados, é possível verificar: (i) que a conduta do médico causou direta e imediatamente a morte do paciente; (ii) que seu comportamento se caracterizou pela negligência e pela imprudência, tendo ele deixado de observar o dever objetivo de cuidado; (iii) que essa conduta deve desencadear, no plano da responsabilidade civil, o dever de reparar aos familiares do paciente os danos materiais e morais sofridos;⁶³ (iv) que essa mesma conduta também deve desencadear, no plano da responsabilidade administrativa, o dever de submeter-se a punição disciplinar, eventualmente imposta pelo Conselho de Medicina, podendo variar da simples advertência em aviso reservado até a cassação do exercício profissional;⁶⁴ (v) que, por fim, essa conduta deve ainda desencadear, no plano da responsabilidade penal, o dever de suportar a pena cominada por lei (detenção, de um a três anos, pela prática de homicídio culposo) e eventualmente aplicada pelo juízo criminal competente.⁶⁵

Note-se, a partir do exemplo, que a finalidade perseguida pelo ordenamento jurídico, ao imputar responsabilidade ao causador do dano injusto pode ser reparatória (e é esse o caso da responsabilidade civil) ou punitiva (responsabilidade administrativa e penal).

11. O presente estudo restringe-se à responsabilidade civil, que corresponde ao dever de reparar o dano injusto, causado a um bem jurídico patrimonial ou extrapatrimonial, em decorrência de determinada conduta ou atividade.

Em regra, esse dever é imputado apenas ao próprio agente causador do dano. Excepcionalmente, porém, ele é imputado a outra pessoa que esteja ligada ao agente por alguma relação jurídica que, a critério do legislador, sirva de justificativa razoável para a extensão da responsabilidade. É o que ocorre nas hipóteses previstas pelo CC, art. 932 (BRASIL, 2002). A extensão do dever a pessoa que não é propriamente a causadora do dano tem o objetivo manifesto de ampliar a proteção jurídica concedida à vítima.

Assim, no exemplo ministrado anteriormente, se, por hipótese, o médico integrasse a equipe de uma clínica particular, dotada de personalidade jurídica própria, sendo seu consultório a ela integrado, poderiam os familiares da vítima optar pela responsabilização desta pessoa jurídica,

63 CC, arts. 948 a 951 (BRASIL, 2002), e CDC, art. 14, § 4.º (BRASIL, 1990).

64 CEM, arts. 1.º e 37, *caput* (CFM, 2018), Lei 3.268/1957, art. 22 (BRASIL, 1957).

65 Código Penal (CP), arts. 18, inciso II, e 121, § 3.º (BRASIL, 1940).

observado o disposto no CC, arts. 932, 933, 934 e 942, parágrafo único (BRASIL, 2002).

12. Delineado o conceito de *responsabilidade civil*, passa-se ao exame da dicotomia *responsabilidade subjetiva* e *responsabilidade objetiva*.

2.3. Responsabilidade subjetiva e responsabilidade objetiva

13. É tradicional a classificação que aparta a *responsabilidade subjetiva* da *responsabilidade objetiva*. O critério distintivo encontra-se no *fundamento que justifica a imputação da responsabilidade*. Na responsabilidade subjetiva, o fundamento reside na *culpa*; na responsabilidade objetiva, consiste no *risco* (TEPEDINO; TERRA; GUEDES, 2020).

14. Nas situações sujeitas ao *regime de responsabilidade subjetiva*, o que justifica a imputação da responsabilidade ao agente é a *conduta culposa* por ele adotada. Em outros termos, somente se atribui ao agente o dever de indenizar, neste regime, quando se verificar: (i) que sua conduta produziu direta e imediatamente o dano injusto e (ii) que tal conduta é caracterizada pela culpa.

Culpa é aqui conceito adotado *em sentido amplo*, para abranger tanto o dolo, quanto a culpa em sentido estrito.

Há dolo quando a conduta voluntária do agente se encontra psicologicamente dirigida à produção do dano injusto. Esse é o denominado *dolo direto*. Há dolo também quando o agente, mesmo não desejando produzir o dano injusto, assume o risco de desencadear tal resultado. Neste segundo caso, o agente atua como se fosse irrelevante a produção ou não do resultado. Para ele é indiferente a ocorrência ou não do dano injusto. Enfim, pouco lhe importa se decorrerá de sua conduta uma lesão injusta à esfera jurídica da vítima. É imprescindível, porém, que a conduta seja voluntária e que o resultado danoso, mesmo não sendo desejado, tenha sido previsto pelo agente. Neste segundo caso, tem-se o *dolo eventual*. Em qualquer das hipóteses, sendo impossível saber o que se passa no íntimo do agente, o dolo deve ser examinado à vista das circunstâncias concretas. Somente estará configurado, quando houver elementos fáticos concretos dos quais se possa inferir o dolo direto ou o dolo eventual.

Por outro lado, há *culpa em sentido estrito* quando a conduta do agente esteja caracterizada pela negligência, pela imprudência ou pela imperícia,

revelando a inobservância de um dever objetivo de cuidado. Ao contrário do que ocorre no dolo, o agente não deseja produzir o dano injusto, tampouco assume o risco de produzi-lo. Entretanto, ao descumprir um dever objetivo de cuidado, por adotar uma conduta negligente, imprudente ou imperita, acaba por produzir o dano não desejado, mas ainda assim injusto. A caracterização da culpa em sentido estrito depende da voluntariedade da conduta e da previsibilidade do dano injusto.

Ao contrário do que ocorre no campo da responsabilidade penal, aqui, na seara da responsabilidade civil, a conduta dolosa e a conduta culposa produzem, a princípio, a mesma repercussão jurídica. Afinal, nos termos do CC, art. 944, *caput*, “[a] indenização mede-se pela extensão do dano” (BRASIL, 2002). O valor da indenização é fixado, portanto, em função do dano injusto sofrido pela vítima, não dependendo, como regra, da maior ou menor reprovabilidade da conduta do agente.⁶⁶ À vista da opção legislativa, é possível perceber que, no ordenamento jurídico pátrio, a responsabilidade civil possui finalidade meramente reparatória.⁶⁷

Uma vez verificada a culpa em sentido amplo, entende-se que a conduta do agente constitui *ato ilícito*, amoldando-se à previsão do art. 186 do Código Civil (BRASIL, 2002). Destaque-se ainda que a conduta culposa representa a *violação de um dever jurídico*, pouco importando se o dever violado lhe foi imposto diretamente por lei, ou se ele decorre de um negócio jurídico (PEREIRA, 2018).

Com a finalidade de permitir um exame mais preciso da culpa, desvinculando-o de qualquer pesquisa acerca dos elementos psíquicos do agente, surgiu um *novo conceito de culpa*, o qual se vem denominando *concepção normativa (ou objetiva) da culpa*. De acordo com essa concepção, o exame da culpa não deve ser realizado em abstrato, sendo necessário observar as circunstâncias do caso concreto, para concluir se a conduta do agente é ou não culposa. À vista das circunstâncias concretas e das condições em que se encontra o agente, deve-se definir qual padrão (ou *standard*) de

66 Destaque-se, porém, que, em caráter excepcional, o juiz poderá, nos termos do art. 944, parágrafo único, reduzir o valor da indenização, quando verificar “excessiva desproporção entre a gravidade da culpa e o dano” (BRASIL, 2002).

67 Anote-se, porém, que a matéria não é pacífica entre os estudiosos. Há inclusive o registro de entendimento diverso no enunciado 379 da IV Jornada de Direito Civil do Conselho da Justiça Federal, que possui o seguinte teor: “O art. 944, *caput*, do Código Civil não afasta a possibilidade de se reconhecer a função punitiva ou pedagógica da responsabilidade civil” (2007).

comportamento deveria ser por ele observado. Sua conduta será ou não culposa, conforme se desvie do padrão de comportamento definido ou a ele se adeque. A tarefa do aplicador do direito, ao determinar o padrão concretamente aplicável, encontra fundamento jurídico no princípio da boa-fé objetiva, e pode servir-se, como instrumento para sua consecução, de diretrizes éticas e deontológicas ditadas por entidades de classe, além de outros mecanismos cuja utilização se revele adequada. Note-se que o padrão de comportamento é um conceito dinâmico, pois varia em função de circunstâncias concretas e condições do agente (TEPEDINO; TERRA; GUEDES, 2020).

Em caráter ilustrativo, observe-se que o padrão de comportamento ao qual deve adequar-se a conduta de um médico nesta segunda década do século XXI é diverso do padrão ao qual deveria adequar-se a conduta de um médico cem anos atrás. Do mesmo modo, podem aplicar-se padrões de comportamento distintos a dois procedimentos médicos similares e contemporâneos, se um deles não é realizado em condições ideais (v.g., procedimento efetuado em um hospital público que não disponha de instalações e equipamentos adequados), enquanto o outro é realizado dentro de um centro de referência em sua área. Enfim, as circunstâncias concretas se alteram no tempo e no espaço, e isso se reflete na definição dos padrões de comportamento, o que revela a dinamicidade do conceito normativo de culpa.

No regime de responsabilidade subjetiva, são *pressupostos* para a imputação da responsabilidade a *conduta culposa do agente*, o *dano injusto* e o *nexo causal*.

15. Nas situações sujeitas ao *regime de responsabilidade objetiva*, o que justifica a imputação da responsabilidade é o *risco* criado por determinada *atividade*. Neste regime, a imputação da responsabilidade independente da caracterização da culpa. Em termos mais precisos, a imputação da responsabilidade objetiva dispensa, para sua verificação, qualquer cogitação acerca da prática de uma conduta culposa.

Este regime baseia-se nas seguintes premissas: (i) certas atividades geram riscos elevados; (ii) não obstante, elas devem ser admitidas, em razão dos benefícios que trazem para a coletividade; (iii) por vezes, os riscos se transformam em danos concretos para uma pessoa ou para um grupo de pessoas; (iv) não seria justo que os benefícios da atividade fossem colhidos por todos, enquanto os prejuízos fossem sofridos apenas por um ou por

alguns; (v) surge, então, a necessidade de socialização do risco; (vi) aquele que explora a atividade de risco, ofertando-a à coletividade, deve suportar a responsabilidade pela reparação dos danos injustos provocados; (vii) afinal, se o explorador da atividade pertence ao setor privado, ele possui condições de considerar, na composição do preço do serviço ou produto ofertado ao público, os valores necessários ao pagamento de eventuais indenizações; (viii) e, se o explorador da atividade for o Estado, torna-se ainda mais evidente a socialização do risco, haja vista a fonte dos recursos que serão utilizados em eventual reparação.

Dessa forma, a responsabilidade objetiva resulta da necessidade de restabelecer-se o equilíbrio rompido pelo advento do dano injusto, sendo regida pelo “[...] princípio *ubi emolumentum, ibi onus* [...]”, segundo o qual “[...] quem tem o proveito deve correr o risco” (MARTINS-COSTA, 1991, p. 45).

Se, no regime de responsabilidade objetiva, a imputação da responsabilidade independe de qualquer consideração acerca da culpa, então não se questiona se há uma conduta lícita ou ilícita, tampouco se há ou não violação de um dever jurídico. Disso se infere que a responsabilidade objetiva pode incidir mesmo que o dano injusto decorra de um ato lícito (MELLO, 2013; TEPEDINO; TERRA; GUEDES, 2020).

Este regime de responsabilidade foi introduzido paulatinamente na ordem jurídica brasileira, à medida que se intensificou o progresso tecnológico e se consolidou a sociedade de massa, fenômenos que trouxeram consigo a ampliação dos riscos e a dificuldade de individualizar as condutas causadoras de danos, além do obscurecimento da identidade do agente causador. Diante de tal cenário, a adoção do regime de responsabilidade objetiva revelou-se uma opção legislativa tendente a assegurar maior proteção à vítima (PEREIRA, 2018; MARTINS-COSTA, 1991).

Primeiro, o regime foi inserido em leis especiais, aplicando-se, especificamente, a determinadas atividades de risco.⁶⁸ A Constituição de 1988 também estabeleceu normas especiais, prevendo, de modo casuístico, algumas hipóteses de aplicação desse regime.⁶⁹ Além disso, a CRFB agregou ao

68 Cite-se, como exemplo, o Código Brasileiro de Aeronáutica (CBA), cujos arts. 256 e 268 estabelecem a responsabilidade objetiva por danos injustos causados a passageiros e a terceiros na superfície (BRASIL, 1986).

69 Cite-se, como exemplo, a regra prevista na CRFB, art. 37, § 6.º, que estabelece a responsabilidade objetiva do Estado por danos injustos causados por seus agentes públicos

sistema os fundamentos genéricos do regime de responsabilidade objetiva, que são os *princípios da solidariedade social e da justiça distributiva*, previstos em seu art. 3.º, incisos I e III (TEPEDINO; TERRA; GUEDES, 2020), bem como o *princípio da primazia do interesse público*, previsto pelo inciso IV do mesmo art. 3.º (MARTINS-COSTA, 1991). Com o Código Civil de 2002, passou a existir, ao lado das disposições especiais, previstas em leis especiais, na CRFB e no próprio Código Civil,⁷⁰ uma previsão de caráter genérico, constante de seu art. 927, parágrafo único, dispositivo consoante o qual a obrigação de reparar o dano, independentemente de culpa, surgirá, não apenas nos casos especificados em lei, mas também “[...] quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem” (BRASIL, 2002).

No regime de responsabilidade objetiva, são *pressupostos* para a imputação da responsabilidade o *exercício habitual de uma atividade de cuja natureza resultem riscos para terceiros*, o *dano injusto* e o *nexo causal* (TEPEDINO; TERRA; GUEDES, 2020).

16. Por mais de um motivo, o *nexo causal*, pressuposto de responsabilização em ambos os regimes, adquiriu especial relevo no estudo da responsabilidade civil.⁷¹

(BRASIL, 1988).

70 Cite-se, como exemplo, a responsabilidade objetiva do empregador pelos danos injustos causados por seus empregados, extraída do CC, arts. 932, inciso III, e 933 (BRASIL, 2002).

71 Em primeiro lugar, o *nexo causal auxilia na determinação do valor da indenização*. Explica-se. O ordenamento jurídico brasileiro atribuiu caráter reparatório à responsabilidade civil; e, em decorrência dessa opção, é o *nexo causal*, e não a culpa, que determina a medida da responsabilidade. Em outros termos, é o *nexo causal* que determina a extensão do dano atribuível ao responsável e, por conseguinte, o montante da indenização, pouco importando a maior ou menor reprovabilidade da conduta (TEPEDINO; TERRA; GUEDES, 2020). Em segundo lugar, o *nexo causal permite a identificação daquele a quem se pode atribuir a responsabilidade*. Explica-se. Antes, com a ampla prevalência do regime de responsabilidade subjetiva, a ideia de culpa era utilizada para limitar a imputação do dever de indenizar apenas ao autor da conduta culposa. Quem não agisse com culpa não poderia ser responsabilizado. Todavia, ocorreu a progressiva introdução no ordenamento jurídico do regime de responsabilidade objetiva, no qual não se pode utilizar a culpa como instrumento de identificação do responsável. Em tal contexto, o *nexo causal* passa a desempenhar esse papel, servindo como instrumento de determinação daquele a quem deve ser atribuída a responsabilidade. Quem não deu causa ao dano não pode ser responsabilizado (SCHREIBER, 2009).

Nexo causal é o vínculo que estabelece uma relação de causa e efeito entre dois fatos: (i) a conduta culposa ou a atividade ensejadora do risco, como antecedente; (ii) o dano injusto, como consequente.

Schreiber alerta que o *sentido jurídico* da expressão *nexo causal* não coincide com o *sentido natural ou lógico* que se atribui à mesma expressão, de modo que “[...] nem tudo que, no mundo dos fatos ou da razão, é considerado como causa de um evento pode assim ser considerado juridicamente [...]” (2009, p. 53). O sentido jurídico é, portanto, mais restrito, impedindo a ampliação desmedida do rol de responsáveis que decorreria do *regressus ad infinitum*. É nesse contexto que entram em cena as teorias que tentam delinear o conceito jurídico do *nexo causal*. Conquanto existam múltiplas concepções teóricas, prevalece a *Teoria da causa direta e imediata*, na compreensão que lhe é dada pela *Subteoria da necessariedade da causa* (TEPEDINO; TERRA; GUEDES, 2020; SCHREIBER, 2009).

A Teoria da causa direta e imediata, consagrada pelo CC, art. 403 (BRASIL, 2002), somente considera causa do dano a conduta ou atividade que se vincule direta e imediatamente ao dano injusto (TEPEDINO; TERRA; GUEDES, 2020). Essa teoria veio restringir o conceito de *nexo causal*, admitindo como causa apenas os acontecimentos que possuam ligação mais próxima com o dano (SCHREIBER, 2009). Assim, há *nexo causal* se a conduta culposa do médico ocasiona dano ao paciente, como nos casos em que desencadeia a morte ou a impotência sexual, porquanto presente a causalidade direta e imediata. Entretanto, em sua formulação original, essa teoria se revelou demasiadamente restritiva, pois representava obstáculo à reparação do dano reflexo, mesmo naqueles casos em que tal dano estivesse intimamente ligado ao dano próximo, sendo também aquele uma consequência necessária do evento antecedente. Nos exemplos anteriores, os danos próximos sofridos pelos pacientes coexistem com danos reflexos sofridos por outras pessoas. No primeiro, sofrem danos reflexos, de natureza material e moral, os familiares do paciente morto; no segundo, também sofre dano reflexo o cônjuge que teve sua vida sexual prejudicada pela impotência de seu consorte.⁷²

72 O primeiro exemplo encontra previsão legal expressa no CC, art. 948 (BRASIL, 2002). O segundo identificado pela expressão *dano sexual* é fornecido por SCHREIBER (2009), com indicação de julgados proferidos por tribunais estrangeiros, reconhecendo o direito à indenização.

Em razão de situações como essas, surgiu uma releitura da Teoria da causa direta e imediata que se denominou *Subteoria da necessariedade causal*, concepção “[...] que entende as expressões *dano direto* e *dano imediato* de forma substancial, como reveladoras de um liame de necessariedade – e não de simples proximidade – entre a causa e o efeito” (SCHREIBER, 2009, p. 60). Assim, se o dano for efeito necessário de determinada causa, por não existir qualquer outra cadeia causal independente que interrompa a anterior, haverá nexo causal e deverá ocorrer imputação de responsabilidade (TEPEDINO; TERRA; GUEDES, 2020).

17. Delineados, em linhas gerais, os dois regimes de responsabilidade que convivem no direito brasileiro, assevere-se que, nas *relações médico-paciente regidas pelo direito privado*, a responsabilidade do profissional é subjetiva. Nesse sentido, o texto normativo do CDC, art. 14, § 4.º, expressamente, exige a verificação de culpa como condição para que seja imputada a responsabilidade aos profissionais liberais (BRASIL, 1990). Assim, para imputar-se ao médico o dever de indenizar, é preciso que sua conduta, além de provocar direta e imediatamente o dano injusto, seja movida por dolo ou esteja qualificada por negligência, imprudência ou imperícia. Vale dizer: é preciso que sua conduta configure a prática de um ato ilícito. De modo ainda mais claro, evidencie-se que o médico só pode ser responsabilizado se, diante das circunstâncias concretas, restar caracterizada a violação de um dever que integre a relação médico-paciente, sendo oportuno observar que aqui estão inclusos os denominados deveres acessórios (ou anexos), decorrentes do princípio da boa-fé. Lembre-se que o exame da conduta médica deve ser orientado pelo conceito normativo de culpa, consoante o qual somente se considera culposa a conduta que se desvia do padrão de comportamento modelado diante das circunstâncias concretas e das condições do médico. Essa consideração se mostra particularmente relevante à vista do problema examinado no presente capítulo, uma vez que o uso da inteligência artificial, como instrumento de apoio à tomada de decisão diante de conflitos éticos, produz inevitáveis reflexos sobre a modelagem do *standard* de comportamento a ser considerado.

Ainda no campo das relações de direito privado, exige equacionamento o exame da *responsabilidade dos hospitais particulares* (e o raciocínio que será desenvolvido também se aplica às *clínicas particulares*). Primeiro, é necessário considerar o *contrato de hospedagem* celebrado com o paciente.

Por ele, o hospital assume a obrigação de fornecer os recursos materiais e humanos necessários à prestação do serviço pelo médico. Trata-se de atividade cujo risco é expressamente reconhecido pelo art. 14, *caput*, do CDC (BRASIL, 1990), aplicando-se, na hipótese de eventual dano injusto (como, v.g., o dano resultante de uma infecção hospitalar que não decorra de caso fortuito), a regra de imputação própria do regime de responsabilidade objetiva. Neste caso, o hospital responde objetivamente por ato próprio, em razão do defeito do serviço prestado. Em segundo lugar, é preciso indagar se o hospital responde por danos causados pelo médico e, em caso positivo, se tal responsabilidade é subjetiva ou objetiva. Neste ponto, distinguem-se duas situações: (i) *se o médico não possui vínculo de subordinação com o hospital*, essa entidade não se responsabiliza pelos danos causados pelo profissional; (ii) por outro lado, *se existe tal vínculo de subordinação*, o hospital responde solidariamente pelos danos injustos causados pelo médico, o que é expressamente previsto pelo CC, art. 932, inciso III, e art. 942, parágrafo único (BRASIL, 2002), assim como pelo CDC, art. 34 (BRASIL, 1990). Se a responsabilidade é solidária, a vítima pode escolher entre apresentar sua pretensão indenizatória apenas contra o hospital, ou apenas contra o médico, ou ainda contra ambos. Neste caso, o hospital não responde por ato próprio, mas sim por ato de terceiro. A responsabilidade do médico continua sendo subjetiva. Entretanto, uma vez comprovada a culpa do médico, não é necessário verificar se, por parte do hospital, houve conduta culposa. Sua responsabilidade é objetiva, como expressamente determina o CC, art. 933 (BRASIL, 2002). É justamente nesse sentido a jurisprudência do STJ, como ficou consignado em acórdão recente, proferido no julgamento do AgInt no AREsp 1.595.158/MG (BRASIL, 2020).

Para findar o exame das relações travadas no âmbito do direito privado, cuide-se ainda da responsabilidade das operadoras de planos de saúde. Note-se que a operadora pode, pelo contrato, assumir a obrigação de prestar serviços médicos e hospitalares, servindo-se de *hospitais e médicos próprios*. Também pode assumir a obrigação de prestar serviços médicos e hospitalares, servindo-se de *hospitais e médicos credenciados*. Em ambos os casos, aplicam-se às operadoras as considerações formuladas acima em relação à responsabilidade dos hospitais e das clínicas particulares.⁷³ Por

73 Varia apenas o fundamento legal da responsabilidade, que não poderá ser o Código de Defesa do Consumidor quando a operadora do plano de saúde for uma entidade de

outro lado, é possível ainda que *não haja vínculo entre a operadora do plano de saúde e o hospital, tampouco entre a operadora e o médico*. Neste caso, o hospital ou o médico são escolhidos livremente pelo paciente, e a operadora assume apenas o dever de reembolso das despesas, dentro dos limites contratuais. Em tal hipótese, nenhuma responsabilidade poderá ser imputada à operadora por eventual dano causado pelo hospital ou pelo médico. É nesse sentido a jurisprudência do STJ, conforme se observa pelo exame do acórdão proferido no REsp 866.371/RS (BRASIL, 2012b).

Por fim, se o dano injusto for causado por médico da rede pública de saúde, o Estado responderá objetivamente, nos termos do art. 37, § 6.º, da CRFB (BRASIL, 1988). O dispositivo repete mandamentos de constituições anteriores que, após longa evolução jurisprudencial e legislativa, passaram a atribuir a responsabilidade civil diretamente ao ente estatal, para proteger a vítima do dano injusto causado por agente público. Vale observar, entretanto, que o Supremo Tribunal Federal (STF) vislumbra no dispositivo não apenas uma proteção à vítima, mas também uma garantia para o próprio agente público, acolhendo o entendimento que ficou conhecido como *Tese da Dupla Garantia*. Isso impõe à hipótese uma peculiaridade que inexistente quando o dano se dá no âmbito das relações privadas. O paciente somente poderá formular sua pretensão indenizatória contra o Estado, jamais contra o médico causador do dano injusto. Eventual responsabilidade civil do médico, que será subjetiva, nos termos do mesmo art. 37, § 6.º (BRASIL, 1988), somente poderá ser apurada em ação de reembolso, a ser ajuizada pelo Estado, caso tenha sido, anteriormente, condenado a indenizar o paciente. Esse entendimento foi consolidado pelo STF no julgamento do RE 1.027.633 (BRASIL, 2019).

18. Examinados os conceitos jurídicos fundamentais para a solução do problema proposto, passa-se, na sequência, ao exame da inteligência artificial e do respectivo uso na prática médica.

autogestão. Esse entendimento, que resulta da súmula 608 do STJ (BRASIL, 2018b), não chega a alterar substancialmente o tratamento dado à questão, pois, consoante o mesmo tribunal, aplica-se a essas entidades as normas do Código Civil, que, nessa matéria, seriam “[...] tão rígidas quanto a legislação consumerista” – REsp 1.644.829/SP (BRASIL, 2017).

3. A INTELIGÊNCIA ARTIFICIAL E SEU USO NA PRÁTICA MÉDICA: CONCEITOS E CONTEXTOS.

3.1. A inteligência artificial: um caminho sem volta

19. A *inteligência humana* ainda é um campo da ciência que não se apresenta como algo conhecido por completo. Todavia, é possível afirmar ser a inteligência e todo o estudo que a ela se refere como importante para o ser humano (RUSSEL; NORVIG, 2010). Ocorre que a mesma situação pode ser experimentada, explicada, sentida e respondida de modo diferente por cada pessoa. Nos dizeres de Alexandre Quaresma (2019, p. 3):

Cada pessoa, cada ser humano, cada indivíduo singular, pode perceber, experimentar, conceituar, definir e representar tais experiências de maneiras diversas entre si, desde pontos de vista díspares, subjetivos e particulares, por vezes até antagônicos e contraditórios, pois somos diversos, somos multiplicidade. Num só termo, somos semelhante-mente diferentes e diferentemente semelhantes. E, principalmente, não seguimos um esquema pré-estabelecido, não estamos limitados a um programa computacional de execução rígida qualquer, e também não somos traduzíveis numa fórmula bem formulada, como diriam os lógicos simbólicos, e muito menos ainda reduzíveis a uma equação ou algoritmo, ou mesmo a um punhado deles, para o profundo desgosto dos matemáticos contemporâneos.

O fato de não haver uma fórmula para descrever a inteligência a ponto de especificá-la, defini-la em esquadros, é importante para se compreender que nem sempre o que se visa com os sistemas computacionais inteligentes é imitar a inteligência humana. Claro, esta é sua grande inspiração, mas não se pode limitar o conceito de inteligência artificial à possibilidade de uma máquina agir conforme ou de modo semelhante à inteligência humana.

20. A *inteligência artificial* (IA) pode ser definida como a possibilidade e capacidade de uma máquina solucionar um problema que lhe foi apresentado pela análise dos algoritmos e o processamento de dados. Nos dizeres de LOBO (2018, p. 4) trata-se a IA de “um ramo da ciência da computação que se propõe a desenvolver sistemas que simulem a capacidade humana na percepção de um problema, identificando seus componentes e, com isso, resolver problemas e propor/tomar decisões”.

Não é possível ignorar a complexidade do problema e reduzir a inteligência artificial a uma abordagem conceitual única. Tendo sido elaborados

diversos conceitos ao longo da história da IA, sem que as variadas concepções sejam excludentes entre si. Nem sempre a inteligência artificial visa alcançar a capacidade da inteligência humana; há máquinas que visam alcançar exatamente o que o cérebro humano não alcançaria sozinho, no mesmo período de tempo.

Um exemplo disso dá-se no uso da IA na cardiologia. Os diagnósticos tradicionais nessa área são normalmente lentos, caros e específicos para o caso em análise. Por outro lado, utilizando-se a IA, é possível alcançar resultados significativos, com melhores resultados dos que a inteligência humana, sozinha, seria capaz de alcançar (SOUZA FILHO, *et al*, 2019).

21. O termo Inteligência Artificial foi criado por John McCarthy (McCARTHY; MINSKY; SHANNON, 2006), mas pode-se afirmar que a IA foi utilizada pela primeira vez por *Alan Turing*, em 1950. O conceito de IA introduzido por Alan Turing inspirava-se na inteligência humana, buscava-se construir uma máquina capaz de pensar. Apesar das mais diversas críticas a essa conceituação, o que se tem atualmente é que as máquinas projetadas para responder questões complexas, a partir de um armazenamento e da interpretação de um volume grande de dados, já funcionam e inclusive superam a própria capacidade humana, mas desde Turing (1950), os estudos em IA ainda não alcançaram a capacidade de expressar sentimentos e sensações, essas ainda são habilidades exclusivamente humanas. As questões apresentadas por Turing ainda estão presentes nos estudos de IA. A máquina, portanto, não pensa ou pensa? Mas que diferença na prática isso geraria? Esta questão vem acompanhando o desenvolvimento destas tecnologias ao longo de sua história.

22. Muitos estudiosos de IA buscaram a elaboração de premissas lógicas capazes de gerar soluções corretas, exatas, advindas da sua interpretação. Para problemas formais e exatos, que podem ser representados em fórmulas matemáticas, essa abordagem de IA é adequada. A lógica, todavia, funciona, na maior parte das vezes, por operações matemáticas que partem de classificação das coisas em conjuntos de elementos que, por terem caracteres semelhantes, acabam por representar numericamente uma palavra, uma operação matemática ou uma interpretação possível.

Por outro lado, é mais complexo representar matematicamente situações informais ou palavras ou situações com maior variação de significados, o que culminou, no campo da IA, na elaboração de outras estruturas

lógicas mais complexas, como a lógica fuzzy, ou difusa, em que a interpretação das palavras com significados variáveis ocorre pela descrição de conjuntos estanques entre si.

23. O estudo da IA pode dividir-se em campos diferentes e para cada um destes há uma abordagem adequada de IA. Segundo Russel e Norvig (2010) esses campos são o Processamento de Linguagem Natural, a Representação de Conhecimento, o Raciocínio Automatizado, o Aprendizado de Máquina, a Visão Computacional e a Robótica. O grande desafio, em todos os campos, é desenvolver a IA com uma racionalidade perfeita, o que implica a capacidade de processar os dados e apresentar as soluções racionalmente adequadas, de modo mais geral possível (RUSSEL; NORVIG, 2010).

É possível desenvolver a inteligência artificial usando diversas metodologias, a depender do objetivo e da aplicação prática perseguida. Nos dizeres de José Damião de Melo e outros dois autores, tem-se a inteligência artificial “como o estudo e a modelagem de sistemas e dispositivos que visem um comportamento que permita simular as capacidades humanas de decidir, raciocinar, tomar decisões e resolver situações em que existam problemas.” (MELO, *et al*, 2012, p. 1).

24. O aprendizado de máquina (AM) é um campo da IA que se desenvolve a partir de dados já acumulados, “é uma área da inteligência artificial voltada à construção de sistemas capazes de induzir hipóteses ou aproximar funções a partir da experiência acumulada em problemas.

O AM revela-se por um processo indutivo *supervisionado* ou *não supervisionado*, que pode gerar soluções compreensíveis ou não, a partir de paradigmas distintos. Desse modo, é importante compreender o processo indutivo utilizado, a possibilidade de compreensão, pelo ser humano, da solução apresentada e, por fim, o paradigma utilizado no AM.

No aprendizado supervisionado, o algoritmo indutor processa uma quantidade de exemplos de treinamento – dados fornecidos – de modo que as classes associadas aos exemplos de treinamento são conhecidas. Por outro lado, “no aprendizado não-supervisionado, o indutor analisa os exemplos fornecidos e tenta determinar se alguns deles podem ser agrupados de alguma maneira, formando agrupamentos” (MONARD; BARANAUSKAS, 2003a, p. 40), que proporcionam à máquina a elaboração de novas soluções a partir dos agrupamentos, de modo não conhecido.

Nesse sentido:

Para a indução derivar conhecimento novo representativo, os exemplos das classes têm que estar bem-definidos e ter uma quantidade suficiente de exemplos, obtendo assim hipóteses úteis para um determinado tipo de problema. Quanto mais exemplos relevantes selecionados para treinamento no indutor, mais bem classificado será o novo conjunto de dados. O objetivo do algoritmo de indução é construir um classificador que possa determinar a classe [a que] um exemplo não rotulado pertence. É possível rotular um novo exemplo devido à generalização (MATOS, *et al*, 2009, p. 7).

A questão, portanto, é saber se a solução apresentada pela máquina é compreensível, se o ser humano consegue compreender como a máquina alcançou aquela solução. Há sistemas, classificados como *caixa-preta*, em que não é possível compreender os caminhos que a máquina trilhou para gerar a solução. Há também sistemas, classificados como *sistemas orientados a conhecimento*, que criam estruturas simbólicas inteligíveis para o ser humano (MONARD; BARANAUSKAS, 2003a).

No processo de AM, podem ser utilizados cinco paradigmas distintos: simbólico, estatístico, baseado em exemplos, conexionista e genético (MONARD; BARANAUSKAS, 2003a). Neste trabalho, trataremos de duas formas de AM, redes neurais artificiais, processada pelo paradigma conexionista, e árvore de decisão, processada pelo paradigma simbólico.

O paradigma simbólico está normalmente associado a um aprendizado de máquina classificado como sistemas orientados a conhecimento, tendo em vista a compreensibilidade da solução apresentada pela máquina. O sistema computacional processa soluções a partir de exemplos e contraexemplos fornecidos de um mesmo conceito. O volume de dados processados por um sistema simbólico é pequeno e, desse modo, ao ser humano é possível o acompanhamento dos caminhos seguidos pela máquina para a tomada de decisão (MONARD; BARANAUSKAS, 2003a).

Por outro lado, o paradigma conexionista é representado tão somente pelas redes neurais artificiais, que “são construções matemáticas simplificadas inspiradas no modelo biológico do sistema nervoso” (MONARD; BARANAUSKAS, 2003a).

25. A partir do desenvolvimento, no campo da IA, das *redes neurais artificiais*, o aprendizado por máquinas tornou-se um processo complexo, inspirado no aprendizado do cérebro humano.

As RNA são sistemas computacionais – inspirados no funcionamento do cérebro humano – cujo processamento se inicia com uma fase de aprendizagem, na qual um conjunto de dados (para os quais já se conhece a resposta) é apresentado, fazendo com que as forças das conexões da rede se alterem de modo a gerar um resultado que seja o mais próximo possível daquele observado nos dados de treinamento (SIQUEIRA-BATISTA, et al, 2014, p. 457).

A primeira fase do AM conjuga os dados fornecidos e os algoritmos para produzir as respostas inteligentes. Pode-se afirmar que os “dados e os algoritmos são os insumos da IA, sem os quais seria inviável desenvolver a tecnologia” (MILHOLLAND, 2019, p. 329). Anote-se que se parte da visão da IA como um campo universal, que pode “tornar possível automatizar tarefas intelectuais e ser aplicável a qualquer atividade, desde áreas de uso geral, como aprendizado e percepção, até atividades específicas como diagnósticos de doenças, jogos e sistemas especialistas” (PEREIRA; SIMÕES, 2006, p. 1).

As primeiras RNAs possuíam uma limitação; elas somente conseguiam gerar soluções gerais, não tinham capacidade de processar resultados excepcionais e muito menos capacidade de criar soluções inovadoras, pela recombinação dos algoritmos (NAVEGA, 2001).

A partir dessa limitação e das evoluções tecnológicas, os estudiosos em IA e as empresas de tecnologia⁷⁴ que produzem modelos computacionais por meio das redes neurais artificiais foram desafiados a desenvolver o aprendizado por máquinas cada vez mais complexo, com uma capacidade cada vez mais ampla de geração de soluções e com maior inteligibilidade⁷⁵. Os sistemas computacionais de IA processam hoje um volume cada vez maior de dados⁷⁶ e, em um sistema de RNA, processam soluções imprevisíveis, à medida que conseguem tomar decisões autônomas ou semiautônomas,

74 Desde o final da década de 1990 até os dias atuais, houve um aumento nas publicações científicas na área de inteligência artificial que, até 2015, chega a 700 % (setecentos por cento) e entre 2015 e 2018 houve um aumento de 113% (cento e treze por cento) de startups que trabalham a partir de inteligência artificial. (SILVA, 2019, p. 43-45).

75 As respostas (saídas) apresentadas pela inteligência artificial, mesmo quando decorrerem de complexas operações matemáticas, devem ser compreensíveis aos destinatários da solução (SILVA, 2019, p. 46)

76 “Quanto mais dados forem inseridos, absorvidos ou tratados pela IA, maior é a capacidade de ‘racionalização’ e processamento desses dados e, por consequência, melhores são os resultados obtidos, beneficiando, em tese, os usuários de tal tecnologia” (MULHOLLAND, 2019, p. 328).

sem necessária interferência humana. Há sistemas de IA que são capazes de aferir a existência de crédito para uma pessoa física ou jurídica, de calcular o tempo de pena no âmbito da execução penal, de direcionar publicidade de consumo aos consumidores e também há sistemas de IA capazes de determinar o tratamento médico mais adequado para o bem-estar do paciente (MULHOLLAND, 2019).

26. A árvore de decisão se processa com poucas condições de resposta para um mesmo problema. Assim, para cada hipótese de resposta, a árvore vai se subdividindo em nós – nós de árvore – com outras condicionantes. Se o problema central for a identificação das condições de saúde do paciente, por exemplo, a resposta pode ser alcançada por um raciocínio binário, em que o paciente pode ser identificado como saudável ou não. Ocorre que, se o paciente não estiver bem, a cada novo atributo acerca do estado da sua saúde, novos nós, com outras possibilidades de resposta, vão se formando a fim de buscar a solução. O desenho do raciocínio sugere o de uma árvore e seus galhos. Trata-se de um formato de AM muito semelhante ao modo como o raciocínio humano processa suas decisões (MONARD; BARANAUSKAS, 2003b).

Mas as divagações metodológicas, matemáticas e computacionais acima tomam sentido quando aplicadas à integração, à resolução de problemas reais. Importante mencionar que atualmente muito se usam as RNAs e as árvores de decisão na prática médica, mas “a indução de regras ou árvores de decisão pode auxiliar médicos a compreender melhor os dados, enquanto uma Rede Neural convencional, mesmo com precisão similar, pode ser muito difícil de ser compreendida por seres humanos.” (MONARD; BARANAUSKAS, 2003b, p. 58).

27. Logo, apresentaremos, a seguir a *utilização da inteligência artificial na prática médica*, saindo do mundo computacional para o cotidiano do trabalho em saúde, a fim de tomar decisões cada vez mais adequadas ao complexo campo das profissões e ações da saúde.

3.2. O uso da inteligência artificial na prática médica

28. A incorporação das biotecnociências não é novidade na área da saúde. Como já disse Schramm (2010, p 195-196) acerca da atual situação:

[...] a biotecnociência tem algo a mais, pois é a tecnociência aplicada a organismos vivos, inclusive aos organismos humanos. Assim sendo, a diferença entre tecnociência e biotecnociência reside no fato de que a biotecnociência manipula sistemas vivos que, contrariamente aos sistemas não vivos, são sistemas autopoieticos, em princípio “renováveis”, graças ao metabolismo [...] e à reprodução [...]

Neste campo amplo e diversificado, percebem-se situações que trazem intensos debates no âmbito técnico e no âmbito ético. Deve-se estar atento que tais avanços, como a genética, a nanotecnologia e a inteligência artificial, o último mote de nosso diálogo, são três técnicas que podem trazer inúmeras benesses para a espécie humana, contudo é preciso estar atento aos perigos e malefícios, uma vez que tais tecnologias podem trazer graves riscos de utilização, podendo afetar, inclusive a sobrevivência da espécie humana (BOSTROM, 2003, p.5).

Assim, pensar em IA na prática médica é também não esquecer que o cuidado ampliado em saúde (AYRES, 2004) e que a integralidade e a humanização são pilares da arte/ciência/prática (BRASIL, 2004; REGO, 2008) e que qualquer intervenção tecnológica deverá contribuir para o aumento da qualidade desta, sempre tendo o ser humano como foco principal. Não se deve considerar em qualquer hipótese que o computador deva substituir a decisão e a prática do médico. O seu uso sempre deverá se dar como apoio, seja em processos de tomada de conduta técnica ou mais hodiernamente nas situações de conflitos éticos (MOTTA, 2020).

29. Diversos *desafios da saúde na contemporaneidade* devem ter soluções implementadas, a fim de se garantir o *acesso à saúde* como direito de todo cidadão no Brasil (BRASIL, 2010).

As novas tecnologias podem contribuir para atenuar os problemas relativos à falta de acesso ao sistema de saúde ou ao seu acesso precário, fortalecendo os sistemas de saúde e os profissionais no enfrentamento de novos desafios. Entretanto, ao se falar da introdução de novas tecnologias, fala-se da necessidade de capacitar as equipes e indivíduos para seu uso, assim como nos processos de educação permanente fortalecer a reflexão acerca dos limites e riscos oriundos da própria utilização (LAPÃO, 2019).

Por conseguinte, os desafios a serem enfrentados passam por diversos campos. O primeiro refere-se exatamente a considerar o paciente o centro da implementação de qualquer nova tecnologia, ou seja, esta deve prover

um melhor cuidado àquele. A este, associam-se, segundo Ribeiro e Oliveira (2010, p. 371-373) outros, como “a incorporação dessas novas tecnologias para diagnóstico e tratamento deve ocorrer após validação robusta de sua eficácia clínica, o uso de ferramentas digitais e algoritmos de decisão pelos pacientes deve ser opção para os pacientes que se sentem capacitados” e, finalmente, “o treinamento interdisciplinar precisará ser realizado, envolvendo profissionais de saúde, engenheiros, cientistas da computação e bioinformáticos, que minimizarão as dificuldades de implementar a nova tecnologia.” Há de se refletir também acerca dos custos e das desigualdades sociais, que podem dificultar ou mesmo impedir o acesso às tecnologias recentemente incorporadas.

30. Pode-se relatar diversos cenários onde a IA possui potencial para auxiliar a prática médica, sem, entretanto, esquecer que médicos e demais profissionais da saúde devem utilizar tais recursos sempre em benefício dos pacientes, a fim de cuidar. Mas o que significa cuidar? Etimologicamente a palavra “cuidado”, de origem latina, provém de *cogitatus*, relativo ao que é pensado e/ou refletido. Na língua portuguesa recebe os significados de atenção, preocupação, zelo ou desvelo que se dedica a alguém ou alguma coisa. Quando chega à área da saúde vem sustentado exatamente por esta relação de responsabilidade, empatia, de interação, algo subjetiva, que perpassa o trabalho em saúde, estabelecendo vínculos, proximidade e significados (CONTATORE, 2017). Assim, como observa Alfaiate (2008, p.13) “O cuidado manifesta-se nos poderes-deveres de proteção e assistência de um sujeito por outro, mediante ações concretas que se sustentam na assunção de uma consciência de responsabilidade pela melhor decisão para esse outro”.

Certamente os usos da IA para diagnóstico e prognóstico já são possíveis em diversas práticas da medicina; é essencial que, independentemente da eficiência de algoritmos, a experiência humana, a empatia, a compaixão e a responsabilidade do ato, do encontro entre médico-paciente não se percam (SANAL, 2019). São humanos cuidando de humanos, mesmo que haja mediação de máquinas, aparelhos, inteligência artificial, algoritmos ou qualquer outra tecnologia.

31. Apresentando os usos já correntes das IAs na prática médica, começar-se-á com Simões e colaboradores, que descreveram a implantação em 2018 em três hospitais portugueses de uma ferramenta de auxílio à tomada de decisão na prescrição de antimicrobianos, a fim de reduzir a ocorrência

de resistência bacteriana aos fármacos nas unidades. Os resultados encontrados pelos pesquisadores são de sucesso e redução das prescrições fora dos parâmetros recomendados com base nos perfis de sensibilidade das bactérias, do uso dos antimicrobianos no hospital e do quadro apresentado pelo paciente (SIMÕES, 2018). Outro uso encontrado foi o realizado pelo Grupo Suíço que acompanha pacientes vivendo com HIV, em uma grande coorte de mais de 12.000 pessoas, observando o desenvolvimento de doença renal crônica através do desenvolvimento de modelo de predição com base em aprendizagem por máquinas (ROTH, 2020).

Já Zhou *et al.* (2019) verificaram o emprego de inteligência artificial, com uso de redes neurais e algoritmos de aprendizagem por máquinas para avaliar lesões hepáticas focais, segmentares ou difusas com sucesso em exames de imagem. Mas o uso em Radiologia das IAs já é importante em diversos cenários como avaliação de radiografias, tomografia computadorizada, ressonância magnética e PET-CAN. O aprimoramento das imagens e a utilização de técnicas de aprendizagem de máquina para a avaliação de exames, inclusive para estadiamento oncológico e verificação de melhora com o tratamento são práticas cada vez mais frequentes, como é apontado no trabalho de Hosny (2018).

Outras áreas da medicina têm se beneficiado do uso de IAs como a cardiologia, conforme relata Souza Filho (2019), a oftalmologia (BALYEN, 2019) na detecção de doenças como a retinopatia diabética ou o glaucoma, a nefrologia com a investigação da nefropatia progressiva por Imunoglobulina A (NIEL, 2019). Os sistemas que auxiliam a tomada de decisão clínica também são utilizados nas áreas de urgência e emergência e apresentam, até agora, grande acurácia diagnóstica quando diante de sintomas e sinais clínicos, permitindo a decisão orientada, assim como o apoio à tomada de decisão terapêutica. Há também a possibilidade de uso da telemedicina, da robótica e das tecnologias de informação, reforçando este quadro de cooperação homem (médico)-máquina (computador) na resolução dos problemas de saúde da população (LOBO, 2018).

32. Ao se debater sobre qualquer questão nova e complexa, percebe-se que sempre haverá vantagens e desvantagens de seu uso. Pensando na inserção das técnicas de apoio computacional não é diferente. As vantagens emergem com grande facilidade à análise: possibilidade de trabalho colaborativo e interdisciplinar, interseção de conhecimentos, aumento da

acurácia diagnóstica em enfermidades, pesquisa *in silico* e sem o uso e abuso de animais, possibilidade de ganho no acesso à saúde, seja a especialistas ou não, pelo teleatendimento ou por cirurgia robótica.

Virando a moeda, podem-se verificar possibilidades que exigem compromisso e grande reflexão, que estão relacionados à criação e à utilização de algoritmos enganosos; falta de ética, confidencialidade e sigilo no manuseio de dados por pesquisadores, criadores de bancos e usuários, possibilitando questões profundas de responsabilidade ética como eugenia e preconceitos diversos, por exemplo com relação a deficiências ou alterações genéticas, racismo e sexismo e, finalmente, a questões legais relacionadas à responsabilidade civil, administrativa e penal daqueles que utilizam ou criam as ferramentas computacionais. Há de se ratificar que os estudos sobre as IAs são muitas vezes alicerçados em dados observacionais oriundos de bancos de dados administrativos ou registros médicos, com potencial para diferentes tipos de vieses e fatores de confusão. Logo as associações raramente atendem a critérios de causalidade. Além disso, muitos dos algoritmos funcionam com base no princípio da “caixa preta”, sem permitir que o usuário da informação saiba os motivos pelos quais um diagnóstico ou recomendação foi gerada. Questões relacionadas à ética, privacidade e segurança das informações ainda estão longe de serem resolvidas (RIBEIRO, 2019).

33. No campo da tomada de decisão ética diante de um conflito ou dilema moral, o auxílio da inteligência artificial ainda não é tão marcado na literatura. A contribuição de Mota e colaboradores, em 2016, traz importantes questões com relação à questão. A abordagem computacional tornou-se uma ferramenta importante para a assistência dos profissionais de saúde com relação ao processo de tomada de decisão em Bioética diante da complexidade dos problemas contemporâneos. As estratégias de uso das IAs para tal têm sido utilizadas por várias áreas do conhecimento desde as engenharias, às finanças, à saúde com foco na clínica, mas por que não utilizar na Bioética? Zoboli (2013, p.389) abre esta discussão do processo de tomada de decisão em Bioética clínica assim:

Nos problemas éticos da clínica é preciso apreciar a situação com o uso de procedimentos sistematizados e organizados para a tomada de decisão, visando diminuir as áreas de incerteza características dos conflitos de valores e deveres descobertos na clínica e chegar a resoluções práticas, prudentes e responsáveis.

O que diminuiria mais a área de incerteza, a área cinzenta, do que o suporte complementar oferecido por uma inteligência artificial, treinada por experts em Bioética e dotada de dados que permitissem acessar às diversas possíveis fontes que permitissem uma tomada de decisão embasada em ferramentas como princípios, correntes, normas éticas e até legais? Na Bioética casuística proposta pela autora no artigo a deliberação é exclusiva de quem toma decisão. A máquina, portanto, oferece suporte e é a autonomia do profissional que guia seu processo final de decisão. Não poderia ser diferente, já que ele é o responsável pelo cuidado do paciente. Assim, no âmbito das decisões éticas é possível a utilização de tecnologias que suportem a decisão. Apesar de ainda restrito pode-se comentar a utilização de casos clínicos simulados para treinar a tomada de decisão ética de profissionais de enfermagem e o desenvolvimento de simulação computacional para apoiar a tomada de decisão dirigida ao planejamento de ações de cuidado aos enfermos em processo de fim de vida, como proposto por Levi em 2011.

34. Para compreender um pouco melhor as proposições de discussão vindouras, é importante que sejam esclarecidos alguns pontos relevantes, a fim de que seja estabelecido um diálogo mais profícuo. Logo, é necessário comentar acerca do que é a Bioética, esta área do conhecimento que se fortaleceu após o século XX e cuja palavra tem sua etimologia oriunda de *bios* e *ethos*, ou seja, vida e costumes (SIQUEIRA-BATISTA, MOTTA, GOMES, 2020). Constitui-se em um saber inscrito no movimento típico das sociedades laicas e plurais contemporâneas, tendo por característica ser uma ética aplicada (SCHRAMM, 2015). É um campo em crescimento, com diversos eixos e diferentes abordagens, que incluem desde a reflexão ética sobre a moralidade das ações que apresentam interseção com as biotecnociências, como a genética, a clonagem, a genômica/proteômica/metabolômica, a nanociência e a nanomedicina, o aprimoramento humano e a extensão da vida, até chegar à robótica e à inteligência artificial (IA), passando pelos problemas não menos importantes de relação com animais e o meio ambiente (SIQUEIRA-BATISTA, 2008). Certamente diversos conflitos bioéticos como os relacionados à alocação de recursos, início e fim da vida são graves problemas que se dão na prática dos profissionais de saúde dentro de seu processo de trabalho.

Cabe comentar o desenvolvimento de ferramenta computacional por Siqueira-Batista e colaboradores na Universidade Federal de Viçosa, que propõe o emprego de algoritmos de aprendizagem de máquina para o

desenvolvimento de sistema computacional de apoio à tomada de decisão em Bioética clínica – envolvendo aspectos atinentes ao processo decisório (*Bio-Oracle = Organizer of the Rational Approach in Computational Learning Bioethics*) Assim, o suporte computacional pode ser utilizado pelo médico no curso de seu processo de tomada de decisão como amparo teórico para o fomento da melhor decisão naquele caso, naquele momento.

Tem surgido a possibilidade de uso de ferramentas de apoio como a anteriormente citada, que utiliza técnicas de IA com uma interface bastante amigável e de fácil interação com não especialistas. O sistema computacional utiliza uma base de dados construída a partir das respostas de especialistas, neste caso em Bioética, aplicando técnicas de aprendizagem de máquina para, dada uma questão sobre os dados, sugerir uma resposta baseada nas informações contidas no conjunto de dados inseridos. Assim, os profissionais, diante de um dilema ou conflito, poderiam lançar mão da “comissão de experts” para ajudá-los a tomar a decisão complexa à beira do leito (MOTTA, 2020). A partir da consulta ao sistema computacional, ele fornece uma resposta probabilística de qual seria a melhor decisão a ser tomada naquele contexto e diante daquele caso, o que funcionaria como uma conversa entre máquina e humano, ou melhor entre humanos e humanos mediada pela inteligência de uma máquina. Dessa interação surgem respostas que apoiarão o humano, a partir de conceitos, princípios e decisões pensadas e tomadas anteriormente por outros humanos, que não estão presentes, mas que determinaram previamente as escolhas, que serão agora processadas pelo sistema, já que permitiram o treinamento da máquina. Máquina esta que tem por base o denominado aprendizado de máquina, ou seja, algoritmos de IA que foram capazes de aprender de forma autônoma a partir daquelas informações que foram previamente fornecidas, e que neste momento podem generalizar o conhecimento adquirido para inferência de novas respostas, que são então fornecidas àquele que consulta (KERSTING, 2018).

35. Fica claro, a partir do exposto até aqui, o desafio da interdisciplinaridade necessária para a atuação na área da saúde na atualidade, já que conhecimentos dos campos da Ciência da Computação, da Matemática, da Engenharia Biomédica, da Medicina, da Ética e do Direito precisam estabelecer diálogos com o objetivo de contribuir para uma atenção mais justa, eficiente e resolutiva, inclusive ampliando-se a interface entre a pesquisa e a prática da medicina.

Torna-se também transparente que as questões éticas despontam, já que as máquinas são treinadas a partir dos humanos e que, para o bem ou mal, elas poderão fornecer respostas que auxiliem outros humanos na tomada de decisão, mas não estão livres de vieses oriundos de quem as programa ou dos dados que propiciam seu treinamento (SHADOEN, 2020). Como dizem Motta, Silva e Siqueira-Batista (2020, p. 1): “... todas as dificuldades – e aporias – do ensino da ética para os membros da espécie *Homo sapiens* não desaparecerão ao se tentar *ensinar ética* a uma máquina...”. Por conseguinte, teremos que lidar com as emergentes questões éticas e legais oriundas desse entrelace entre humanos-máquinas-humanos.

4. AS FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL COMO SUPORTE NA TOMADA DE DECISÃO NA PRÁTICA MÉDICA: DE QUEM É A RESPONSABILIDADE?

4.1. Tomada de decisão na prática médica

36. A concentração e o esforço deste capítulo se darão exatamente no entrelaçar dos conceitos outrora construídos. Para tal será necessário adentrar alguns processos que se dão diariamente na prática da Medicina, muitas vezes sem o devido rigor de reflexão e de metacognição pelos profissionais. Sobretudo quando se fala em tomada de decisão ética há defasagem, tanto no estabelecimento do problema moral quanto em seu reconhecimento, assim como o processo e os porquês que levaram ao resultado final passam despercebidos aos profissionais de saúde, havendo uma cegueira na detecção, no diagnóstico, que impede certamente o adequado plano terapêutico e a cura, como é mostrado por Siqueira-Batista e colaboradores ao estudar as situações éticas oriundas da Atenção Primária à Saúde (SIQUEIRA-BATISTA *et al.*, 2015). Começar-se-á, contudo, um passo antes com a discussão do *conceito de tomada de decisão*.

37. A *tomada de decisão* é na verdade um processo cognitivo de avaliação e julgamento, a fim de selecionar entre opções e avaliar oportunidades, chegando-se a uma resposta final. A decisão é exatamente esta resposta, que gera ações que relacionam o âmbito técnico e ético na prática clínica (ZOBOLLI, 2013). Para uma adequada decisão é preciso que haja a compreensão do quanto é multifatorial e que diversos fatores influenciarão este processo, a saber: sociais, legais e institucionais envolvidas no caso, os objetivos dos

profissionais; os padrões da boa prática; costumes da comunidade; as regras legais; políticas de saúde e até a própria formação dos profissionais de saúde. Portanto, cada decisão técnica na clínica é também permeada por uma escolha ética, mesmo que não percebida pelo tomador de decisão.

A tomada de decisão é uma competência essencial para a prática cotidiana do médico, que abrange uma série de habilidades que se integram (CARVALHO, 2015). Torna-se interessante então que a decisão seja tomada tendo por base a consideração a diversos aspectos, ou ainda devendo ser consideradas todas as alternativas possíveis, com base no uso de todas as informações disponíveis, avaliando-se cuidadosamente riscos e benefícios, de forma sólida e embasada e que sempre leve em consideração o interesse e a beneficência, mas sempre também considere a autonomia dos indivíduos, com foco no usuário (MOTTA, *et al*, 2016).

38. Para que se visualize a complexidade do campo observe-se que são inúmeras as possibilidades de *problemas éticos* na atualidade. Pode-se perceber problemas relacionados ao fim da vida como eutanásia e suicídio assistido, mas também do início da vida como reprodução assistida, eugenia, clonagem e aborto, passando por inúmeras outras facetas como a ética em pesquisa, a alocação de recursos, enfim uma infinidade de situações em que profissionais de saúde muitas vezes sozinhos na sua atuação. Diante desta situação têm sido propostas alternativas de suporte para o auxílio dos profissionais que estão na linha de frente dos cuidados à saúde. Logo, diante dessas possibilidades podem ser utilizadas, segundo Motta (2016), tanto a criação e consulta a comissões de (bio)ética, quando possível, quanto a aplicação da abordagem computacional. A discussão central se foca exatamente neste ponto, já que diante de uma decisão que se dá mediante o apoio de um sistema computacional a quem caberá a responsabilidade tanto eticamente quanto em termos legais com relação aos aspectos civis?

Quem será responsável em caso de lesão ou resultado que não dê certo? Quem toma a decisão é o sujeito que está cuidando certamente. Não há dispensa desta responsabilização, já que a ferramenta é de suporte e que a autonomia do profissional é plena, mas poderá ser responsabilizado o criador do suporte computacional? Ou aqueles que treinaram a máquina? Ou ainda os que preencheram pesquisas ou forneceram dados para elas? Essa e outras questões serão examinadas a seguir com base na legislação vigente no Estado brasileiro.

4.2. Análise da responsabilidade civil pela tomada de decisão na prática médica

39. Foi explicitado na primeira seção desta pesquisa que a responsabilidade civil médica é multifacetada, pois que a aplicação das regras jurídicas e, portanto, das soluções jurídicas variam conforme se dá, na prática, a relação médico-paciente. Assim, a responsabilidade médica pela tomada de decisão ética com apoio computacional será analisada individualmente nas situações já narradas por ocasião da primeira abordagem feita acerca da responsabilidade civil médica.

40. Antes de proceder à análise da responsabilidade civil médica, importante considerar que a relação médico-paciente não mais se sustenta com a ideia do médico como único detentor do conhecimento, dono de decisões incontestáveis. “Fundado no atual conceito de autonomia privada, construiu-se a ideia da maior participação do paciente nas decisões que envolvam a saúde de seu corpo e de sua alma” (RESENDE, 2018, p. 94). Assim, na *relação médico-paciente* da atualidade, pode-se afirmar que convivem a *autonomia* do paciente em relação à sua personalidade e a *autonomia profissional* do médico que, com base em seus conhecimentos profissionais, se propõe a elaborar a solução médica mais adequada para o cuidado com seu paciente.

Concomitante às autonomias que compõem essa relação, relevante demarcar a *vulnerabilidade* do paciente nos ambientes em que ocorrem a relação médico-paciente.

No contexto do consultório, há pacientes que buscam, com o atendimento médico, analisar as condições de sua saúde, permitindo a prevenção quanto às mais diversas doenças, ou pacientes que estão em tratamento, situação em que o médico acompanha sua evolução. Perceba que nesse cenário, a vulnerabilidade é caracterizada pela distância entre o conhecimento técnico do médico e do seu paciente, e isso, via de regra, não interfere na capacidade do paciente de exercer sua autonomia.

Todavia, observa Frederico Ferri de Rezende:

[...] o ambiente hospitalar contribui sobremaneira para agravar a vulnerabilidade do paciente. A pessoa que procura atendimento no hospital, em regra, já se encontra fragilizada, não só em virtude da patologia a que está acometida, mas também por tudo que envolve a procura pelos serviços de saúde [...] qual é a doença? Haverá internação? Por quanto

tempo? Tem cura? Os medicamentos são caros? Como farei com o trabalho, os estudos, os compromissos já anteriormente assumidos? (RESENDE, 2018, p. 96).

Apontar essa distinção quanto ao grau de vulnerabilidade, irá permitir ao leitor compreender melhor os meandros da aplicação da responsabilidade, a fim de relacionar o aspecto das conviventes autonomias no contexto da relação médico-paciente, das condições de vulnerabilidade do paciente e as obrigações médicas nos distintos ambientes.

41. Desse modo, analisar-se-á, adiante, a responsabilidade civil por danos decorrentes do uso do sistema computacional na tomada de decisão ética pelo médico, considerando, primeiro, *a relação privada com o paciente*, no âmbito do seu consultório; segundo, *a mesma relação, no ambiente hospitalar ou clínico particular*; terceiro, *a relação de direito público* que surge no âmbito da rede pública de saúde.

42. No ambiente de consultório, há decisões a serem tomadas durante o ciclo de vida das pessoas – dos pacientes atendidos – que envolvem conflitos éticos. A questão ética no tratamento de pacientes nessas circunstâncias envolve alguns elementos como a necessidade ou não de se ministrar algum medicamento ou terapia; a escolha entre realizar o tratamento/medicamento que possa prolongar a vida do paciente e, por outro lado, a análise do bem-estar do paciente, da qualidade de vida do paciente; a análise, nas circunstâncias concretas, da observância da vontade do paciente e a divergência dessa manifestação volitiva com o desejo dos familiares ou com a prescrição de tratamento pelo médico. A tomada de decisão ética no ambiente do consultório médico envolve conflitos morais que devem ser superados buscando alcançar o bem-estar do paciente (princípio da beneficência).

Como já foi dito, a relação entre o médico e o paciente que se desenvolve no âmbito do consultório médico possui natureza contratual e impõe ao médico o conjunto de deveres delineado anteriormente, no item 2.1.

Essa relação jurídica inicia-se, necessariamente, com a prestação de informações. Assim, tem-se que a obrigação de informar é “mais do que um simples elemento formal, afeta a essência do negócio, pois a informação repassada ou requerida integra o conteúdo do contrato” (MARQUES, et al, 2019, p. 337). No caso da relação médico-paciente, em virtude da vulnerabilidade do paciente, a informação deve ser prestada do modo claro e da forma mais cuidadosa possível.

The relationship between doctor and patient is often uneven: the first has knowledge concerning the problem afflicting the last. The professional is responsible for the diagnosis, prognosis and treatment options. On the other hand, the patient only has the ability to decide based on a clear and true communication performed by the doctor. (MOTTA, et al, 2016, p. 331)

A informação prestada pelo médico ao paciente é que lhe permite o exercício de autonomia e, por consequência, a tomada de decisões sobre os rumos de sua terapia, tratamento e saúde. Desse modo, a informação prestada de modo inadequado pode gerar um dano ao paciente que, por não ter compreendido adequadamente as implicações da decisão, pode também não compreender as consequências daquela decisão na sua própria vida. Por mais que a vulnerabilidade do paciente seja menor neste ambiente (se comparada com sua vulnerabilidade no ambiente hospitalar), é importante considerar o paciente como vulnerável e a compreensão da sua doença e dos efeitos de qualquer tratamento a que venha ser submetido é essencial para a autonomia e para o correto cumprimento da obrigação de informar.

Se um médico resolve instalar no computador de seu consultório o software *Bio-Oracle* como auxílio computacional para tomada de decisão de cunho ético, algumas obrigações adicionais surgem no âmbito da relação jurídica. Primeiramente, o médico passa a ter a obrigação de capacitar-se para compreender o sistema e, assim, adotá-lo em sua prática médica, a ausência de capacitação implica na não compreensão da decisão sugerida pelo software e, por decorrência, em aumento do risco de dano ao paciente pelo médico. Ato contínuo, amplia-se a obrigação de informar, pois que deve o médico explicar que está se apoiando em um sistema computacional e justificar-se diante da sua decisão de acompanhar o aconselhamento da máquina ou não.

O *Bio-Oracle* leva em consideração a expertise em Bioética de profissionais que, diante do caso apresentado, apontam que decisão tomariam na situação ética apresentada na prática médica. Percebe-se que o intuito do uso da IA é exatamente melhorar as condições de saúde e bem-estar do paciente. A Bioética, apesar de fazer parte essencial da formação do médico, foi incluída nos currículos somente a partir de 2001 com a publicação das Diretrizes Curriculares Nacionais dos cursos de graduação em Medicina (BRASIL, 2001). Assim, anteriormente os conceitos de ética ficavam restritos à disciplina de ética e deontologia e desde sempre foram desvalorizados em relação aos aspectos técnicos

na escola médica (REGO, GOMES, SIQUEIRA-BATISTA, 2008). Muito pelo contrário, como apontam Motta e outros autores, ainda é comum que os médicos tomem decisões éticas sem sequer consultar a família ou o paciente, muitas vezes tomando somente para si a responsabilidade da decisão (MOTTA, et al, 2016). Tem-se, diante disso, a situação do médico que não possui conhecimentos em Bioética e irá se apoiar na tomada de decisão indicada pelo sistema computacional como um tiro no escuro e, por outro lado, quando o médico detiver conhecimentos específicos sobre os princípios da Bioética, poderá seguir ou não a recomendação da máquina, ficando obrigado a, quando decidir refutar a sugestão computacional, fundamentar suas razões.

O apoio computacional gera, então, uma ampliação das responsabilidades médicas e o paciente pode suportar danos de uma tomada de decisão inadequada ou de uma prestação de informação não feita ou mal feita. Perceba que, na tomada de decisão com apoio computacional, a máquina não é a tomadora de decisão médica, ela inclusive amplia as chances de êxito na prática médica. Assim, a responsabilidade é do médico e não da pessoa que criou ou que controla o software.

Como também já foi mencionado neste estudo, a obrigação do médico é, via de regra, caracterizada como uma obrigação *de meio*. Nas situações da prática médica que demandam tomada de decisão ética, é muito pouco provável que um médico se comprometa a um resultado preciso, porque é exatamente a nebulosidade do diagnóstico, das condições de saúde do paciente é que tornam a decisão de cunho ético (SIMPKIN, *at al*, 2016).

Assim, caracteriza-se a responsabilidade do médico na tomada de decisão ética com apoio computacional como subjetiva. A responsabilização civil demandará prova dos elementos *dano ao paciente, conduta culposa do médico e nexo de causalidade* entre esses dois elementos.

O dano ao paciente pode ocorrer pela prescrição de um tratamento/ medicamento que venha a piorar seu quadro clínico ou, ainda que não modifique o quadro clínico, implique em piora na qualidade de vida do paciente. Neste caso, a piora na saúde do paciente pode impactar o patrimônio do paciente – dano material, mas necessariamente impacta a sua saúde, o que implica dano extrapatrimonial. Perceba que, caso a decisão quanto ao tratamento tenha partido do sistema computacional, a tomada de decisão é do profissional e não do controlador do sistema, em todas as situações. O fato de o médico ter simplesmente seguido, ou seguido porque concorda

com a decisão processada pelo sistema computacional ou, ainda, se ele, não concordando com a decisão sugerida pelo sistema, tenha elaborado sua própria decisão, não altera o fato da decisão ser sua e não da máquina.

O dano decorre do cumprimento do contrato entre o médico e o paciente e, se configurados estiverem os outros elementos configuradores da responsabilidade civil, ela deve ser imputada ao médico. E daí surge a questão: mas e se ocorrer falha no processamento dos dados e o sistema computacional indicar uma decisão equivocada? Ainda assim a responsabilidade é do médico? A resposta é sim, porque a escolha em apoiar-se em um sistema computacional foi do médico de modo que, perante a contraparte da relação jurídica, o paciente, ele responde pelos danos decorrentes da sua prestação de serviços. Como já mencionado, a decisão perante o paciente pertence ao médico!

Todavia, uma vez constatada a falha do sistema computacional, o médico pode, pela via regressa, tomar as medidas cabíveis em face do titular do software ou mesmo em face do seu programador, sejam eles pessoa física ou jurídica. Quanto mais autonomia tiver o sistema computacional, maior será a probabilidade de não se necessitar de programador para o sistema, o sistema inteligente pode tornar-se capaz de tomar decisões de casos novos a partir da análise dos dados já presentes no sistema e aqueles novos dados que porventura ele consiga captar sem o programador. Neste caso, como não há reconhecimento jurídico nos sistemas processados por IA como sujeitos de direito, responde a pessoa – natural ou jurídica – em nome de quem estiver registrado o sistema operacional. Quando, porém, há a figura do programador e a falha decorrer da sua atividade, repõe este pelos danos ocasionados (MAGRANI, *et al*, 2019).

Na primeira situação, quando o sistema é operado sem a figura do programador, o sistema processado pela IA figura-se como um produto adquirido pelo médico ou pela clínica médica. Assim, caso o produto seja defeituoso, responde o titular do sistema pelo defeito do produto. Na segunda situação, o programador realiza um serviço e também responde pela prestação de serviços defeituosa. Para Migrani e outros autores (2019), e virtude da dificuldade de se provar a conduta culposa e mesmo de se construir de modo firme uma prova da relação de causalidade, a melhor solução é considerar a responsabilidade pelos danos decorrentes do defeito do produto

– sistema de IA – e do serviço – programação do sistema de IA – como objetiva, que independe da prova do elemento culpa.

O dano é, todavia, apenas um dos elementos configuradores da responsabilidade. O dano deve decorrer de uma conduta culposa do médico. Como já mencionado, o médico deve agir de modo cuidadoso no exercício da sua profissão. A falta de cuidado do médico pode decorrer da utilização do sistema computacional de apoio à tomada de decisão sem a devida e prévia capacitação; da tomada de decisão que implique em contrariedade à autodeterminação do paciente e, de modo bastante frequente, por descumprimento do dever de informação. Nas duas primeiras circunstâncias, o dano ao paciente decorre de escolha profissional não cuidadosa. Na terceira circunstância o médico, além de descumprir o dever profissional de informar ao paciente, desrespeita também a autonomia do paciente, a possibilidade do paciente de decidir sobre sua própria saúde.

Desse modo, caso a obrigação de informação seja devidamente cumprida, com anotação na ficha do paciente ou inclusive assinatura, pelo paciente e esclarecimento, pelo médico, do Termo de Consentimento Livre e Esclarecido (TCLE),⁷⁷ o dano, consistente na piora das condições de bem-estar do paciente, não guarda relação de causalidade com a conduta médica, além de não ser possível vislumbrar na conduta médica o elemento culpa normativa, já que teria agido com o devido cumprimento da sua obrigação de cuidado. Dizendo com outras palavras, caso o médico informe ao paciente adequadamente acerca das suas condições de saúde, dos possíveis tratamentos, da sua qualidade de vida após a submissão a cada um dos tratamentos possíveis, além de informar sobre o uso do suporte computacional e as justificativas para que o médico adira ou não à decisão sugerida pela máquina, possibilitando ao paciente o exercício de sua autonomia, não haverá, consequentemente, os elementos culpa e nexo de causalidade, a piora na saúde do paciente não terá como causa direta e imediata a conduta médica.

Importante mencionar que, por força da aplicação do art. 14 do CDC e diante da vulnerabilidade do paciente, acima descrita, caberá ao médico provar que informou adequadamente ao paciente, o que reitera a necessidade médica de registrar suas ações na ficha ou prontuário do paciente ou no TCLE.

77 Segundo Stacioli (2004, p. 48 *apud* RESENDE, 2018, p. 95), o consentimento informado serve para “tornar explícita a vontade do paciente, na qual recusa ou aceita o tratamento – e sua forma – com base nas informações recebidas e seu *ethos* crítico”.

43. A relação entre o médico e o paciente pode se dar fora do ambiente do consultório médico, em uma clínica ou hospital privado. Neste caso, pode o médico ser empregado ou prestador de serviços do hospital, mantendo com este uma outra relação contratual, de emprego ou de prestação de serviços. Essa informação é importante, já que, neste ambiente, o paciente busca o atendimento da empresa médica ou hospitalar e, portanto, realiza o contrato de prestação de serviços com a clínica. Ademais, é provável que o paciente possua um contrato com uma operadora de plano de saúde, que pode manter convênio com o médico e com o hospital/clínica, ou, lado outro, pode o plano manter seus próprios médicos e hospitais/clínicas.

Trata-se de uma circunstância mais complexa, pois que, para além da relação entre o paciente e o médico, há uma série de outras relações jurídicas que a permeiam, implicando o que os estudiosos do direito contratual designam contratos conexos ou coligados. “O termo coligação contratual, em sentido amplo, significa uma ligação, um vínculo entre relações jurídicas contratuais diferentes que conformam uma operação econômica unificada” (BERGSTEIN, 2017, p. 2). É importante compreender esse fenômeno e seus efeitos na análise da responsabilidade civil que aqui se propõe a fazer.

O paciente, quando procura a prestação de serviços da clínica/hospital privado, realiza com a clínica o contrato de prestação de serviços considerado de adesão⁷⁸, de modo que a clínica, por meio de cláusulas contratuais previamente elaboradas por ela, responsabiliza-se por uma prestação de serviço médico de qualidade visando cuidar da saúde do paciente. Todavia, isso não significa que não haja relação jurídica com o médico que porventura venha atender o paciente e acompanhá-lo em seu tratamento. Por fim, muitas vezes, o que possibilitou ao paciente procurar um ambiente privado para tratar-se é a existência prévia de convênio daquele estabelecimento médico com o plano de saúde do paciente. Está tudo interligado, todos os envolvidos – clínica/hospital, médico e plano de saúde –, de modo conexo, contribuem para a realização da prestação de serviços.

Neste ambiente, a existência de um sistema computacional de apoio à tomada de decisão ética é uma decisão da clínica/hospital, do mesmo

78 A esse respeito, tem-se que os contratos de adesão, por determinarem previamente as cláusulas contratuais em favor do sujeito mais forte da relação (neste caso a clínica/hospital provado), acabam na maioria das vezes restringindo os direitos do paciente e ampliando sua vulnerabilidade.

modo que os demais aparelhos que dão suporte à prática médica. O sistema computacional compõe, na perspectiva desse estudo, o contrato de hospedagem descrito no item 2.1. Entende-se, desse modo, que caberá à empresa capacitar as pessoas que poderão utilizar a tecnologia. O fato de não haver o treinamento impacta diretamente no risco da atividade, ampliando-o.

Por outro lado, o serviço relativo ao cuidado com o paciente continua sendo prestado pelo médico ou pela equipe de saúde que o acompanha. Estes possuem a obrigação de informar ao paciente todas as decisões relativas à sua saúde, inclusive a decisão de tomar uma decisão com suporte computacional. A realização de uma anamnese de qualidade, com prestação de informações claras e precisas é que permite ao paciente exercitar sua autodeterminação, cabe ao médico e à sua equipe respeitar a vontade do paciente, quando houver condições de manifestação. O fato de o sistema computacional ter sido instalado no ambiente da clínica/hospital ter sido uma decisão da empresa, não significa que o médico não possua responsabilidade profissional quando do seu uso. Profissionalmente, o médico deve, por prudência, apenas se apoiar na decisão computacional quando conseguir compreender a sua prática médica.

Feita essa análise da complexidade e conexidade da relação contratual de prestação de serviços médico-hospitalares, passa-se à compreensão as responsabilidades decorrentes. Antes, porém, vale registrar que, para além da vulnerabilidade aumentada do paciente nesse ambiente (comparado com o ambiente de consultório), há uma vulnerabilidade jurídica decorrente da conexidade contratual acima descrita. Questiona-se, portanto: se o paciente sofrer algum dano, de quem é a responsabilidade?

A responsabilidade pela prestação de serviços realizada no âmbito da clínica/hospital é *objetiva*, já que não se pode falar, em relação a esta empresa, de conduta culposa e sim da atividade de risco que ela exerce, nos termos do que se explanou no item 2.1, acima.

A responsabilidade objetiva depende da configuração dos seguintes elementos: dano, atividade e nexo de causalidade entre a atividade e o dano ocasionado. Quanto ao elemento dano, não há distinção do dano que o paciente possa sofrer no ambiente da clínica/hospital e no ambiente do consultório: o dano é consistente em uma piora no quadro clínico do paciente ou na sua qualidade de vida ou ainda na eventual morte do paciente, que irá gerar o dano reflexo em relação aos entes familiares da pessoa falecida. O

que é importante destacar aqui é o que irá gerar o nexo de causalidade para a imputação da responsabilidade.

No caso da responsabilidade objetiva, o nexo causal não possui um conceito distinto, continua sendo o liame entre dois elementos de responsabilidade civil. Todavia, o que irá determinar a relação de causalidade é a falha no dever de segurança imposto ao prestador de serviços. Assim, a falha ou a falta de qualquer elemento que compõe o *contrato de hospedagem* implica responsabilização da clínica/hospital por ato próprio. Assim, a falta de capacitação dos profissionais que forem operar o sistema computacional para decisão médica pode configurar falha na prestação de serviços e gerar o nexo de causalidade, a ausência de um equipamento necessário para assegurar o bem-estar do paciente também, a ausência de um medicamento necessário no tratamento, todas essas situações fariam configurar a responsabilidade da empresa médico-hospitalar.

O dano ao paciente deve decorrer da falha na prestação de serviços ou da falha na conduta dos profissionais de modo que, se os equipamentos e medicamentos tivessem sido disponibilizados adequadamente; se o médico ou a equipe médica tiver agido adequadamente, o dano não ocorreria. Se a piora no quadro clínico do paciente for uma decorrência natural do seu quadro clínico, não haverá relação de causalidade.

E as operadoras de planos de saúde? Respondem? Se o estabelecimento médico for de sua titularidade, as operadoras respondem pela falha na prestação de um serviço seu, nos termos já mencionados em relação à responsabilidade da clínica/hospital. Por outro lado, quando a clínica/hospital e o médico que atender ao paciente forem conveniados, haverá responsabilidade solidária também das operadoras de planos de saúde. Cabe destacar, todavia, que se entende, neste trabalho, que o sistema computacional compõe o conjunto de equipamentos hospitalares do contrato de hospedagem e, desse modo, quando houver alguma falha técnica no sistema computacional, a responsabilidade é exclusiva da clínica, não incidindo essas regras de solidariedade que foram explanadas.

44. Quanto à responsabilidade civil pela *tomada de decisão ética feita com suporte computacional no ambiente de um hospital público*, a caracterização do dano não é diferente do que já foi tratado nas hipóteses mencionadas anteriormente. Quanto à atividade, trata-se de um serviço público e, por essa razão, o Estado responde pelos danos de modo *objetivo*, não se

podendo falar em solidariedade entre o Estado e o médico. Deste modo, havendo relação de causalidade entre o dano ocasionado e a atividade hospitalar o Estado responderá perante o paciente pelos danos. Não pode o paciente, na ocasião do dano ter sido ocasionado pela conduta médica, acionar judicialmente o médico.

45. Assim a utilização de uma ferramenta computacional de suporte para a tomada de decisão não exime o médico de suas responsabilidades de cunho ético, civil, administrativo e, inclusive penal como já apontado neste estudo. Como responsável pelo paciente e integrando uma relação regrada por princípios éticos e legais o médico deverá estar atento para a incorporação de tecnologias sempre em prol do bem-estar daquele que cuida. Nenhuma tecnologia, contudo, deve substituir seu papel de protagonista na condução do caso e na atenção integral ao paciente.

5. CONCLUSÕES

46. Este trabalho tratou da responsabilidade civil do médico por ocasião da tomada de decisão ética realizada com apoio computacional. Para tanto, foram desenvolvidos, nos dois primeiros capítulos, a base teórica acerca da responsabilidade civil no Direito Brasileiro, da inteligência artificial e da tomada de decisão ética na prática médica. Assim, o último capítulo foi dedicado à discussão, ao entrelaçamento das pesquisas teóricas. Alguns resultados importantes foram alcançados de modo a solucionar os problemas de pesquisa propostos.

47. A relação médico-paciente é caracterizada como um contrato atípico em que médico cuida e acompanha o paciente a fim de assegurar-lhe o tratamento de saúde adequado. Nesta relação foi destacada a vulnerabilidade do paciente e, partindo dessa premissa, foram elencadas as obrigações médicas no âmbito dessa relação.

48. O médico, antes visto como um detentor de conhecimentos inquestionáveis e, por isso, tomador das melhores decisões, hoje é desafiado a tomar suas decisões de maneira motivada, devendo prestar as melhores informações ao paciente, respeitando suas autodeterminações.

Algumas decisões médicas, em que estão presentes conflitos éticos, demandam um esforço importante do profissional, já que este estará diante de conflitos éticos. Nesse caso, ficou claro que a decisão deve ser guiada

pelo cuidado com a saúde e bem-estar do paciente. A melhor decisão é aquela que melhor preserva a saúde do paciente e, ao mesmo tempo, a que respeita o seu exercício de autonomia.

49. A inteligência artificial foi estudada neste ponto, discutiu-se a responsabilidade civil quando as decisões deste tipo são tomadas com apoio em sistemas computacionais inteligentes. Concluiu-se que a decisão, mesmo quando tomada com este apoio, é uma decisão do médico e não da máquina. Quando o sistema inteligente sugere uma melhor decisão, ainda assim cabe ao médico, com base em seus conhecimentos, avaliar se aquela é a melhor decisão para o caso em exame.

Todavia, quando a decisão sugerida pelo sistema computacional e acompanhada pelo paciente não tiver sido adequada para preservar a saúde e bem-estar do paciente, percebeu-se a ocorrência de um dano e a partir daí é que se discutiu a imputação da responsabilidade civil na prática médica, tendo sido analisada a responsabilidade civil do médico nos diversos ambientes.

No consultório médico, a responsabilidade médica é subjetiva e somente pode ser imputada ao médico se provados os elementos conduta culposa do médico, dano ao paciente e o nexo causal entre a conduta e o dano. Neste caso foi frisada a necessidade do médico de capacitar-se para o uso do sistema computacional e de informar ao paciente que a decisão está sendo tomada com suporte computacional.

No ambiente de uma clínica ou hospital privado, a responsabilidade pode ser atribuída a diversos atores, que solidariamente podem responder por eventual dano ao paciente. A responsabilidade da clínica ou hospital privados é do tipo objetiva e ela pode ser imputada à empresa uma vez comprovado que a atividade, por ter sido defeituosa, provocou dano ao paciente. A imputação da responsabilidade surge da relação direta e necessária da atividade com o dano ocasionado ao paciente. Quando o dano decorre do atendimento médico, o paciente pode responsabilizá-lo, mas do modo visto acima, a responsabilidade é subjetiva. Provada a conduta culposa e sua relação de causalidade com o dano ao paciente, o médico responde, mas a empresa para a qual o médico estava trabalhando – clínica médica, plano de saúde ou ambas – respondem objetivamente perante ao paciente pela conduta médica. Destacou-se, neste estudo, que o sistema

computacional integra o contrato de hospedagem e, que, portanto, cabe à clínica toda a responsabilidade pela capacitação dos usuários do sistema.

No ambiente de um hospital público, a responsabilidade civil é objetiva e quem responde, perante o dano ocasionado ao paciente, é o Estado. O que distingue a responsabilidade do ente público da responsabilidade da empresa hospitalar privada é a não possibilidade, na primeira situação, de a vítima do dano acionar o médico pelo dano.

Por fim, destacou-se que o dano ao paciente pode decorrer de uma falha no sistema computacional e, neste caso, o titular do software ou o seu programador respondem objetivamente pela falha do produto ou do serviço prestado.

50. Conclui-se, por tudo o que se expôs, que o uso da inteligência artificial na prática médica é um caminho sem volta, ela traz muito mais benefícios que malefícios. Quando este uso é adotado para a tomada de decisões éticas, o sistema computacional auxilia o profissional e pode tornar a tomada de decisão mais benéfica ao paciente.

7. REFERÊNCIAS

ALFAIATE, Ana Rita. Autonomia e cuidado. In: PEREIRA, Tânia da Silva; OLIVEIRA, Guilherme de (coordenadores). **O cuidado como valor jurídico**. Rio de Janeiro: Forense, 2008.

AYRES, José Ricardo. O cuidado, os modos de ser (do) humano e as práticas de saúde. **Saúde Soc.** 2004;13(3):16-29.

BERGSTEIN, Laís. Conexidade contratual, redes de contratos e contratos coligados. **Revista de Direito do Consumidor**, vol. 109, p. 159 – 183, jan./fev., 2017.

BOSTROM, N. A history of transhumanist thought. **J. Evol. Technol.**, v.14, n.1, p.1-25, 2005. Disponível em: <https://www.nickbostrom.com/papers/history.pdf>. Acesso em 26 jun 2020.

BRASIL. **Código Brasileiro de Aeronáutica (1986)**: Lei 7.565, de 19.12.1986. Disponível em: < http://www.planalto.gov.br/ccivil_03/LEIS/L7565.htm#art324>. Acesso em: 15 ago 2020.

BRASIL. **Código Civil (2002)**: Lei 10.406, de 10.01.2002. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/2002/L10406compilada.htm>. Acesso em: 12 ago 2020.

BRASIL. **Código de Defesa do Consumidor (1990)**: Lei 8.078, de 11.09.1990. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm>. Acesso em: 13 ago 2020.

BRASIL. **Código Penal (1940)**. Decreto-lei 2.848, de 07.12.1940. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm>. Acesso em: 13 ago 2020.

BRASIL. Conselho Nacional de Educação, Câmara de Educação Superior. **Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina**. Resolução CNE/ CES Nº. 4, Brasília, 7 de novembro, 1-6, 2001.

BRASIL. **Constituição da República Federativa do Brasil**. São Paulo: Revista dos Tribunais; 2000.

BRASIL. **Lei 3.268, de 30.09.1957**. Dispõe sobre os Conselhos de Medicina e dá outras providências. Disponível em: <http://www.planalto.gov.br/ccivil_03/LEIS/L3268.htm>. Acesso em: 13 ago 2020.

BRASIL. Ministério da Saúde. **Política Nacional de Humanização**: documento base para gestores e trabalhadores do SUS. Brasília: Ministério da Saúde; 2004. 54 p.

BRASIL. Superior Tribunal de Justiça. AgInt no AREsp 1.595.158/MG. Relator: Ministro Raul Araújo, Quarta Turma. **Diário da Justiça Eletrônico**, Brasília, 01 jun 2020. Disponível em: <https://ww2.stj.jus.br/processo/julgamento/eletronico/documento/mediado/?documento_tipo=integra&documento_sequencial=109736057&registro_numero=201902968495&peticao_numero=202000025453&publicacao_data=20200601&formato=PDF>. Acesso em: 12 ago 2020.

BRASIL. Superior Tribunal de Justiça. AgRg nos EDcl no AREsp 328.110/RS. Relator: Ministro Luís Felipe Salomão, Quarta Turma. **Diário da Justiça Eletrônico**, Brasília, 25 set. 2013a. Disponível em: <https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencial=1265389&num_registro=201301100134&data=20130925&formato=PDF>. Acesso em: 12 ago 2020.

BRASIL. Superior Tribunal de Justiça. REsp 819.008/PR. Relator: Ministro Raul Araújo, Quarta Turma. **Diário da Justiça Eletrônico**, Brasília, 29 out 2012a. Disponível em: <https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencial=1170346&num_registro=200600298640&data=20121029&formato=PDF>. Acesso em: 13 ago 2020.

BRASIL. Superior Tribunal de Justiça. REsp 866.371/RS. Relator: Ministro Raul Araújo, Quarta Turma. **Diário da Justiça Eletrônico**, Brasília, 20 ago 2012b. Disponível em: <https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencial=1134556&num_registro=200600634485&data=20120820&formato=PDF>. Acesso em: 15 ago 2020.

BRASIL. Superior Tribunal de Justiça. REsp 1.046.632/RJ. Relator: Ministro Marco Buzzi, Quarta Turma. **Diário da Justiça Eletrônico**, Brasília, 13 nov. 2013b. Disponível em: <https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencial=1266564&num_registro=200800757680&data=20131113&formato=PDF>. Acesso em: 12 ago 2020.

BRASIL. Superior Tribunal de Justiça. REsp 1.644.829/SP. Relatora: Ministra Nancy Andrighi, Terceira Turma. **Diário da Justiça Eletrônico**, Brasília, 23 fev 2017. Disponível em: <https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencial=1574879&num_registro=201602094080&data=20170223&formato=PDF>. Acesso em: 12 ago 2020.

BRASIL. Superior Tribunal de Justiça. Súmula 608. Segunda Seção. **Diário da Justiça Eletrônico**, Brasília, 17 abr 2018b. Disponível em: <http://www.stj.jus.br/docs_internet/SumulasSTJ.pdf>. Acesso em: 13 ago 2018.

BRASIL. Supremo Tribunal Federal. RE 1.027.633/SP. Relator: Ministro Marco Aurélio, Pleno. **Diário da Justiça Eletrônico**, Brasília, 06 dez 2019. Disponível em: <<http://redir.stf.jus.br/paginadorpub/paginador.jsp?docTP=-TP&docID=751571868>>. Acesso em: 12 ago 2020.

CONSELHO FEDERAL DE MEDICINA (BRASIL). **Código de Ética Médica**: Resolução CFM 2.217, de 01.11.2018. Disponível em: <<https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2018/2217>>. Acesso em: 19 ago 2020.

CONTATORE, AO; MALFITANO APS, BARROS NF. Care process in the health field: ontology, hermeneutics and teleology. **Interface (Botucatu)**. v. 21, n. 62, p. 553-63, 2017.

COUTO E SILVA, Clóvis do. **A obrigação como processo**. Rio de Janeiro: Editora FGV, 2006.

GONÇALVES, Carlos Roberto. **Direito civil brasileiro**, Volume 4: responsabilidade civil. 5. ed. São Paulo: Saraiva, 2010.

HOSNY, Ahmed; PARMAR, Chintan; QUACKENBUSH, John; Schwartz, Lawrence H.; *et al.* Artificial intelligence in radiology. **Aerts Nat Rev Cancer**. 2018; 18(8), p. 500–510. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6268174/>. Acesso em 23 june 2020.

JORNADA DE DIREITO CIVIL, IV. Organização Ministro Ruy Rosado de Aguiar Jr. Brasília: Conselho da Justiça Federal, 2007.

KERSTING, K. Machine learning and artificial intelligence: two fellow travelers on the quest for intelligent behavior in machines. *Front. Big Data*. v.1, n.6, 2018.

LAPÃO, Luís Velez. Inteligência artificial: parceira ou inimiga do médico? Einstein (São Paulo). 2019;17(2):eED4982. Disponível em: http://dx.doi.org/10.31744/einstein_journal/2019ED4982. Acesso em: 26 june 2020.

LEVI, BH; HEVERLEY, SR; VERDE, MJ. Accuracy of a decision aid for advance care planning: simulated end-of-life decision making. **J Clin Ethics**. 2011;22(3), p. 23-38.

LOBO, Luis Carlos. Inteligência Artificial, o futuro da Medicina e a Educação Médica. **Revista Brasileira de Educação Médica**, n. 42 (3), p. 3-8, 2018.

MCCARTHY, John; MINSKY, Marvin L.; ROCHESTER, Nathaniel; SHANNON, Claude E. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. **AI Magazine**, Volume 27, number 4, 2006, p. 12-14.

MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael. Novas perspectivas sobre ética e responsabilidade de inteligência artificial. In: FRAZÃO, Ana; MULHOLLAND, Caltlin. **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil, 2019, p. 115-147.

MARQUES, Cláudia Lima; BENJAMIN, Antônio Herman V.; MIRAGEM, Bruno. **Comentários ao Código de Defesa do Consumidor**. 6 ed. Ver., atual. E ampl. São Paulo: Thomson Reuters Brasil, 2019.

MARTINS-COSTA, Judith. Os Fundamentos da Responsabilidade Civil. **Revista Trimestral de Jurisprudência dos Estados**, v. 93, São Paulo: Editora Velenich, 1991, p. 29-52.

MATOS, Pablo Freire; LOMBARDI, Leonardo de Oliveira; CIFERRI, Ricardo Rodrigues; PARDO, Thiago Alexandre Salgueiro; CIFERRI, Cristina Dutra de Aguiar; VIEIRA, Marina Teresa Pires. **Relatório Técnico “Conceitos sobre aprendizado por máquinas”**. Disponível em: <https://sites.icmc.usp.br/tas-pardo/TechReportUFSCar2009b-MatosEtAl.pdf>. Acesso em: 20 ago. 2020.

MELLO, Celso Antônio Bandeira de. **Curso de direito administrativo**. 30. ed. São Paulo: Malheiros, 2013.

MELO, José Damião de; BRITO, Patrick Henrique da Silva; COSTA, Evandro de Barros. Protótipo de Sistema Especialista com o Shell Expert Sinta aplicado ao domínio do

Licenciamento Ambiental. In: **VII CONPEDI**, 2012, Palmas, Tocantins. Disponível em: <http://propi.ifto.edu.br/ocs/index.php/connepi/vii/paper/view-File/3250/950>. Acesso em: 15. Jun. 2020.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In: REZENDE, S. O. (Ed.). **Sistemas inteligentes: fundamentos e aplicações**. São Carlos: Manole, 2003a. p. 89-114.

MONARD, M. C.; BARANAUSKAS, J. A. Indução de regras e árvores de decisão. In: REZENDE, S. O. (Ed.). **Sistemas inteligentes: fundamentos e aplicações**. São Carlos: Manole, 2003b. p. 57-74.

MOTTA, Oswaldo Jesus Rodrigues da. Inteligência artificial e Bioética: apoio computacional à tomada de decisão ética do enfermeiro, em questões relacionadas ao fim de vida, na unidade de terapia intensiva. 2020. 225 f. Tese (Doutorado em Bioética, Ética Aplicada e Saúde Coletiva), **Universidade Federal do Rio de Janeiro**, Rio de Janeiro. 2020.

MOTTA, Oswaldo Jesus Rodrigues da; SILVA, Eugênio; SIQUEIRA-BATISTA, Rodrigo. “Ensinar” ética às máquinas dotadas de inteligência artificial. **Jornal da Ciência**. 27 de julho de 2020. Disponível em: <http://www.jornaldaciencia.org.br/edicoes/?url=http://jcnoticias.jornaldaciencia.org.br/30-ensinar-etica-as-maquinas-dotadas-de-inteligencia-artificial/>. Acesso em: 3 aug, 2020.

MOTTA, Luis Claudio de Souza; OLIVEIRA, Lucas Nicolau de; SILVA, Eugenio; SIQUEIRA-BATISTA, Rodrigo. Tomada de decisão em (bio)ética clínica: abordagens contemporâneas. **Revista Bioética**, v. 24, n. 2, p. 304-14, 2016. Disponível em: https://revistabioetica.cfm.org.br/index.php/revista_bioetica/article/view/1146/1514. Acesso em: 3 aug, 2020.

MULHOLLAND, Caitlin. Responsabilidade civil e processos decisórios autônomos em sistemas de inteligência artificial (IA): autonomia, imputabilidade e responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Revista dos Tribunais, 2019, p. 325-348.

NAVEGA, Sérgio C. Inteligência artificial: presente, passado e futuro. In: **Anais do Infoimagem**, Cenadem, out. 2001.

NIEL, Olivier; BASTARD, Paul. Artificial Intelligence in Nephrology: Core Concepts, Clinical Applications, and Perspectives. **Am J Kidney Dis**. 2019; 74(6), p. 803-810.

PEREIRA, Caio Mário da Silva. **Responsabilidade civil**. 12. ed. Rio de Janeiro: Forense, 2018.

PEREIRA, Mario; SIMÕES, Marco. (2006). **Um Framework para Apoio ao Ensino de Algoritmos de Busca em Árvore de Soluções para Problemas de Inteligência Artificial**. Disponível em: https://www.researchgate.net/profile/Mario_Pereira5/publication/255822275_Um_Framework_para_Apoio_ao_Ensino_de_Algoritmos_de_Busca_em_Arvore_de_Solucoes_para_Problemas_de_Inteligencia_Artificial/links/00463520bf96977947000000.pdf. Acesso em: 23 jun. 2020.

QUARESMA, Alexandre. Inteligências artificiais e o problema da consciência. **Paakat: Revista de Tecnología y Sociedad**, año 9, n. 16, marzo-agosto 2019.

REGO, Sergio; GOMES, Andréia Patrícia; SIQUEIRA-BATISTA, Rodrigo. Bioética e humanização como temas transversais na formação médica. *Revista Brasileira de Educação Médica*, 32 (4), p. 482-491, 2008. Disponível em: https://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-55022008000400011. Acesso em: 20 jun. 2020.

RESENDE, Frederico Ferri de. A vulnerabilidade do paciente no ambiente hospitalar e a preservação da autonomia privada do médico a partir do exercício

da objeção de consciência. In: SÁ, Maria de Fátima Freire de; NAVES, Bruno Torquato de Oliveira; SOUZA, Iara Antunes de. **Direito e Medicina: autonomia e vulnerabilidade em ambiente hospitalar**. São Paulo: Editora Foco, 2018, p. 93-102.

RIBEIRO, Antonio Luiz e OLIVEIRA, Gláucia Maria Moraes de. Rumo a uma cardiologia centrada no paciente e orientada a dados. **Arq. Bras. Cardiol.** v.112, n.4, p. 371-73, 2019.

RIGNEL, Diego Gabriel de Sousa; CHENCI, Gabriel Pupin; LUCAS, Carlos Alberto. Uma introdução a lógica fuzzy. **Revista Eletrônica de Sistemas de informação e Gestão Tecnológica**. 2011; 1 (1), p. 17-28. Disponível em: http://www.logicafuzzy.com.br/wp-content/uploads/2013/04/uma_introducao_a_logica_fuzzy.pdf. Acesso em 26 jun 2020.

ROTH, Jan A; RADEVSKI, Gorjan; MARZOLINI, Catia; RAUCH, Andri; *et al.* Cohort-Derived Machine Learning Models for Individual Prediction of Chronic Kidney Disease in People Living With Human Immunodeficiency Virus: A Prospective Multicenter Cohort Study **The Journal of Infectious Diseases**. 2020; 1–11. Disponível em: <https://academic.oup.com/jid/article/doi/10.1093/infdis/jiaa236/5835004>. Acesso em: 22 jun 2020.

RUSSEL, Stuart; NORVIG, Peter. **Artificial intelligence**. 3 ed. New Jersey: Pearson Education, 2010.

SANAL, M.G.; PAUL, K; KUMAR, S; GANGULY, NK. Artificial Intelligence and Deep Learning: The Future of Medicine and Medical Practice. **J Assoc Physicians India**. 2019; 67(4):71-73.

SCHRAMM, F. R. **Três Ensaios de Bioética**. Rio de Janeiro: Editora Fiocruz, 2015.

SCHREIBER, Anderson. **Novos paradigmas da responsabilidade civil: da erosão dos filtros de reparação à diluição dos danos**. 2. ed. São Paulo: Atlas, 2009.

SHADOWEN, N. **How to prevent bias in machine learning**. Disponível em: <<https://becominghuman.ai/how-to-prevent-bias-in-machine-learning-fb-d9adf1198>> Acesso em 02 jul 2020.

SIMPKIN, Arabella L.; SCHWARTZSTEIN, Richard M. Tolerating Uncertainty: the Next Medical Revolution? **The New England Journal of Medicine**, 375;18 nejm.org, november 3, 2016.

SIQUEIRA-BATISTA, R. Quantum bioethics: ethics for all beings. **Ciencia & Saude Coletiva**, v. 13, p. 1013-1015, 2008.

SIQUEIRA-BATISTA, Rodrigo; GOMES, Andréia Patrícia; MOTTA, Luís Claudio de Souza; RENNÓ, Lucas; LOPES, Tulio Correia, *et al.* (Bio)ética e Estratégia Saúde da Família: mapeando problemas. **Saúde e Sociedade**, v.4, n.1, p.113-128, 2015. Disponível em: <https://dx.doi.org/10.1590/S0104-12902015000100009>. Acesso em 31, jul, 2020.2025

SIQUEIRA-BATISTA, Rodrigo; MOTTA, Oswaldo Jesus Rodrigues da; GOMES, Andréia Patrícia. Sob as bênçãos de Jano: (bio)ética – passado, presente e futuro. 2020. No prelo

SILVA, Nilton Correia. Inteligência Artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters, 2019, p. 35-52.

SIMÕES, AS; MAIA, MR; GREGÓRIO, J; COUTO, I; *et al.* Participatory implementation of an antibiotic stewardship programme supported by an innovative surveillance and clinical decision-support system. **J Hosp Infect**. 2018; 100(3), p. 257-264.

SIQUEIRA-BATISTA, Rodrigo; GOMES, Andréia Patrícia; MAIA, Polyana Mendes; COSTA, Israel Teoldo da; PAIVA, Alcione Oliveira de; CERQUEIRA, Fábio Ribeiro. Modelos de tomada de decisão em Bioética clínica: apontamentos para a abordagem computacional. **Revista Bioética**, 22 (3), p. 456-461, 2014.

SOUZA FILHO, Erito Marques de; FERNANDES, Fernando de Amorim; SOARES, Celine Lacerda de Abreu; SEIXAS, Flavio Luiz; *et al.* Inteligência Artificial em Cardiologia: Conceitos, Ferramentas e Desafios – “Quem Corre é o Cavalo, Você Precisa ser o Jôquei”. **Arq Bras Cardiol**, 114 (4), p. 718-725, 2019. Disponível em: <https://www.scielo.br/pdf/abc/v114n4/0066-782X-abc-114-04-0718.pdf>. Acesso em: 11. jun. 2020.

TEPEDINO, Gustavo; TERRA, Aline de Miranda Valverde; GUEDES, Gisela Sampaio da Cruz. **Fundamentos do direito civil**, Volume 4: responsabilidade civil. Rio de Janeiro: Forense, 2020.

TOLEDO, Francisco de Assis. **Princípios básicos de direito penal**. 4. ed. São Paulo: Saraiva, 1991.

TURING, Allan. Computing Machinery and Intelligence. **Mind**, 49, p. 433-460, 1950. Disponível em: <https://www.csee.umbc.edu/courses/471/papers/turing.pdf>. Acesso em: 26 jun 2020.

ZHOU, Li-Qiang; WANG, Jia-Yu; YU, Jia-Yu; WU, Ge-Ge; *et al.* Artificial intelligence in medical imaging of the liver. **World J Gastroenterol**. 2019; 14; 25(6), p. 672-682. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6378542/>. Acesso em: 26 jun 2020.

ZOBOLI, Elma. Tomada de decisão em Bioética clínica: casuística e deliberação moral. **Rev. bioét.** (Impr.). v. 21, n.3, p. 389-96, 2013. Disponível em: <https://www.scielo.br/pdf/bioet/v21n3/a02v21n3.pdf>. Acesso em: 3 aug, 2020.

DIÁLOGO 01

SOBRE CIÊNCIA E MATEMÁTICA – QUESTÕES EPISTEMOLÓGICAS E ÉTICAS

Mario Livio

Eugênio Silva

Rodrigo Siqueira-Batista

Traduzido e revisado por:

Matheus Moura Novelli

Beatriz Gomes Batista

1. PRÓLOGO

Mario Livio – eminente astrofísico israelense-estadunidense – é um renomado escritor, autor de *best-sellers* sobre Ciência e Matemática. Nascido em Bucareste, na Romênia, migrou com a família para Israel aos cinco anos de idade. Diplomou-se em Física e em Matemática pela Universidade Hebraica de Jerusalém, obteve o mestrado em Ciências (Física de partículas) no Instituto Weizmann, e o doutorado em Astrofísica teórica na Universidade de Tel Aviv. No período de 1981 a 1991 foi professor de Física no Instituto de Tecnologia de Israel, antes de assumir uma posição no *Space Telescope Science Institute*, que opera o famoso Telescópio Espacial Hubble, no qual trabalhou por mais de 20 anos. Dr. Livio publicou mais de 400 artigos científicos sobre temas como cosmologia, buracos negros, explosão de supernovas, além de planetas extra-solares e o surgimento de vida no universo (LIVIO, 2020a).

O cientista popularizou a Astronomia e a Matemática por meio de livros, palestras, artigos e aparições na televisão para discussão de assuntos científicos sobre os mais diversos temas. É, também, autor de sete livros científicos populares, como *“The Golden Ratio”*, o qual discute a misteriosa relação entre os números e a natureza da realidade e lhe rendeu o Prêmio Peano, no ano de 2003, e o *International Pythagoras Prize*, no

ano de 2004 (LIVIO, 2008). A obra intitulada “*Is God a Mathematician?*”, em que Livio investiga se a Matemática é uma invenção ou uma descoberta, inspirou o programa “*The Great Math Mystery*” que foi nomeado para o prêmio EMMY, em 2016, organizado pela *International Academy of Television Arts & Sciences* (LIVIO, 2010). Além disso, seu livro “*Brilliant Blunders*” apresenta a vida de Charles Darwin, William Thomson, Linus Pauling, Fred Hoyle e Albert Einstein, descrevendo os erros e as realizações destes geniais cientistas, durante suas respectivas trajetórias intelectuais, foi selecionado pelo *The Washington Post* como um dos melhores livros do ano de 2013 (LIVIO, 2013, 2020a).

Durante as últimas três décadas, Mario Livio ministrou centenas de palestras em todo o mundo, como no *Hayden Planetarium*, em Nova York e na *Royal Astronomical Society*, em Londres. Em 2009, o Conselho da Associação Americana para o Avanço da Ciência (AAAS) o elegeu como membro dessa associação, com destaque para suas “*contribuições distintas à Astrofísica por meio de pesquisas em estrelas e galáxias e por meio da interpretação e comunicação da Ciência e Matemática ao público*” (HUBBLE, 2009). Seu último livro “*Galileo and the Science Deniers*” foi publicado em maio de 2020 e mostra uma nova interpretação sobre a vida e as descobertas de Galileu Galilei em contextos científicos e sociais modernos, discutindo sobre como foi desafiado por pessoas que negavam a ciência e acabou sendo acusado de heresia (LIVIO, 2020a, 2020b).

2. OS DIÁLOGOS

O presente diálogo acadêmico – construído com os professores Eugênio Silva e Rodrigo Siqueira Batista, com a cooperação dos estudantes Beatriz Gomes Batista e Matheus Moura Novelli – tem por objetivo principal a construção de conversações em torno de determinados pontos do pensamento do professor Mario Livio, especialmente no âmbito de questões éticas e políticas evocadas pelo avanço das Ciências Exatas e da Natureza: Astrofísica, Computação, Lógica e, claro, Matemática. Dialoga-se sobre a possibilidade da existência de outras formas de vida inteligente no Universo, os desafios éticos e epistemológicos da desigualdade – relativos aos avanços das civilizações –, a capacidade de raciocínio do *Homo sapiens* e a Inteligência Artificial, além de temas presentes em suas famosas obras,

como a negação da ciência – discutida no livro “*Galileo and the Science Deniers*” (LIVIO, 2020b) – e as relações entre Natureza e Matemática, como brilhantemente abordado nas obras “*The Golden Ratio* (LIVIO, 2008) e “*Is God A Mathematician?*” (LIVIO, 2010).

Rodrigo Siqueira-Batista: O famoso Paradoxo de Fermi faz referência a uma das questões mais intrigantes da ciência contemporânea: a possibilidade da existência de outras formas de vida – inteligente – no Universo. Ao refletir que somente a Via Láctea deve albergar cerca de um bilhão de planetas com as dimensões da Terra, e considerando o número de galáxias existentes, parece ser altamente provável que o *Homo sapiens* não esteja sozinho no Universo. O senhor poderia discorrer sobre essa questão, ponderando sobre os seguintes pontos: (i) Em que estágio estão as pesquisas em busca de vida fora do Sistema Solar? (ii) Quais são as conjecturas atuais sobre as formas de vida que porventura existam no Universo, em termos de sua estrutura química (ou seja, há possibilidade de outras modalidades de matéria viva para além das “cadeias carbônicas”)? (iii) Ao ser formulada, seriamente, a possibilidade de existência de formas inteligentes de vida em outros planetas, qual seria o *compromisso ético* dos seres humanos em relação a estes *outros* viventes?

Mario Livio: Existem muitas muitas respostas factíveis para o Paradoxo de Fermi, o que significa que não sabemos realmente a resposta. É possível que exista um “filtro” – um grande obstáculo – que não permite que as formas de vida cheguem a uma fase muito complexa (“inteligente”), nem permaneçam nessa fase por muito tempo. Este filtro pode ter estado no nosso passado (o que significaria que somos uma rara exceção tendo conseguido ultrapassá-lo), ou no nosso futuro (o que significaria que não sobreviveríamos por muito mais tempo). Também é possível que a vida biológica seja uma fase relativamente curta na evolução de civilizações inteligentes, e que seja substituída por uma vida do tipo Inteligência Artificial (IA), que não temos nenhuma ideia de como procurar, nem mesmo está confinada à superfície de um planeta. No sistema solar, os locais mais promissores para formas de vida (primitivas) são Marte, Encélado, Europa e, talvez, Titã. O carbono é único em termos de sua capacidade de formar cadeias muito complexas, portanto, é provável que toda a vida comece inicialmente sendo

baseada no carbono. É provável que nas próximas duas décadas encontraremos bioassinaturas de algumas formas de vida extra-solares simples ou sejamos capazes de colocar restrições mais significativas sobre o quão raras essas formas de vida são. As questões éticas são complicadas, porque se a vida inteligente realmente existe em outras partes da Galáxia (não podemos realmente pesquisar outras galáxias para a vida), ela pode ser muito mais avançada do que nós.

Eugênio Silva: Ao considerar a vida inteligente é importante destacar a capacidade de raciocínio organizado, um dos atributos da razão, a qual juntamente com uma forma de linguagem extremamente evoluída, parecem ser as principais características que distinguem os seres humanos dos outros animais. É claro que essas capacidades trouxeram inúmeros benefícios para as civilizações, em termos da qualidade e da expectativa de vida das pessoas. Por outro lado, esse mesmo intelecto, humano, vem provocando guerras, agindo em nome da ganância, fomentando desigualdades e miséria, promovendo a destruição do meio ambiente. Diante disso, surge a pergunta: os benefícios proporcionados pela razão humana são capazes de suplantarem seus malefícios? Ou, em outros termos, é ainda possível dar crédito ao “projeto civilizatório” da espécie humana?

Mario Livio: Não há dúvida de que o raciocínio e a ciência trouxeram enormes vantagens à vida humana. Também é verdade, entretanto, que esses mesmos avanços também produziram armas mortais e resultaram em enormes desigualdades. Acho que se essas desigualdades não forem tratadas de forma adequada, isso pode produzir um daqueles “filtros” que impedirão a civilização de existir por muito tempo. Outros riscos, como pandemias, mudanças climáticas, desafios cibernéticos, ameaças nucleares, entre outros, também podem limitar a vida das civilizações.

Eugênio Silva: O senhor, um estudioso da vida e da obra de Galileu Galilei, conhece bem os extremismos que a negação da Ciência pode provocar. Afinal, o sábio italiano por pouco não foi literalmente queimado em uma fogueira, por tentar divulgar a nova ciência. Hoje, quase 400 anos depois da morte de Galileu, em uma sociedade altamente tecnológica – ao se considerar os avanços da Robótica, da Nanotecnologia e da Inteligência

Artificial, por exemplo –, observa-se um contingente significativo de pessoas que se negam a dar crédito à ciência. Apenas para ilustrar, não raro se encontram aqueles que ainda duvidam da viagem do ser humano à Lua, que contestam a eficácia das vacinas e, até mesmo, indivíduos que chegam a afirmar que a Terra é plana. Evidentemente, esse estranho posicionamento tem uma série de consequências no âmbito da sociedade, o que pode ser notado, por exemplo, no atual negacionismo da pandemia de COVID-19, por alguns, cujas nefastas repercussões para o controle da doença têm sido observadas. Como o senhor analisaria esse cenário, de um posto de vista científico e ético-político?

Mario Livio: Infelizmente, a negação da ciência ainda é galopante, 400 anos depois que Galileu travou essa batalha e venceu! Nunca é uma boa ideia apostar contra o julgamento da ciência. Fazê-lo quando a vida humana (como no caso da pandemia COVID-19) ou o futuro da biosfera (como no caso das mudanças climáticas) está em jogo é absolutamente inaceitável. No entanto, estudos mostram que é extremamente difícil mudar opiniões fortes sobre adultos, mesmo que você os confronte com fatos claros. Portanto, a única esperança é começar com crianças pequenas e introduzir uma apreciação da ciência por meio do sistema educacional. Isso não significa que todos devem se tornar cientistas, mas que todos devem compreender que a ciência é uma parte extremamente importante da cultura humana.

Eugênio Silva e Rodrigo Siqueira-Batista: Com os mais recentes avanços no campo da Inteligência Artificial, pesquisadores têm sido encorajados a pensar que uma IA Forte talvez não esteja restrita, apenas, aos filmes de ficção científica, ou seja, há possibilidade de que máquinas inteligentes – e autoconscientes – se tornem uma realidade. De fato, à medida que os mistérios da consciência são desvendados, especialmente com alguns resultados recentes da neurobiologia e de modelos propostos a partir da Mecânica Quântica, alguns acreditam que esse conhecimento pode contribuir para a construção de agentes computacionais dotados de uma inteligência autoconsciente. Por outro lado, a modelagem de uma IA consciente estaria, inevitavelmente, alicerçada em sistemas formais, os quais, segundo os Teoremas da Incompletude de Gödel, têm suas limitações. Diante disso,

em seu ponto de vista, a construção de máquinas conscientes será apenas uma questão de tempo ou as atuais limitações lógico-matemáticas serão impeditivas para tal empreendimento?

Mario Livio: Eu pessoalmente não acho que os Teoremas da Incompletude de Gödel serão o principal obstáculo. A questão será se as máquinas de IA terão consciência e autoconsciência da mesma forma que os humanos, ou se permanecerão o que no mundo da IA eles chamam de “zumbis” (sem experiência consciente). Se sistemas suficientemente complexos têm a consciência como uma propriedade emergente, então não vejo razão para que não superem os humanos.

Eugênio Silva: As íntimas relações entre Natureza e Matemática são inegáveis, aspecto que o senhor aborda excepcionalmente bem em suas obras. A observação atenta destas relações permite afirmar a existência de alguns números que se apresentam como verdadeiros guias para a explicação de muitos dos fenômenos naturais. Dentre esses números, talvez o mais intrigante seja a seção áurea (Φ) que é subjacente a uma infinidade de padrões observados na natureza. Não menos curiosos, há números como π (π) e o número de Euler (e) que, também, estão presentes em vários fenômenos. Ao se considerar, igualmente, as diversas constantes presentes em muitas equações da Física, essa lista se torna ainda mais extensa. Nesse amplo conjunto de números intrigantes, observa-se uma característica em comum: todos são irracionais. Desde que foram descobertos, ainda nos tempos pitagóricos, os números irracionais sempre causaram estranheza pelas suas características de inexatidão e de não periodicidade. A partir dessas breves conjecturas, emergem duas perguntas: (1) Haveria algo racional – ou irracional – no fato dos padrões da Natureza se basearem nesses números “estranhos”? (2) Seria possível conjecturar sobre a existência de uma *Matemática mais profunda* que permitisse incluir essas “aparentes esquisitices” dos números irracionais, em uma estrutura simétrica de números simples e “mais bem comportados”?

Mario Livio: Embora números irracionais como π (π) apareçam em algumas das equações matemáticas que usamos para descrever as leis da natureza, a questão mais profunda é se podemos encontrar uma explicação

para os valores das constantes da natureza (por exemplo, os pontos fortes das interações fundamentais, as proporções das massas das partículas fundamentais, entre outros). Os físicos originalmente esperavam que no contexto de uma “Teoria de Tudo”, que deveria unificar a Relatividade Geral com a Mecânica Quântica, todos esses valores encontrariam uma explicação. Muitos físicos agora não têm tanta certeza. Foi sugerido que nosso universo pode ser apenas um membro de um enorme conjunto de universos – um “multiverso” –, e que os valores das constantes (e talvez até mesmo as próprias Leis da Física) podem ser diferentes em distintos universos. Segundo essa visão, nos encontramos em um universo em que os valores e as leis permitiram o surgimento da complexidade e da vida, enquanto muitos universos (talvez a maioria) não seriam capazes de produzir essa complexidade.

Rodrigo Siqueira-Batista: Sobre as simetrias, cabe mencionar que o senhor aborda o tema – nas artes, nas humanidades e nas ciências exatas, por exemplo – de forma bastante clara e bem organizada, em *The equation that couldn't be solved*. Os pontos de partida para assinalar a beleza das simetrias são as contribuições, à Matemática, dos geniais Évariste Galois e Niels Henrik Abel. Uma das questões, intrigantes, que podem ser propostas a partir das abordagens trabalhadas em seu ensaio, é a seguinte: onde “moram” as simetrias? Na mente humana ou em uma realidade “lá fora”?

Mario Livio: Parece que as leis que caracterizam nosso universo (especificamente as leis de conservação) derivam do conceito matemático de simetria. É muito difícil separar o que observamos e os conceitos matemáticos que inventamos para descrever/explicar nossas observações. Não há dúvida de que o conceito de simetria provou ser extremamente produtivo em nossas tentativas de decifrar a natureza. Muitas das simetrias parecem estar defeituosas em algum nível, mas ainda não teríamos falado sobre simetrias defeituosas, se não houvesse algumas simetrias para começar. No geral, em seu nível mais fundamental, o universo representa certas relações, e essas relações, se você preferir, constituem alguma forma de matemática. Nesse aspecto, nosso universo é certamente matemático. Isso ainda não explica totalmente o enorme sucesso do tipo particular de matemática que os humanos criaram para explicar fenômenos fundamentais (por exemplo, a precisão alcançada na Eletrodinâmica Quântica).

Eugênio da Silva e Rodrigo Siqueira-Batista: Esta última questão, de algum modo, é afim à temática tratada em seu livro *Is God a Mathematician?* – aliás, uma belíssima obra, de leitura extremamente agradável. O tema central do ensaio é a discussão acerca da ‘natureza’ da Matemática, ou seja, se este campo do conhecimento diz respeito ao âmbito da invenção ou da descoberta. Ao longo da leitura, nota-se que o senhor tenta se manter imparcial em relação à questão debatida, adotando a estratégia de apresentar os principais argumentos de ambas as perspectivas, a partir das investigações de outros estudiosos. Diante disso, é inevitável que se apresente a seguinte pergunta: de acordo com o seu ponto de vista, a Matemática é uma invenção ou uma descoberta?

Mario Livio: Ao fazer a pergunta desta forma (ou seja, “A Matemática é uma invenção ou uma descoberta?”), dá-se a impressão de que tem de ser uma ou outra e que não existe uma terceira possibilidade. Acho que a Matemática é uma combinação intrincada de invenções e descobertas. Especificamente, inventamos os conceitos e, em seguida, descobrimos as relações entre esses conceitos. Por exemplo, inventamos o conceito de triângulos ideais (inspirados por coisas que vemos na natureza) e, em seguida, descobrimos vários teoremas sobre triângulos. Ou inventamos o conceito de números imaginários (como a raiz quadrada de -1) e, em seguida, descobrimos todos os teoremas sobre números complexos. Eu sei que muitos matemáticos pensam na Matemática como uma descoberta, mas alguns (por exemplo, o falecido Michael Atiyah) concordam com a opinião que acabei de apresentar (ou melhor, eu concordo com eles). Então, para resumir, na minha humilde opinião, a Matemática não é nem uma invenção nem uma descoberta, mas sim uma intrincada combinação das duas.

3. EPÍLOGO

A oportunidade – única – de estabelecer fecunda conversação acadêmica com um cientista e escritor da estatura do Professor Mario Livio foi, realmente, bastante salutar – tendo em vista o compartilhamento do vasto conhecimento do notável astrofísico, em termos dos segredos já desvelados e dos enigmas que a ciência ainda não foi capaz de solucionar –, na perspectiva temática do presente livro: as reflexões bioéticas sobre as

Neurociências, a Inteligência Artificial, a Nanotecnologia e a Robótica. A genuína possibilidade de existência de vida extraterrestre – tendo em vista o paradoxo de Fermi (NASA, 2021) e as atuais discussões em termos, principalmente, da Astrobiologia (GALANTE, 2016) e da Astroecologia (MENDONÇA, 2014) – aumenta, sobremaneira, a responsabilidade ética do *Homo sapiens* com o Planeta Terra e, de forma mais ampla, com o Universo como um todo (SIQUEIRA-BATISTA et al., 2021). Nestes termos, as propostas éticas mais abrangentes – como a Ética da Terra (LEOPOLD, 1989) e a *(Bio)ética para todos os seres* (SIQUEIRA-BATISTA, 2020) – podem prefigurar interessantes abordagens para a condução das relações com uma eventual vida alienígena. Destaca-se, ainda, nesse contexto, o comentário do Professor Livio, sobre adicionais desafios vindouros, com destaque para aqueles relativos à possibilidade de que a “*vida inteligente em outras partes da Galáxia*” seja muito mais avançada, quando comparada à espécie humana; as repercussões desse cenário – para o bem e para o mal – necessitarão, certamente, de profunda reflexão.

A despeito desse *olhar para fora*, as questões bioéticas terráqueas, inscritas nas relações que o *H. sapiens* estabelece com seus semelhantes e com as demais formas de existência do Planeta Terra, seguem como significativos problemas que demandam enfrentamento, sob o risco – como explicitado pelo Prof. Livio – de emergência de um “*daqueles ‘filtros’ que impedirão a civilização de existir por muito tempo*”. Tal preocupação pode ser perscrutada nas próprias origens da Bioética – tanto ao se considerar a proposta de Jahr (1927), quanto a de Potter (2016) – dimensionadas em termos das relações entre ética e ciência, particularmente importantes em um contexto, atual, de reaparecimento de ideologias fascistas, em diferentes partes do mundo, as quais têm se articulado a um marcante negacionismo científico. A sensível aposta do Prof. Livio, quase freireana, é pedagógica – afinal, a “*Educação não transforma o mundo. Educação muda pessoas. Pessoas transformam o mundo*” (FREIRE, 1987, p. 87) –, ao se reconhecer que “*a única esperança é começar com crianças pequenas e introduzir uma apreciação da ciência por meio do sistema educacional*”.

A atenção para as questões da Terra impõe a preocupação com o presente, mas, igualmente, com o porvir. Nesse sentido, as implicações bioéticas instadas pelas novas tecnologias – incluindo a Inteligência Artificial e a Robótica (MOTTA et al., 2020; NASCIMENTO et al., 2018) – apontam

para um cenário que pode se tornar marcadamente distópico, quiçá como nos clássicos de ficção científica, “2001: Uma Odisseia no Espaço” (KUBRICK, 1968) e “Blade Runner” (SCOTT, 1982). As responsabilidades humanas com os robôs dotados de IA – estes serão, ou não, dignos de consideração ética? (ver Capítulo 4) – e os desafios de ensinar ética às máquinas (MOTTA *et al.*, 2020) devem merecer atenção das sociedades humanas, tendo em vista o *melhor convívio* com este outro, *maquínico*, que se anuncia. Questões epistemológicas propostas pela IA – como as dúvidas sobre se “os Teoremas da Incompletude de Gödel serão o principal obstáculo” para o surgimento de uma IA autoconsciente, ou não – são igualmente relevantes, especialmente ao se considerar um contexto no qual “*sistemas suficientemente complexos têm a consciência como uma propriedade emergente*”, quer sejam constituídos de carbono ou de silício.

A natureza da Matemática, a “*rainha das ciências*” (GARBI, 2006), permanece como uma questão chave no âmbito da teoria do conhecimento, belamente trabalhada nas obras do Prof. Mario Livio. A questão da *estranha recorrência* dos números irracionais na realidade física – por exemplo, π (π) – é abordada em termos da “*questão mais profunda*” sobre a possibilidade de se “*encontrar uma explicação para os valores das constantes da natureza (por exemplo, os pontos fortes das interações fundamentais, as proporções das massas das partículas fundamentais, entre outros)*”. No bojo dessa mesma temática, a interessante apreciação proposta para a compreensão das simetrias, reconhece as relações matemáticas fundamentais que parecem *povoar* a realidade, aspecto que permite enunciar que, em certo sentido, o “*universo é certamente matemático*”, de um modo muito próprio, dada a premissa inteligentemente formulada pelo Prof. Livio de que “*a Matemática não é nem uma invenção nem uma descoberta, mas sim uma intrincada combinação das duas*”.

Desde a Antiguidade, mulheres e homens buscam explicações racionais para os fenômenos naturais, mas, a própria Natureza parece teimar em não revelar seus mistérios facilmente – afinal, como vaticinado por Heráclito, ela *ama ocultar-se* (BORHEIM, 2008) – mas, somente à custa da inquebrantável determinação e do tenaz entusiasmo de incansáveis cientistas, os quais teimam em não desistir. É precisamente sobre estas questões – as quais, em última análise, convergem para a fantástica aventura do conhecimento – que foram construídas as presentes reflexões. As respostas serenamente

formuladas pelo Prof. Mario Livio às espinhosas questões propostas, denotam a sabedoria de um pensador que, indubitavelmente, meditou (quicá no sentido cartesiano) ao longo de *muitas luas* sobre as temáticas trabalhadas nas presentes conversações. Oportuno assinalar a significativa aprendizagem dos seus interlocutores – consubstanciada na abordagem proposta, a cada pergunta, pelo eminente astrofísico –, em platônicas *belas conversas* (PLATÃO, 1988) acerca da realidade, da experiência humana de existir e da elegância matemática da vida.

4. REFERÊNCIAS

- BORNHEIM, G.A. **Os filósofos pré-socráticos**. São Paulo: Cultrix, 2008.
- FREIRE, P. **Pedagogia do Oprimido**. Rio de Janeiro: Paz e Terra, 1987.
- GALANTE, D.; SILVA, E. P.; RODRIGUES, F.; HORVATH, J. E.; AVELLAR, M. G. B. **Astrobiologia: uma ciência emergente**. [S.l: s.n.], 2016.
- GARBI, G. G. **A Rainha das Ciências**. São Paulo: Editora Livraria da Física, 2006.
- HUBBLE Site: NASA and STScI [internet]. **STSCI senior astrophysicist Mario Livio elected AAAS fellow**. Baltimore: Maryland; 2009 [acesso em abril de 2021]. Disponível em: <https://hubblesite.org/contents/news-releases/2009/news-2009-34.html>
- JAHR, F. Bio-Ethic: eine umschau über die ethischen. Beziehungen des menschen zu tier und pflanze. **Kosmos. Handweiser für Naturfreunde**, v. 24, n. 1, p. 2-4, 1927.
- KUBRICK, S. **2001: A Space Odyssey**. EUA/Reino Unido, 1968.
- LIVIO, M. **The golden ratio: The story of phi, the world's most astonishing number**. Crown, 2008.
- LIVIO, M. **Is God a mathematician?** Simon and Schuster, 2010.
- LIVIO, M. **Brilliant blunders: From Darwin to Einstein-colossal mistakes by great scientists that changed our understanding of life and the universe**. Simon and Schuster, 2013.
- LIVIO, M. **Biography**. [internet]. New York; 2020a [acesso em abril de 2021]. Disponível em: <https://www.mario-livio.com/about>

- LIVIO, M. **Galileo: And the Science Deniers**. Simon and Schuster, 2020b.
- MENDONÇA, M. de S., Jr. Spatial ecology goes to space: metabiospheres. **Icarus**, v. 232, p. 157, 2014.
- MOTTA, O. J. R.; SILVA, E.; SIQUEIRA-BATISTA, R. Ethical education and artificial intelligence. **International Journal of Development Research**, v. 10, p. 43111-43112, 2020.
- NASCIMENTO, E.; SILVA, E.; SIQUEIRA-BATISTA, R. The ‘use’ of sex robots: a bioethical issue. **Asian Bioethics Review**, v. 10, p. 231-240, 2018.
- NASA. National Aeronautics and Space Administration. **Putting the ethics into planetary protection**. Disponível em: <<https://astrobiology.nasa.gov/news/putting-the-ethics-into-planetary-protection>> Acesso em 30 abr 2020.
- PLATÃO. **Cármides**. Tradução de Francisco Oliveira. Coimbra: I.N.I.C, 1988.
- POTTER, V. R. **Bioética: ponte para o futuro**. São Paulo: Edições Loyola; 2016.
- SCOTT, R. **Blade runner**. Los Angeles: Warner Home Video, 1982.
- SIQUEIRA-BATISTA, R. **(Bio)ética para todos os seres: prêmio**. In: CASTRO, J. C.; NIEMEYER-GUIMARÃES, M.; SIQUEIRA-BATISTA, R. Caminhos da Bioética - Volume III. 1ed.Teresópolis: Editora Unifeso, 2020, p. 257-276.
- SIQUEIRA-BATISTA, R.; MOTTA, O. J. R.; GOMES, A. P. **Sob as bênçãos de Jano: (Bio)ética - passado, presente e futuro**. In: SOEIRO, A. C. V.; PALHETA NETO, F. X. Ensaio em bioética e cuidado em saúde. Belém: GCOM Gráfica e Editora, 2021, v. 1, p. 23-60.

DIÁLOGO 02

BIOÉTICA, TECNOLOGIAS EMERGENTES E ÉTICA DAS VIRTUDES

Robert Sparrow

Elen Nas

1. APRESENTAÇÃO

Desde que a discussão ética acerca das tecnologias emergentes vem se intensificando nos últimos anos, como situar o debate em temas que estão em movimento? As definições de inteligência artificial (IA) vêm mudando ao longo dos anos, de acordo com as possibilidades de seu desenvolvimento em dado momento histórico. Compreensões genéricas como *Strong AI* ou *Weak AI* ganharam novas subdivisões e sinônimos, dependendo de sua funcionalidade. Por exemplo, enquanto ‘IA fraca’ (*Weak AI*) pode ser sinônimo de ‘IA estreita’ (*Narrow AI*), *Artificial General Intelligence* (AGI) está entre a ‘IA Estreita’ e a ‘Superinteligência’, esta última comparada ao que seria uma ‘IA Forte’. Estes são termos que uma simples busca no Google oferece definições simples, capazes de dar ao público leigo informações preliminares.

As revoluções nas comunicações trouxeram desafios ao ser humano, na forma de lidar com o mundo, consigo mesmo e com o outro. Buscar respostas e informações de forma proativa, sem simplesmente “comprar” as histórias como estão sendo vendidas é uma atitude possível com as tecnologias digitais, e é bem diferente do primeiro momento dessas revoluções em que as tecnologias analógicas foram criando um encantamento e condicionamento resultando em uma relação passiva com a mídia.

A expansão do acesso à internet, os *smartphones*, e o amplo acesso às ferramentas de edição digital de imagens criaram possibilidades para contar histórias, construir e manipular narrativas com grande facilidade.

Finalmente, chegamos ao ponto em que a ação e a responsabilidade individuais são expostas aos holofotes. Um momento propício para revisar tradições éticas ancestrais, relacionadas às virtudes pessoais. Como a noção básica de ética presente em nossa cultura, as revoluções tecnológicas são produto da epistemologia ocidental. Assim, revisar a ética baseada nas virtudes contribui potencialmente para os ‘ajustes’ necessários ao enfrentamento dos desafios atuais.

Os desenvolvimentos técnico-científicos são comumente recebidos com entusiasmo. E é de fato fascinante que uma pessoa que perdeu um membro do corpo possa substituí-lo por um dispositivo técnico operando com a mesma funcionalidade, deixando inclusive de ser um ‘deficiente físico’ para tornar-se um ‘super eficiente’ físico como é o caso no uso das pernas biónicas, que se tornam grande vantagem no esporte (SAVULESCU, 2016). Assim, são muitos os exemplos onde a tecnologia entra como agregador no desenvolvimento humano e solucionadora de problemas. Porém a rápida revolução tecnológica das últimas quatro décadas reservou pouco espaço para reflexão (NICOLACI-DA-COSTA, 2002), não somente em seus setores industriais de produção como também dentro da esfera científica e ambiente acadêmico.

As discussões sobre os impactos da era industrial e a mecanização do mundo ocorrem a mais de um século enquanto o vigor da indústria corre alheio as críticas intelectuais. O século 20, seja em seu início ou na fase pós guerras, é povoado com vozes de entusiasmo e vozes críticas às tecnologias, porém, é no século 21 que o debate acerca dos valores morais se revigora e busca tráfegar para além das academias de filosofia e ciências humanas, alcançando mais diretamente a formação dos profissionais envolvidos nos novos planos de biotecnologia: setores da saúde, engenharia e computação.

A transversalidade dos debates é ainda um “trabalho em processo” que constitui um grande desafio desde que os saberes são treinados para atuar em separado e não para atuar de forma colaborativa. A colaboração e troca entre os saberes envolve conflitos de conhecimentos prévios, as crenças que eles carregam, assim como a dificuldade de manter a pró-atividade e receptividade fora da “zona de conforto”. As consequências positivas são os potenciais frutos da interdisciplinaridade. Quanto às consequências negativas, a humanidade, até aqui, conhece bem: disputas de ego e vaidade, competitividade, jogos de poder e guerras.

Devido à complexidade dos desafios que envolve pensar a ética para os projetos tecnológicos a bioética emerge. Trata-se de um campo plural, onde não há coesão das percepções e propostas. Como exemplo, as linhas de pesquisa do Programa de Pós Graduação em Bioética, Ética Aplicada e Saúde Coletiva (PPGBIOS) tem uma cobertura ampla dos temas “Bioética, Educação e Cuidados com a Saúde; Educação, Ética em Pesquisa em Humanos; Ética Animal e Ambiental; Direitos Básicos, Justiça Social e Políticas Públicas; Biotecnologia, Biopolítica e Biodireito”⁴ com trabalhos específicos que envolve profissionais e conhecimentos de diferentes áreas do saber.

Esta série de entrevistas, portanto, busca dar voz aos pesquisadores dos temas relacionados aos estudos da Bioética em relação às tecnologias emergentes.

A presente entrevista expõe o ponto de vista de um pesquisador que iniciou sua carreira acadêmica com atenção ao debate ético sobre as tecnologias, apresentando os aspectos da vida e trajetória que raramente aparecem nos artigos e livros. Via de regra, os aspectos da vida sensível estão embutidos no ambiente acadêmico-científico. Nas suas respostas, Robert Sparrow, irá narrar como foi sua formação e como ela influenciou os argumentos éticos eleitos para mediar os debates presentes nos seus trabalhos.

2. O ENTREVISTADO

Robert Sparrow é professor de Filosofia da Universidade Monash, em Melbourne, Australia, pesquisador chefe do Centro de Excelência para Ciência de Eletromateriais do Australian Research Council e professor adjunto do Center for Human Bioethics da Monash University. Toda sua pesquisa envolve o debate ético em torno das novas tecnologias. Sua produção conta com mais de cem artigos e capítulos de livros, e seus trabalhos vem sendo citados em milhares de artigos científicos⁵. Seus interesses de pesquisa estão na filosofia política e ética aplicada, que resultam em discussões que vão desde “aprimoramento humano”, à robótica em suas mais diversas aplicações, até temas como teoria da guerra justa. Ele é também membro do Comitê Técnico de Ética para Robótica do IEEE e um dos fundadores do Comitê Internacional para Controle de Armas Robótica.

3. ÉTICA DAS VIRTUDES

A palavra ‘virtude’ aparece com frequência nos escritos de Sparrow. Em geral ela trata da virtude das coisas, em um olhar onde tudo tem uma virtude em si. Digamos então que uma ou mais virtudes estão impregnadas em cada coisa que existe no mundo.

Por exemplo, ele menciona que os idosos, se ao invés de enfermeiros, são assistidos por robôs, estão sendo privados do reconhecimento e respeito de outro ser humano (SPARROW, 2016a). Se definirmos ‘reconhecimento e respeito’ como empatia, as questões defendidas nos seus artigos sugerem que o cuidado e assistência ao outro requer a empatia pela sua própria virtude. Desse modo, se os robôs são seres fabricados, com capacidades sensoriais e cognitivas limitadas, fazer com que substituam o cuidado de um ser-humano por outro ser-humano é privar ambos os lados (pacientes e cuidadores) de experimentar e exercitar suas virtudes.

A outra menção às virtudes diz respeito mais diretamente a ética das virtudes como uma visão ética com características específicas que podem se diferenciar de outras perspectivas como a deontológica ou utilitarista.

A ética das virtudes não atende diretamente a um conjunto de regras, mas à moralidade das ações em uma perspectiva de primeira pessoa, isto é, diz respeito ao carácter do indivíduo no exercício das virtudes ou na negação das mesmas através dos vícios. Por exemplo, se o ato sexual com uma boneca-robô carece de consentimento e do entendimento das necessidades e vontade da parceira, podendo até ser atribuído a tal ato a representação de um estupro, a ideia de que o ato sexual com tais robôs é moralmente errado baseia-se na ética das virtudes (SPARROW, 2017). Isto porque a ética das virtudes requer o exercício prático da sabedoria moral que está baseado não necessariamente na razão, mas no entendimento sobre como as emoções expressam as virtudes (o carácter) de um indivíduo (PETTIGROVE, 2018).

Desse modo, Sparrow refuta a ideia de que o sexo com robôs nos dá a oportunidade de expressar virtudes ou vícios (no caso, os vícios como pedofilia, estupro e outros comportamentos agressivos na esfera íntima) já que a virtude de amar um robô pode não ser suficiente para tornar alguém virtuoso. Por outro lado, acreditar que a prática de certos vícios com o robô poderia colaborar para erradicar tais práticas entre humanos é menos plausível do que elas se tornarem um reforço a tais vícios. (SPARROW, 2017)

Enquanto a ética das virtudes é o fundamento para a compreensão da ética, não apenas no Ocidente (filosofia grega da antiguidade) como em sociedades do oriente nos ensinamentos de Mencius e Confúcio (PETTIGROVE, 2018), na tradição moderna a divisão mente-corpo é mais demarcada, desde o “penso, logo, existo” do filósofo francês René Descartes (1596 - 1650) até a ‘razão pura’ defendida pelo filósofo alemão Immanuel Kant (1724 – 1804).

A ética das virtudes, vista pelas lentes contemporâneas têm o potencial de reunificar mente e corpo assim como agregar o valor das emoções para a compreensão da ética, não através do mero controle das emoções e necessidades do corpo, mas do entendimento que tais emoções fazem parte de um todo integrado que não está dissociado do debate ético. Sendo assim, antes de se perguntar se robôs-enfermeiros podem ser úteis ou se robôs como parceiros sexuais potencialmente solucionam alguns problemas, pergunta-se como a pessoa que está sendo cuidada irá se sentir, ou que virtudes ou vícios estão sendo estimulados através de invenções tais como robôs sexuais.

Em outras palavras, como a interação humano-computador e a interação humano-robô contribui ou favorece na formação e vivência das virtudes constituintes e essenciais no caráter do indivíduo de modo que este possa melhor responder aos seus próprios anseios e aos de sua comunidade.

4. ENTREVISTA

E.N.: Quando você decidiu ser filósofo e o que o motivou a especializar-se na ética da ciência e da tecnologia?

R.S.: Antes de começar a universidade, eu queria ser astrofísico. Meu pai era físico e eu amava ciências e matemática. Minha mãe era professora de inglês e de história e nossa casa sempre foi muito argumentativa. Meus pais ensinaram meus irmãos e eu que tínhamos o direito de esperar uma razão que justificasse a organização das instituições sociais e também que tínhamos deveres com outras pessoas; ensinaram ideias que exigiam um envolvimento com o mundo. Portanto, eu sempre fui alguém que queria entender por que as coisas eram do jeito que eram. Na escola, eu gostava de debater e conversar sobre política com meus amigos. Eu li o livro *Animal*

Liberation de Peter Singer aos dezessete anos e, como resultado, me tornei vegetariano. Isso provocou muita discussão.

Quando cheguei à universidade para estudar ciências, eu já estava pensando nas “grandes questões” que tendem a obcecar um certo tipo de adolescente: livre arbítrio; a natureza de uma sociedade justa; psicologia moral; e o “problema do mal”. No final do meu primeiro ano, eu senti que os módulos de matemática e física que eu estudava estavam me ensinando técnicas poderosas para resolver problemas que admitiam formalização, mas que a filosofia que eu estudava estava me ajudando a abordar vários aspectos das minhas angustias morais e existenciais. Embora eu tivesse iniciado a graduação estudando ciências, tive a sorte de poder creditar minhas matérias de matemática em um diploma de artes e, no meu segundo ano em diante, não estudei nada além de filosofia, história e filosofia da ciência.

Essa é a estória mas também acho que houve uma certa quantidade de “ajuste” psicológico envolvido. De certa forma, se importar com a filosofia é como uma neurose. Você não consegue parar de se preocupar com coisas que outras pessoas normalmente não se incomodariam. O estilo de raciocínio, pelo menos na filosofia analítica, às vezes pode ser um pouco robótico. Posso pensar em outros filósofos dos quais isso é mais verdadeiro do que é em mim, mas eu mentiria se dissesse que não era verdade para mim também.

Quanto à ética tecnológica... Meu primeiro amor, que continua sendo um interesse central, foi a filosofia política. Eu me matriculei em um PhD na Universidade Nacional Australiana, com Bob Goodin e Philip Pettit como meus supervisores. Pensei em escrever algo sobre anarquismo e a teoria democrática, mas fiquei interessado no debate “liberal-comunitário” (DELANEY, 1994, SWIFT, 1992), que era popular naqueles dias, e no feminismo interseccional de Iris Marion Young também (YOUNG, 1990). Então minha tese acabou sendo sobre a filosofia política do multiculturalismo. Ao longo do caminho — em parte como forma de não escrever minha tese — escrevi artigos sobre a ética da engenharia planetária (“A ética da terraformação”), robôs para idosos (“A marcha dos cães-robô”) e o status moral das inteligências artificiais (“o teste de triagem de Turing”) (SPARROW, 1999, 2002, 2004).

Cada um desses trabalhos envolvia uma exploração de um problema filosófico clássico por meio de um dilema ético contemporâneo — ou futuro

plausível — apresentado por uma nova tecnologia. O artigo sobre terraformação é realmente sobre ética da virtude e o escrevi porque eu estava interessado em debates sobre organismos geneticamente modificados, onde os mesmos argumentos poderiam ser aplicados. O artigo sobre cães-robô é sobre a natureza do bem-estar, enquanto o artigo do Teste de Triagem de Turing baseia-se na antropologia filosófica de Wittgenstein (GAITA, 1991). Muitos dos meus artigos possuem a mesma forma: eu gosto de escrever para possivelmente mostrar como uma questão genuinamente filosófica é relevante para um problema ético.

Eu tive a sorte de terminar meu doutorado assim que o governo australiano financiou o Centro de Filosofia Aplicada e Ética Pública (CAPPE), o que significava que eles precisavam de oito ou nove pesquisadores jovens em um prazo relativamente curto. Eu fui contratado para trabalhar com tópicos bioéticos, o que eu acabei fazendo.

Trabalhar com a ética aplicada depois de fazer um doutorado em outro campo é, na verdade, um pouco de tradição na Austrália. A filosofia acadêmica tende a operar no pressuposto de que a filosofia “real” é a metafísica, a filosofia da mente e a filosofia da linguagem. Os filósofos às vezes menosprezam a bioética e a ética aplicada (às vezes com uma boa justificativa, pois é preciso dizer que a qualidade média dos trabalhos nesses campos é indubitavelmente menor do que nos principais jornais de filosofia), também como a filosofia política. No entanto, os graduandos estão muito mais interessados em filosofia política do que em metafísica; portanto, mais empregos surgem no primeiro campo. Os departamentos costumam ensinar os módulos de ética nos cursos de enfermagem, engenharia, ciência da computação e negócios, de modo que precisam de pessoas que possam ensinar ética aplicada. Por fim, existem mais fontes de financiamento de pesquisa para pessoas que trabalham com ética aplicada do que para pessoas que trabalham — no que tradicionalmente se pensa ser — nas áreas mais “centrais” da filosofia. Posso pensar em pelo menos meia dúzia de pessoas da minha geração que concluíram doutorados em metafísica, filosofia da mente ou filosofia da linguagem e depois tiveram que desenvolver subitamente um interesse pela filosofia política para garantir uma posição. Pessoalmente, eu gostaria que a profissão chegasse a um acordo com esse fato e fizesse um esforço maior para incentivar bons alunos a estudar ética aplicada e filosofia política.

Quando eu comecei a trabalhar no CAPPE, percebi que algumas das questões nos debates sobre multiculturalismo, sobre os quais eu havia escrito na tese, eram diretamente relevantes para a ética do diagnóstico genético pré-implantação (SPARROW, 2011a). Também pude recorrer à minha pesquisa de tese para escrever um artigo sobre a controvérsia do implante coclear (SPARROW, 2005), que me incentivou a sentir que eu tinha algo a contribuir para o campo. Dolly, a ovelha, foi clonada assim que eu comecei, o que provocou um novo conjunto de debates éticos. Depois de escrever alguns artigos sobre paternidade genética (SPARROW, 2006, 2009), comecei a me interessar mais por debates em bioética. O fato de que não seria necessário competir com Aristóteles, Kant ou Hume ao escrever sobre novas tecnologias também ajudou: é muito mais difícil ter ideias novas em debates com histórias mais longas.

Deixei meu emprego no CAPPE, que era uma posição de contrato, para assumir uma posição de titularidade na Universidade de Wollongong, ensinando filosofia política e ética aplicada. Mais uma vez, tive sorte quando um emprego titular abriu não muito tempo depois no (então) Centro de Bioética Humana, da Universidade de Monash, para o qual me inscrevi e fui nomeado. O centro foi cofundado por Peter Singer e Helga Kuhse em 1981: eles realizaram grande parte de seu trabalho influente sobre a ética da fertilização *in vitro* e a tomada de decisões no final da vida útil (OAKLEY, 2009). O centro tinha duas características distintivas que o tornou atraente como um local de trabalho. Foi baseado na Faculdade de Artes e não na Faculdade de Medicina, onde geralmente estão localizados os centros de bioética (dizem que a Faculdade de Medicina estava preocupada com o fato de Singer encorajar seus alunos a se oporem à vivissecção e/ou experimentação animal: aparentemente isto também é a razão pela qual foi chamado de Centro de Bioética *Humana*). E estava — e ainda está — explicitamente comprometida a uma bioética filosófica e secular, ao invés de empírica ou religiosa: quando cheguei, ainda estava co-localizada com o departamento de filosofia. Depois de trabalhar no centro, principalmente na ética do diagnóstico genético pré-implante, por alguns anos, mudei para o departamento de filosofia para ter a oportunidade de ensinar filosofia política e ética ambiental, que ainda são minhas paixões.

E.N.: Como você vê essas questões sendo relacionadas à bioética e quão importante você acha que é para o campo da bioética responder às preocupações sobre o uso de tecnologias?

R.S.: A bioética não existiria se não por novas tecnologias. Foi o poder que as novas tecnologias médicas estavam concedendo sobre questões de vida e morte que criou a oportunidade retórica para os filósofos criarem um novo campo e, mais cinicamente, gerar novos fluxos de renda. Os primeiros debates sobre a definição de morte encefálica (TRUOG, POPE, e JONES, 2018) foram o resultado do advento da ventilação mecânica. A invenção da inseminação artificial pelo doador e a fertilização *in vitro* colocaram novas questões éticas que as pessoas poderiam insistir que eram “bioéticas” e não filosóficas. Por exemplo, o fato do segundo filho nascido após a fertilização *in vitro* ter sido concebido em um laboratório na Monash teve um papel importante na geração de apoio para o estabelecimento do Centro de Bioética Humana na Monash (SINGER e KUHSE, 2006). O progresso científico e tecnológico, em transplante de órgãos, clonagem humana, implantes médicos, diagnóstico genético pré-implante e agora edição de genoma, tem impulsionado o campo desde então.

De muitas maneiras, penso, o foco na tecnologia é infeliz, na medida em que, em nível global, a maioria das pessoas enfrenta desafios relacionados ao nascimento e à morte que não são realmente o produto da tecnologia. Os debates éticos sobre a edição do genoma em seres humanos são fascinantes, por exemplo, mas o número de pessoas que provavelmente terão os genomas editados na próxima década é minúsculo se comparado ao número de pessoas que enfrentam insegurança alimentar ou não têm acesso aos cuidados básicos de saúde ou que provavelmente migrarão como resultado das mudanças climáticas. Se alguém estivesse realmente preocupado com quem vive e morre e como vive e morre, ficaria mais preocupado com questões políticas e econômicas, especialmente com a justiça da distribuição da riqueza e o acesso a oportunidades. Certamente, alguma bioética tem se preocupado com essas questões, como, por exemplo, nos debates sobre justiça na distribuição do acesso à saúde (DANIELS, 1985). A parte da literatura que também pode ser descrita como “ética médica” também trata de questões no relacionamento entre médicos e pacientes que afetam a vida de quem entra em um hospital ou consultório.

Dito isto, há três razões pelas quais a bioética precisa se atentar à tecnologia. Primeiro, algumas das tecnologias com as quais a bioética tradicionalmente se preocupa são agora tão difundidas que os problemas que elas levantam realmente afetam um grande número de pessoas em todo o mundo. A fertilização *in vitro* é o caso óbvio em questão. As questões bioéticas sobre a melhor forma de realizar testes de drogas durante pandemias seriam outra. Assim, algumas vezes as questões de ética em tecnologia são de grande importância prática. Segundo, na medida em que as pessoas costumam propor soluções tecnológicas para problemas sociais, políticos e éticos, é importante ter um entendimento adequado do que a tecnologia pode ou não fazer e das maneiras pelas quais ela pode influenciar nossas opções, apesar das nossas melhores intenções. Finalmente, em terceiro lugar, cada vez mais com debates sobre novas tecnologias ocorrem debates políticos mais amplos — por exemplo, sobre relações de gênero, relações raciais, a importância relativa da natureza e da criação, liberdade individual versus bem público e o escopo adequado da democracia política. Tendo como outro exemplo, a literatura sobre aprimoramento genético humano é um dos lugares em que ideias sociais darwinistas perigosas, bem como reivindicações sobre a genética da raça, estão sendo reabilitadas no momento. Os resultados desses debates, por sua vez, têm implicações significativas para os arranjos políticos e econômicos em geral. Qualquer pessoa interessada no estado do futuro deve atentar-se aos debates sobre tecnologia por esse motivo.

E.N.: Como você se dirige a uma questão específica que, em sua opinião, requer discussão ética e bioética?

R.S.: Eu não tenho certeza de que sou a pessoa certa para responder essa pergunta — ou, pelo menos, a resposta que tenho para dar, se estou sendo honesto, não é uma da qual me orgulho especialmente. Eu costumo escrever onde posso, ou onde encontro um novo ângulo filosófico, ou onde sei que posso estar “à frente da curva” e publicar algo que as pessoas possam prestar atenção. Ocasionalmente, os meus motivos são contrários: acho que muitas vezes há um grão de verdade em argumentos impopulares ou difíceis, que pode ser útil tentar desvendar. Essas motivações são pessoais e/ou intelectuais. Às vezes, minhas motivações são pragmáticas. O sistema universitário australiano é super-gerenciado e fortemente “medido”

(CONNELL, 2019), e os acadêmicos estão sob muita pressão para publicar muito, em bons periódicos e para obter “impacto”. Escrevi artigos porque achei que poderia escrevê-los rapidamente, ou colocá-los em bons periódicos, ou porque eram o tipo de coisa que eu pensei que poderia ser citada.

Embora eu defendesse, pelo menos até certo ponto, a ideia de que os acadêmicos são justificados em trabalhar no que acham interessante, acho que há um argumento forte a ser feito para a ideia de que, em um mundo caracterizado por pobreza e injustiça tão difundidas, os acadêmicos também possuem a obrigação de se concentrar nas questões de maior importância prática. Infelizmente, muito pouco que eu escrevi se enquadra nessa categoria. Além disso, achei difícil redirecionar minha pesquisa para caminhos que eu poderia defender mais facilmente ao abordar questões sociais ou políticas importantes. Em parte, isso ocorre porque há um problema estrutural quando se trata de incentivos à publicação acadêmica. Os periódicos — assim como a mídia, na medida em que os filósofos deveriam se preocupar com isso — recompensam a novidade em vez da verdade ou importância. Você lutará para publicar um trabalho argumentando que as pessoas têm o direito à água potável ou que elas devem receber água potável. Ambas as coisas são verdadeiras e importantes. No entanto, por serem verdadeiras e importantes, elas já foram ditas antes e é difícil encontrar algo novo para adicionar ao tópico. Consequentemente, é improvável que os “árbitros” julguem que é uma contribuição útil ao conhecimento. Um artigo sobre robôs sexuais é outra história! Quando escrevi sobre esse assunto, quase nada havia sido publicado, então era relativamente fácil dizer algo novo. Infelizmente, essa dinâmica tende a incentivar os filósofos a escreverem sobre questões cada vez mais obscuras e exageradas, apenas porque é onde é possível progredir com mais facilidade. Fico sempre impressionado quando as pessoas conseguem encontrar uma nova opinião sobre um problema que é realmente importante, simplesmente porque é muito difícil fazê-lo.

E.N.: Por exemplo, você escreveu sobre veículos sem motorista, sexo com robôs, raça e robótica, robôs sociais em geral, robôs para cuidados com a saúde e assim por diante. O que chamou sua atenção para essas questões e como você vê o impacto das discussões éticas no desenvolvimento dessas tecnologias?

R.S.: Eu comecei a escrever sobre robôs porque eles se prestam a experimentos de pensamento que podem iluminar questões filosóficas clássicas, especialmente sobre o significado do corpo, a distinção entre a realidade e a ilusão e sobre o conceito de pessoa. Até recentemente, quase toda a literatura sobre robótica era dessa natureza — até porque a ideia do robô era popular muito antes de existirem robôs que realmente fizessem algo remotamente útil ou interessante. Estórias e argumentos sobre robôs eram realmente estórias e argumentos sobre o que significa ser humano. As pessoas ainda entendem isso errado — como, por exemplo, quando confundem as “três leis da robótica” de Asimov como uma proposta séria de ética do robô, e não como um mecanismo para gerar tensão narrativa. Assim, meus artigos sobre robôs sexuais (SPARROW, 2017) e outros robôs sociais (SPARROW, 2016, 2006) são realmente artigos sobre a ética da pornografia e da fantasia e sobre o bem-estar e a base social do respeito próprio, respectivamente. Vários dos meus trabalhos sobre robótica também são explorações em ética da virtude (SPARROW, 2013, 2017, 2020).

No entanto, a robótica percorreu um longo caminho desde que comecei a escrever sobre robôs. Quando comecei a escrever sobre eles no final da década de 1990, os robôs realmente estavam presentes apenas na fabricação, e qualquer reflexão ética sobre seu papel foi, na maioria das vezes, incluída em um debate sobre “automação”. Embora ainda seja o caso em que a grande maioria dos robôs seja usada na fabricação, agora parece provável que veremos robôs em papéis que terão impactos sociais e políticos mais complexos e levantarão questões éticas mais interessantes. Por exemplo, enquanto a data exata em que eles serão totalmente realizados ainda é estranhamente controversa (porque alguém poderia pensar que isso seria relativamente fácil de discernir do atual e controverso estado da arte); a tecnologia do veículo sem motorista avançou ao ponto em que devemos esperar que, em um futuro não muito distante, a maioria dos carros novos seja capaz de dirigir a si mesmos. Dado que as cidades modernas são projetadas em torno do automóvel, as implicações sociais, econômicas e geográficas provavelmente serão dramáticas. Como quero escrever sobre questões importantes, tenho escrito sobre veículos autônomos (SPARROW e HOWARD, 2017, 2020). Da mesma forma, se menos urgente, se o projeto de robótica social decolar, o modo como os robôs representam as pessoas se tornará importante pelas mesmas razões pelas quais as representações de

raça e gênero em outras mídias são importantes. Meus artigos sobre robôs sexuais e a política racial dos robôs (SPARROW, 2019, 2020b) abordam esse assunto. Como os militares são os principais financiadores da pesquisa em robótica e como os drones agora são amplamente utilizados e os sistemas de armas autônomos provavelmente serão amplamente utilizados em um futuro não muito distante, também escrevi muito sobre robótica militar (Veja, por exemplo, (SPARROW, 2007, 2009b & 2009c); 2012, 2015, 2016b; (SPARROW e LUCAS, 2016); (SPARROW, MCLAUGHLIN e HOWARD, 2019). Além disso, como mencionei acima, argumentos sobre novas tecnologias são onde debates mais amplos sobre o tipo de mundo que queremos criar para nós mesmos no futuro estão ocorrendo — e isso é especialmente verdade nos argumentos sobre robôs. Meus artigos sobre robôs sexuais e sobre robótica para idosos são, em parte, contestações sobre a política de representação e as relações intergeracionais nesse nível.

Os debates sobre ética são absolutamente centrais no campo da robótica. É amplamente reconhecido que as questões éticas e políticas permanecem como uma grande barreira para a introdução da tecnologia de veículos autônomos (LIN, 2016). Existe uma campanha global pedindo a proibição do desenvolvimento de sistemas de armas autônomos (ROBOTS, 2020), em grande parte motivada por preocupações éticas, das quais tive um pequeno papel em provocar através do meu cargo como um dos cofundadores da Comitê Internacional para Controle de Armas Robóticas. Nos últimos anos, vimos engenheiros se organizarem para protestar contra o financiamento militar de pesquisas sobre visão computacional (SHANE e WAKABAYASHI, 2018) e realidade aumentada (WONG, 2019). As discussões sobre robôs sexuais sempre tratam suas implicações para as relações de gênero e frequentemente discutem a possibilidade de as pessoas quererem comprar e vender robôs sexuais infantis (HILL, 2014; MORIN, 2016). As pessoas que constroem robôs sexuais se envolvem nesses argumentos para defender a ideia de que existe um mercado legítimo para seus robôs. Da mesma forma, os engenheiros que pesquisam robôs educacionais ou para atendimento a idosos precisam responder às preocupações de que os robôs possam levar as pessoas a serem mais isoladas socialmente ou vulneráveis à fraudes e manipulações. De maneira mais geral, na medida em que o arquétipo do robô é uma “pessoa mecânica”, que pode executar tarefas que as pessoas tradicionalmente realizavam, argumentos sobre a ética

de substituir pessoas por robôs em vários papéis nunca estiveram longe do centro das discussões entre roboticistas (WIENER, 1950).

Questões éticas também podem ter um impacto muito real na viabilidade comercial de tecnologias. Se as pessoas sentirem que uma empresa as está explorando ou espionando através de seus produtos, elas podem recorrer a uma tecnologia diferente. Se um algoritmo obtém a reputação de ser sexista, racista ou de alguma forma tendencioso, os consumidores podem boicotá-lo. Lidar mal com uma questão ética, por exemplo, optar por priorizar o lucro ou a participação de mercado em vez da segurança, pode levar os reguladores a intervir ou mesmo a proibir o uso de uma tecnologia. Obviamente, nada disso acontece sempre, mas pesquisadores e fabricantes ignoram as considerações éticas por sua conta e risco.

E.N.: De todos os temas sobre dos quais você escreveu, há um em particular que você acha mais essencial que os outros? Se sim, por quê? Porque é algo que você teve prazer em investigar, ou porque é importante para a sociedade, ou ambos?

R.S.: Eu luto para poder nomear um tema porque, como espero ser óbvio pelas minhas observações acima, sou motivado por interesses intelectuais e políticos e acho que ambos são importantes por diferentes razões. As duas coisas que acredito serem mais essenciais estão realmente ausentes da lista de meus projetos de pesquisa. Eu escrevi um pouco sobre aprimoramento humano (hipotético) via PGD ou edição de genoma ao longo da minha carreira (SPARROW, 2007b, 2010, 2011b, 2011a, 2012b, 2019c). Mais recentemente, escrevi sobre inteligência artificial (SPARROW e HATHERLEY, 2019b, 2020c).

O debate sobre o aprimoramento humano refere-se a questões profundas sobre a relação entre cultura e liberdade, que são intelectualmente fascinantes e têm sido uma obsessão minha desde o segundo ano dos meus estudos de doutorado. Quais são as implicações para argumentos sobre o valor da liberdade e a atitude adequada em relação à mudança cultural se — como eu diria (SPARROW, 2011a) é o caso — individualizarmos e avaliarmos as opções com referência a uma “concepção do bem” (uma descrição do que faz a vida humana correr bem)? Se o que não podemos fazer às vezes é constitutivo das coisas que podemos fazer, como devemos avaliar o valor de novas opções? Se nossos projetos inevitavelmente dependem

de entendimentos particulares de conceitos que são necessariamente sociais, poderíamos às vezes ser justificados em “policiar” significados? Essas são questões em que, acredito, o pensamento conservador tem algo valioso para contribuir... embora minha própria política esteja longe de ser conservadora. Apesar de existir pouco perigo de aprimoramento humano significativo — pelo menos por meios genéticos — no futuro próximo, essas questões, que aparecem de forma gritante no debate sobre a ética do aprimoramento humano, são relevantes — de fato, vitais — para uma variedade muito maior de questões políticas, e é por isso que acho que vale a pena escrever sobre elas.

Contudo, tenho pesquisado e escrito sobre inteligência artificial porque essa é uma tecnologia que realmente tem o potencial de mudar o mundo. Algumas das questões éticas que cercam a IA — por exemplo, a questão do status moral de inteligências artificiais gerais hipotéticas — são intelectualmente fascinantes (GUNKEL, 2012; SPARROW, 2004). No entanto, existem várias questões éticas mais urgentes, se não sempre tão interessantes, sobre a IA e suas aplicações que são importantes porque as respostas a elas afetarão a todos nós. Por exemplo, o equilíbrio entre fabricantes, juízes e legisladores, entre privacidade e benefício público, quando se trata do uso de grandes volumes de dados, terão implicações dramáticas nas relações entre indivíduos, corporações e governos no futuro. Saber como e quando confiar na IA também é vital nas aplicações em medicina, engenharia, guerra e finanças. As perguntas sobre a natureza e a prevalência de “preconceitos” na IA são grandes quando você percebe quanto poder será concedido a esses sistemas sobre nós em um futuro não muito distante. Infelizmente, suspeito que muitos dos resultados prováveis dos desenvolvimentos da IA sejam distópicos, o que torna pensar sobre essas questões agora duplamente importante, se quisermos ter alguma chance de garantir que essa tecnologia seja usada para melhorar, ao invés de reduzir nossa liberdade e bem-estar.

E.N.: Há alguma questão referente à ética e à tecnologia que foi resolvida?

R.S.: Há um sentido em que a resposta a esta pergunta é necessariamente não. Se sabemos o que fazer, não enfrentamos uma questão ética. Assim, se alguém pensa em alguma questão ética sobre tecnologia, então, por definição, ela deve estar aberta.

No entanto, há outro sentido em que acho que a resposta à sua pergunta é claramente sim. Há questões éticas sobre as quais as pessoas costumavam discutir, mas não discutem mais. Para dar apenas um exemplo, na maioria das vezes, as pessoas não se opõem à ideia de morte cerebral ou ao transplante de órgãos. Além disso, acredito que é possível falar sobre progresso ético, inclusive no âmbito de nossas relações com a tecnologia. Por exemplo, o fato de casais solteiros e do mesmo sexo serem capazes de acessar tratamentos de infertilidade, incluindo fertilização *in vitro*, na Austrália, me parece um progresso óbvio em relação às restrições arbitrárias que foram originalmente impostas ao acesso dos conservadores sociais. Além das questões que eu acredito que foram (na maior parte) resolvidas, existem inúmeras questões que são melhor compreendidas como resultado do trabalho de filósofos e bioeticistas. Por exemplo, as implicações de uma preocupação com a “abertura do futuro” de nossas crianças para decisões sobre sua genética são agora relativamente claras como resultado de debates sobre a ética da clonagem, diagnóstico genético pré-implantação e aprimoramento humano. Temos uma compreensão muito melhor das diferentes ideias de “paternidade/maternidade” e seu significado relativo como resultado de discussões sobre a ética das tecnologias de reprodução assistidas: foi impressionante ver a rapidez com que os bioeticistas foram capazes de mapear as questões no debate sobre tecnologias de substituição mitocondrial por conta disso. Da mesma forma, quando se pensa na ética dos robôs sexuais, é possível se basear nos achados de uma longa história de discussão, por acadêmicas feministas e outras, e das questões éticas levantadas pela pornografia em outras mídias.

E.N.: Como você veria a contribuição da bioética para a compreensão das tecnologias emergentes?

R.S.: Na verdade, não me identifico como “bioeticista” ou mesmo “eticista”, embora eu seja frequentemente apresentado como ambos por outras pessoas em contextos públicos. Minha lealdade disciplinar — como ela é — é a filosofia e eu gostaria de pensar que ainda sou um filósofo, mesmo trabalhando em contextos aplicados. Acho que o rigor intelectual que a filosofia pode (às vezes) trazer para as discussões é incrivelmente importante em um momento histórico em que as pessoas parecem dispostas a creditar a agência histórica às tecnologias e a conceder poder àqueles que as fabricam.

Às vezes, os filósofos servem como “moscas intelectuais”, fazendo perguntas difíceis sobre significados e valores que as pessoas — especialmente pessoas poderosas — relutam em enfrentar. Há também um corpo de trabalho rico na filosofia da tecnologia (ver, por exemplo, VERBEEK, 2011; WINNER, 1989) que tem muito a oferecer aos debates sobre novas tecnologias, e que eu gostaria que as pessoas se atentassem.

Quanto à bioética.... Eu acho que tem algumas coisas a oferecer além daquelas oferecidas pela filosofia. A melhor bioética reconhece que somos criaturas corporificadas, que somos vulneráveis como consequência dessa corporificação e que florescemos e sofremos de — e por causa de — relacionamentos com aqueles que estão próximos a nós e com nossa sociedade em geral. Essas verdades são desesperadamente importantes nas discussões de novas tecnologias, que muitas vezes procedem como se fôssemos criaturas de puro espírito celebrando contratos livres com outras pessoas para usar produtos que caíam do céu. Da mesma forma, como a grande maioria dos desafios da saúde são problemas que enfrentamos juntos e que só podem ser abordados coletivamente, a (boa) bioética é orientada para políticas públicas. O fato de que é bom, às vezes até essencial, que a comunidade tome uma decisão coletiva sobre como abordar uma questão ética é outra lição que precisamos aprender quando se trata de fazer escolhas sobre tecnologias emergentes. Quando temos essas conversas, é essencial que não nos limitemos apenas aos tecnólogos. Existe uma rica tradição da filosofia da bioética da qual podemos recorrer ao longo do projeto de moldar um futuro em que a tecnologia sirva nossos valores — ao invés de agir como curto-circuito.

5. AGRADECIMENTOS

A Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES), que atraves do Programa de Doutorado Sanduiche no Exterior (PDSE) facilitou a aproximacao do trabalho dos autores durante estadia na Monash University. A Stefania Salvador P. Montenegro, diplomada em Medicina, Faculdade Dinamica do Vale do Ipiranga (FADIP), pela valiosa e gentil colaboracao na traducao.

6. NOTAS

A presente entrevista foi realizada dentro da investigação do tema da tese “Robótica, Inteligência Artificial e Bioética: debates contemporâneos”, desenvolvida junto ao Programa de Pos-Graduação em Bioética, Ética Aplicada e Saúde Coletiva (PPGBIOS), da Universidade Federal do Rio de Janeiro (UFRJ), Fundação Oswaldo Cruz (Fiocruz), Universidade do Estado do Rio de Janeiro (UERJ) e Universidade Federal Fluminense (UFF).

7. REFERÊNCIAS

CONNELL, R. **The Good University: What Universities Actually Do and Why It's Time for Radical Change**. Melbourne: Monash University Publishing, 2019.

DANIELS, N. **Just Health Care**. Cambridge & New York: Cambridge University Press, 1985.

DELANEY, C. F. **The Liberalism-communitarianism Debate: Liberty and Community Values**. Lanham: Rowman & Littlefield, 1994.

GAITA, R. **Good and Evil: An Absolute Conception**. London: MacMillan, 1991.

GUNKEL, D. J. **The Machine Question: Critical Perspectives on AI, Robots, and Ethics**. Cambridge, Massachusetts: MIT Press, 2012.

HILL, K. Are child sex-robots inevitable? **Forbes**, 14 jul. 2014.

JULIAN SAVULESCU. Doping scandals, Rio and the future of human enhancement. **Bioethics**, v. 30, n. 5, 2016.

KUHSE, P. S. AND H. 1980-2005: Bioethics then and now. **Monash Bioethics Review**, v. 25, n. 1, 2006.

LIN, P. Why ethics matters for autonomous cars. In: MARKUS MAURER, J. CHRISTIAN GERDES, BARBARA LENZ, AND H. W. (Ed.). **Autonomous Driving: Technical, Legal and Social Aspects**. Berlin, Heidelberg: Springer, 2016. p. 69–85.

MORIN, R. Can child dolls keep paedophiles from offending? **The Atlantic**, 11 jan. 2016.

NICOLACI-DA-COSTA, A. M. Revoluções tecnológicas e transformações subjetivas. **Psicologia: Teoria e Pesquisa**, v. 18, n. 2, p. 193–202, 2002.

OAKLEY, J. Monash University Centre for Human Bioethics. In: GRAHAM OPPY, NICK TRAKAKIS, LYNDIA BURNS, S. G. AND F. L. (Ed.). . **A Companion to Philosophy in Australia and New Zealand**. Melbourne: Monash ePress, 2009. p. 304–307.

PETTIGROVE, R. H. AND G. “**Virtue Ethics**”, 2018. (Nota técnica).

ROBOTS, C. TO S. K. **A Growing Global Coalition**. Disponível em: <<https://www.stopkillerrobots.org/about/>>. Acesso em: 12 mar. 2020.

SPARROW, ROBERT, AND GEORGE LUCAS, J. When robots rule the waves? **Naval War College Review**, v. 69, n. 4, p. 49–78, 2016.

SPARROW, ROBERT, ROBERT MCLAUGHLIN, AND M. H. Naval robots and rescue. **International Review of the Red Cross**, 2019.

SPARROW, ROBERT, AND L. S. In the hands of machines? The future of aged care. **Minds and Machines**, v. 16, p. 141–161, 2006.

SPARROW, ROBERT, AND M. H. When human beings are like drunk robots: driverless vehicles, ethics, and the future of transport. **Transportation Research Part C**, v. 80, p. 206–215, 2017.

SPARROW, ROBERT, AND M. H. Make way for the wealthy? Autonomous vehicles, markets in mobility, and social justice. **Mobilities**, 2020.

SPARROW, R. The ethics of terraforming. **Environmental Ethics**, v. 21, n. 3, p. 227–245, 1999.

SPARROW, R. The march of the robot dogs. **Ethics and Information Technology**, v. 4, n. 4, p. 305–318, 2002.

SPARROW, R. The Turing triage test. **Ethics and Information Technology**, v. 6, n. 4, p. 203–213, 2004.

SPARROW, R. Defending Deaf Culture: The case of cochlear implants. **The Journal of Political Philosophy**, v. 13, n. 2, p. 135–152, 2005.

SPARROW, R. Cloning, parenthood, and genetic relatedness. **Bioethics**, v. 20, n. 6, p. 308–318, 2006.

SPARROW, R. Killer Robots. **Journal of Applied Philosophy**, v. 24, n. 1, p. 62–77, 2007a.

SPARROW, R. Procreative beneficence, obligation, and eugenics. **Genomics, Society and Policy**, v. 3, n. 3, p. 43–59, 2007b.

SPARROW, R. Therapeutic cloning and reproductive liberty. **The Journal of Medicine and Philosophy**, v. 34, n. 2, p. 102–108, 2009a.

SPARROW, R. Predators or plowshares? Arms control of robotic weapons. **IEEE Technology and Society**, v. 28, n. 1, p. 25–29, 2009b.

SPARROW, R. Building a better warbot: Ethical issues in the design of unmanned systems for military applications. **Science and Engineering Ethics**, v. 15, n. 2, p. 169–187, 2009c.

SPARROW, R. Better than men? Sex and the therapy/enhancement distinction. **Kennedy Institute of Ethics Journal**, v. 20, n. 2, p. 115–144, 2010.

SPARROW, R. Liberalism and eugenics. **Australasian Journal of Philosophy**, v. 89, n. 3, 2011a.

SPARROW, R. A not-so-new eugenics: Harris and Savulescu on human enhancement. **Hastings Center Report**, v. 41, n. 1, p. 32–42, 2011b.

SPARROW, R. ‘Just say no’ to drones. **IEEE Technology and Society**, v. 31, n. 1, 2012a.

SPARROW, R. Human enhancement and sexual dimorphism. **Bioethics**, v. 26, n. 9, p. 464–475, 2012b.

SPARROW, R. War without virtue? In: STRAWSER, B. J. (Ed.). **Killing By Remote Control**. Oxford and New York: Oxford University Press, 2013. p. 84–105.

SPARROW, R. Twenty seconds to comply: Autonomous Weapon Systems and the recognition of surrender. **International Law Studies**, v. 91, p. 699–728, 2015.

SPARROW, R. Robots in aged care: a dystopian future? **AI & society**, v. 31, n. 4, p. 445–454, 2016a.

SPARROW, R. Robots and respect: Assessing the case against Autonomous Weapon Systems. **Ethics and International Affairs**, v. 30, n. 1, p. 93–116, 2016b.

SPARROW, R. Robots, Rape and Representation. **International Journal of Social Robotics**, p. 1–13, 2017.

SPARROW, R. Do robots have race? Race, Social Construction, and HRI. **IEEE Robotics and Automation Magazine**, 2019a.

SPARROW, R. The promise and perils of AI in medicine. **International Journal of Chinese and Comparative Philosophy of Medicine**, v. 17, n. 2, p. 79–109, 2019b.

SPARROW, R. Yesterday's child: How gene editing for enhancement will produce obsolescence – and why it matters. **American Journal of Bioethics**, v. 19, n. 7, p. 6–15, 2019c.

SPARROW, R. Virtue and vice in our relationships with robots: Is there an asymmetry and how might it be explained? **International Journal of Social Robotics**, 2020a.

SPARROW, R. Robotics has a race problem. **Science, Technology, & Human Values**, v. 45, n. 3, p. 538–560, 2020b.

SPARROW, R. AND J. H. High hopes for 'Deep Medicine'? AI, economics, and the future of care. **Hastings Center Report**, v. 50, n. 1, p. 14–17, 2020c.

SWIFT, S. M. AND A. **Liberals and Communitarians**. Oxford, UK; Cambridge USA: Blackwell, 1992.

TRUOG, ROBERT D., THADDEUS MASON POPE, AND D. S. J. The 50-Year Legacy of the Harvard Report on Brain Death. **Journal of the American Medical Association**, v. 320, p. 335–36, 2018.

VERBEEK, P.-P. **Moralizing Technology: Understanding and Designing the Morality of Things**. Chicago & London: University of Chicago Press, 2011.

WAKABAYASHI, S. S. AND D. 'The Business of War': Google Employees Protest Work for the Pentagon. **The New York Times**, 4 abr. 2018.

WIENER, N. **The Human Use of Human Beings: Cybernetics and Society**. Boston: Houghton Mifflin, 1950.

WINNER, L. **The Whale and the Reactor a Search for Limits in an Age of High Technology**. Chicago: University of Chicago Press, 1989.

WONG, J. C. "We won't be war profiteers": Microsoft workers protest \$480m army contract. **The Guardian**, 23 fev. 2019.

YOUNG, I. M. **Justice and the Politics of Difference**. Princeton, N.J.: Princeton University Press, 1990

DIÁLOGO 03

BIOPOLÍTICA E OS VALORES OCULTOS NOS SISTEMAS

Geoffrey Bowker

Elen Nas

1. APRESENTAÇÃO

Este texto é parte de uma série de ‘diálogos’ que decidi fazer enquanto pesquisava sobre ética em tecnologias emergentes. A inteligência artificial (IA) e a robótica têm sido discutidas, com seu crescente desenvolvimento há pelo menos cinco décadas. Se considerarmos as referências históricas, esses temas já existem há mais de um século, ou até mais se buscarmos as raízes desses conceitos.

No entanto, apesar de esses temas serem “tópicos da moda” nas últimas décadas, ainda são escassas as pesquisas que os relacionam com a ética e a bioética. Por isso, esta iniciativa dá voz a pesquisadores que estão percorrendo este caminho desde longa data, e ouvir suas perspectivas. Pois, enquanto as discussões éticas são comumente teóricas e especulativas, vale-nos saber como elas estão associadas com a experiência, e de que maneiras vêm influenciando o científico-acadêmico e com reais impactos para as mudanças necessárias.

A partir da formação em Ciências Sociais, Artes e Design, o interesse e a experiência com as tecnologias me trouxeram para o campo da Bioética, para endereçar, com maior propriedade as discussões éticas.

Assim, parte da minha pesquisa de doutorado, realizada no exterior, foi observar como os estudiosos do campo da tecnologia, sociedade e ética se engajam com os temas atuais das tecnologias e entender o que os moveu nesta direção.

Identifiquei o trabalho do Professor Geoffrey Bowker com afinidades nas discussões da Bioética, com uma abordagem crítica voltada a entender os mecanismos de classificação no ambiente científico e nos sistemas de conhecimento.

Essas afinidades conectam ideais de justiça social e respeito às diferentes formas de estar no mundo: culturas e microculturas, valorizando narrativas e a diversidade de saberes. Além disso, o questionamento sobre a padronização de certas características físicas no discurso científico, as classificações utilizadas na biomedicina, são, no seu trabalho, alvo de um exame que busca estimular o pensamento crítico no ensino de ciências.

As Ciências Sociais, especialmente a Antropologia, ajudam-nos a compreender a diversidade de crenças, narrativas e verdades. É, portanto, uma disciplina que confronta verdades únicas e universais. Para acolher os modos de compreensão diversos, advindos destas perspectivas, é necessário confrontar antigos dogmas sobre como entender e fazer ciência. Pois, de uma perspectiva ética, acolher a alteridade e a pluralidade é fundamental. Permanece, portanto, um desafio a ser posto em prática, quando a maioria das especialidades científicas é geralmente endêmica e manifesta dificuldade em lidar com as incertezas.

Descobrimos nas nossas interações (eu e o Professor Bowker) a convergência com a área de Design, que tem uma procura constante pela criação de futuros e realidades possíveis através da materialização de ideias, conceitos, projetos e produtos. A partir deste ponto de interseção surgem as discussões epistemológicas, onde Design e Ciência apontam para preocupações éticas, procurando materializá-la através de ações. São questões que encontram os questionamentos sobre linguagem, conhecimento, controle e como a criatividade poderia fluir no cenário de interação humano-computador. Estas tensões entre teoria e prática nestas áreas têm sido um combustível para encontrar e criar novas soluções.

Assim, combinar a Informática com as discussões sobre Antropologia, Bioética, Design e História nos trouxe um diálogo frutífero, ilustrado neste texto. Entendemos que a onipresença da computação no dia a dia de todos vem resultando em impactos nem sempre tangíveis e que, portanto, carecem exercitar o pensamento.

2. INTRODUÇÃO

A bioética começa a tomar corpo como campo de estudo através da iniciativa do médico Van Rensselaer Potter nos anos 70, na época, Diretor Assistente do Laboratório McArdle para Pesquisa do Câncer da Escola de Medicina da Universidade de Wisconsin. Ele identificou a necessidade de desenhar um manual de sobrevivência humana e, em sua percepção, o caminho para atingir esse objetivo seria integrando valores, ou seja, perspectivas éticas, com a biologia (WHITEHOUSE, 2003). Portanto, o papel da bioética é antever os possíveis danos e difundir os “alertas sobre o progresso insustentável, principalmente nos sistemas de saúde” (WHITEHOUSE, 2003), além de intervir na pesquisa e na prática médica por meio de recomendações e anúncios.

A bioética pré-anunciada a “um profundo compromisso com o futuro” (POTTER, 1962) aproxima as práticas médicas das Ciências Sociais. Isso representa, não apenas preocupações com a justiça social e com o meio ambiente, mas também implica a plena compreensão dos sistemas de conhecimento e como eles apoiam a biopolítica.

Mais complexa do que desejável, a biopolítica é um conceito do filósofo francês Michel Foucault (1926-1984) apresentou no curso que ministrou no College de France, de 1978/79.

“Um novo tipo de racionalidade na arte de governar,” (FOUCAULT, 2008, p.25) a biopolítica é uma compreensão mais ampla de um sistema que interliga produções de verdade, por meio da linguagem e da ciência.

Ela está embutida na epistemologia que é funcional na operação do sistema industrial moderno e liberal, que, por exemplo, cria verdades baseadas nos interesses do sistema financeiro.

Segundo Foucault, há um marco histórico (século XVIII) onde as práticas de governo controlam vidas e corpos da população em geral por meio de rótulos e classificações em racionalidade ordenada aplicado à saúde, higiene, natalidade, longevidade, raça, gênero e assim por diante (FOUCAULT, 2008, p.431).

Enquanto a biopolítica está relacionada à política institucional estruturada, o biopoder é a ampliação do conceito quando o controle dos corpos não está apenas relacionado ao formato de cima para baixo, mas atinge a vida como uma prática de baixo para cima (FOUCAULT 1993).

Vigilância e vergonha social representam os sistemas de poder. Empoderada por projetos como o Panóptico (FOUCAULT, 1979), a vigilância constante cria uma possibilidade imperiosa de punição. Naturaliza a automação dos corpos e sua mecanização. Ou seja, um corpo não é livre para se expressar, ele vive cerceado por uma série de regras. Em última instância, epidemias e pandemias “deram origem a projetos disciplinares” (FOUCAULT, 1979), que se espalharam como sistemas de micropoder e controle, quando as pessoas passam a se policiar umas às outras.

Além disso, o controle do corpo e do comportamento se dá por meio do conhecimento. Como apontado por Donna Haraway, “a biologia conta contos sobre origens, sobre a gênese e a natureza” (HARAWAY, 1991, p.72); assim, algumas verdades do mundo biológico são construídas por interpretação tendenciosa. A imunologia, por exemplo, incorporou “um mapa biopolítico dos principais sistemas de ‘diferença’.” (HARAWAY, 1991, p.1-2). Em outras palavras, é um sistema que faz distinções entre o suposto “natural” e os “outros inadequados”. Uma ideia territorial de guerra e invasões. Como se o corpo estivesse separado do meio ambiente e livre da influência de diversos seres. No entanto, se considerarmos que um corpo tem mais bactérias do que células, a hostilidade geral em relação a microrganismos, como as bactérias, parece ligeiramente distorcida.

Alinhado a tal perspectiva crítica, o professor Geoffrey Bowker começou a pesquisar como os sistemas de classificação aparecem na história e se espalharam durante o século XIX para as universidades se enraizando no cerne de todas as disciplinas (G. C. BOWKER, 2014).

A hierarquia do conhecimento privilegia as ciências da saúde e da natureza, guiadas pela noção de objetividade, enquanto as ideias positivistas que procuram transferir para as ciências humanas algum senso de objetividade, são ainda presentes de modos implícitos. Impor objetividades a corpos subjetivos é um fator obstrutivo para a boa fluência da bioética.

Assim, a compreensão de como a medicina deve incorporar a justiça social e as preocupações ambientais na assistência à saúde desafiam seu principal papel social, quando ‘cuidar’ dos corpos também significa ‘controlá-los’. Para analisar como os valores estão embutidos no conhecimento e, consequentemente, nos sistemas computacionais, é necessário ampliar o repertório de possíveis interpretações sobre enunciados, conceitos e palavras. Ou seja, dar atenção aos estudos culturais não apenas entre diferentes etnias ou outros pontos

de diferença, como religião ou nacionalidade, mas também às “identidades”, que na sociedade ocidental vêm se multiplicando como células desordenadas.

Em seus trabalhos, Bowker também chama a atenção para a forma como a ciência cria mitos com ferramentas como “sistemas de classificação” para definir as coisas em vez de descrevê-las. Ele revela as exigências morais associadas a cada etapa da observação da pesquisa. Esses são valores, interpretações, desejos, projeções e necessidades que são externas ao objeto.

O preconceito embutido expressa como a separação entre sujeito-objeto e quaisquer outras formas dualistas é um axioma distante dos fatos.

3. ENTREVISTADO

Geoffrey Bowker, enquanto professor da Universidade da Califórnia em Irvine, criou o Laboratório de Valores em Design, também chamado *Evoke Lab*.

Ele tem um Ph.D. da Universidade de Melbourne, Austrália, em História e Filosofia da Ciência, e fez sua pesquisa de pós-doutorado na École des Mines em Paris.

Começou a trabalhar como Professor de Cyberscholarship na iSchool, Universidade de Pittsburgh, e também foi Diretor Executivo e Professor Regis e Dianne McKenna no Centro de Ciência, Tecnologia e Sociedade da Universidade da Califórnia Santa Clara.

Sua pesquisa tem sido sobre o uso da web e de outros recursos digitais em um conjunto de disciplinas, sempre trabalhando com acadêmicos em ambientes multidisciplinares para descobrir maneiras pelas quais novas formas de conhecimento estão sendo (ou poderiam ser) geradas pelo uso criativo desses recursos digitais.

Além disso, ele examinou como a infraestrutura de informações envolve olhar para os sistemas de classificação em mudança, como na medicina. Ele abordou, como uma questão analítica central, como os cientistas nas várias ciências contribuem para o tema da biodiversidade, e como as estruturas de dados e práticas em uso afetam a comunicação científica.

Ele publicou um livro sobre gerenciamento de informações e geofísica industrial na Schlumberger, “Science on the Run”, e também outro livro com Susan Leigh Star, “Sorting Things Out: Classification and its Consequences” com a MIT Press em 1999. Grande parte de seus artigos científicos publicados enfatizam as características sociais e organizacionais

das cyber-infraestruturas científicas emergentes. Uma publicação mais recente, “All Knowledge Is Local”, está disponível em Comunidades de Aprendizagem: Um Jornal Internacional de Aprendizagem em Contextos Sociais em 2010. Geoffrey Bowker faz parte do conselho editorial da Sociedade da Informação, Informação e Organização, Metaciência e Estudos Sociais da Ciência.

Sua proposta de missão para o *Laboratório de Valores em Design* foi treinar pesquisadores na ampla gama de disciplinas necessárias (incluindo Informática, Ciência da Computação, Design e Estudos de Ciência e Tecnologia) com o objetivo de produzir novas formas de sistemas de informação e tecnologia que expressem e desempenhem os valores sociais e éticos.

Também o *Evoke Lab* começou com um olhar voltado para as questões e preocupações sociais, integrando teoria social, arte, ciência da computação e design, para criar novas tecnologias e artefatos que evocam o pensamento, o debate, a discussão e a consideração de um mundo que valorize o bem-estar individual e social. Com o objetivo de facilitar a execução de projetos que variam de pequenos experimentos com inovações a estudos em larga escala sobre questões importantes como “big data”, privacidade, justiça social e autoexploração por meio da tecnologia. A ideia do laboratório foi de criar uma comunidade aberta de acadêmicos e designers interdisciplinares para se divertir, aprender e crescer juntos.

4. ENTREVISTA

E.N. “Como você começou a dar atenção especial aos valores no design, e você acha que o uso de tecnologias emergentes está mudando os valores humanos?”

G.B. “A atenção aos valores no design veio quando começamos a olhar para os sistemas de classificação para abordar as questões de valores políticos, sociais e culturais. No início, não queríamos ir para o caso fácil, como ir para a psicologia, onde é realmente óbvio onde estão os valores. Queríamos pegar casos mais difíceis, como a classificação médica ou a classificação de vírus porque isso mostra como a política se exerce por meio do design, Assim, quando você uma nova tecnologia é criada, uma nova humanidade está a caminho.

É impossível pensar em novas tecnologias que não mudem os valores humanos.”

E.N. “Como você acha que os valores estão inseridos em nossos sistemas de computacionais?”

G.B. “Gosto de pensar nisso geologicamente. A primeira camada nos sistemas computacionais são a arquitetura. Isso não mudou desde 1947. Há valores embutidos na forma como os programas computacionais são organizados. Eles revelam crenças sociais. Frequentemente, você vê como a história do computador reflete a forma como as organizações funcionam. De Babbage⁷⁹ vemos essa ideia de usar a mão de obra mais barata possível como base para a industrialização mecanizada.

A organização tem tudo a ver com entradas e saídas. A linguagem de programação tem todos os níveis de valores. Um nível é arquitetônico, o segundo nível é a linguagem, o terceiro nível são os aplicativos com os quais lidamos todos os dias. E eles têm muitos valores políticos e sociais.”

E.N. “Como você acha que as pessoas começaram a se dar conta disso? Quando os cientistas da computação e desenvolvedores começaram a se dar conta dessas questões?”

G.B. Primeiramente, desde 2003, eu e outros colegas começamos a organizar uma série de workshops em diferentes universidades como UCI, Universidade de Santa Clara (UCSC), Helsinque, NYU e Stanford. Isso reuniu cientistas da computação com cientistas sociais, e também pessoas dos estudos de negócios, publicidade e humanidades em geral, para discutir e trabalhar em projetos para explorar os valores embutidos em sistemas computacionais.

Hoje em dia, aqueles por trás do Facebook, Twitter e assim por diante são forçados a reconhecer que estão propagando certas crenças e valores.

O Google, por exemplo, mostrou a natureza do racismo embutido nas pesquisas. Isso foi mencionado no trabalho de Safiya Noble (NOBLE, 2018).

79 Babbage foi um matemático, filósofo, inventor e engenheiro mecânico que escreveu sobre diversos , philosopher, inventor and mechanical engineer que escreveu sobre uma série de questões, incluindo uma visão sobre “On the Economy of Machinery and Manufactures” (1832), onde discute a organização da produção industrial. Fonte: <https://www.britannica.com/biography/Charles-Babbage> and https://en.wikipedia.org/wiki/Charles_Babbage#Declinarians,_learned_societies_and_the_BAAS

Mas, no final das contas, apresentar preocupações éticas é como jogar Whac-a-mole⁸⁰ com o Google. Eles consertam uma parte e outra questão com viés político aparece. Portanto, houve um reconhecimento crescente por parte dos empresários do Vale do Silício sobre como eles estavam incorporando valores sociais em suas máquinas e programas. Talvez, em vez de lidar com cada caso, eles precisem lidar como um todo.

Fizemos muitos projetos colocando essas pessoas juntas. Mas pela experiência com o projeto maior e mais longo, percebemos que após 8 anos eles (da Ciência da Computação) não estavam realmente ouvindo o que (nós das Humanidades) dizíamos. Eles não conseguiam entender por que deveriam falar sobre ética/valores. Afirmavam frequentemente que estavam apenas fazendo a arquitetura e os valores ocorrem em camadas acima da infraestrutura. Mas nós defendemos que as maneiras com as quais as infraestruturas são concebidas contêm valores.

Marshall McLuhan argumentou que a tecnologia se desenvolve principalmente para usos militares e os outros usos vêm posteriormente (McLUHAN M, FIORE Q, 1968). Usamos guerras como forma de testar e desenvolver tecnologia. O DARPA⁸¹ foi o centro do desenvolvimento da Internet. Eles tinham preocupações com a segurança interna, e o que se quer é uma comunidade distribuída, que armazene conhecimento em diferentes lugares. Este é um diagrama inicial sobre como a Internet deve funcionar. Portanto, a conexão entre o desenvolvimento militar e da computação foi absolutamente essencial.

No sentido dos valores, estou sempre brincando de “catch up” porque os computadores também pertencem ao trabalho cooperativo, com design participativo. Na Escandinávia, se você deseja desenvolver aplicativos de computador, precisa consultar os trabalhadores, também seus sindicatos, ao mesmo tempo. Portanto, um novo aplicativo deve ser desenvolvido em conjunto por cientistas da computação, designers e assistentes sociais.

Para mim, tem sido um grande objetivo realizar isso na prática.

80 “O termo “Whac-a-mole” (ou “Whack-a-mole”) é usado coloquialmente para denotar uma tarefa repetitiva e fútil: cada vez que uma tarefa é concluída ou um problema é resolvido, ainda outra tarefa/problema aparece em outro lugar.” Fonte: <https://en.wikipedia.org/wiki/Whac-A-Mole>

81 Defense Advanced Research Projects Agency — Agência de Projetos de Pesquisa Avançada de Defesa (traduzido para o português) (<https://www.darpa.mil>)

E.N. “Você e alguns outros pesquisadores já pensaram em estratégias para abordar as pessoas que são mais resistentes a pensar em valores e tecnologia?”

G.B. “Um pós-doutorando sob minha supervisão abriu uma empresa no Vale do Silício chamada *Ethical Resolve*⁸². Começamos a trabalhar com um grupo chamado *Planet Labs*⁸³ e a ideia deles era colocar um floco de satélites no espaço. Eles tinham cerca de 200 satélites do tamanho de uma geladeira. Com a pressão sobre a discussão de grupos antinucleares na Inglaterra, eles perceberam que teriam problemas éticos, então contrataram uma consultoria ética. O *Ethical Resolve* também treinou a equipe de IA do Facebook em torno das questões de valores.

Também estive envolvido com a criação do *Data & Society* no grupo liderado por Danah Boyd⁸⁴ da NYU e que tem sido uma organização verdadeiramente engajada, trabalhando entre a indústria e as ciências sociais.

Eles reúnem ativistas políticos que questionam o quanto os dados reproduzem ou contestam valores sociais. Nós organizamos workshops e ambientes onde todas essas pessoas (grupos diversos) possam trabalhar juntas. O objetivo foi gerar integração e diálogos entre diferentes áreas.

Porque não sabemos as respostas e não existe uma fórmula para ordenar as questões mais complexas, o desafio é o debate, que é a comunicação entre diferentes tipos de conhecimento e objetivos. No dia a dia, muitos problemas diferentes surgiram e não temos nenhuma solução perfeita. Precisamos criar engajamento, e só podemos fazer isso se continuarmos trabalhando por alguns anos. Essa seria a estratégia de maior sucesso.

E.N. “Como você acha que os valores embutidos na arquitetura podem mudar nossa percepção do tempo e de estar no mundo?”

G.B. “O curso que estou ministrando agora, ‘Life at the Femtosecond’, está trabalhando para perceber que muitas decisões estão sendo tomadas muito mais rápido do que qualquer pessoa tem a chance de processá-las. Meu argumento fundamental é que precisamos de um novo tipo de política. A velha política sustenta que tudo gira em torno da esfera pública com a crença de que este é um espaço aberto para discutir questões políticas. Mas agora a política

82 <https://ethicalresolve.com/index.html>

83 <https://www.planet.com>

84 <https://www.danah.org>

está acontecendo dentro dos sistemas computacionais. Então, o que precisamos é de um conjunto de estratégias políticas, ações e discursos que possam envolver todos os níveis temporais trabalhando dentro dos computadores.

Quando Michel Foucault fala sobre a noção de democracia, a ascensão do Estado liberal, ele diz que não é realista falar de democracia como uma coisa que se desenvolve ao longo do tempo. Da ágora dos gregos, à forma como funciona na Escandinávia, ao sistema parlamentar britânico, à democracia que temos hoje.

A democracia é um índice diferente pelo que a tecnologia tem. Portanto, não existe “coisa” chamada democracia. Essa é uma ideia estúpida. O que há são valores que sempre fazem parte da era da mídia, o que é o processamento da informação, e que muda de regime técnico para regime técnico, de um grupo técnico para outro grupo técnico.

No início do século vinte, você olha para o campo da comunicação. Isso é tudo sobre comunicação de massa e como controlamos a comunicação de massa? Os jornais e a televisão foram uma revolução na comunicação em sua época. Agora precisamos de cientistas políticos que sejam capazes de entrar dentro dos sistemas computacionais porque é lá que ocorrem as lutas políticas.

E.N. “Certo, somos solicitados a tomar decisões muito rapidamente; e os “termos de serviço” de todos os softwares e aplicativos que exigem que concedamos acesso aos seus serviços?”

G.B. “Anos atrás, você não conseguiria ler os “termos de serviço” do *Facebook* sem “atingir o tempo limite”.

Essas regras dão boas ideias sobre como deve ser: se pudéssemos visualizar facilmente cada recurso e verificar aqueles que você está dando permissão. Mas esse tipo de organização não é do interesse de empresas como o *Facebook*, porque tudo o que elas querem são suas informações. As decisões de design podem tornar isso muito melhor. Os europeus estão se saindo muito melhor do que os americanos nesse aspecto agora.

Por exemplo, a proposta de regulamentação Europeia considera o “direito de saber” quando os algoritmos tomam uma decisão sobre você. Isso é uma coisa extremamente difícil de decidir: como o algoritmo sabe disso? Por que eles colocaram você nesta classificação em vez de outra classificação? Por isto, o direito de saber é extremamente desafiador.”

E.N. “Da citação, *“O primeiro tema principal é ver a onipresença da classificação e padronização. Os esquemas e padrões de classificação saturam literalmente os mundos em que vivemos”* (STAR, 1999), como você acha que a padronização e a classificação necessárias para a maioria dos procedimentos científicos, a organização de dados e a coleta em sistemas de computadores inteligentes podem nos desviar do tempo presente?”

G.B. “Os indivíduos ficam presos em classificações geridas por estratégias de marketing. Agora eles têm acesso a todas as suas pesquisas e, a partir da sua navegação, pensam que sabem quem você é. Então, eles fornecem informações para incentivá-lo a ser mais como eles o veem.”

E.N. “Sobre *“Sorting Things Out”* (STAR, 1999) o livro de sua coautoria, você acha que os padrões e classificações que estão imbricados em nossas vidas se tornam mais invisíveis com a onipresença da computação?”

G.B. “Há um número infinito de padrões e classificações trabalhando em mim. Esta é uma realidade alternativa, e não posso reconhecer essa realidade alternativa como falsa porque tudo o que vejo é essa realidade alternativa.”

E.N. “Já na Califórnia, é possível pedir às empresas que compartilhem as informações que têm sobre você. Mas imagino que se eu perguntar a eles sobre o acesso à informação, possivelmente irão me dar o que eu já dei a eles com os conteúdos que postei, quando o que eu quero são os resultados das informações analisadas. Pois somente com acesso a tais informações poderei entender melhor como os sistemas de classificação me lêem e operam em mim. Mas como obter os dados processados do seu perfil?”

G.B. “Essa é uma boa pergunta. Acho que para isso precisamos de novas políticas. A revolução tecnológica aconteceu, mas não precisamos aceitar a forma atual da realidade.”

E.N. “Como você vê o processo histórico do mundo moderno, mais precisamente, os aspectos na mudança sobre a percepção do tempo, as mudanças no corpo humano e na sensação de estar no mundo?”

G.B. “Há a abordagem marxista que vai explicar como, por que, e o que vai ser diferente com a industrialização. Eu prefiro visões contemporâneas como da Donna Haraway, que fala de ciborgues como identidade e

processo de criação. Isso é o entendimento conceitual sobre como mesclar humanos e máquinas.”

E.N. “Você acha que a interação homem-computador reforça a sensação de que os seres humanos são algo além da natureza?”

G.B. “Sim. Isso potencialmente reforça a ideia do excepcionalismo.”

E.N. “Na ciência e na tecnologia nem sempre é possível prever os impactos negativos das conquistas destinadas a solucionar problemas ou preencher lacunas. Tomando o exemplo da saúde, o que você acha que poderia ser uma solução para evitar o uso indevido de conhecimento e dados?”

G.B. “O uso indevido sempre é possível, mas acho que devemos continuar coletando e usando dados. O problema começa e, conforme começamos a aplicar o controle a ele, o controle se torna o problema, não o fluxo livre de dados.

Para o caso médico, sim, eles devem ser capazes de coletar os dados. Genericamente, quanto mais dados, melhor. Como eles usam os dados e como são capazes de analisá-los, é o que precisamos controlar. Não estamos absolutamente certos ou errados eticamente, precisamos de um sistema político em que possamos confiar.

Portanto, a questão não é sobre regras e regulamentos. A questão é sempre como processar e como devemos lidar com esses dados que estão chegando.

É sobre como podemos usá-lo e por que não podemos usá-lo. Eu, genericamente, critico o conceito de dados. Você prefere um mundo que seja auto-consciente ou que não seja auto-consciente?

Do que se trata esse conceito de tecnogênese Hayles (HAYLES, 2012), em que humanos e máquinas coexistem?

Se eu olhar para isso, quero a melhor comunicação possível. Por exemplo, minha pergunta sobre COVID-19 é por que temos esse conceito de destruir o forasteiro. Compartilhamos informações e, porque compartilhamos informações, somos capazes de coabitar.

E.N. “Em relação ao controle dos corpos por meio de dados médicos, como você está vendo hoje aquelas coisas que investigou no passado e como está sua percepção dessas questões desde então, com mais de vinte anos de rápido aumento do desenvolvimento tecnológico? Estou pensando nesta citação dos “múltiplos corpos do prontuário”: *“Nos concentramos no prontuário porque*

este artefato ocupa um nicho central nesta rede: é onde muitas das tarefas dos enfermeiros e médicos começam, terminam, e são coordenadas, onde as inscrições se acumulam e onde se desenvolvem os espaços e tempos específicos que descreveremos. O registro não se limita a espelhar os corpos que mapeia, argumentamos - mas também não os determina. Para enfatizar o papel ativo dos artefatos sem cair no determinismo tecnológico, utilizamos o termo “mediação”: o registro medeia as relações que organiza os corpos que se configuram por meio dele.” (traduzido pelos autores) (MB e G. BOWKER, 1997)

G.B. “Isso me lembra Michel Serres⁸⁵ na ideia crítica de que a natureza tem que ser controlada, e também Foucault no “Nascimento da Clínica”: não existe corpo médico, essa é uma categoria equivocada.

O mundo é feito de corpos que se interpenetram em todos os níveis, desde o nível macro até o nível muito micro.

A dicotomia mente-corpo é um erro fundamental de Descartes e do Iluminismo.

E.N. “Abordando a discussão do homem como máquina, existem muitos desenvolvimentos na ciência ocidental orientados por esta metáfora. Quando essa metáfora começa a ameaçar os valores humanos, ou seja, sua cultura e tradições?”

G.B. “Uma das coisas ruins que existe é dizer que uma máquina é fundamentalmente diferente das pessoas, como dizer que uma máquina é fundamentalmente mecânica. Pessoas e máquinas estão juntas no mundo definindo moralidade, definindo valores humanos.

Existem tentativas de fazer a participação política parecer um jogo. O jogo político que está acontecendo agora está prestes a nos transformar em máquinas e os humanos estão se tornando cada vez mais irrelevantes.

A questão é que os valores estão dentro das máquinas e de nossa infraestrutura e isso estabelece uma relação entre nós, máquinas e infraestrutura.

Há uma decisão que precisamos tomar agora, abordando nossos computadores e tecnologia da informação com relação à ética e valores.”

85 Sobre Michel Serres: <https://muse.jhu.edu/article/739471>

E.N. “Voltemos a comentar mais sobre sistemas de classificação. Você pensa em alguns exemplos para citar, como estratégias diferentes e criativas que podem propor outras alternativas para modificar isso?”

G.B. “Onde os valores são integrados? Existem exemplos de reconhecimento de imagem que excluem automaticamente bebês do sexo masculino. Eles bloqueiam isso. É o mesmo com as mulheres que bebem cerveja, não é reconhecido. Paul Ohm⁸⁶ um acadêmico jurídico do Colorado explica que existem cerca de dezesseis níveis cujos valores entram na construção de algoritmos.

Os valores já estão lá no momento em que o reconhecimento está sendo feito.

Na Grã-Bretanha, nos primeiros dias, um grupo colocava palavras-chave que soavam como “terrorista”, e isso era uma estratégia política.

Sobre o uso da imagem da mulher com o objetivo de as envergonhar, eu penso que a melhor resposta a certa moralidade é a desafiando, mostrando mais do que é considerado tabu.

Mas toda a ideia de privacidade é manter segredos de coisas que talvez não devêssemos estar fazendo.

Daniel Howe, um artista de Hong Kong, criou a obra chamada “track me not”⁸⁷ como um aplicativo para o Google.

A ideia era que sempre que você pesquisasse algo no Google, o aplicativo deslizaria a pesquisa para diferentes questões e, no final, eles não saberiam o que você estava procurando.

Isso foi criticado no campo dos valores com o argumento de que cada uma dessas consultas custa dinheiro e energia. Genericamente, foi uma ideia muito boa confundir os sistemas de classificação para bloquear a maneira como eles desejam rastrear cada indivíduo.”

E.N. “Sobre o controle extremo dos dados: você acha que esse controle pode ser usado para suprimir o direito à informação? Em outras palavras, você acha que a regulação dos dados pode ser mal interpretada?”

86 Ohm P. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization’(2010). UCLA Law Review.; 57:1701. (Disponível em: <https://www.uclalawreview.org/pdf/57-6-3.pdf>)

87 <https://trackmenot.io>

G.B. “A informação quer ser livre e os dados devem ser livres. A regulamentação da privacidade é altamente problemática e se torna um conceito inútil.

Estou continuamente sendo microgerenciado por programas como Facebook, Google e assim por diante. Eu gostaria de impedi-los de fazer isso.

Em um bom cenário, não seria um problema compartilhar dados. “

E.N. “Agora estamos vivendo um momento sem precedentes com uma pandemia e enquanto estamos todos nos tornando virtuais, como você acha que isso está mudando nossa percepção do tempo?”

G.B. “Agora? Esta é uma boa pergunta. Para mim, genericamente quando me movo pelo mundo eu mudo minha atenção para fora, pois estou conhecendo pessoas diferentes, personagens diferentes, em culturas e ambientes diferentes, em momentos diferentes. Agora eu tenho que lidar muito mais nos últimos meses, com a integração de quem sou socialmente, quem sou espiritualmente, quem sou politicamente, quem sou academicamente. São categorias influenciadas pelos espaços, mas não consigo mais separar isso.

Costumava ser tipo, eu iria trabalhar na universidade de manhã e deixaria algumas células de mim para trás. Eu não posso mais deixá-los para trás porque não há para onde ir. Fico na mesma sala 16 horas por dia. Todas as partes de mim, integrando ou reintegrando o mesmo espaço.

Isso tem sido muito interessante e me causa um bom impacto, que é reintegrar partes do eu que são tão fáceis de separar. Tipo: “este é o meu tempo na estrada”, “este é o meu tempo na universidade”, “este é o meu tempo de fazer isto”, “este é o meu tempo de fazer aquilo”. Não consigo mais separar esses momentos.

Me inspira o exemplo do *Decameron*⁸⁸, de Boccaccio, onde um grupo foge da peste e se encontra fora de Veneza para criar e contar histórias uns aos outros.

Nossas gerações não têm sido boas para pensar o fim da vida, pensar na morte. Genericamente, acho que parte do que é feito é que todos os computadores e tecnologias estão tentando nos convencer de que somos basicamente imortais, estamos basicamente no presente e está ficando cada vez mais rápido.

88 “O *Decameron* compreende um grupo de histórias unidas por uma história principal. Quando a narrativa do quadro se abre, 10 jovens (sete mulheres e três homens) fogem da Florença atingida pela peste para uma vila encantadora nas proximidades de Fiesole. Cada um rege por um dia e estabelece estipulações para os contos diários a serem contados por todos os participantes, resultando num acervo de 100 peças. (Fonte: <https://www.britannica.com/topic/Decameron>)

E a grande questão é entender a (i)mortalidade.

O que eu adoraria seria se realmente aproveitássemos isso como uma oportunidade para reconhecer que estamos todos juntos nisso.

Incluindo bactérias, vírus, plantas, animais não humanos e assim por diante.

Esta experiência está fazendo isso por mim, mas não estou convencido de que esteja fazendo o mesmo pelos outros”.

5. REFERÊNCIAS

Bowker, G. C. Emerging Configurations of Knowledge Expression. In and K. A. F. Tarleton Gillespie, Pablo J. Boczkowski (Ed.), **Media Technologies: Essays on Communication, Materiality, and Society** (pp. 99–118). The MIT Press. 2014.

Bowker, M. B. and G. The multiple bodies of the medical record: Toward a sociology of an artifact. **Sociological Quarterly**, 38 (3), 513–537. 1997.

Donna J. Haraway. **Simians, Cyborgs, and Women: The Reinvention of Nature**. Routledge. 1991.

Foucault, M. **Discipline and Punish: The Birth of the Prison** (Vintage Bo). A Division of Random House. 1979.

Foucault, M. **Microfísica do Poder**. Graal. 1993.

Foucault, M. **O nascimento da Biopolítica**. Martins Fontes. 2018.

Hayles, N. K. **How we think: Digital media and contemporary technogenesis**. University of Chicago Press. 2012.

McLuhan M, Fiore Q, A. J. **War and peace in the global village**. Bantam Books. 1968.

NAS, Elen. **Ethics of Emerging Technologies: An Interview with Geoffrey C. Bowker**. Engaging Science, Technology, and Society 8.2, pp. 176-180. 2022.

Noble, S. U. **Algorithms of oppression: How search engines reinforce racism**. nyu Press. 2018.

Potter, V. Bridge to the future: the concept of human progress. **Land Economics**, 38(1), 1–8. 1962.

Star, G. C. B. and S. L. **Sorting Things Out: Classification and Its Consequences**. The MIT Press. 1999.

Whitehouse, P. The rebirth of bioethics: extending the original formulations of Van Rensselaer Potter. **American Journal of Bioethics**, 38(1), 26–31. 2003.

COLABORADORES

Ademir Nunes Ribeiro Júnior

Professor, Coordenador Pedagógico e membro do Núcleo Docente Estruturante, Curso de Enfermagem da Universidade Evangélica de Goiás. Doutorando no Programa de Pós-Graduação em Bioética, Ética Aplicada e Saúde Coletiva, Universidade Federal do Rio de Janeiro. Mestre em Ciências da Saúde, Universidade Federal de Viçosa. Bacharel em Enfermagem, Universidade Federal de Viçosa.

Adenauer Girardi Casali

Professor Adjunto do Instituto de Ciência e Tecnologia da Universidade Federal de São Paulo. Doutor em Fisiologia, Università degli Studi di Milano (Milão, Itália). Mestre em Física, Universidade de São Paulo. Bacharel em Física, Universidade Federal de Santa Maria.

Adilson Costa da Silva

Professor Adjunto do Programa de Engenharia Nuclear da Universidade Federal do Rio de Janeiro. Mestre em Ciências em Engenharia Nuclear, Universidade Federal do Rio de Janeiro. Doutor em Ciências em Engenharia Nuclear, Universidade Federal do Rio de Janeiro. Licenciado em Física, Universidade Federal Rural do Rio de Janeiro.

Ana Luisa Lima Grein

Graduada em História e Mestranda em Filosofia pela Universidade Federal de São João Del-Rei.

Anderson Luiz do Vale Fonseca

Mestrando em Filosofia pela Universidade Federal de São João Del Rei.
Graduado em Literaturas de Língua Portuguesa, Universidade Estácio de Sá.

Andréia Patrícia Gomes

Professora Associada do Núcleo de Bioética e Ética Aplicada, Universidade Federal do Rio de Janeiro. Docente Permanente do Programa de Pós-graduação em Bioética, Ética Aplicada e Saúde Coletiva pela Escola Nacional de Saúde Pública Sérgio Arouca, Fundação Oswaldo Cruz. Ex-professora Titular do Curso de Graduação em Medicina de Doenças Infecciosas e Parasitárias, Centro Universitário Serra dos Órgãos (2001-2009). Doutora em Ciências (Saúde Pública), Escola Nacional de Saúde Pública Sérgio Arouca, Fiocruz. Mestre em Medicina Tropical pelo Instituto Oswaldo Cruz, Fiocruz. Especialista (Residência Médica) em Doenças Infecciosas e Parasitárias, Universidade Federal do Rio de Janeiro. Diplomada em Medicina, Universidade Federal do Rio de Janeiro. Diplomada em Direito, Universidade Federal de Viçosa.

Beatriz Gomes Batista

Diplomada em Direito Internacional e Europeu pela Rijksuniversiteit Groningen (Holanda). Diplomada pelo United World College Costa Rica com diploma International Baccalaureate bilíngue (Inglês-Espanhol).

Bruno Carlos da Cunha Costa

Professor do Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro. Doutor em Informática, Universidade Federal do Rio de Janeiro. Mestre em Informática, Universidade Federal do Rio de Janeiro. Bacharel em Ciência da Computação, Centro Universitário Serra dos Órgãos.

Daniel Louzada Fernandes

Professor Adjunto do Departamento de Informática, Universidade Federal de Viçosa. Pós-doutorando em Meteorologia Aplicada, Universidade Federal de Viçosa. Doutor e Mestre em Ciência da Computação, Universidade Federal de Viçosa. Bacharel em Sistemas de Informação, Faculdade do Espírito Santo.

Débora Fernandes Pessoa Madeira

Professora de Direito Civil, Universidade Federal de Viçosa. Ex-Chefe do Departamento de Direito da Universidade Federal de Viçosa. Mestre em Direito Privado, Pontifícia Universidade Católica de Minas Gerais. Especialista em Direito Civil e Processual Civil, Universidade Presidente Antônio Carlos. Graduada em Direito pela Universidade Federal de Viçosa.

Fábio Ribeiro Cerqueira

Professor associado do Departamento de Engenharia de Produção – Escola de Engenharia de Petrópolis, Universidade Federal Fluminense. Professor colaborador no Programa de Pós-graduação em Ciência da Computação do Departamento de Informática - Universidade Federal de Viçosa. Doutorado em Informática Biomédica, University for Health Sciences, Medical Informatics and Technology, Áustria. Mestrado em Ciência da Computação, Universidade Estadual de Campinas. Especialização em Bioinformática, Universidade de Lisboa, Portugal. Graduação em Informática, Universidade Federal de Viçosa.

Fabíola Alves Alcântara

Professora da Universidade dos Guararapes. Pós doutoranda no Programa de Ciências da Saúde da Universidade Federal de Viçosa. Doutora e Mestre em Bioética, Ética Aplicada e Saúde Coletiva, Universidade Federal do Rio de Janeiro. Especialista em Fisioterapia em Oncologia pela Associação Brasileira de Fisioterapia em Oncologia. Especializada (Residência Multiprofissional) em Saúde do Adulto e Idoso pela Universidade Federal de Sergipe, e em Cuidados Paliativos, pelo Instituto de Medicina Integral Prof. Fernando Figueira. Especialista em Inteligência Artificial e Visão Computacional, Universidade Federal de Viçosa. Especialista em Saúde 4.0: Gestão, tecnologia e inovação, Pontifícia Universidade Católica do Paraná. Bacharel em Fisioterapia, Faculdade Estácio FaSe de Sergipe.

Francisco de Paula Barretto

Professor Adjunto do Instituto de Humanidades, Artes e Ciências Prof. Milton Santos. Coordenador do Laboratório de Interatividade, Computação e Novas Interfaces (ICON), Universidade Federal da Bahia. Doutor em Arte e Tecnologia, com ênfase em Criatividade Computacional e Inteligência

Artificial, Universidade de Brasília. Mestre em Arte e Tecnologia, com ênfase em Inteligência Artificial, Universidade de Brasília. Bacharel em Ciência da Computação, Centro Universitário de João Pessoa.

Gabriel Resende Machado

Pesquisador em Ciência da Computação, Instituto ECOA (PUC-Rio). Mestre em Sistemas e Computação, Instituto Militar de Engenharia (IME-RJ). Bacharel em Ciência da Computação, Centro Universitário Serra dos Órgãos (UNIFESO).

Geisla Teles Vieira

Professora Titular da Escola de Medicina, Faculdade Dinâmica do Vale do Piranga. Doutora em Bioquímica Aplicada, Universidade Federal de Viçosa. Mestre em Ciências Farmacêuticas, Universidade Federal de Ouro Preto. Diploma (Licenciatura e Bacharelado) em Ciências Biológicas, Universidade Federal de Ouro Preto.

Geoffrey Charles Bowker

Professor da Faculdade de Informação e Ciência da Computação, Universidade da Califórnia, Irvine. PhD em História e Filosofia da Ciência, Universidade de Melbourne, Austrália. Especializado em História Britânica do Século XIX, Australian National University. Diplomado em História, Australian National University.

Gláucio Inácio da Silveira

Professor Adjunto do Departamento de Direito, Universidade Federal de Viçosa. Mestre em Direito Processual, Universidade do Estado do Rio de Janeiro. Bacharel em Direito, Universidade Federal de Viçosa.

Gustavo Leal Toledo

Professor Associado do Departamento de Filosofia e Métodos da Universidade Federal de São João Del-Rei. Doutor em Filosofia, Pontifícia Universidade Católica do Rio de Janeiro. Mestre em Filosofia, Pontifícia Universidade Católica do Rio de Janeiro. Graduado em Filosofia, Universidade do Estado do Rio de Janeiro.

José Abdalla Helayël Neto

Pesquisador Titular, Centro Brasileiro de Pesquisas Físicas. PhD in Physics, International School for Advanced Studies. MPh in Physics pela in Physics pela International School for Advanced Studies. MSc Pontifícia Universidade Católica do Rio de Janeiro. Bacharel em Física, Pontifícia Universidade Católica do Rio de Janeiro.

Leimin Tian

Pesquisadora no Human-Robot Interaction group, School of Electrical and Computer Systems Engineering, Monash University. PhD in Informatics, com ênfase em Computação Afetiva, School of Informatics, University of Edinburgh. MSc in Artificial Intelligence (with Distinction), School of Informatics, University of Edinburgh. B.Eng in Intelligent Science and Technology, School of Computer Science, Beijing University of Posts and Telecommunications.

Leonardo Vasconcelos Guaurino de Oliveira

Sócio do Escritório Vasconcelos Guaurino e Guapyassú Vianna Advogados Associados. Doutorando em Direito Privado e Médico, Universidade Autônoma de Lisboa. Mestre em Direito Privado e Médico, Universidade Autônoma de Lisboa. Especialista em Direito Civil-Constitucional, Universidade do Estado do Rio de Janeiro. Diplomado em Direito, Universidade Federal do Estado do Rio de Janeiro.

Lindisley Ferreira Gomides

Professora Titular da Escola de Farmácia, Faculdade Dinâmica do Vale do Piranga. Docente Permanente do Programa de Mestrado em Ensino de Ciências da Saúde e do Ambiente (PROCISA) da Faculdade Dinâmica do Vale do Piranga. Doutora em Biologia Celular, Universidade Federal de Minas Gerais. Mestre em Fisiologia e Farmacologia, Universidade Federal de Minas Gerais. Farmacêutica Generalista, Centro Universitário do Leste de Minas Gerais.

Maria Aglaé Tedesco Vilar do

Professora da Escola da Magistratura do Tribunal de Justiça do Estado do Rio de Janeiro. Doutora em Bioética, Ética Aplicada e Saúde Coletiva pelo

Programa de Pós-Graduação em Bioética, Ética Aplicada e Saúde Coletiva (PPGBIOS) em associação da Universidade do Estado do Rio de Janeiro, Universidade Federal do Rio de Janeiro, Universidade Federal Fluminense e Fundação Oswaldo Cruz. Mestre em Pesquisa e Planejamento Urbano e Regional, Instituto de Pesquisa e Planejamento Urbano e Regional. Diplomada em Direito, Universidade Federal do Rio de Janeiro. Presidente do Fórum Permanente de Biodireito, Bioética e Gerontologia da Escola da Magistratura do Estado do Rio de Janeiro- EMERJ. Presidente do Núcleo de Pesquisas em Bioética e Saúde Social da EMERJ.

Mario Livio

Ex-Astrofísico do Space Telescope Science Institute (1991-2015). Ex-Professor de Física, Techion-Israel Institute of Technology (1981-1991). Ph.D. in theoretical astrophysics at Tel-Aviv University. M.S. degree in theoretical particle physics at the Weizmann Institute. Bacharelado em Física e Matemática, Universidade Hebraica de Jerusalém.

Matheus Moura Novelli

Diplomado em Medicina na Universidade Federal de Viçosa. Técnico em Edificações pelo Instituto Federal do Espírito Santo.

Mathias Viana Vicari

Professor de Física, Xadrez e Robótica, Colégio Nossa Senhora do Carmo (Viçosa-MG). Doutor em Meteorologia Aplicada, Universidade Federal de Viçosa. Mestre em Física com ênfase no Ensino de Física, Universidade Federal de Viçosa. Diplomado em Física, Universidade Federal de Viçosa. Diplomado em Matemática, Universidade Estácio de Sá.

Mirna Peçanha Brito

Professora Adjunta, Universidade Federal de Viçosa. Doutora em Bioética, Ética aplicada e Saúde Coletiva, Universidade Federal do Rio de Janeiro. Mestre em Ciências da Saúde Universidade Federal de Viçosa. Especialista em Pneumologia Infantil, Sociedade Brasileira de Pediatria. Especialista em Alergia e Imunopatologia Faculdade de Ciências Médicas da Santa Casa de São Paulo.

Omar Paranaíba Vilela Neto

Professor Associado do Departamento de Ciência da Computação, Universidade Federal de Minas Gerais. Doutor em Engenharia Elétrica, com ênfase em Nanotecnologia, Pontifícia Universidade Católica do Rio de Janeiro. Mestre em Engenharia Elétrica, com ênfase em Métodos de Apoio à Decisão, Pontifícia Universidade Católica do Rio de Janeiro. Diplomado em Engenharia de Computação, Pontifícia Universidade Católica do Rio de Janeiro. Bolsista de Produtividade do CNPq.

Poliana Aparecida Corrêa de Oliveira

Professora do Departamento de Computação, Centro Federal de Educação Tecnológica de Minas Gerais. Doutoranda em Ciência da Computação da Universidade Federal de Minas Gerais. Mestre em Informática, Pontifícia Universidade Católica de Minas Gerais. Diplomada em Sistemas de Informação, Pontifícia Universidade Católica de Minas Gerais.

Renan Albuquerque Marks

Professor da Faculdade de Computação, Universidade Federal de Mato Grosso do Sul. Doutor em Ciência da Computação, Universidade Federal de Minas Gerais. Mestre em Ciência da Computação, Universidade Federal de Mato Grosso do Sul. Diplomado em Ciência da Computação, Universidade Federal de Mato Grosso do Sul.

Renzo Comolatti

Pesquisador no Center for Sleep and Consciousness, Universidade de Wisconsin-Madison (EUA). Doutor em Ciências Cognitivas pelo programa Mind, Brain e Reasoning da Universidade de Milão (Itália). Mestrado em Engenharia Biomédica na Universidade Federal de São Paulo, campus São José dos Campos. Bacharel em Ciências Moleculares, Universidade de São Paulo.

Robert Sparrow

Professor do Departamento de Filosofia, Monash University. PhD em Filosofia na Australian National University, Bacharel em Filosofia, University of Melbourne.

Rômulo Cristóvão de Souza

Professor adjunto no Departamento de Tecnologias da Informação e Educação em Saúde (DTIES), Faculdade de Ciências Médicas da Universidade do Estado do Rio de Janeiro. Doutor em Ciências Médicas, Universidade do Estado do Rio de Janeiro. Mestre em Saúde Coletiva, Universidade Federal do Rio de Janeiro. Graduação em Enfermagem, Faculdade de Enfermagem de Teresópolis.

Ronaldo Ribeiro Goldschmidt

Professor Associado, Instituto Militar de Engenharia. Doutor em Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro. Mestre em Sistemas e Computação, Instituto Militar de Engenharia. Bacharel em Matemática, Universidade Federal Fluminense.

Stefânia Salvador Pereira Montenegro

Diplomada em Medicina na Faculdade Dinâmica do Vale do Piranga (FADIP).

Suzete Venturelli

Professora Titular, Universidade Anhembi Morumbi/Programa de pós-graduação em Design e Universidade de Brasília/Programa de Pós-Graduação em Artes Visuais. Doutora em Artes e Ciências da Arte, Université Paris 1 Panthéon Sorbonne, França. Mestrado em Esthétique et Science de l'Art, Université Paris 1 Pantheon Sorbonne, França. Graduada em Licenciatura em Desenho e Plástica, Universidade Mackenzie em São Paulo. Pesquisadora do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Vanderson Esperidião Antonio

Professor Associado do Departamento de Medicina e Enfermagem, Universidade Federal de Viçosa. Doutor em Bioética, Ética Aplicada e Saúde Coletiva, Universidade Federal Fluminense. Mestre em Ciências Morfológicas, Universidade Federal do Rio de Janeiro. Residência Médica em Cirurgia Geral, Hospital Central da Polícia Militar do Estado do Rio de Janeiro. Diplomado em Medicina, Centro Universitário Serra dos Órgãos.

