

UNIVERSIDADE FEDERAL DE JUIZ DE FORA/MG
FACULDADE DE DIREITO
CHRISTIANO LEONEL QUEIROZ

**INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS EM DIAGNÓSTICOS E
TRATAMENTOS MÉDICOS: O Colapso do Consentimento na Saúde Algorítmica**

Juiz de Fora
2025

CHRISTIANO LEONEL QUEIROZ

**INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS EM DIAGNÓSTICOS E
TRATAMENTOS MÉDICOS: O Colapso do Consentimento na Saúde Algorítmica**

Dissertação apresentada ao programa de Pós-graduação *Stricto Sensu* em Direito da Faculdade de Direito da Universidade Federal de Juiz de Fora, como requisito parcial para obtenção do grau de Mestre no Mestrado em Direito e Inovação, sob a orientação do Prof. Dr. Sergio Marcos Carvalho de Ávila Negri.

Juiz de Fora

2025

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Queiroz, Christiano Leonel.

INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS EM DIAGNÓSTICOS E TRATAMENTOS MÉDICOS: O Colapso do Consentimento na Saúde Algorítmica / Christiano Leonel Queiroz. -- 2025.

Orientador: Sergio Marcos Carvalho de Ávila Negri
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Faculdade de Direito. Programa de Pós-Graduação em Direito, 2025.

1. Inteligência Artificial. 2. Neurotecnologias. 3. Diagnóstico Médico. 4. Opacidade e Viés Algorítmicos. 5. Consentimento Livre e Esclarecido. 6. Neurodireitos. 7. Direitos Fundamentais. I. Negri, Sergio Marcos Carvalho de Ávila, orient. II. Título.

CHRISTIANO LEONEL QUEIROZ

**INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS EM DIAGNÓSTICOS E
TRATAMENTOS MÉDICOS: O Colapso do Consentimento na Saúde Algorítmica**

Dissertação
apresentada ao
Programa de Mestrado
em Direito da
Universidade Federal
de Juiz de Fora como
requisito parcial à
obtenção do título de
Mestre em Direito.
Área de concentração:
Direito e Inovação

Aprovada em 18 de dezembro de 2025.

BANCA EXAMINADORA

Doutor Sérgio Marcos Carvalho de Ávila Negri - Orientador
Universidade Federal de Juiz de Fora

Doutor Bruno Farage da Costa Felipe
Faculdade Metodista Granbery

Doutor Daniel Bucar Cervasio
Universidade do Estado do Rio de Janeiro

Juiz de Fora, 11/12/2025.



09/01/2026, às 11:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **BRUNO FARAGE DA COSTA FELIPE, Usuário Externo**, em 05/02/2026, às 11:27, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Daniel, Usuário Externo**, em 10/02/2026, às 19:58, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Christiano Leonel Queiroz, Usuário Externo**, em 13/02/2026, às 15:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2789303** e o código CRC **F8C6BC6D**.

DEDICATÓRIA

Dedico este trabalho a quem acredita que a tecnologia só tem sentido quando amplia direitos. Que esta pesquisa some na construção de IA's e Neurotecnologias a serviço da vida, da igualdade e da autonomia dos pacientes.

AGRADECIMENTOS

A Deus, pela graça que me sustentou quando as forças faltaram. Conviver com TDAH severo é, muitas vezes, “escrever contra a maré” — entre pausas forçadas, recomeços e silêncios. Concluir esta dissertação é, para mim, mais do que um título: é a prova de que limites podem virar caminho quando a fé e a persistência caminham juntas.

Sempre sonhei cursar o mestrado na UFJF, mas, por muito tempo, esse sonho pareceu distante. Até que Deus colocou no meu caminho dois anjos, Álvaro e Sandra, que insistiram comigo quando eu hesitava e abriram a porta por onde tive coragem de passar.

Aos meus pais, Mauro e Rosely, meu amor e a minha gratidão mais profunda. Mesmo sem terem tido as oportunidades que eu tive — ambos com o ensino fundamental —, foram vocês que me deram tudo o que realmente importa: trabalho honesto, caráter e a certeza de que estudo transforma destino. Este trabalho também é de vocês. À minha irmã, Patrícia, pelo apoio de sempre. Obrigado por cada palavra de incentivo e por me lembrar, nos dias difíceis, quem eu sou.

À minha esposa, Priscila: mais do que essencial, você foi o chão e o fôlego desta caminhada. Quando o TDAH me dispersava, era você quem me trazia de volta ao texto; quando o desânimo batia, você me lembrava o propósito. Segurou a rotina e as urgências, amparou meus silêncios, aparou ansiedades e celebrou cada página como vitória. Confiou quando eu duvidava, insistiu quando eu quase desisti e me ensinou que cuidado também é método. Este título carrega a sua assinatura invisível. Obrigado por amar com paciência, coragem e presença — sem você, esta dissertação não existiria.

Ao Professor Sergio Negri, pela orientação generosa e pela paciência com meu ritmo, nem sempre linear — obrigado por transformar exigência acadêmica em oportunidade de crescimento humano e intelectual.

Aos amigos que o caminho me deu — Álvaro Xavier, Danilo e Adriana (cunhada) — pelos respiros necessários, pelas conversas que acalmaram a mente e pela presença indireta no processo de pesquisa. À minha sogra, Aparecida, pelo apoio incansável nesses últimos dois anos, principalmente, e pelo cuidado carinhoso com nossos cinco filhos de quatro patas — Marley, Mel, Sol, Simba e Nina — que trouxeram leveza a cada página escrita.

À UFJF — instituição pública, gratuita e de excelência —, à Faculdade de Direito e ao Programa de Pós-Graduação em Direito e Inovação, na pessoa de seus professores e servidores: obrigado por manterem viva a missão de uma Universidade que forma, pesquisa e serve ao país.

Por fim, agradeço a todos aqueles que, à sua maneira, estiveram presentes nesta jornada — com uma mensagem, um café, um abraço ou um silêncio compartilhado. Cada gesto foi tijolo no caminho até aqui. Esta conquista é nossa.

“The ‘mechanical’ medical doctor (MD) is certainly not for tomorrow and certainly not desirable. The doctor of the future should be ‘augmented’, better equipped, and well informed to prevent, analyze, decide, and treat disease with empathy and the human touch. The aim will be to improve diagnoses, observations, therapeutic choices, and outcomes”

RESUMO

Esta dissertação analisa o uso de sistemas de inteligência artificial (IA) e de neurotecnologias em diagnósticos e tratamentos médicos, com foco na forma como suas características técnicas – especialmente a opacidade algorítmica, a “dupla caixa-preta” entre redes neurais e circuitos cerebrais e os vieses estruturais de dados – tensionam a tutela dos direitos fundamentais à saúde, à igualdade, à privacidade de dados e à integridade mental. O objetivo central é demonstrar que a convergência entre opacidade e discriminação algorítmica produz uma erosão progressiva da autonomia do paciente e conduz ao colapso funcional do modelo clássico de Consentimento Livre e Esclarecido, tal como concebido para a medicina analógica.

Adota-se uma abordagem jurídico-dogmática e interdisciplinar, baseada em pesquisa bibliográfica, análise de documentos regulatórios e estudo comparado entre Brasil, União Europeia e Estados Unidos. A partir de casos paradigmáticos de IA diagnóstica e de interfaces cérebro-computador, mapeiam-se, em primeiro lugar, as potencialidades clínicas e operacionais das tecnologias (aumento de acurácia, personalização terapêutica e eficiência dos serviços) e, em segundo lugar, os riscos decorrentes da opacidade, dos vieses e da possibilidade de manipulação da decisão clínica e do comportamento do paciente por meio de *nudges* e *dark patterns* digitais.

A proposição defendida é que, em ambiente de saúde inteligente e neurotecnológica, o consentimento do paciente, isoladamente, deixa de ser instrumento suficiente de proteção, exigindo a transição para um paradigma de autodeterminação efetiva ancorado em uma “tutela estrutural”. Essa tutela envolve deveres positivos de transparência, explicabilidade, auditoria de vieses, governança de neurodados e reconhecimento de salvaguardas específicas para a esfera mental (neurodireitos), bem como um regime de responsabilidade compartilhada entre desenvolvedores, profissionais de saúde, instituições e reguladores. Conclui-se que apenas uma arquitetura normativa e institucional robusta, que combine consentimento, direitos fundamentais e mecanismos preventivos de governança tecnológica, é capaz de compatibilizar inovação em IA e neurotecnologias com a dignidade e a igualdade dos pacientes na era da saúde algorítmica.

Palavras-chave: Inteligência Artificial. Neurotecnologias. Diagnóstico Médico. Opacidade e Vieses Algorítmicos. Consentimento Livre e Esclarecido. Neurodireitos. Direitos Fundamentais.

ABSTRACT

This dissertation examines the use of artificial intelligence (AI) systems and neurotechnologies in medical diagnosis and treatment, focusing on how their technical characteristics – especially algorithmic opacity, the “double black box” between neural networks and brain circuits, and structural data biases – strain the protection of the fundamental rights to health, equality, data privacy, and mental integrity. The central aim is to show that the convergence of opacity and algorithmic discrimination progressively erodes patient autonomy and leads to the functional collapse of the traditional informed consent model as designed for analog medicine.

The dissertation adopts a legal-dogmatic and interdisciplinary approach, grounded in bibliographic research, analysis of regulatory documents, and comparative study across Brazil, the European Union, and the United States. Drawing on paradigmatic cases of diagnostic AI and brain-computer interfaces, it first maps the clinical and operational potential of these technologies (increased accuracy, therapeutic personalization, and service efficiency) and, second, the risks arising from opacity, bias, and the possibility of manipulating clinical decision-making and patient behavior through digital nudges and dark patterns.

The proposition defended is that, in an intelligent and neurotechnological healthcare environment, patient consent alone is no longer a sufficient instrument of protection, requiring a transition to a paradigm of effective self-determination anchored in “structural guardianship”. This protection entails positive duties of transparency, explainability, bias auditing, neurodata governance, and the recognition of specific safeguards for the mental sphere (neuro-rights), as well as a regime of shared responsibility among developers, health professionals, institutions, and regulators. It concludes that only a robust normative and institutional architecture – one that combines consent, fundamental rights, and preventive mechanisms of technological governance – can reconcile innovation in AI and neurotechnologies with patient dignity and equality in the age of algorithmic health care.

Keywords: *Artificial Intelligence. Neurotechnologies. Medical Diagnosis. Algorithmic Opacity and Bias. Free and Informed Consent. Neuro-rights. Fundamental Rights.*

LISTA DE ILUSTRAÇÕES

| | | |
|-----------|---|-----|
| Figura 1 | – Capturas de tela do visualizador com suporte de IA | 60 |
| Figura 2 | – Comparação do alinhamento entre imagens e relatórios | 63 |
| Figura 3 | – Classificação AI-Risk e previsão do risco cardiovascular (parte 1) | 67 |
| Figura 4 | – Classificação AI-Risk e previsão do risco cardiovascular (parte 2) | 68 |
| Figura 5 | – Visão geral do <i>AlphaMissense</i> | 71 |
| Figura 6 | – Fluxograma da Metodologia Adotada | 76 |
| Figura 7 | – IA em TBP (<i>Total Body Photography</i>) – “Fotografia de Corpo Inteiro”.. | 79 |
| Figura 8 | – <i>Gráficos de dispersão dos valores SHapley Additive exPlanations (SHAP)</i> | 82 |
| Figura 9 | – Estratificação do <i>Sepsis ImmunoScore</i> para Sepsis-3 em todas as coortes | 83 |
| Figura 10 | – Localização dos eletrodos e configuração da decodificação da fala | 89 |
| Figura 11 | – Interface do usuário no modo de conversação | 90 |
| Figura 12 | – Uso da neuroprótese para fala iniciada pelo participante | 91 |
| Figura 13 | – Mecanismos de estimulação cerebral profunda | 94 |
| Figura 14 | – <i>Diagrama esquemático da ECP (DBS) para o tratamento da DA grave</i> . | 95 |
| Figura 15 | – Esquema de funcionamento da Neuroprótese | 100 |
| Figura 16 | – Esquema da estrutura do modelo <i>DANMLP</i> | 106 |
| Figura 17 | – Resultados da visualização das regiões cerebrais discriminatórias identificadas pelo <i>DANMLP</i> para alguns indivíduos com DA e DCL | 108 |
| Figura 18 | – <i>Design, tecnologia e implantação do BSI</i> | 111 |
| Figura 19 | – <i>A BSI que restaura o controle natural da marcha</i> | 111 |
| Figura 20 | – Desenho do estudo, eletrodo do dispositivo de neuroestimulação e modos de estimulação, e localização dos eletrodos | 120 |
| Figura 21 | – Resultados do estudo e desfechos de eficácia da análise combinada | 121 |
| Figura 22 | – Implante Neuralink N1 | 126 |
| Figura 23 | – Macaco Pager jogando <i>pong</i> com sua mente | 126 |
| Figura 24 | – <i>Design of visual prostheses</i> | 129 |

LISTA DE TABELAS

| | | |
|----------|--|-----|
| Tabela 1 | – Panorama das Aplicações de IA na Medicina (2023-2025) | 57 |
| Tabela 2 | – Estratificação de risco para sepse segundo o <i>Sepsis ImmunoScore</i> | 84 |
| Tabela 3 | – Panorama das Aplicações de Neurotecnologias na Medicina (2023-2025) | 87 |
| Tabela 4 | – Matriz de Impacto da Opacidade e do Viés sobre a Autonomia e o Consentimento Livre e Esclarecido | 191 |
| Tabela 5 | – Matriz de Requisitos Complementares: <i>MDR/IVDR vs. AI Act</i> para <i>MDAI</i> de Alto Risco | 224 |
| Tabela 6 | – Principais Categorias e Princípios dos Neurodireitos | 237 |
| Tabela 7 | – Quadro Comparativo das Abordagens Regulatórias para IA e Neurotecnologias | 261 |

LISTA DE ABREVIATURAS E SIGLAS

| | |
|-----------|--|
| ABNT | Associação Brasileira de Normas Técnicas. |
| AI | <i>Artificial Intelligence</i> (ver IA). |
| AI Act | Regulamento Europeu de Inteligência Artificial (UE). |
| AI/ML | <i>Artificial Intelligence / Machine Learning</i> (Inteligência Artificial / Aprendizado de Máquina). |
| AI-DSF | <i>Artificial Intelligence-Enabled Device Software Function</i> (função de software de dispositivo habilitada por IA). |
| ANPD | Agência Nacional de Proteção de Dados (Brasil). |
| ANVISA | Agência Nacional de Vigilância Sanitária (Brasil). |
| BCI | <i>Brain-Computer Interface</i> (Interface Cérebro-Computador). |
| BDP | <i>Breakthrough Devices Program</i> (Programa de Dispositivos Inovadores da FDA). |
| CBER | <i>Center for Biologics Evaluation and Research</i> (FDA). |
| CDC (BR) | Código de Defesa do Consumidor (Lei nº 8.078/1990). |
| CDER | <i>Center for Drug Evaluation and Research</i> (FDA). |
| CDRH | <i>Center for Devices and Radiological Health</i> (FDA). |
| CREMERS | Conselho Regional de Medicina do Estado do Rio Grande do Sul |
| CFM | Conselho Federal de Medicina |
| CFR | <i>Code of Federal Regulations</i> (EUA). |
| DBS | <i>Deep Brain Stimulation</i> (Estimulação Cerebral Profunda). |
| De Novo | Via de classificação para dispositivos inéditos (FDA). |
| ECoG | Eletrocorticografia. |
| EEG | Eletroencefalografia. |
| EHR | <i>Electronic Health Record</i> (ver PEP). |
| EUA | Estados Unidos da América. |
| EU/UE | União Europeia. |
| FDA | <i>Food and Drug Administration</i> (EUA). |
| FD&C Act | <i>Federal Food, Drug, and Cosmetic Act</i> (EUA). |
| FDORA | <i>Food and Drug Omnibus Reform Act</i> (2022, EUA). |
| GMLP | <i>Good Machine Learning Practice(s)</i> (boas práticas de ML). |
| GDPR/RGPD | <i>General Data Protection Regulation</i> / Regulamento Geral sobre a Proteção de Dados (UE). |
| HIPAA | <i>Health Insurance Portability and Accountability Act</i> (EUA). |

| | |
|---------------|--|
| IA | Inteligência Artificial. |
| IDE | <i>Investigational Device Exemption</i> (isenção para investigação clínica de dispositivo, FDA). |
| IEC | <i>International Electrotechnical Commission</i> . |
| IMDRF | <i>International Medical Device Regulators Forum</i> . |
| ISO | <i>International Organization for Standardization</i> . |
| IVD | <i>In Vitro Diagnostic</i> (dispositivo para diagnóstico in vitro). |
| IVDR | <i>In Vitro Diagnostic Regulation</i> (Regulamento (UE) 2017/746). |
| LGPD | Lei Geral de Proteção de Dados (Lei nº 13.709/2018, Brasil). |
| MDR | <i>Medical Device Regulation</i> (Regulamento (UE) 2017/745). |
| ML | <i>Machine Learning</i> (Aprendizado de Máquina). |
| NIST | <i>National Institute of Standards and Technology</i> (EUA). |
| OCDE | Organização para a Cooperação e Desenvolvimento Econômico. |
| OCP | <i>Office of Combination Products</i> (FDA). |
| OMS/WHO | Organização Mundial da Saúde / <i>World Health Organization</i> . |
| OS (ANVISA) | Orientação de Serviço. |
| PCCP | <i>Predetermined Change Control Plan</i> (plano pré-determinado de controle de mudanças). |
| PEP | Prontuário Eletrônico do Paciente (equivalente a EHR). |
| PMA | <i>Premarket Approval</i> (aprovação pré-comercialização, FDA). |
| PL 2.338/2023 | Projeto de Lei nº 2.338/2023 (Marco Legal da IA no Brasil). |
| Q-Sub | <i>Q-Submission</i> (programa de pré-submissão da FDA). |
| QMS | <i>Quality Management System</i> (Sistema de Gestão da Qualidade). |
| RDC | Resolução da Diretoria Colegiada (ANVISA). |
| RE (ANVISA) | Resolução-RE (ANVISA). |
| SaMD | <i>Software as a Medical Device</i> (Software como Dispositivo Médico). |
| SUS | Sistema Único de Saúde (Brasil). |
| TPLC | <i>Total Product Life Cycle</i> (ciclo de vida total do produto). |
| UDI | <i>Unique Device Identification</i> (Identificação Única de Dispositivos). |
| UNESCO | Organização das Nações Unidas para a Educação, a Ciência e a Cultura. |
| XAI | <i>Explainable Artificial Intelligence</i> (IA Explicável). |
| XAI4BCI | XAI para BCI. |
| 510(k) | <i>Premarket Notification</i> (notificação prévia por equivalência substancial, FDA). |

SUMÁRIO

| | | |
|--------------|---|-----------|
| 1 | INTRODUÇÃO | 19 |
| 2 | INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS EM DIAGNÓSTICOS MÉDICOS: FUNDAMENTOS E CONTEXTO | 32 |
| 2.1 | EVOLUÇÃO HISTÓRICA DA IA E DAS NEUROTECNOLOGIAS | 33 |
| 2.1.1 | Evolução da IA em Aplicações Médicas | 34 |
| 2.1.1.1 | <i>Dos Primórdios Conceituais à Primeira Onda de IA na Saúde: A Era dos Sistemas Especialistas</i> | 34 |
| 2.1.1.2 | <i>Os "Invernos da IA" e a Transição para Novas Abordagens</i> | 38 |
| 2.1.1.3 | <i>A Revolução do Aprendizado de Máquina (ML) e a Ascensão do Aprendizado Profundo (DL)</i> | 40 |
| 2.1.1.4 | <i>IA Generativa, Grandes Modelos de Linguagem (LLMs) e as Novas Fronteiras</i> | 45 |
| 2.1.2 | Evolução das Neurotecnologias em Aplicações Médicas | 47 |
| 2.1.2.1 | <i>Primeiros Registros Elétricos do Cérebro Humano (EEG)</i> | 48 |
| 2.1.2.2 | <i>Surgimento da Neuroimagem Clínica (Tomografia e Ressonância Magnética)</i> .. | 48 |
| 2.1.2.3 | <i>Neuroestimulação e Neuropróteses: Restaurando Funções Perdidas</i> | 51 |
| 2.1.2.4 | <i>Interfaces Cérebro-Computador Modernas e IA na Neurotecnologia</i> | 54 |
| 2.2 | APLICAÇÕES TECNOLÓGICAS EM DIAGNÓSTICOS MÉDICOS: POTENCIALIDADES E RISCOS | 56 |
| 2.2.1 | A Inteligência Artificial como Ferramenta de Aumento da Capacidade Diagnóstica e Terapêutica | 57 |
| 2.2.1.1 | <i>Detecção de Câncer de Mama em Mamografias: O Caso do Estudo PRAIM</i> | 59 |
| 2.2.1.2 | <i>Patologia Digital e os Foundation Models: A Revolução do Prov-GigaPath</i> | 62 |
| 2.2.1.3 | <i>Medicina Preditiva na Cardiologia: A Ferramenta de IA de Oxford (ORFAN)</i> .. | 66 |
| 2.2.1.4 | <i>Decifrando o Genoma: O AlphaMissense e a Classificação de Variantes Genéticas</i> | 70 |
| 2.2.1.5 | <i>Democratizando o Diagnóstico: IA na Retinopatia Diabética</i> | 75 |
| 2.2.1.6 | <i>A Lupa Digital: IA na Triagem de Câncer de Pele</i> | 78 |
| 2.2.1.7 | <i>Correndo Contra o Tempo: IA na Detecção e Manejo da Sepsis</i> | 81 |
| 2.2.2 | As Neurotecnologias e a Fronteira da Mente Humana | 86 |
| 2.2.2.1 | <i>Interfaces Cérebro-Computador (BCI): Restaurando a Fala em Pacientes com ELA</i> | 89 |
| 2.2.2.2 | <i>Estimulação Cerebral Profunda (DBS) na Doença de Alzheimer</i> | 94 |

| | | |
|--------------|---|------------|
| 2.2.2.3 | <i>Neuroprótese para a Doença de Parkinson: Recuperando a Marcha</i> | 99 |
| 2.2.2.4 | <i>Neuroimagem e IA no Diagnóstico Precoce de Alzheimer: O Método DANMLP</i> | 105 |
| 2.2.2.5 | <i>A “Ponte Digital” cérebro-coluna (BSI) para Lesão Medular</i> | 110 |
| 2.2.2.6 | <i>Neuroestimulação em Malha Fechada para Epilepsia Refratária</i> | 119 |
| 2.2.2.7 | <i>BCIs Invasivas Neuralink – (Telepathy NI e Blindsight)</i> | 125 |
| 2.2.3 | Conclusão | 132 |
| 3 | OPACIDADE, INTEGRIDADE E O DIREITO À EXPLICAÇÃO EM SISTEMAS DE SAÚDE INTELIGENTES | 133 |
| 3.1 | O PROBLEMA DA “DUPLA CAIXA-PRETA”: OPACIDADE ALGORÍTMICA E A INESCRUTABILIDADE NEURAL | 134 |
| 3.1.1 | Análise da “Caixa-Preta” Algorítmica em IA Diagnóstica | 134 |
| 3.1.2 | Introdução e Aprofundamento do Conceito de “Dupla Caixa-Preta” em Neurotecnologias | 135 |
| 3.1.3 | Implicações da Dupla Opacidade para a Confiança, Validação Clínica e Atribuição de Nexo Causal | 137 |
| 3.2 | O DESAFIO DA EXPLICABILIDADE (XAI): DA JUSTIFICAÇÃO DE PREDIÇÕES À COMPREENSÃO DE AÇÕES NEUROTECNOLÓGICAS | 138 |
| 3.2.1 | Apresentação do Campo da IA Explicável (XAI) e suas Técnicas | 138 |
| 3.2.1.1 | <i>Métodos Orientados a Características (Feature-oriented methods)</i> | 139 |
| 3.2.1.2 | <i>Métodos Globais (Global methods)</i> | 140 |
| 3.2.1.3 | <i>Modelos de Conceito (Concept Models)</i> | 140 |
| 3.2.1.4 | <i>Modelos Substitutos (Surrogate Models)</i> | 140 |
| 3.2.1.5 | <i>Métodos Locais Baseados em Pixels (Local, pixel-based methods)</i> | 141 |
| 3.2.1.6 | <i>Métodos Centrados no Humano (Human-centric Methods)</i> | 141 |
| 3.2.2 | A Persistência da Opacidade e os Limites da Explicabilidade | 142 |
| 3.2.2.1 | <i>Complexidade Inerente dos Modelos de ML</i> | 143 |
| 3.2.2.2 | <i>Limitações das Metodologias XAI</i> | 143 |
| 3.2.2.3 | <i>Compromisso entre Desempenho e Transparência (Performance vs. Transparency Trade-off)</i> | 143 |
| 3.2.2.4 | <i>Lacuna entre Especialistas e Usuários Finais</i> | 144 |
| 3.2.2.5 | <i>Necessidade de Supervisão Humana e Confiança Limitada</i> | 144 |
| 3.2.2.6 | <i>Ambiguidade na Definição de Interpretabilidade</i> | 144 |
| 3.2.3 | Explicação e Lacunas de Explicabilidade em Neurotecnologias: da Justificação de Predições à Compreensão de Ações (XAI4BCI) | 145 |

| | | |
|---------|--|-----|
| 3.2.3.1 | <i>Justificação de predições (IA diagnóstica “estática”)</i> | 145 |
| 3.2.3.2 | <i>Compreensão de ações (neurotecnologias interativas/fechadas)</i> | 145 |
| 3.3 | O DIREITO À EXPLICAÇÃO EM SAÚDE E SUAS NOVAS DIMENSÕES JURÍDICAS | 147 |
| 3.3.1 | Análise do Direito à Explicação à Luz da LGPD (Art. 20), do GDPR (Art. 22) e do AI Act Europeu (Arts. 13 e 14) | 147 |
| 3.3.2 | A Fusão do Direito à Explicação com o Consentimento Livre e Esclarecido no Contexto da Saúde | 151 |
| 3.3.3 | O Desafio do Consentimento Livre e Esclarecido em Populações Vulneráveis Diante da “Dupla Caixa-Preta” | 153 |
| 4 | VIESES, DISCRIMINAÇÃO E MANIPULAÇÃO NA SAÚDE DIGITAL E NEUROTECNOLÓGICA | 156 |
| 4.1 | FONTES MÚLTIPLAS DE VIÉS NA FRONTEIRA TECNOLÓGICA DA SAÚDE | 156 |
| 4.1.1 | Viés de Representação e Amostragem de Dados | 157 |
| 4.1.2 | Viés Histórico e de Medição (Proxy) | 158 |
| 4.1.3 | Viés de Anotação e Rotulagem | 159 |
| 4.1.4 | Viés Cognitivo e de Automação na Prática Clínica | 160 |
| 4.1.5 | Viés em Neurotecnologias e Interfaces Cérebro-Computador | 161 |
| 4.2 | ARQUITETURA DA MANIPULAÇÃO: NUDGES E DARK PATTERNS | 164 |
| 4.3 | RISCOS DE DISCRIMINAÇÃO: DAS INIQUIDADES DIAGNÓSTICAS À “NEURODISCRIMINAÇÃO” | 166 |
| 4.3.1 | A Discriminação Algorítmica como Violação do Princípio Constitucional da Isonomia | 166 |
| 4.3.2 | A Materialização do Dano: Iniquidades em Saúde e Diagnósticos Falhos ... | 168 |
| 5 | A EROSÃO DA AUTONOMIA E O COLAPSO DO CONSENTIMENTO LIVRE E ESCLARECIDO: REPERCUSSÕES DA OPACIDADE E DO VIÉS NA SAÚDE INTELIGENTE E NEUROTECNOLÓGICA | 170 |
| 5.1 | AUTONOMIA E CONSENTIMENTO: PERSPECTIVAS BIOÉTICAS E JURÍDICAS | 171 |
| 5.2 | A OPACIDADE COMO HIPÓTESE DE AUSÊNCIA DE CONSENTIMENTO: O RISCO EPISTÊMICO E O DIREITO À COMPREENSÃO | 175 |
| 5.2.1 | O Consentimento para o Inexplicável: A “Dupla Caixa-Preta” e a Inexistência da Informação como Hipótese de Nulidade | 176 |

| | | |
|--------------|---|------------|
| 5.2.2 | A Integridade Psicológica e a Limitação da Autonomia em Terapias Moduladoras | 178 |
| 5.2.2.1 | <i>Análise dos Riscos à Integridade Física</i> | 180 |
| 5.2.2.2 | <i>Análise dos Riscos à Integridade Psicológica: a Modulação da Personalidade e o Debate sobre o “Direito à Continuidade Psicológica”</i> | 181 |
| 5.2.2.3 | <i>Conexão com o Debate Constitucional Brasileiro: a PEC 29/2023 e a Positivção da “Proteção à Integridade Mental”</i> | 183 |
| 5.2.3 | O Risco Epistêmico como Novo Elemento do Dever de Informar | 183 |
| 5.3 | VIÉS ALGORÍTMICO E MANIPULAÇÃO: A CORRUPÇÃO DA LIBERDADE DE ESCOLHA E OS LIMITES DA AUTONOMIA | 185 |
| 5.3.1 | Escolha Desigual: Como o Viés Algorítmico Prejudica a Autonomia de Grupos Vulneráveis | 185 |
| 5.3.2 | <i>Nudges e Dark Patterns</i> na Relação Paciente-Médico | 187 |
| 5.3.3 | O Consentimento sob Viés Sistêmico: Informação Materialmente Defeituosa e Invalidez | 188 |
| 5.4 | NEURODISCRIMINAÇÃO E O CONSENTIMENTO COAGIDO | 190 |
| 6 | SALVAGUARDAS JURÍDICAS AO CONSENTIMENTO: TRANSPARÊNCIA, EQUIDADE E <i>ACCOUNTABILITY</i> | 193 |
| 6.1 | TRANSPARÊNCIA E EXPLICABILIDADE COMO PRINCÍPIOS FUNDAMENTAIS PARA APOIAR A TOMADA DE DECISÃO INFORMADA | 194 |
| 6.2 | FORTALECENDO O PROCESSO DE CONSENTIMENTO: DO FORMULÁRIO AO DIÁLOGO HUMANIZADO | 197 |
| 6.3 | DIREITO AO CONSENTIMENTO LIVRE E ESCLARECIDO E O DIREITO DE RECUSA | 201 |
| 6.4 | SALVAGUARDAS REGULATÓRIAS E ÉTICAS ADICIONAIS | 205 |
| 6.4.1 | Avaliação rigorosa e certificação de algoritmos | 206 |
| 6.4.2 | Monitoramento pós-mercado e auditorias contínuas | 209 |
| 6.4.3 | Guias éticos e comitês de supervisão institucionais | 212 |
| 6.4.4 | Cultura de responsabilização compartilhada | 213 |
| 6.4.5 | Novos direitos para salvaguardar a integridade mental e a identidade pessoal frente a tecnologias neuroinvasivas | 215 |

| | | |
|---------|--|------------|
| 6.5 | GOVERNANÇA DUPLA DA IA EM DISPOSITIVOS MÉDICOS: A COMPLEMENTARIDADE ENTRE O AI ACT E AS REGULAMENTAÇÕES MDR/IVDR NA UNIÃO EUROPEIA | 217 |
| 6.5.1 | A Contribuição Complementar do AI Act: Transparência e Supervisão Humana | 221 |
| 6.5.1.1 | <i>Transparência Ampliada (Questões nº 11, 14 e 15 do AIB/MDCG)</i> | <i>221</i> |
| 6.5.1.2 | <i>Supervisão Humana (Questões nº 18, 19 e 36 do AIB/MDCG)</i> | <i>222</i> |
| 6.5.1.3 | <i>Responsabilidade (Accountability) - (Questões nº 16 e 17 do AIB/MDCG)</i> | <i>223</i> |
| 6.5.2 | A Confluência para o Consentimento Livre e Esclarecido do Paciente | 224 |
| 6.5.2.1 | <i>Pré-condições Técnicas (Questões nº 22 e 23 do AIB/MDCG)</i> | <i>224</i> |
| 6.5.2.2 | <i>Informação e Compreensão (Questões nº 14 e 15 do AIB/MDCG)</i> | <i>225</i> |
| 6.5.2.3 | <i>Controle e Autonomia (Questões nº 18, 19 e 20 do AIB/MDCG)</i> | <i>225</i> |
| 6.6 | SÍNTESE PARA UMA GOVERNANÇA PROATIVA E MULTISSETORIAL | 226 |
| 7 | REGULAÇÃO DE INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS NA SAÚDE DIGITAL: UM PANORAMA COMPARATIVO | 230 |
| 7.1 | A GÊNESE DOS NEURODIREITOS: DO DEBATE ACADÊMICO À POSITIVAÇÃO JURÍDICA | 230 |
| 7.1.1 | O Chamado Acadêmico por Novas Garantias | 231 |
| 7.1.2 | O Pioneirismo do Chile: A Constitucionalização da Proteção Cerebral | 232 |
| 7.1.3 | A Abordagem Principiológica da Espanha | 233 |
| 7.1.4 | Principais Categorias e Princípios dos Neurodireitos | 237 |
| 7.1.4.1 | <i>Justificativa e Desafios para a Implementação dos Neurodireitos</i> | <i>240</i> |
| 7.2 | O MOSAICO REGULATÓRIO BRASILEIRO | 242 |
| 7.2.1 | A Proposta de Blindagem Constitucional: PEC 29/2023 | 242 |
| 7.2.2 | A Construção de um Microsistema Legal para Neurodireitos: PL 2.174/2023 | 243 |
| 7.2.3 | A Tentativa de Adaptação da LGPD: PL nº 1.229/2021 e PL nº 522/2022 ... | 243 |
| 7.2.4 | Anteprojeto de Reforma do Código Civil | 244 |
| 7.2.5 | A Regulação da IA: PL 2.338/2023 (Marco Legal da IA) | 245 |
| 7.3 | O CONSENTIMENTO LIVRE E ESCLARECIDO NA FRONTEIRA DA IA E DA NEUROTECNOLOGIA | 247 |
| 7.3.1 | A Distinção Fundamental: Consentimento para Dados vs. Consentimento para o Ato Médico | 247 |
| 7.3.2 | A Governança Ética na Prática Clínica e na Pesquisa no Brasil | 248 |

| | | |
|-------|---|-----|
| 7.4 | A GOVERNANÇA TÉCNICA EM SAÚDE: A ATUAÇÃO DAS AGÊNCIAS REGULADORAS | 251 |
| 7.4.1 | A ANVISA e a Regulação de <i>Software as a Medical Device</i> (SaMD) | 251 |
| 7.4.2 | A FDA e a Regulação de Dispositivos com IA/ML e Interfaces Cérebro-Computador (BCI) | 255 |
| 7.5 | ANÁLISE COMPARATIVA E CONSIDERAÇÕES FINAIS | 261 |
| 8 | CONCLUSÃO | 265 |
| | REFERÊNCIAS | 269 |

1 INTRODUÇÃO

A expressão “inteligência artificial” tem sido alvo de críticas conceituais por parte de diversos estudiosos, que a consideram um rótulo inadequado ou exagerado. Autores como Kate Crawford, Melanie Mitchell, Emily Bender e Shannon Vallor argumentam que os sistemas de IA atuais não possuem consciência, cognição ou intencionalidade nos termos em que essas faculdades se manifestam em humanos. Em vez disso, tais sistemas operam identificando correlações estatísticas e padrões em grandes volumes de dados, sem entendimento ou significado inerente. Crawford (2021), por exemplo, enfatiza que a chamada IA “não é nem artificial nem inteligente”, mas sim um conjunto de tecnologias fundamentadas em recursos materiais e trabalho humano, desprovido de autonomia cognitiva genuína. De modo semelhante, Mitchell (2019) salienta que, por trás da aparente sofisticação de algoritmos avançados, esses programas não compreendem de forma humana os dados que processam nem os resultados que produzem (ou seja, carecem de *common sense* e contexto). Autores como Bender et al. (2021) e Vallor (*apud* PASTERNAK, 2024) sustentam que esses sistemas apenas espelham padrões humanos, sem qualquer compreensão genuína. Essas perspectivas convergem na ideia de que a “inteligência” da IA é fundamentalmente diferente da inteligência humana, sendo uma habilidade automatizada de manipular símbolos e informações sem vivência, consciência ou intencionalidade equivalente à nossa.

Importa frisar que essa crítica ao termo *IA* não é uma questão meramente semântica, mas traz implicações práticas e políticas profundas. A forma como definimos e entendemos a “inteligência” nas máquinas molda nossas expectativas e decisões sobre seu desenvolvimento, uso e governança (EBSWORTH *et al.*, 2023). Por exemplo, atribuir a esses sistemas o *status* de inteligências autônomas tende a gerar confiança exagerada em sua objetividade e capacidade supostamente neutra, levando instituições a delegar decisões sensíveis a algoritmos opacos (EBSWORTH *et al.*, 2023). MITCHELL (2019) adverte que o aspecto mais perigoso dos sistemas de IA é justamente a propensão de confiarmos demais neles e lhes conferirmos autonomia excessiva sem plena consciência de suas limitações. Esse deslumbramento pode mascarar os vieses e erros embutidos nos modelos – falhas muitas vezes derivadas dos dados de treinamento e das lógicas de classificação simplistas – sob a aparência de infalibilidade técnica, dificultando a responsabilização por decisões automatizadas injustas (BENDER *et al.* 2021). Ademais, o mito de uma IA plenamente inteligente e independente serve para ocultar os fatores humanos e materiais envolvidos. CRAWFORD (2021) observa que o uso acrítico do termo “inteligência” para descrever esses sistemas obscurece tudo o que está por trás da

tecnologia, incluindo a extração massiva de dados, o labor humano precário e os altos custos energéticos necessários para fazê-la funcionar. Longe de ser neutra, a IA atual tende a reforçar estruturas de poder existentes, uma vez que exige investimentos e infraestruturas acessíveis apenas a grandes corporações e Estados, alinhando-se a interesses dominantes (a IA funciona, nas palavras de Crawford, como um “registro de poder”).

Em suma, repensar criticamente a noção de “inteligência artificial” é fundamental não apenas por rigor conceitual, mas para informar políticas e práticas mais responsáveis. Compreender que esses sistemas são ferramentas estatísticas avançadas – e não mentes conscientes – permite desenvolver abordagens de regulação e ética tecnocientífica que levem em conta suas verdadeiras capacidades e limitações, prevenindo tanto a sobrevalorização acrítica quanto a delegação imprudente de decisões de alto impacto a máquinas que, afinal, não pensam nem entendem como nós (MITCHELL 2019).

Cientes dessas reservas terminológicas – segundo as quais “inteligência artificial” carece de precisão técnica e pode induzir à superestimação de capacidades algorítmicas (CRAWFORD, 2021; MITCHELL, 2019) – optamos, contudo, por manter a expressão Inteligência Artificial (IA) ao longo desta dissertação. A decisão é metodologicamente pragmática: o termo está consagrado no vocabulário legislativo, regulatório e midiático, além de figurar nos principais foros internacionais de saúde digital. Ao adotar a denominação corrente, preservamos a inteligibilidade do debate para profissionais do direito, da medicina, formuladores de políticas públicas e demais leitores não especialistas, sem perder de vista as limitações conceituais apontadas pela literatura crítica. Assim, sempre que mencionarmos “IA”, referir-nos-emos a sistemas computacionais baseados em aprendizado de máquina e algoritmos estatísticos avançados, reconhecendo que tais dispositivos não possuem consciência, mas fornecem, de modo automatizado, análises preditivas que podem apoiar – e não substituir – o julgamento clínico humano.

Partindo dessa delimitação conceitual e adotando a expressão em seu sentido pragmático, observa-se que a adoção de sistemas de Inteligência Artificial (IA) e de neurotecnologias na medicina desponta entre as inovações de maior potencial transformador na saúde contemporânea. Algoritmos de aprendizado profundo, nutridos por extensos repositórios de dados clínicos, prometem revolucionar a prática médica ao proporcionar diagnósticos e tratamentos mais céleres, acurados e personalizados. Eric Topol (2019) salienta que o *deep learning* é capaz de “radicalmente remodelar os sistemas de saúde”, pois detecta padrões intrincados em imagens e dados genômicos que escapariam à observação humana, elevando a precisão em áreas como radiologia, patologia e dermatologia. Paralelamente, o autor adverte

que a IA deve atuar como parceira no esforço de “trazer de volta o cuidado” à medicina – isto é, liberar profissionais de tarefas mecânicas e fortalecer a conexão humana – e não substituir o elemento humano (TOPOL, 2019). Nessa perspectiva, sua concepção de *Deep Medicine* (“Medicina Profunda”) sustenta que a tecnologia, bem aplicada, devolve aos médicos a “dádiva do tempo”, aliviando a carga burocrática e ampliando o espaço para a relação paciente-médico e para decisões clínicas mais humanizadas. Grandes empresas de tecnologia igualmente impulsionam essa transformação: o Google, por exemplo, investe em algoritmos capazes de interpretar exames com desempenho equivalente ao de especialistas e alcança marcos científicos expressivos – seu *AlphaFold2* previu com alta precisão a estrutura de praticamente todas as proteínas humanas conhecidas (feito que rendeu um Prêmio Nobel aos pesquisadores do *DeepMind*), e uma de suas redes neurais médicas atingiu 91% de acurácia em questões do exame de licenciamento médico norte-americano (GOOGLE, 2025). Tais avanços ilustram o imenso potencial inovador da IA na saúde, alimentando expectativas de maior eficiência nos serviços e até de concretização do direito fundamental à saúde, uma vez que ampliam o acesso a diagnósticos de qualidade e colaboram para terapias mais personalizadas.

Paralelamente à evolução da IA, emergem as neurotecnologias como fronteira igualmente promissora e disruptiva no campo médico. Neurotecnologia pode ser entendida como o conjunto de dispositivos e técnicas que permitem a interação direta com o sistema nervoso – a exemplo de interfaces cérebro-computador (*BCIs - Brain-Computer Interfaces*), implantes neurais e métodos avançados de neuroimagem (NICOLÁS-ALONSO; GÓMEZ-GIL, 2012; OECD, 2019). Essas ferramentas já possibilitam, por exemplo, a restauração de funções motoras em pacientes com paralisia por meio de “pontes neurais” que conectam o cérebro à medula espinhal, contornando lesões (METZGER *et al.*, 2022). Em janeiro de 2024, a empresa Neuralink realizou seu primeiro implante cerebral em ser humano, reacendendo debates sobre privacidade mental e autonomia cognitiva (REUTERS, 2024; LAVAZZA *et al.*, 2025). Tecnologias desse tipo podem revolucionar o tratamento e o diagnóstico de condições neurológicas e psiquiátricas: já se testam *BCIs* para permitir a comunicação de pessoas com síndrome do encarceramento ou ELA, e algoritmos de IA capazes de decodificar sinais elétricos cerebrais que detectam precocemente crises epiléticas ou indícios de declínio cognitivo (MOSES *et al.*, 2021; ZHANG *et al.*, 2024). Os benefícios potenciais são evidentes – desde melhorar a qualidade de vida de pacientes com doenças neurodegenerativas até auxiliar no diagnóstico diferencial de transtornos mentais. Não surpreende, portanto, que organismos internacionais vejam essas inovações com ambição: segundo a UNESCO (2023), tais tecnologias têm o potencial de alterar “o próprio núcleo do que significa ser humano”, exigindo

esforços robustos e multidisciplinares de regulação. Em suma, IA e neurotecnologia abrem uma janela de oportunidades inédita na medicina diagnóstica, mas, ao mesmo tempo, projetam desafios éticos e sociais que demandam atenção imediata.

Assim, embora as perspectivas técnico-científicas sejam alvissareiras, elas vêm acompanhadas de desafios éticos, jurídicos e de engenharia até então inéditos. Entre esses desafios, destaca-se a opacidade algorítmica – a célebre “caixa-preta” da IA –, expressão cunhada por Frank Pasquale (2015) para designar modelos cujos processos decisórios internos permanecem em larga medida indecifráveis, mesmo para especialistas. Quando um sistema complexo de *machine learning* (“aprendizado de máquina”) emite uma recomendação diagnóstica ou prognóstica, raramente se compreende de modo transparente como se chegou àquela conclusão. Essa lacuna de explicabilidade dificulta tanto a identificação quanto a correção de falhas, podendo minar a confiança de médicos e pacientes em decisões automatizadas. Ademais, a obscuridade interna dos algoritmos pode ocultar problemas de qualidade: um modelo treinado em bases de dados enviesadas pode apresentar bom desempenho médio e, ao mesmo tempo, cometer erros sistemáticos em subgrupos específicos, sem que isso se torne prontamente perceptível. W. Nicholson Price II (2017), levanta a inquietante pergunta: “como saber se a dose de insulina sugerida por um *software* realmente está correta?” – interrogação que ilustra o risco à segurança do paciente quando se depende de “caixas-pretas” carentes de mecanismos robustos de auditoria. Nesse contexto, ganha relevo a máxima de Lawrence Lessig (2011) de que “o código é lei”: se a arquitetura algorítmica regula comportamentos, a ausência de transparência e de explicações sobre o código pode inviabilizar direitos fundamentais. Acrescente-se, ainda, a preocupação com privacidade e responsabilização na chamada *black-box medicine* (“medicina de caixa-preta”): a coleta maciça de dados sensíveis e o emprego de algoritmos proprietários turvam os limites da proteção de dados e da atribuição de culpa. A literatura jurídica indica que os marcos regulatórios tradicionais da área médica podem revelar-se inadequados frente a tecnologias opacas e em constante evolução, exigindo novas abordagens para garantir *accountability*. Roger Allan Ford e W. Nicholson Price II (2016), por exemplo, sustentam que a ascensão dessa medicina de caixa-preta demanda garantias adicionais de privacidade ao paciente e de responsabilização dos desenvolvedores, pois a incapacidade de explicar um resultado clínico não exclui a responsabilidade por eventuais erros cometidos.

Um desafio igualmente urgente reside nos vieses algorítmicos e na consequente discriminação automatizada. Como os algoritmos de IA aprendem a partir de registros históricos, eles tendem a reproduzir preconceitos e distorções já incrustados nesses dados.

Cathy O’Neil (2020) advertiu que inúmeros modelos de *Big Data* operam como verdadeiros “algoritmos de destruição em massa”, capazes de amplificar desigualdades pré-existentes em distintos domínios. No âmbito da saúde, isso significa que, se grupos populacionais específicos – diferenciados por raça, gênero ou condição socioeconômica – estiverem sub-representados nos conjuntos de treinamento ou se tiverem historicamente menor acesso a cuidados de qualidade, a IA médica poderá reproduzir e até intensificar disparidades diagnósticas e terapêuticas (NORORI *et al.*, 2021). Alguns casos ilustram esse risco: ferramentas de triagem automatizada já foram flagradas subestimando a gravidade de enfermidades em determinados grupos étnicos, por disporem de poucos exemplos relativos a essas populações ou por incorporarem critérios enviesados durante o aprendizado de máquina (OBERMEYER *et al.*, 2019). Tais falhas contrariam o princípio da igualdade no atendimento e revelam que, sem salvaguardas adequadas, a adoção da IA pode violar deveres de não discriminação e de universalidade de acesso à saúde. Além disso, a escala global de difusão dessas tecnologias tende a agravar os efeitos dos vieses.

Kate Crawford (2021) assinala que os sistemas de IA se inserem em vastas estruturas de poder no eixo da materialidade – um autêntico “atlas”, que vai dos minerais extraídos para o *hardware* até as colossais bases de dados que alimentam os algoritmos – de modo que desigualdades estruturais podem permear todo o ciclo tecnológico. Nesse sentido, a IA não constitui um artefato neutro: ela reflete as condições sociais, econômicas e políticas de seu desenvolvimento e uso, podendo solidificar assimetrias de poder já existentes. A literatura crítica, portanto, reforça a necessidade premente de incorporar princípios éticos e de direitos humanos no *design* e na implementação desses sistemas, sob pena de a tecnologia aprofundar injustiças em vez de corrigi-las.

No caso das neurotecnologias, os riscos éticos tornam-se ainda mais sensíveis. Se algoritmos de IA lidam com grandes volumes de dados pessoais, as neurotecnologias lidam potencialmente com neurodados – registros de atividade elétrica, hemodinâmica ou metabólica do cérebro –, isto é, informações que tocam o âmago da vida mental. Um implante cerebral profundo ou um *headset* de eletroencefalografia de alta densidade pode, em tese, inferir estados emocionais, intenções ou mesmo conteúdos rudimentares de pensamento (YUSTE *et al.*, 2021). Por isso, a privacidade mental desponta como preocupação central: Stefano Rodotà (2008) já advertia, de modo visionário, que, na sociedade da vigilância, a coleta ubíqua de dados mina a esfera privada se não houver limites jurídicos claros. Transposto ao domínio neurotecnológico, isso implica reconhecer que dados neurais exigem salvaguardas legais tão ou mais rigorosas que aquelas previstas para outros dados sensíveis, como os genéticos ou biomédicos (IENCA;

ANDORNO, 2017). A União Europeia começa a refletir essa exigência: o Regulamento (UE) 2024/1689, conhecido como AI Act, classifica como “sistemas de alto risco” as aplicações de IA integradas a dispositivos médicos implantáveis ou a tecnologias capazes de decodificar sinais cerebrais, sujeitando seus desenvolvedores a obrigações estritas de avaliação de risco, governança de dados e supervisão humana (UNIÃO EUROPEIA, 2024). Em paralelo, o Regulamento (UE) 2017/745 sobre Dispositivos Médicos já enquadra implantes neurais ativos na categoria de risco máximo (*Classe III*), obrigando a demonstração de segurança clínica e de proteção de dados (UNIÃO EUROPEIA, 2017a). Nesse cenário, o Consentimento Livre e Esclarecido (CLE) torna-se pedra de toque: como garantir que o paciente compreenda e autorize, livre de coação, a captação, armazenamento e eventual decodificação de sua atividade cerebral? E de que modo prevenir a exploração comercial ou securitária de neurodados sem permissão explícita? Esses dilemas mostram que a fronteira neurotecnológica expande o debate sobre privacidade para além da proteção de dados pessoais, alcançando a própria inviolabilidade da mente (UNESCO, 2023).

Portanto, paralelamente à evolução da IA, a difusão de dispositivos neurotecnológicos impõe ao Direito a tarefa de formular respostas adequadas em vários planos normativos. Na saúde mediada por tecnologia, diversos direitos fundamentais entram em tensão. O direito à igualdade e à não discriminação exige que ninguém seja prejudicado por vieses implícitos em modelos preditivos, sob pena de ofensa à dignidade humana (ANGWIN *et al.*, 2016). Já o direito à privacidade e à proteção de dados pessoais ganha novo relevo, pois a IA depende de grandes volumes de informações sensíveis – genéticas, clínicas e comportamentais coletadas por *wearables* (“dispositivos vestíveis”). No caso das neurotecnologias, essa proteção alcança a própria dimensão mental, pois registros de atividade cerebral integram o que se convencionou chamar de privacidade mental (IENCA; ANDORNO, 2017). No Brasil, essa tutela foi reforçada pela Lei Geral de Proteção de Dados Pessoais – LGPD (Lei 13.709/2018), que positivou princípios de finalidade, necessidade, transparência e segurança, prevendo bases legais específicas – entre elas o consentimento do titular ou a tutela da saúde – para tratar dados sensíveis (BRASIL, 2018). Nesse cenário, destaca-se o princípio da autodeterminação informativa, difundido por Danilo Doneda, que confere ao indivíduo o poder de governar seus próprios dados (DONEDA, 2021). Transposto à IA médica, isso significa que o paciente deve poder decidir se seus registros clínicos serão usados no treinamento de algoritmos e recusar diagnósticos exclusivamente automatizados sem supervisão humana. Essa exigência conecta-se ao direito à informação e ao consentimento livre e esclarecido: o paciente precisa saber que uma IA ou um dispositivo neurotecnológico participa de sua avaliação clínica, compreender

suas limitações e manter o direito de questionar ou buscar segunda opinião ou, ainda, recusar a intervenção médica (PLOUG; HOLM, 2020; 2023). Em relatório recente, o Conselho da Europa realçou a importância de preservar a confiança na relação paciente-médico diante das tecnologias inteligentes. Entre as recomendações, propôs adotar o termo “inteligência aumentada” (*augmented intelligence*) para enfatizar o caráter assistencial da IA, que deve coexistir – e não suplantar – a decisão clínica humana (MITTELSTADT, 2021). O mesmo documento conclama os legisladores a estabelecer critérios de qualidade, explicabilidade e *accountability* nos sistemas de IA em saúde, garantindo justificativas claras para recomendações algorítmicas e delimitando, de forma transparente, quem responde por eventuais falhas (MITTELSTADT, 2021). Assim, a IA opera como suporte ampliando a capacidade diagnóstica do profissional, mas o predomínio humano na decisão clínica permanece como condição ética e jurídica indispensável.

Diante de todos esses desafios, surge uma indagação decisiva: estará o ordenamento jurídico brasileiro aparelhado para enfrentar essa realidade em acelerada transformação? Constata-se que o País ainda carece de legislação específica que discipline o emprego da IA e das neurotecnologias na saúde. A jurisprudência pátria tampouco consolidou precedentes robustos acerca de temas como erro médico algorítmico, responsabilidade por decisões automatizadas ou violações de privacidade mental oriundas de dispositivos neurotecnológicos. Essa lacuna normativa e jurisprudencial projeta considerável insegurança jurídica para profissionais de saúde, pacientes, desenvolvedores e instituições, que carecem de parâmetros claros sobre os limites e garantias envolvendo a adoção dessas tecnologias de vanguarda. A doutrina especializada sublinha a tensão existente entre fomentar a inovação e assegurar proteção: se a ausência de regulação pode deixar indivíduos desassistidos e direitos fundamentais desprotegidos, por outro lado uma rigidez excessiva tem o potencial de tolher avanços benéficos e desestimular pesquisas. Impõe-se, pois, a construção de paradigmas regulatórios flexíveis e responsivos, aptos a alinhar o estímulo ao progresso tecnológico com salvaguardas sólidas de segurança, eficácia e respeito aos direitos fundamentais.

Nesse sentido, Price II (2017) defende que agências reguladoras de saúde – a exemplo da FDA nos Estados Unidos e da ANVISA no Brasil – adotem procedimentos adaptativos e colaborativos na avaliação de algoritmos médicos, exigindo transparência dos desenvolvedores, auditorias independentes e monitoramento contínuo de desempenho. Afinal, marcos estáticos de regulação tendem a se tornar incompatíveis com sistemas de aprendizado dinâmico que evoluem com o tempo. No terreno da responsabilidade civil, discute-se de que forma atribuir culpa quando um diagnóstico gerado por IA se mostrar equivocado ou quando um dispositivo

neurotecnológico causar danos imprevistos. Fruzina Molnár-Gábor (2020) sustenta que a inserção da IA na relação terapêutica exige repensar o modelo tradicional de responsabilidade, de modo a combinar a culpa profissional do médico, que utiliza a ferramenta, com a responsabilidade pelo produto, atribuída ao desenvolvedor ou fornecedor. Essa articulação, possivelmente por meio de um regime graduado ou compartilhado, busca evitar as zonas cinzentas em que falhas do sistema não resultem em responsabilização efetiva de nenhum dos envolvidos. Em última instância, o desafio reside em harmonizar a inovação algorítmica e neurotecnológica com a tutela dos direitos fundamentais dos pacientes, assegurando que o avanço tecnológico se concretize sem sacrificar a justiça, a segurança e a dimensão humana do ambiente hospitalar.

Vale notar que não apenas o Brasil, mas diversos países e organismos internacionais vêm construindo respostas normativas a esses desafios. A União Europeia acaba de aprovar o Regulamento (UE) 2024/1689 – AI Act (Lei de Inteligência Artificial), que impõe a sistemas de IA utilizados em saúde obrigações de transparência, avaliação de risco e supervisão humana por se enquadrarem na categoria de “alto risco” (UNIÃO EUROPEIA, 2024). Paralelamente, o Regulamento (UE) 2017/745 sobre Dispositivos Médicos já inclui implantes cerebrais ativos e *software* de diagnóstico algorítmico na classe III, sujeitando-os a certificação clínica rigorosa (UNIÃO EUROPEIA, 2017a). Nos Estados Unidos, a *Food and Drug Administration* (FDA) adotou um modelo de *Predetermined Change Control Plans - PCCPs* (“Plano Pré-aprovado de Modificações Futuras”), permitindo que algoritmos “aprendizes” de imagiologia ou de BCIs recebam autorização condicional acompanhada de monitoramento pós-mercado e atualizações contínuas (FDA, 2025). A Organização Mundial da Saúde (OMS) publicou, em 2021, seis princípios orientadores para IA em saúde – transparência, inclusão, responsabilidade, equidade, segurança e promoção de direitos humanos – recomendando avaliações de impacto e supervisão humana em todos os ciclos de vida do sistema (OMS, 2021). No campo da neurotecnologia, o Chile desponta como pioneiro: a Lei 21.383/2021 inseriu na Constituição a proteção expressa da integridade mental e da privacidade dos dados neurais, condicionando qualquer aplicação neurotecnológica ao respeito à atividade cerebral e às informações dela derivadas (CHILE, 2021). Trata-se de medida visionária, que antecipa cenários antes restritos à ficção científica, em que dispositivos poderiam ler ou modular conteúdos mentais. Iniciativas semelhantes florescem em outras jurisdições. A Declaração de León sobre Neurotecnologia Europeia (Espanha, 2023) propõe salvaguardas para privacidade mental e identidade pessoal; no Brasil, tramita a Proposta de Emenda Constitucional 29/2023, que busca incluir a proteção à integridade mental e a transparência algorítmica no rol dos direitos fundamentais (BRASIL,

2023). Todos esses movimentos sinalizam o reconhecimento, em nível global, de que a incorporação de IA e neurotecnologias à medicina exige arcabouços jurídicos inovadores e vinculantes, para além de diretrizes éticas voluntárias.

No âmago desse debate surge a proposta de reconhecer (ou não) novos direitos específicos – os neurodireitos – para salvaguardar a esfera mental diante do avanço das neurotecnologias. O conceito remete a garantias que complementam os direitos já consagrados, pois tais tecnologias podem acessar ou interferir em processos mentais íntimos. Rafael Yuste, coordenador do *Morningside Group*, propôs ampliar o rol de direitos humanos com, ao menos, quatro salvaguardas básicas: (i) *O direito à privacidade mental*, que protege dados neurais contra decodificação ou uso não consentido; (ii) *O direito à identidade pessoal* (ou “continuidade psicológica”), para impedir que intervenções tecnológicas perturbem o senso de identidade; (iii) *O direito à liberdade cognitiva*, garantindo autonomia decisória sem coerção ou manipulação neurotecnológica; e (iv) *O direito ao acesso equitativo* às inovações de aumento cognitivo ou sensorial, evitando o agravamento de desigualdades (YUSTE *et al.*, 2017). Mais recentemente, o grupo acrescentou um *quinto pilar*, voltado à *proteção contra vieses algorítmicos e discriminação em neurotecnologias*, assegurando que *BCIs* ou implantes não introduzam injustiças na cognição ou no comportamento dos usuários (YUSTE *et al.*, 2021).

Esses princípios vêm catalisando discussões em fóruns internacionais. Enquanto alguns juristas questionam se é necessário cunhar direitos “novos” ou apenas reinterpretar os já existentes (LIGTHART; BUBLITZ; ALEGRE, 2023), categorias como privacidade mental e liberdade cognitiva já inspiram iniciativas legislativas concretas – vide a emenda constitucional chilena de 2021 que protege a integridade mental e os dados neuronais (CHILE, 2021). O objetivo comum é assegurar que, mesmo em cenário no qual dispositivos possam “ler” ou “escrever” no cérebro, permaneçam intocados os elementos essenciais da condição humana: autonomia de pensamento, intimidade psíquica, identidade e dignidade de cada pessoa.

Diante desse cenário de oportunidades extraordinárias, mas também de riscos consideráveis, torna-se imprescindível adotar uma perspectiva interdisciplinar e jurídico-crítica na análise do emprego de IA e neurotecnologias em diagnósticos e tratamentos médicos. A presente pesquisa propõe-se a escrutinar tais inovações à luz de seus principais desafios – em especial a opacidade sistêmica e a discriminação algorítmica – e de suas repercussões na autonomia do paciente. Busca-se, assim, contribuir para harmonizar o progresso tecnológico com a salvaguarda de valores constitucionais, demonstrando que é possível enfrentar os dilemas técnicos sem sacrificar a justiça, a segurança e a dimensão humana do cuidado em saúde.

Nesse contexto, o objetivo geral desta dissertação é demonstrar de que modo a opacidade e os vieses de sistemas de IA e neurotecnologias aplicados à saúde colocam em crise o modelo clássico de Consentimento Livre e Esclarecido e propor salvaguardas jurídicas estruturais aptas a reconstruir um patamar adequado de proteção da autonomia e da integridade mental do paciente na era da saúde algorítmica.

De forma mais específica, buscam-se os seguintes objetivos: (i) *Mapear as principais aplicações de IA e neurotecnologias em diagnósticos e tratamentos médicos*, destacando suas potencialidades clínicas e riscos correlatos; (ii) *Analisar o fenômeno da opacidade algorítmica* – inclusive sob a perspectiva do que eu denominei chamar de “dupla caixa-preta” – e suas implicações para a confiança, a validação clínica e o direito à explicação em saúde; (iii) *Examinar as múltiplas fontes de viés e discriminação algorítmica na saúde digital e neurotecnológica* e seus impactos sobre grupos vulneráveis e sobre o princípio constitucional da isonomia; (iv) *Investigar como opacidade e viés contribuem para a erosão da autonomia do paciente e para o colapso funcional do Consentimento Livre e Esclarecido* em contextos de diagnóstico e terapia mediados por tecnologia; e (v) *Identificar e sistematizar salvaguardas jurídicas mínimas* – como deveres de transparência, explicabilidade, auditoria de vieses, governança de neurodados e reconhecimento de neurodireitos – necessárias à reconstrução da eficácia protetiva do consentimento e à promoção de uma governança proativa da IA e das neurotecnologias na saúde.

Para atingir esse propósito, parte-se de uma abordagem metodológica qualitativa, de caráter exploratório e descritivo-analítico. Foram realizadas extensa pesquisa bibliográfica e análise documental, abarcando literatura nacional e internacional de Direito, Bioética, Ciência da Computação e Neurociências, bem como documentos normativos (leis, projetos de lei, regulamentações setoriais e diretrizes de órgãos como a UE, OMS etc.). Adotou-se também o método comparativo, examinando-se modelos regulatórios estrangeiros e casos ilustrativos, a fim de extrair lições e boas práticas aplicáveis ao contexto brasileiro. Trata-se, portanto, de um estudo teórico e aplicado, que alia revisão de doutrina e jurisprudência com a discussão de proposições normativas.

A pergunta de pesquisa que orienta este trabalho, rearticulada para refletir a crise paradigmática que constitui o cerne da investigação, pode ser sintetizada nos seguintes termos: **de que maneira as características operacionais intrínsecas da inteligência artificial e das neurotecnologias – notadamente a opacidade algorítmica e o viés sistêmico – provocam a erosão da autonomia do paciente e o colapso funcional da doutrina do Consentimento Livre e Esclarecido na prática médica? Diante dessa crise, quais novos paradigmas**

jurídicos e de governança, incluindo a positivação de neurodireitos, são necessários para reconstruir as salvaguardas da dignidade e da autodeterminação humana?

Essa formulação reflete, portanto, o abandono de uma perspectiva de mera gestão de risco em favor de uma análise de crise paradigmática. O argumento que sustenta este trabalho não é que os riscos da IA podem ser simplesmente “mitigados” pelo arcabouço ético-jurídico vigente. Pelo contrário, sustenta-se que as propriedades intrínsecas dessas tecnologias – a opacidade que corrói o esclarecimento e o viés que corrompe a liberdade – tornam a doutrina do Consentimento Livre e Esclarecido funcionalmente inadequada. A questão inevitavelmente se desloca de “como podemos ‘consertar’ o sistema para lidar com a IA e as neurotecnologias?” para “como a IA e as neurotecnologias ‘quebram’ o sistema do Consentimento e que novo sistema precisamos construir?”. Essa abordagem posiciona a dissertação como uma obra de análise crítica e proposição normativa, que primeiro investiga as causas do colapso para, em seguida, avaliar as respostas necessárias para sua superação.

Por delimitação metodológica, circunscrevemos a análise ao consentimento prestado por pessoas adultas e plenamente capazes, com finalidade diagnóstica e de tratamentos de saúde. Não examinaremos, portanto, o assentimento de menores nem as hipóteses de representação ou substituição decisória, isto é, situações em que o consentimento é fornecido por terceiro em razão de incapacidade do titular. Também, não abordaremos diretamente o consentimento para a proteção de dados pessoais, embora tenhamos que fazer uso de suas normas (nacional e internacional) para fundamentar as salvaguardas jurídicas desenvolvidas no decorrer do texto e, especificamente, no Capítulo 5, frente às lacunas normativas sobre o CLE em si.

Assim, para responder a essa problemática, a presente dissertação estrutura-se em seis capítulos centrais, que constroem uma linha argumentativa coesa, além desta Introdução e da Conclusão.

O Capítulo 2 funciona como o prólogo da crise, estabelecendo suas condições de possibilidade tecnológicas. Apresentam-se os fundamentos e o contexto histórico-evolutivo da IA e das neurotecnologias, demonstrando que o colapso do consentimento não é um acidente, mas uma consequência direta da trajetória tecnológica em direção a sistemas cada vez mais poderosos, autônomos e indecifráveis. Traça-se a evolução da IA em saúde – dos sistemas especialistas baseados em regras e relativamente transparentes dos anos 1970, como o MYCIN, à ascensão do *deep learning* contemporâneo, caracterizado por sua complexidade e opacidade inerentes. Em paralelo, detalha-se a evolução das neurotecnologias, de ferramentas de observação (EEG, MRI) a intervenções diretas e invasivas no cérebro (BCIs, DBS), que intensificam os desafios à privacidade e à integridade mental.

O Capítulo 3 ergue o primeiro pilar da dissertação, argumentando que a opacidade ataca diretamente o pressuposto “esclarecido” do consentimento. Aprofunda-se o problema da opacidade algorítmica, examinando a noção de “caixa-preta” tanto em IA quanto em dispositivos neurotecnológicos. A dissertação avança ao propor o conceito de “dupla caixa-preta”, que emerge na intersecção da IA com a neurociência: um algoritmo opaco que analisa a fonte de dados mais inescrutável de todas, o cérebro humano. Essa dupla camada de indecifrábilidade cria um “abismo probatório” na atribuição de responsabilidade e corrói a confiança, tornando o dever de informação do médico uma tarefa quase impossível de ser cumprida. Ao consentir com um diagnóstico ou terapia baseada em uma “caixa preta” ou em uma “dupla caixa-preta”, o paciente está, na prática, consentindo com o inexplicável, o que esvazia o ato de seu conteúdo informativo e o transforma em um mero cheque em branco.

O Capítulo 4 constitui o segundo pilar da dissertação, sustentando que o viés e a manipulação atacam o pressuposto “livre” do consentimento. A análise desloca-se da opacidade (o processo) para o viés (a substância), identificando como as desigualdades sociais são absorvidas e amplificadas por algoritmos treinados com dados que refletem as iniquidades do mundo real. São exploradas as múltiplas fontes de viés – de bases de treino desequilibradas a escolhas de modelagem – e suas consequências, como diagnósticos sistematicamente falhos para grupos vulneráveis. Adicionalmente, o Capítulo investiga como a arquitetura da tecnologia, por meio de *nudges* e *dark patterns*, pode explorar vieses cognitivos para manipular sutilmente as escolhas de médicos e pacientes, e introduz o risco emergente da “neurodiscriminação”. Uma escolha baseada em informação de qualidade desigual ou influenciada por uma arquitetura enganosa não é genuinamente autônoma. A verdadeira força destrutiva, contudo, reside na sinergia entre opacidade e viés: a opacidade funciona como um escudo que permite ao viés operar de forma oculta e incontestável, enquanto o viés torna a opacidade socialmente tóxica, transformando uma “caixa-preta”, que comete erros aleatórios, em uma arma de injustiça que comete erros sistemáticos contra populações vulneráveis.

No Capítulo 5, coração analítico da dissertação, a proposição central é plenamente desenvolvida. Demonstra-se como a confluência da opacidade (que torna o consentimento não esclarecido) e do viés (que o torna não livre) leva ao colapso funcional da doutrina do Consentimento Livre e Esclarecido. Argumenta-se que o modelo tradicional, centrado na escolha individual e na divulgação de informações, é insuficiente para proteger o paciente em um ambiente tecnológico opaco e potencialmente manipulador. Propõe-se, então, no Capítulo 6, a transição para um paradigma de autodeterminação efetiva e de uma “tutela estrutural”, que desloca o foco da escolha individual para a governança da tecnologia e a garantia da integridade

do ambiente decisório. Este Capítulo, portanto, articula a consequência final da crise: a inadequação do principal mecanismo de proteção da autonomia do paciente, justificando a necessidade de buscar novas soluções normativas.

Finalmente, o Capítulo 7 é enquadrado como a resposta prescritiva à crise diagnosticada. Diante do colapso funcional do consentimento como salvaguarda única, o Capítulo explora os novos paradigmas regulatórios que emergem globalmente para reconstruir a proteção da pessoa. Analisa-se a abordagem baseada em direitos fundamentais, com o pioneirismo do Chile e as propostas no Brasil (PEC 29/2023); o modelo preventivo e baseado em risco da União Europeia (AI Act e MDR); e a abordagem setorial e reativa dos Estados Unidos (FDA). O Capítulo conclui o arco argumentativo ao passar da análise do problema para a avaliação crítica das soluções, posicionando a dissertação como uma contribuição não apenas para o diagnóstico da crise, mas também para o debate sobre como construir um futuro onde a inovação tecnológica e a dignidade humana possam coexistir.

2 INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS EM DIAGNÓSTICOS MÉDICOS: FUNDAMENTOS E CONTEXTO

A inteligência artificial (IA), concebida como o estudo e a construção de agentes que percebem seu ambiente e executam ações, emergiu como um domínio de pesquisa e desenvolvimento com o potencial de remodelar fundamentalmente diversos setores da sociedade. No cerne dessa transformação reside a capacidade de tais agentes de implementar funções complexas, mapeando sequências de percepções em ações por meio de variadas representações, como sistemas de produção, agentes reativos, planejadores condicionais em tempo real, redes neurais e sistemas de teoria de decisão (RUSSELL; NORVIG, 2021).

No campo da saúde, em especial nos diagnósticos e tratamentos médicos, a IA deixou de ser apenas promessa para se tornar componente efetivo da prática clínica. Combinando grandes volumes de dados clínicos e crescente capacidade de processamento, esses sistemas vêm sendo empregados para identificar padrões associados a enfermidades, apoiar a formulação de hipóteses diagnósticas, estratificar riscos e otimizar fluxos de trabalho assistenciais, com a expectativa de diagnósticos mais rápidos, precisos e individualizados, bem como de maior eficiência na alocação de recursos e ampliação do acesso a serviços de qualidade (UNIÃO EUROPEIA, 2024).

Em paralelo, as neurotecnologias – entendidas como o conjunto de dispositivos e técnicas que estabelecem interface direta ou indireta com o sistema nervoso – avançaram de forma contínua. Dos primeiros eletroencefalogramas clínicos às técnicas de neuroimagem moderna, como tomografia computadorizada e ressonância magnética funcional, e às interfaces cérebro-computador experimentais, até os atuais implantes neurais de alta densidade, observa-se uma aproximação progressiva entre tecnologia digital e neurociência (VIDAL, 1973). Essa combinação permite registrar e interpretar sinais elétricos e metabólicos do cérebro com crescente resolução, abrindo espaço para aplicações diagnósticas sofisticadas: algoritmos de IA acoplados ao EEG já identificam padrões característicos de epilepsia ou de transtornos neurodegenerativos em estágios iniciais, antecipando o diagnóstico em relação ao aparecimento de sintomas clínicos (LI *et al.*, 2025); interfaces cérebro-computador possibilitam a comunicação de pacientes em estado de paralisia completa e, em perspectiva, poderão monitorar estados cognitivos para identificar alterações sutis indicativas de *delirium* ou declínio cognitivo incipiente (LIU *et al.*, 2025).

Nesse cenário, IA e neurotecnologias estabelecem uma complementaridade que está no núcleo do problema investigado nesta dissertação. A IA oferece a capacidade analítica para

extrair padrões complexos de grandes massas de dados clínicos; as neurotecnologias acrescentam uma fonte de informação particularmente sensível – a própria atividade elétrica e funcional do cérebro –, permitindo explorar dimensões até então inacessíveis do funcionamento mental e neural (YUSTE et al., 2017; IENCA; ANDORNO, 2017). Essa convergência potencializa avanços como o diagnóstico precoce de doenças neurodegenerativas, o monitoramento contínuo de condições neurológicas e a reabilitação motora ou cognitiva assistida por IA (TOPOL, 2019).

Ao mesmo tempo, ela aproxima duas esferas de alta complexidade e limitada transparência: de um lado, a opacidade algorítmica de modelos de aprendizado de máquina cuja lógica interna é de difícil reconstrução, mesmo para especialistas (PASQUALE, 2015); de outro, a inescrutabilidade do cérebro humano, cuja dinâmica e plasticidade seguem apenas parcialmente compreendidas (KANDASWAMY; YUSTE, 2022). A convergência dessas duas formas de opacidade – aqui designada “dupla caixa-preta” – gera um risco epistêmico singular e amplia a assimetria de poder entre desenvolvedores, profissionais de saúde e pacientes, fragilizando os pilares da autodeterminação e da tomada de decisão consciente no contexto médico (O’NEIL, 2016), questão que será aprofundada em capítulo específico.

2.1 EVOLUÇÃO HISTÓRICA DA IA E DAS NEUROTECNOLOGIAS

Antes de examinar casos paradigmáticos de aplicação contemporânea, é necessário situar a inteligência artificial e as neurotecnologias em uma trajetória histórica que permita compreender como essas ferramentas se tornaram centrais na prática médica. A evolução da IA – da fase simbólica e dos sistemas especialistas às abordagens conexionistas, ao aprendizado de máquina e ao aprendizado profundo, culminando na atual onda de modelos generativos – e o desenvolvimento paralelo das neurotecnologias – dos primeiros registros eletroencefalográficos e técnicas de neuroimagem às interfaces cérebro-computador e implantes neurais modernos – revelam uma sedimentação gradual de capacidades técnicas que hoje viabilizam diagnósticos cada vez mais automatizados, preditivos e personalizados (RUSSELL; NORVIG, 2021; MARIN, 2024; YUSTE *et al.*, 2017). Reconstruir esse percurso, como se fará neste item 2.1, é condição prévia para compreender não apenas o potencial clínico dessas tecnologias, mas também a gênese dos problemas de opacidade, viés e redistribuição de poder decisório que, mais adiante, se mostrarão decisivos para a análise da autonomia do paciente e do colapso funcional do modelo tradicional de Consentimento Livre e Esclarecido.

2.1.1 Evolução da IA em Aplicações Médicas:

A intrincada jornada da Inteligência Artificial (IA) no setor da saúde é uma narrativa de ambição científica, avanços tecnológicos incrementais, períodos de reavaliação e recentes disrupções exponenciais. Compreender essa evolução não é meramente um exercício historiográfico; é fundamental para contextualizar os debates contemporâneos sobre as potencialidades diagnósticas da IA, bem como para fundamentar a análise crítica dos desafios éticos, legais e sociais que emergem com sua crescente penetração na prática médica. Este subitem se propõe a delinear esse percurso, desde as concepções inaugurais da IA e suas primeiras incursões na medicina, passando pela era dos sistemas especialistas, os desafios dos “invernos da IA”, a ascensão transformadora do aprendizado de máquina e do aprendizado profundo, até as mais recentes fronteiras abertas pela IA generativa e pelos Grandes Modelos de Linguagem (LLMs), sem descuidar da evolução paralela das preocupações com seus impactos sobre os direitos fundamentais.

2.1.1.1 *Dos Primórdios Conceituais à Primeira Onda de IA na Saúde: A Era dos Sistemas Especialistas*

A concepção de máquinas capazes de simular a inteligência humana, embora com raízes filosóficas e literárias mais antigas, ganhou contornos científicos em meados do século XX. O trabalho seminal de Alan Turing (1950), notadamente seu artigo de 1950, *Computing Machinery and Intelligence*, é frequentemente citado como um marco, ao propor o “Jogo da Imitação” (posteriormente conhecido como Teste de Turing) como um critério para aferir a capacidade de uma máquina exibir comportamento inteligente indistinguível do humano. Esta provocação intelectual alimentou as discussões que culminariam, em 1955, na proposta para o *Dartmouth Summer Research Project on Artificial Intelligence*, onde John McCarthy e seus colegas formalmente cunharam o termo “Inteligência Artificial” (NASCIMENTO *et al.*, 2024). O objetivo audacioso era explorar maneiras de fazer as máquinas “usarem a linguagem, formarem abstrações e conceitos, resolverem os tipos de problemas agora reservados aos humanos e melhorarem a si mesmas” (RUSSELL; NORVIG, 2021).

Nesse período inicial, a abordagem dominante era a da IA simbólica, também conhecida como *GOF AI - Good Old-Fashioned AI* (“A boa e velha IA”). Como explicam Stuart Russell e Peter Norvig (2021) em sua obra canônica *Artificial Intelligence: A Modern Approach*, esta vertente da IA se baseava na hipótese do sistema de símbolos físicos, que postula que esse

sistema (de símbolos físicos) tem os meios necessários e suficientes para a ação inteligente geral. Essencialmente, o foco era na representação do conhecimento através de símbolos e na manipulação desses símbolos por meio de regras lógicas e heurísticas para realizar inferências e resolver problemas. A crença fundamental era que o raciocínio inteligente poderia ser formalizado e implementado em máquinas.

A medicina, com sua tradição de conhecimento estruturado (taxonomias de doenças, protocolos de tratamento) e a natureza muitas vezes dedutiva do raciocínio diagnóstico, apresentou-se como um campo de aplicação natural para a IA simbólica. Foi nesse terreno que floresceram os sistemas especialistas, talvez a mais emblemática aplicação da IA na saúde durante essa primeira onda. Conforme resgata Heimar de Fátima Marin, já na década de 1970, esses sistemas buscavam replicar a expertise de médicos especialistas em domínios bem definidos (MARIN, 2024).

Um dos exemplos mais notórios é o *MYCIN*, desenvolvido na Universidade de Stanford no início dos anos 1970. Seu propósito era diagnosticar infecções bacterianas sanguíneas e recomendar o tratamento antibiótico mais adequado (MARIN, 2024). A arquitetura do *MYCIN* era representativa dos sistemas especialistas da época: continha uma base de conhecimento com cerca de 600 regras de produção no formato “SE (condição), ENTÃO (ação/conclusão)”, extraídas de especialistas em doenças infecciosas e da literatura médica. Essas regras representavam o conhecimento heurístico dos médicos. Havia também um motor de inferência que utilizava uma estratégia de *backward chaining* (“encadeamento reverso”), partindo de uma hipótese diagnóstica e buscando as evidências (sintomas, resultados de exames) que a confirmassem ou refutassem (WINSTON, 1984; VILLANI; RONDEPIERRE, 2020). O *MYCIN* introduziu uma inovação importante ao lidar explicitamente com a incerteza: o sistema atribuía “fatores de certeza” às regras e conclusões, permitindo expressar o grau de confiança de cada diagnóstico (RUSSELL; NORVIG, 2021). Além disso, incorporava um subsistema de explicação que possibilitava ao usuário compreender o motivo de cada pergunta e o caminho lógico seguido até determinada conclusão – um precursor ainda rudimentar do que hoje se denomina explicabilidade na IA (MARIN, 2024).

Outro sistema especialista influente foi o *INTERNIST-1*, desenvolvido na Universidade de Pittsburgh, por Harry Pople e Jack Myers, e posteriormente evoluído para o *Quick Medical Reference (QMR)*. Destinado a diagnósticos em medicina interna, o *INTERNIST-1* possuía uma vasta base de conhecimento que cobria centenas de doenças e milhares de manifestações clínicas, utilizando um processo iterativo de geração e teste de hipóteses para chegar a um diagnóstico diferencial (GALA *et al.*, 2024). Outros sistemas notáveis do período incluíram o

PUFF, para interpretação de testes de função pulmonar, e o *ONCOCIN*, para gerenciamento de protocolos de quimioterapia para pacientes com câncer (SHORTLIFFE; SEPÚLVEDA, 2018).

Apesar desses desenvolvimentos pioneiros e do desempenho promissor em domínios restritos, os sistemas especialistas enfrentaram obstáculos significativos que limitaram sua adoção em larga escala na prática clínica (MILLER, 1994). O processo de aquisição de conhecimento – extrair, formalizar e codificar o conhecimento tácito e explícito dos especialistas médicos – era extremamente trabalhoso, demorado e propenso a vieses e inconsistências. A manutenção e atualização dessas complexas bases de regras também eram desafiadoras, dado o rápido avanço do conhecimento médico. A rigidez desses sistemas, baseados em regras predefinidas, dificultava sua adaptação a casos atípicos ou à evolução da prática médica. A integração com os sistemas de informação hospitalar existentes era precária, e a interface com o usuário, muitas vezes, não era amigável, resultando em baixa aceitação por parte dos profissionais de saúde (BALICER; COHEN-STAVI, 2020). Além disso, a capacidade de lidar com a incerteza e a complexidade dos dados do mundo real, que frequentemente são incompletos, ruidosos ou contraditórios, permanecia uma limitação importante. Frank Pasquale (2020), em sua análise sobre a regulação de novas tecnologias, implicitamente aponta para como essa primeira geração de IA, embora menos “opaca” que os modelos atuais, já levantava questões sobre a confiança e a responsabilidade nas decisões automatizadas.

A materialização dessas ideias na medicina deu-se, proeminentemente, através dos sistemas especialistas. Além do *MYCIN* e do *INTERNIST-1/QMR*, já mencionados, outros sistemas pioneiros ilustram a diversidade de aplicações e os desafios enfrentados. O *DENDRAL*, desenvolvido na Universidade de Stanford a partir de meados da década de 1960, por Edward Feigenbaum, Bruce Buchanan e Joshua Lederberg, embora focado na análise química (inferir a estrutura molecular de compostos orgânicos a partir de dados de espectrometria de massa), é crucial por ter sido um dos primeiros sistemas a demonstrar que a IA poderia atingir desempenho do nível de um perito em tarefas científicas complexas e por ter estabelecido muitos dos princípios da engenharia do conhecimento (LEDERBERG *et al.*, 1987). Seu sucesso influenciou diretamente o desenvolvimento de sistemas especialistas em outros domínios, incluindo a medicina, ao provar a viabilidade de codificar conhecimento especializado em regras formais.

No campo estritamente médico, o *CASNET - Causal ASSociational NETWORK* (“Rede Causal Associativa”), desenvolvido na Universidade Rutgers, por Saul Amarel e Casimir Kulikowski, a partir do início dos anos 1970, focava no diagnóstico e tratamento do glaucoma. Sua arquitetura era baseada em uma rede causal que modelava as relações entre observações

clínicas, estados patofisiológicos intermediários e categorias diagnósticas, permitindo um raciocínio mais complexo sobre a progressão da doença (KULIKOWSKI *et al.*, 1982). O sistema conseguia gerar diagnósticos, prognósticos e recomendações terapêuticas, e sua abordagem baseada em modelos causais representava uma tentativa de ir além das simples regras de produção, buscando uma compreensão mais profunda dos mecanismos da doença. Sua performance foi validada em estudos clínicos, mostrando concordância significativa com oftalmologistas especialistas (WEISS *et al.*, 1978).

Outro sistema relevante foi o *PIP - Present Illness Program* (“Programa de Doenças Atuais”), desenvolvido no MIT por Peter Szolovits e Stephen Pauker, também nos anos 1970. O *PIP* era projetado para auxiliar no diagnóstico de doenças renais, especificamente na tomada da história da doença atual. Ele interagia com o médico, fazendo perguntas sobre o paciente e utilizando um modelo de conhecimento que combinava *frames* (estruturas de dados que representam conceitos e suas propriedades) e uma rede semântica para organizar o conhecimento médico sobre edema e suas causas (PAUKER *et al.*, 1976). O *PIP* era notável por sua tentativa de modelar o processo de raciocínio clínico de forma mais flexível, adaptando suas questões com base nas informações já fornecidas.

Segundo Stuart Russell e Peter Norvig (2021), apesar do avanço conceitual que representaram, os sistemas especialistas da primeira geração esbarraram em limitações significativas que frearam sua adoção disseminada: (i) *Aquisição de Conhecimento* - O processo de extrair, formalizar e validar o conhecimento dos especialistas humanos era o principal gargalo. Era uma tarefa intensiva, demorada e que muitas vezes resultava em bases de regras incompletas, inconsistentes ou com vieses dos especialistas. O conhecimento teórico para resolver problemas precisava ser mapeado, de sua forma geral, para “receitas de bolo” específicas para o domínio, um processo detalhado e intensivo; (ii) *Manutenção e Escalabilidade* - Manter atualizadas as vastas bases de conhecimento em face da rápida evolução da medicina era um desafio constante. Além disso, escalar esses sistemas para cobrir domínios mais amplos ou integrar diferentes tipos de conhecimento mostrava-se computacionalmente intratável. O entusiasmo inicial de que bastaria dispor de *hardware* mais rápido para “escalar” os sistemas de IA foi gradualmente refreado, quando se constatou que, na prática, a ampliação para problemas maiores ou mais complexos não se sustentava. Em sistemas probabilísticos baseados em regras, o crescimento do conjunto de regras tornava cada vez mais frequentes interações indesejadas entre elas, de modo que os fatores de ponderação precisavam ser ajustados toda vez que novas regras eram incorporadas, comprometendo a escalabilidade e a previsibilidade do sistema; (iii) *Fragilidade e Inflexibilidade (Brittleness)* - Os sistemas

baseados em regras tendiam a falhar de forma abrupta quando confrontados com situações não previstas em sua base de conhecimento (o “penhasco de conhecimento” - *knowledge cliff*). Sua capacidade de lidar com dados ruidosos, incompletos ou contraditórios – comuns na prática médica – era limitada (WINSTON, 1984), pois não era fácil enunciar conhecimento informal nos termos formais de lógica, em particular quando o conhecimento era menos de 100% certo; (iv) *Integração e Usabilidade* - A integração desses sistemas aos fluxos de trabalho e sistemas de informação hospitalares existentes era tecnicamente complexa. As interfaces com o usuário eram frequentemente pouco intuitivas, exigindo treinamento específico e, por vezes, gerando resistência por parte dos profissionais de saúde, que viam os sistemas mais como um fardo do que como um auxílio (TOPOL, 2019); e (v) *Validação Clínica* - A avaliação rigorosa da eficácia e segurança desses sistemas em cenários clínicos reais era um processo complexo e nem sempre realizado de forma sistemática. A maioria das pesquisas focava na avaliação da IA em ambientes artificiais, resultando em uma lacuna de conhecimento sobre sua eficácia em cenários clínicos reais (GUANDALINI *et al.*, 2024). A falta de validação externa rigorosa e o monitoramento contínuo do desempenho do *software* são essenciais, já que algoritmos podem se tornar tendenciosos ou desviar ao longo do tempo quando implantados em contextos diferentes daqueles em que foram criados (NOGAROLI, 2023; GUANDALINI *et al.*, 2024).

2.1.1.2 Os “Invernos da IA” e a Transição para Novas Abordagens

O entusiasmo inicial que caracterizou os primórdios da IA, incluindo suas aplicações médicas, arrefeceu consideravelmente em dois períodos distintos, conhecidos como os “invernos da IA”¹. O primeiro ocorreu entre meados da década de 1970 e o início dos anos 1980, e o segundo se estendeu do final dos anos 1980 até meados da década de 1990. Esses períodos foram marcados por uma significativa redução no financiamento para pesquisa em IA, um ceticismo crescente quanto à viabilidade de se atingir uma inteligência artificial geral com as abordagens predominantes, e uma reorientação dos esforços para problemas mais delimitados e com potencial de aplicação mais imediato (TOOSI, 2021). As causas desses “invernos” foram multifatoriais, incluindo a superestimação das capacidades da tecnologia da época, exemplificada pelo relatório *Lighthill* no Reino Unido (1973), que criticou duramente o estado da pesquisa em IA, a dificuldade em escalar os sistemas especialistas para problemas

¹ Análise histórica mostrando que sistemas pioneiros de IA em medicina, embora influentes na pesquisa, tinham uso clínico limitado; muitos serviram mais para explicar raciocínios médicos do que para decidir casos reais, e a euforia inicial foi seguida de frustração, contribuindo para o “inverno da IA” (KULIKOWSKI, 2019).

mais complexos e a competição por recursos com outras áreas emergentes da computação, como o desenvolvimento de computadores pessoais (NASCIMENTO *et al.*, 2024).

No entanto, a pesquisa em IA na área da saúde não cessou completamente durante esses períodos. Como observa Heimar de Fátima Marin (2024), houve uma persistência na exploração de aplicações, embora talvez com um foco renovado em áreas mais pragmáticas da informática médica, como a melhoria dos sistemas de registro de pacientes e o desenvolvimento de bases de dados clínicas (RUSSELL; NORVIG, 2021). A digitalização da informação médica, iniciada com sistemas como o *MEDLARS - Medical Literature Analysis and Retrieval System* (“Sistema de Análise e Recuperação de Literatura Médica”) nos anos 1960, continuou, ainda que lentamente, a pavimentar o caminho para o uso futuro de grandes volumes de dados. O estudo “Desafios do uso da inteligência artificial nos diagnósticos de saúde: uma revisão integrativa” (SILVA *et al.*, 2024) indica que, mesmo com percalços, a semente da IA na saúde permaneceu, aguardando condições mais favoráveis para florescer.

Foi também nesse interregno que começaram a ganhar força as abordagens conexionistas, baseadas em redes neurais artificiais, que, embora propostas inicialmente nos anos 1940 e 1950 (com o *Perceptron* de Frank Rosenblatt), haviam sido largamente abandonadas após críticas como as de Minsky e Papert em seu livro “*Perceptrons*” (1969) – (TOPOL, 2019). O ressurgimento do interesse pelas redes neurais, impulsionado por novos algoritmos de treinamento como o *Backpropagation* (“Retropropagação”), popularizado nos anos 1980, e a crescente percepção das limitações da IA puramente simbólica para lidar com problemas do mundo real, ricos em ambiguidades e padrões complexos, prepararam o terreno para a próxima grande transformação. A capacidade das redes neurais de aprenderem a partir de exemplos, sem a necessidade de regras explícitas, começava a ser redescoberta e reavaliada (KULIKOWSKI, 2019).

Os “invernos da IA” foram caracterizados não apenas pela retração de financiamentos, mas também por um intenso debate interno sobre as direções futuras do campo. A crítica de Minsky e Papert aos *Perceptrons*, em 1969, por exemplo, teve um impacto profundo no desenvolvimento das redes neurais, retardando seu progresso por quase duas décadas (MINSKY; PAPERT, 1969). No entanto, a pesquisa médica, embora impactada pela redução geral de recursos para a IA “pura”, continuou a avançar em frentes paralelas, como a informática médica, a bioestatística e a epidemiologia computacional, que se beneficiaram da crescente digitalização de dados, mesmo que os sistemas de IA propriamente ditos fossem menos proeminentes. Heimar de Fátima Marin destaca que, apesar da retração no uso explícito de “técnicas de IA” na saúde, a dificuldade maior residia na integração desses sistemas aos

registros eletrônicos do paciente, que ainda estavam em estágio inicial (MARIN, 2024). Essa observação sugere que a infraestrutura de dados e a interoperabilidade já eram identificadas como gargalos críticos muito antes da explosão do *big data*.

Durante esse período, a comunidade médica também começou a desenvolver uma apreciação mais crítica dos sistemas de IA. A necessidade de validação clínica rigorosa, a compreensão das limitações dos modelos e as implicações éticas de seu uso começaram a ser discutidas com mais seriedade, ainda que de forma incipiente em comparação com os debates atuais. O trabalho de Lawrence Lessig sobre como o “código” (*software*) funciona como uma forma de lei, moldando e restringindo comportamentos, oferece uma lente para entender como a própria arquitetura desses primeiros sistemas de IA, mesmo que não intencionalmente, já impunha certas lógicas e limitações à prática médica (LESSIG, 2011).

2.1.1.3 A Revolução do Aprendizado de Máquina (ML) e a Ascensão do Aprendizado Profundo (DL)

A superação dos “invernos da IA” e a subsequente explosão de interesse e aplicação da IA, especialmente na saúde, estão intrinsecamente ligadas à consolidação e ao avanço exponencial do Aprendizado de Máquina (*machine learning - ML*). Como definido por Russell e Norvig (2021), o *ML* é um campo que confere aos computadores a habilidade de aprender sem serem explicitamente programados para uma tarefa específica, identificando padrões e fazendo previsões a partir de dados. Segundo os autores, essa mudança de paradigma – de sistemas baseados em conhecimento codificado (*IKBS - Intelligent Knowledge-Based Systems*) para sistemas baseados em dados completos – foi fundamental.

As primeiras aplicações de *ML* na medicina, utilizando algoritmos como árvores de decisão, regressão logística, “Máquinas de Vetores de Suporte” (*SVMs - Support Vector Machines*) e Redes Bayesianas, começaram a surgir com mais consistência nos anos 1990 e 2000 (RUSSELL; NORVIG, 2021). Esses métodos eram capazes de analisar conjuntos de dados clínicos e identificar fatores de risco para doenças, auxiliar na classificação de pacientes ou prever desfechos clínicos com maior flexibilidade do que os sistemas especialistas (PAIXÃO *et al.*, 2022). Por exemplo, modelos de regressão logística foram amplamente utilizados para desenvolver escores de risco cardiovascular, enquanto árvores de decisão ajudavam a delinear caminhos diagnósticos. O estudo de GALA *et al.* (2024) sobre o papel da IA na cardiologia menciona o uso de *SVMs* e *Random Forests* em análises preditivas. No entanto, muitos desses algoritmos de *ML* “clássico” ainda dependiam de um processo trabalhoso de “engenharia de

características” (*feature engineering*), onde especialistas humanos precisavam selecionar e transformar as variáveis mais relevantes dos dados brutos para alimentar os modelos, o que ainda representava um gargalo (SILVA *et al.*, 2024).

No campo da saúde, os algoritmos de Regressão Logística, por exemplo, tornaram-se ferramentas padrão para modelar a probabilidade de um evento binário (*e.g.*, presença ou ausência de doença, sobrevida) com base em um conjunto de preditores. Estudos pioneiros os utilizaram para desenvolver escores de risco para doenças cardiovasculares (como o escore de *Framingham*, embora suas bases sejam anteriores à popularização massiva do *ML*) e para identificar fatores prognósticos em oncologia (HAENSSLE *et al.*, 2018). As Árvores de Decisão e seus conjuntos, como *Random Forests*, ganharam popularidade por sua interpretabilidade (no caso de árvores simples) e robustez. Foram aplicadas na estratificação de risco de pacientes, na identificação de subgrupos de doenças e no auxílio diagnóstico em diversas especialidades (GALA *et al.*, 2024). As “Máquinas de Vetores de Suporte” (*SVMs*), conhecidas por sua capacidade de encontrar hiperplanos ótimos de separação em espaços de alta dimensão, foram exploradas para classificação de tumores, análise de dados de expressão gênica e diagnóstico de transtornos neurológicos (GALA *et al.*, 2024). Redes Bayesianas, que modelam relações de probabilidade entre variáveis, foram usadas para construir modelos diagnósticos e prognósticos que poderiam incorporar conhecimento prévio e lidar com incertezas (RUSSELL; NORVIG, 2021). Cada um desses algoritmos, com suas particularidades, contribuiu para familiarizar a comunidade médica com a ideia de que os dados clínicos poderiam ser explorados de formas novas e poderosas para gerar *insights* e apoiar decisões. Ainda que, em muitos cenários clínicos, os ganhos de desempenho sobre métodos estatísticos clássicos tenham sido modestos e dependentes do contexto, a eficácia dos modelos permaneceu fortemente vinculada à engenharia de características, sobretudo antes da difusão do *deep learning*.

O verdadeiro catalisador da revolução da IA na saúde, particularmente em diagnósticos por imagem e outras tarefas de reconhecimento de padrões complexos, foi o advento e a maturação do “Aprendizado Profundo” (*Deep Learning - DL*). O *DL* é uma subcategoria do *ML* que utiliza Redes Neurais Artificiais (ANNs) com múltiplas camadas de processamento (as “camadas profundas”) para aprender representações de dados com múltiplos níveis de abstração (LITJENS *et al.*, 2017). Conforme Eric Topol descreve vividamente em *Deep Medicine*, o *DL*

“reviveu” a IA, demonstrando capacidades que antes pareciam inatingíveis, especialmente na interpretação de *dados perceptuais*² (TOPOL, 2019).

Por outro lado, as Redes Neurais Convolucionais (*CNNs - Convolutional Neural Network*), em particular, mostraram-se extraordinariamente eficazes na análise de imagens médicas. Inspiradas no córtex visual humano, as *CNNs* aplicam filtros (convoluções) para aprender automaticamente características hierárquicas das imagens – desde bordas e texturas em camadas iniciais até formas e objetos complexos em camadas mais profundas – permitindo, por exemplo, a detecção de nódulos pulmonares em tomografias computadorizadas (TCs), a classificação de lesões cutâneas em dermatoscopia, a identificação de retinopatia diabética em exames de fundo de olho, ou a análise de lâminas de patologia digital (PRICE II, 2017; ZHOU *et al.*, 2021). Diversos estudos, como o já mencionado de HAENSSLE *et al.* (2018) sobre melanoma³ e outros revisados por SILVA *et al.* (2024) sobre diagnóstico de doenças pulmonares⁴, demonstraram a performance de *CNNs* igualando ou superando radiologistas, patologistas e dermatologistas em tarefas específicas.

A aplicação de *CNNs* em radiologia tornou-se um dos campos mais emblemáticos do sucesso do *DL*. Estudos de validação multiplicaram-se.

Um trabalho de CIRESAN *et al.* (2012) já demonstrava a capacidade de *CNNs* para detecção de mitoses em imagens de histopatologia de câncer de mama, alcançando performance competitiva em desafios internacionais.

No diagnóstico de câncer de pulmão a partir de Tomografias Computadorizadas, estudos como o de ARDILA *et al.* (2019), do *Google Health*, mostraram que modelos de *DL* poderiam alcançar ou superar a performance de radiologistas na detecção de malignidade em nódulos pulmonares.

Na oftalmologia, o trabalho de GULSHAN *et al.* (2016), também do *Google*, demonstrou alta sensibilidade e especificidade de um algoritmo de *DL* para detecção de retinopatia diabética e edema macular em retinografias, validado em múltiplos conjuntos de dados com diferentes etnias.

² Este é o ponto crucial do argumento de Topol. Ele descreve que o grande salto do *Deep Learning* foi sua performance “super-humana” em tarefas de percepção, como a visão computacional. Ele usa exaustivamente exemplos da radiologia (interpretação de tomografias), patologia (análise de lâminas) e dermatologia (análise de lesões de pele) para mostrar que é exatamente na interpretação de dados perceptuais complexos e não estruturados que o *DL* se destacou, superando os métodos anteriores e, em alguns casos, os próprios especialistas humanos.

³ Também disponível em: BRITO, Marcos Toscano Siebra. *Governança antecipatória da inteligência artificial no setor de saúde: o que dizem os policy papers internacionais*. Nota Técnica IPEA, n. 135, maio 2024.

⁴ Também disponível em: RUSSELL; NORVIG, 2021.

O artigo “*Diagnosing and categorizing of pulmonary diseases using Deep learning conventional Neural network*” (REDDY; KHANAA, 2023) exemplifica a aplicação de *CNNs* para um espectro de doenças pulmonares, ilustrando a versatilidade da arquitetura.

Além das *CNNs*, outras arquiteturas de *DL* ganharam espaço. As Redes Neurais Recorrentes (*RNNs*) e suas variantes, como *LSTMs* (*Long Short-Term Memory*) e *GRUs* (*Gated Recurrent Units*), encontraram nicho na análise de dados sequenciais. Por exemplo: (i) *No processamento de sinais de ECG (eletrocardiograma)*, estudos como o de HANNUN *et al.* (2019) mostraram que *RNNs* poderiam detectar uma variedade de arritmias, com acurácia ao nível de um cardiologista, a partir de dados de ECG ambulatorial, para predição de crises epiléticas; e (ii) *No campo do Processamento de Linguagem Natural (PLN) médico*, *RNNs* foram aplicadas para extrair informações de notas clínicas não estruturadas em Prontuários Eletrônicos do Paciente (PEPs), como extrair informações clínicas, prever risco de reinternação ou identificar eventos adversos a medicamentos (OMS, 2023); identificar menções a doenças, medicamentos e eventos adversos, ou para prever reinternações hospitalares com base no histórico textual do paciente (PRICE II *et al.*, 2019).

As “Redes Adversárias Generativas” (*GANs - Generative Adversarial Networks*), embora menos diretamente diagnósticas (por serem generativas), tiveram um papel importante na evolução da IA médica. Introduzidas por Ian Goodfellow e colaboradores em 2014, consistem em duas redes neurais – uma geradora e uma discriminadora⁵ – que competem entre si, permitindo a geração de dados sintéticos de alta qualidade. Na medicina (YI *et al.*, 2019), *GANs* têm sido usadas para: (i) *Aumento de Dados (data augmentation)* - Gerar imagens médicas sintéticas, mas realistas, para expandir conjuntos de treinamento, especialmente para doenças raras, onde os dados são escassos; (ii) *Anonimização e Privacidade* - Modificar imagens médicas para remover informações identificáveis, preservando a utilidade para pesquisa; e (iii) *Tradução entre Modalidades de Imagem (image-to-image translation)* - Por exemplo, aprender um mapeamento que gera uma Tomografia Computadorizada sintética a partir de uma Ressonância Magnética (ou o inverso), preservando a anatomia e reproduzindo o contraste da modalidade alvo para aplicações como planejamento de radioterapia, reduzindo exames e radiação desnecessários.

As arquiteturas *Transformer*, outro modelo de aprendizado profundo e que inicialmente foram concebidas para Processamento de Linguagem Natural (PLN), rapidamente demonstraram sua potência em outros domínios. Seus mecanismos de autoatenção, que

⁵ O Gerador tenta criar dados falsos (ex: uma imagem sintética de um tumor), enquanto o Discriminador tenta diferenciar os dados falsos do Gerador dos dados reais. Essa competição força o Gerador a se tornar cada vez melhor na criação de dados realistas (GOODFELLOW, Ian *et al.*, 2014).

permitem ao modelo ponderar a importância de diferentes partes da entrada de dados, provaram ser eficazes também na análise de imagens médicas (com os *Vision Transformers - ViTs*) e na integração de dados multimodais. Essa versatilidade pavimentou o caminho para os *LLMs* (DOSOVITSKIY *et al.*, 2021).

Essa revolução do *DL*, no entanto, não ocorreu no vácuo. Ela foi profundamente dependente da evolução da infraestrutura de dados e do poder computacional. A transição de prontuários em papel para Prontuários Eletrônicos do Paciente (PEPs), embora gradual e enfrentando desafios de adoção e padronização em muitos países, começou a criar os grandes repositórios de dados clínicos longitudinais necessários para treinar modelos de *ML* e *DL* (PRICE II, 2019). O desenvolvimento de padrões de dados em saúde – como o *HL7 (Health Level Seven)*, voltado à troca de informações clínicas e administrativas (HL7 INTERNATIONAL, 2023), e o *DICOM (Digital Imaging and Communications in Medicine)*, direcionado a imagens médicas (NEMA, 2024) – foi crucial para viabilizar, ainda que de modo parcial, a interoperabilidade entre esses dados. O *SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms)*, como uma terminologia clínica global abrangente, buscou fornecer uma linguagem comum para registros de saúde, facilitando a agregação e análise de dados (SNOMED INTERNATIONAL, 2023). Apesar desses avanços, a ausência de interoperabilidade técnica e semântica plena entre diferentes PEPs e a persistente fragmentação em silos institucionais continuam a dificultar a formação de conjuntos de dados realmente representativos e em larga escala. Nesse contexto, a União Europeia instituiu o Espaço Europeu de Dados de Saúde (*EHDS*) para reduzir a fragmentação, promover o reuso seguro de dados e fortalecer a continuidade do cuidado (UNIÃO EUROPEIA, 2024). Portugal, inclusive, discutiu, no *White Paper “Inteligência Artificial na Saúde em Portugal: Regulamentação, Impactos e Perspetivas de Futuro”*, como este Regulamento europeu visa superar esses desafios históricos de fragmentação (PORTUGAL, 2025).

O armazenamento e processamento de imagens médicas foram revolucionados pelos sistemas *PACS (Picture Archiving and Communication System)*, que permitiram o arquivamento digital e o acesso rápido a Radiografias, Tomografias Computadorizadas, Ressonâncias Magnéticas etc. (CHOPLIN; BOEHME *et al.*, 1992). Grandes iniciativas de pesquisa, como o *The Cancer Genome Atlas (TCGA)* e o *UK Biobank* (BLAY *et al.*, 2020), disponibilizaram para a comunidade científica conjuntos de dados multimodais (genômicos, de imagem, clínicos) de dezenas ou centenas de milhares de indivíduos, impulsionando descobertas e o desenvolvimento de novos modelos de IA (WORLD ECONOMIC FORUM, 2024).

A capacidade de treinar redes neurais profundas com milhões de parâmetros só se tornou prática com os avanços no poder computacional. O uso de *GPUs* (*Graphics Processing Units*), que oferecem paralelização massiva de cálculos, acelerou drasticamente os tempos de treinamento. A disponibilidade de serviços de computação em nuvem democratizou o acesso a essa infraestrutura, permitindo que pesquisadores e *startups*, mesmo sem grandes recursos próprios, pudessem treinar modelos complexos (MARIN, 2024).

A par da evolução dos algoritmos e da infraestrutura, a forma de avaliar e comparar os sistemas de IA médica também progrediu. Inicialmente, muitos estudos apresentavam resultados em conjuntos de dados proprietários e pequenos, com métricas de avaliação variadas, dificultando a comparação e a replicação. A comunidade científica passou a exigir maior rigor, com a ênfase na validação externa (testar o modelo em dados de populações ou instituições diferentes daquelas usadas no treinamento), no uso de métricas padronizadas (precisão, sensibilidade, especificidade, *F1-score*, curva de precisão-*recall*) e na publicação de *datasets* de *benchmark* públicos. Desafios como o *ILSVRC - ImageNet Large Scale Visual Recognition Challenge* (“Desafio de reconhecimento visual em grande escala da ImageNet”), embora não específico da medicina, impulsionaram o desenvolvimento de *CNNs*. Na área médica, iniciativas como o *ISIC Challenge* para classificação de lesões de pele (HAENSSLE *et al.*, 2018), o *LUNA16 (LUng Nodule Analysis)* para detecção de nódulos pulmonares, e os desafios propostos pela *RSNA (Radiological Society of North America)* em diversas áreas da radiologia, forneceram plataformas para comparar algoritmos de forma transparente e estimular a inovação (NAN *et al.*, 2023). O relatório *European AI Standards – Technical Standardization and Implementation Challenges under the EU AI Act* (KILIAN *et al.*, 2025) aborda a importância crescente da padronização técnica e das métricas de avaliação sob a ótica do *AI Act* europeu.

2.1.1.4 IA Generativa, Grandes Modelos de Linguagem (LLMs) e as Novas Fronteiras

A mais recente e, possivelmente, mais disruptiva onda na evolução da IA é protagonizada pela IA generativa e, em especial, pelos Grandes Modelos de Linguagem (*LLMs - Large Language Models*). Esses modelos, como o GPT-3 e seus sucessores (*e.g.*, GPT-5, Gemini), são redes neurais do tipo *Transformer*, pré-treinadas em quantidades massivas de dados textuais e, cada vez mais, multimodais (BROWN *et al.*, 2020). Eles demonstram uma capacidade sem precedentes de compreender, gerar, resumir e traduzir linguagem humana com notável fluência e coerência, além de realizarem tarefas de raciocínio e resolução de problemas em domínios

para os quais não foram especificamente treinados (*zero-shot* ou *few-shot learning*) – (ZHAO *et al.*, 2023).

Na área da saúde, o potencial dos *LLMs* está sendo intensamente explorado. Aplicações emergentes incluem: (i) *Suporte à Documentação Clínica* - Geração automática de resumos de prontuários, redação de notas de evolução, e elaboração de laudos preliminares a partir de achados de exames (THIRUNAVUKARASU *et al.*, 2023); (ii) *Auxílio à Decisão Clínica e Diagnóstico Diferencial* - Capacidade de processar descrições de casos clínicos complexos, levantar hipóteses diagnósticas e sugerir exames complementares com base na literatura médica (THIRUNAVUKARASU *et al.*, 2023). Pesquisas recentes exploram sistemas de conversação multiagente, nos quais vários *LLMs* simulam uma junta médica para debater casos e aprimorar a acurácia diagnóstica, especialmente para doenças raras (KIM *et al.*, 2024); (iii) *Revisão e Síntese de Literatura Médica* - *LLMs* podem analisar e sintetizar grandes volumes de publicações científicas, auxiliando pesquisadores e clínicos a se manterem atualizados. Essa capacidade tem o potencial de transformar a prática da medicina baseada em evidências, embora exija validação cuidadosa (SALLAM, 2023); (iv) *Interação com Pacientes* - Desenvolvimento de *chatbots* mais sofisticados para triagem de sintomas, fornecimento de informações de saúde, e suporte ao autocuidado e adesão a tratamentos (XIE *et al.*, 2025); e (v) *Educação Médica* - Criação de ferramentas interativas para treinamento de estudantes e profissionais de saúde, permitindo a simulação de casos clínicos e cenários de tomada de decisão (THIRUNAVUKARASU *et al.*, 2023).

O relatório do World Economic Forum (2024) sobre ensaios clínicos inteligentes (*Intelligent Clinical Trials*) já projeta um impacto significativo da IA generativa em todas as fases do desenvolvimento terapêutico, desde a identificação de alvos e o desenho de moléculas até o recrutamento de pacientes e a análise de dados de ensaios. A capacidade dos *LLMs* de integrar e processar dados multimodais – combinando informações textuais de prontuários com dados de imagem, laboratoriais e genômicos – é vista como um passo crucial em direção a uma IA médica mais “generalista” e capaz de um raciocínio clínico mais próximo da complexidade da prática real (BRANDÃO, 2024).

A influência dos *LLMs* já se faz sentir de forma concreta, como indicado no relatório do Google sobre seu impacto na saúde (GOOGLE, 2025), que descreve como modelos como o *Med-PaLM 2* estão sendo avaliados para responder a questões médicas com alta qualidade e segurança, e como a IA generativa está sendo usada para sumarizar informações para médicos e pacientes. O potencial para acelerar a descoberta de medicamentos e otimizar ensaios clínicos é outro campo promissor, como detalhado no *white paper* do World Economic Forum (2024),

que aponta para a capacidade da IA generativa de auxiliar no desenho de protocolos, na identificação de biomarcadores e na análise de dados complexos de ensaios.

A pesquisa de KIM *et al.* (2024) sobre sistemas de “agentes de tomada de decisões médicas” (*MDAgents - Medical Decision-making Agents*), para diagnóstico médico é um exemplo de como os *LLMs* estão sendo levados além das interações de um único agente, buscando replicar a colaboração e o debate que ocorrem em equipes médicas multidisciplinares para resolver casos complexos. Essa abordagem “multiagente” é uma das tendências que visa mitigar algumas limitações dos *LLMs* individuais, como a propensão a erros factuais ou “alucinações”. Similarmente, o trabalho de GENOVESE *et al.* (2025) explora como os *LLMs* podem ser ferramentas poderosas para transformar a prática da medicina baseada em evidências, automatizando parte do processo de revisão sistemática e síntese de conhecimento científico, mas também alerta para a necessidade de validação e curadoria humana desses resultados.

Finalmente, ainda que formulada antes da recente explosão dos modelos de IA generativa, a primeira regra proposta por Frank Pasquale (2020), em *New Laws of Robotics*, nunca se mostrou tão pertinente. Também citado por TOPOL (2019), Pasquale ressalta a importância de preservar a expertise humana em face da automação, um tema que já ecoava nas discussões sobre o papel dos primeiros sistemas de IA, estabelecendo que estes sistemas devem complementar os profissionais de saúde, e não substituí-los (NEGRI, 2021). Conforme argumentam especialistas franceses, “*the ‘mechanical’ medical doctor (MD) is certainly not for tomorrow and certainly not desirable. The doctor of the future should be ‘augmented’, better equipped, and well informed to prevent, analyze, decide, and treat disease with empathy and the human touch. The aim will be to improve diagnoses, observations, therapeutic choices, and outcomes*”⁶ – ou seja, os médicos do futuro devem ser apoiados pela IA, sem perder de vista a dimensão humana do cuidado (NORDLINGER; VILLANI; RUS, 2020).

2.1.2 Evolução das Neurotecnologias em Aplicações Médicas:

No campo nas neurotecnologias, a crescente evolução ocorrida – especialmente nos últimos anos – tem transformado o diagnóstico e tratamento de doenças neurológicas. Desde os primeiros registros elétricos do cérebro humano (EEG - eletroencefalograma) até interfaces

⁶ Em uma tradução livre seria: o médico “mecânico” certamente não é para o futuro e certamente não é desejável. O médico do futuro deve ser “aumentado”, melhor equipado e bem informado para prevenir, analisar, decidir e tratar doenças com empatia e o toque humano. O objetivo será melhorar os diagnósticos, as observações, as escolhas terapêuticas e os resultados.

cérebro-computador (BCI - *Brain-Computer Interface*) alimentadas por inteligência artificial, observamos uma progressão impressionante de ferramentas exclusivamente aplicadas à medicina.

2.1.2.1 *Primeiros Registros Elétricos do Cérebro Humano (EEG)*

Os primórdios da neurotecnologia médica remontam à descoberta da atividade elétrica cerebral. No final do século XIX, estudos iniciais em animais, realizados por Luigi Galvani, já indicavam correntes elétricas no cérebro (PICCOLINO, 2006); porém, foi apenas em 1924 que o psiquiatra alemão Hans Berger realizou o primeiro registro de eletroencefalografia (EEG) em um ser humano. Berger, utilizando eletrodos no couro cabeludo e um galvanômetro, identificou os ritmos cerebrais, incluindo as ondas alfa de aproximadamente 10 Hz. Cinco anos depois, ele publicou seus achados em 1929, cunhando o termo “elektroenzephalogramm” e descrevendo as ondas “alpha” e “beta” (TUDOR *et al.*, 2005). Esse avanço revolucionou a neurologia ao fornecer uma janela não-invasiva para a atividade cerebral, algo impensável numa era em que os diagnósticos neurológicos dependiam de procedimentos arriscados como punção lombar, pneumoencefalografia ou ventriculografia.

O EEG (eletroencefalograma) rapidamente mostrou seu valor clínico, especialmente no estudo da epilepsia. Já na década de 1930, neurologistas como Frederic e Erna Gibbs nos EUA identificaram os padrões de *spike and wave* (“pico e onda”) associados a crises epiléticas. Assim, a eletroencefalografia tornou-se uma ferramenta fundamental para diagnosticar distúrbios convulsivos e outras condições, permitindo distinguir fenômenos fisiológicos de patológicos no cérebro vivo (TUDOR *et al.*, 2005). Nas décadas seguintes, aperfeiçoamentos técnicos (*e.g.* amplificadores valvulados, múltiplos canais de registro) tornaram o EEG mais acessível. Até cerca de 1970, quando outras técnicas de imagem ainda não existiam, o EEG reinou como principal método de investigação funcional do cérebro. Mesmo hoje, quase um século após Berger, o EEG permanece relevante no diagnóstico de epilepsias, no monitoramento do coma e de doenças degenerativas, embora sua importância tenha diminuído relativamente após o advento das modernas técnicas de neuroimagem (TUDOR *et al.*, 2005).

2.1.2.2 *Surgimento da Neuroimagem Clínica (Tomografia e Ressonância Magnética)*

Enquanto o EEG fornecia informações temporais sobre a atividade cerebral, a medicina carecia de métodos para visualizar estruturalmente o cérebro de pacientes vivos sem cirurgia.

Raymond Schulz (2021) relata que nas primeiras décadas do século XX, recorreram-se a técnicas invasivas: em 1918, a pneumoencefalografia de Dandy introduzia ar no espaço liquorístico para tornar o cérebro visível em radiografias, um procedimento doloroso e arriscado. Somente em 1971 a situação mudou dramaticamente com a invenção da tomografia computadorizada (TC). O engenheiro britânico Godfrey Hounsfield, trabalhando na EMI, construiu o primeiro scanner de TC clínico para encéfalo e, em 1º de outubro de 1971, obteve as primeiras imagens tomográficas do cérebro de um paciente (SCHULZ *et al.*, 2021). A TC do crânio provou-se revolucionária: pela primeira vez foi possível obter cortes transversais detalhados do cérebro, não invasivamente, revelando lesões internas com precisão milimétrica.

O impacto foi imediato e profundo. Já em 1972, Hounsfield e outros cientistas apresentavam os resultados a neurocirurgiões e radiologistas, que ficaram impressionados com o nível de detalhe – comparável, diziam, às imagens de autópsia do cérebro (TUDOR *et al.*, 2005). Em poucos anos, tomógrafos espalharam-se pelo mundo, transformando o diagnóstico neurológico. A Nobel Assembly at Karolinska Institutet (1979) descreveu a TC como “*um avanço revolucionário na diagnosis radiológica de doenças neurológicas*”, reduzindo drasticamente a necessidade de exames anteriores como a encefalografia com contraste pneumático. De fato, a TC conseguia demonstrar diretamente tumores, hemorragias, hidrocefalia e outras patologias cerebrais, estimando posição, tamanho e natureza das lesões com precisão inédita. Além disso, por não causar dor ou grande desconforto, podia ser utilizada mesmo em pacientes graves e monitorar a evolução de tratamentos. Em reconhecimento a essa invenção, Hounsfield (juntamente com Allan Cormack, pelos fundamentos matemáticos) recebeu o Prêmio Nobel de Medicina de 1979 (Nobel Assembly at Karolinska Institutet, 1979).

Quase simultaneamente, outro salto tecnológico ampliou as possibilidades de visualização do cérebro: a Ressonância Magnética (RM). Diferentemente da TC (baseada em raios X), a RM surgiu do fenômeno físico da Ressonância Magnética Nuclear, aplicado à obtenção de imagens anatômicas. Em 1973, o químico Paul Lauterbur demonstrou a geração de imagens bidimensionais usando gradientes de campo magnético (LAUTERBUR, 1973). Nos anos seguintes, Peter Mansfield aperfeiçoou técnicas de análise de sinal e varredura rápida, tornando viável a RM para uso médico (MANSFIELD *et al.*, 1977). Durante a década de 1980, os primeiros equipamentos de RM foram instalados em hospitais; já em 2003, Lauterbur e Mansfield seriam laureados com o Nobel de Medicina por essa conquista. A RM revelou-se particularmente valiosa por gerar imagens de alta resolução dos tecidos moles sem usar radiação ionizante – ideal para o sistema nervoso central. Por volta de 2002, mais de 22 mil aparelhos de RM estavam em uso e realizavam 60 milhões de exames anuais mundialmente (Nobel

Assembly at Karolinska Institutet, 2003). A técnica não apenas substituiu uma série de exames invasivos, reduzindo riscos ao paciente, como também elevou a acurácia diagnóstica em muitas doenças (Nobel Assembly at Karolinska Institutet, 2003). Lesões cerebrais que antes exigiam procedimentos agressivos passaram a ser diagnosticadas por RM de forma rápida e indolor, beneficiando imensamente a neurologia clínica.

Outro desenvolvimento importante em neuroimagem médica foi a Tomografia por Emissão de Pósitrons (PET), que adicionou uma dimensão funcional/metabólica às imagens cerebrais. Em meados dos anos 1970, pesquisadores como Michel Ter-Pogossian, Michael Phelps e Edward Hoffman desenvolveram os primeiros scanners PET dedicados ao cérebro humano (PORTNOW *et al.*, 2013). Em 1975, Phelps e colegas introduziram o modelo PETT (Tomografia por Emissão de Pósitrons Transaxial), capaz de detectar os fótons gerados por radiofármacos emissores de pósitrons no corpo (PHELPS *et al.*, 1975). Em 1976, a fluorodesoxiglicose marcada com flúor-18 (FDG)⁷ foi sintetizada, permitindo mapear o metabolismo da glicose no cérebro humano. Logo, estudos pioneiros demonstraram que o PET poderia identificar áreas de alta ou baixa atividade metabólica cerebral, correlacionando-se com funções ou patologias. Por exemplo, no Hospital da Universidade da Pensilvânia, Reivich e Alavi usaram FDG-PET em 1976 para visualizar o consumo de glicose no cérebro de voluntários, inaugurando uma nova era na pesquisa de neurociências e neurologia clínica (ALAVI *et al.*, 2002). Paralelamente, o grupo de Ter-Pogossian em Washington University consolidou o primeiro PET plenamente bem-sucedido para uso humano, tornando possível obter imagens funcionais não-invasivas do cérebro (ALAVI *et al.*, 2002).

Nos anos 1980, o PET difundiu-se em centros de pesquisa e, posteriormente, em grandes hospitais, revelando-se valioso para diagnosticar tumores cerebrais, acidentes vasculares e demências, ao mostrar anomalias no metabolismo regional da glicose. Estudos comprovaram, por exemplo, que o PET com FDG (Fluordesoxiglicose) diferenciava tecido tumoral ativo de necrose, por radioterapia, melhor do que a TC ou RM convencionais (ALAVI *et al.*, 2002). Assim, TC, RM e PET – cada qual a seu modo – consolidaram-se como pilares da neuroimagem médica: a TC e a RM oferecendo detalhamento anatômico de lesões e a PET revelando aspectos funcionais/químicos do cérebro vivo. Essas tecnologias revolucionaram a neurologia,

⁷ É um radiofármaco usado na tomografia por emissão de pósitrons (PET), uma modalidade de imagiologia médica. Quimicamente, trata-se de 2-desoxi-2-(¹⁸F)fluoro-D-glicose, um análogo da glicose, com o isótopo radioativo emissor de pósitrons flúor-18 em substituição do grupo hidroxila normal na posição 2 da molécula de glicose. Após a injeção de ¹⁸F-FDG no paciente, uma máquina de PET pode formar imagens da distribuição da FDG pelo corpo. Estas imagens podem então ser avaliadas por um médico de medicina nuclear ou radiologista para diagnóstico de várias doenças. Disponível em: <https://pt.wikipedia.org/wiki/Fluordesoxiglicose>.

neurocirurgia e psiquiatria, possibilitando diagnósticos antes inconcebíveis e guiando intervenções com maior segurança.

2.1.2.3 Neuroestimulação e Neuropróteses: Restaurando Funções Perdidas

Além de “fotografar” o cérebro, a tecnologia também passou a intervir diretamente nele para tratar doenças. Uma das estratégias é a neuroestimulação elétrica – enviar pulsos controlados ao sistema nervoso para modular sua atividade. Essa ideia remonta a experimentos de meados do século XX (por exemplo, cirurgias como Wilder Penfield já estimulavam o córtex exposto para mapear funções na década de 1940) – (PENFIELD *et al.*, 1950). Entretanto, o uso terapêutico crônico de estimulação cerebral só ganhou destaque décadas depois. Um marco fundamental ocorreu em 1987, quando o neurocirurgião francês Alim-Louis Benabid descobriu acidentalmente que estimular certas estruturas profundas do cérebro em alta frequência podia suprimir os tremores da doença de Parkinson (BENABID *et al.*, 1987). Durante uma cirurgia, Benabid inseriu um eletrodo no tálamo de um paciente parkinsoniano; ao elevar a frequência de 50 Hz para 100 Hz, observou que o tremor intenso do paciente cessou quase instantaneamente (BENABID, 2016). Essa descoberta fortuita inaugurou a era da Estimulação Cerebral Profunda (*DBS - Deep Brain Stimulation*). Diferentemente da lesão ablativa (talidotomia) usada até então, a *DBS* controla sintomas motores de Parkinson sem destruir tecido, modulando circuitos neurais de forma reversível. A técnica rapidamente evoluiu de experimentação para prática clínica: nos anos 1990, Benabid e colaboradores implantaram eletrodos de estímulo permanente em núcleos subtalâmicos e talâmicos de pacientes com Parkinson e outros distúrbios do movimento, obtendo melhora significativa em tremores, rigidez e bradicinesia (BENABID *et al.*, 1987). Em 1997, a *DBS* de alta frequência recebeu aprovação regulatória para Parkinson e tremor essencial, tornando-se terapia consagrada. Hoje, dezenas de milhares de pacientes no mundo portam esses “*marcapassos cerebrais*”, que lhes devolvem controle motor e qualidade de vida (CHEN, 2019). A *DBS* também está sendo aplicada em epilepsia refratária, distonias e até em transtornos psiquiátricos como transtorno obsessivo-compulsivo, com resultados promissores.

Outra técnica de neuroestimulação destacada é a Estimulação Magnética Transcraniana (EMT), um método *não invasivo* introduzido em 1985 por Anthony Barker e colegas (1985). A EMT utiliza campos magnéticos breves para induzir correntes elétricas focais no cérebro. No primeiro experimento, Barker aplicou pulsos magnéticos sobre o córtex motor de voluntários, obtendo contrações musculares no membro correspondente (BARKER *et al.*, 1985). Essa

demonstração provou ser possível modular a atividade cortical sem cirurgia. Nos anos seguintes, a EMT foi refinada e acabou aprovada para tratamento de depressão maior resistente (usando estimulação repetitiva do córtex pré-frontal) e para reabilitação de sequelas motoras pós-AVC, entre outras indicações. Embora menos potente e duradoura que a DBS, a EMT representa um importante avanço por permitir neuromodulação sem implante, em sessões ambulatoriais. Tecnologias aparentadas, como a estimulação transcraniana por corrente contínua (*tDCS - Transcranial Direct Current Stimulation*), também têm sido exploradas em reabilitação motora e cognitiva (SILVA *et al.*, 2018).

Em paralelo à estimulação elétrica/magnética para modular funções existentes, emergiu o campo das neuropróteses implantáveis, destinado a restaurar funções perdidas do sistema nervoso. O pioneiro nessa área foi o cirurgião William House, que realizou implantes cocleares monoeletrodo em pacientes surdos no final dos anos 1960 (ESHLAGHI *et al.*, 2012). Esses dispositivos estimulavam diretamente o nervo auditivo, conferindo alguma percepção de som. Posteriormente, em 1978, o médico e pesquisador australiano Graeme Clark implantou com sucesso o primeiro implante coclear multicanal em um paciente adulto⁸. Com múltiplos eletrodos inseridos na cóclea, codificando diferentes frequências sonoras, essa prótese auditiva permitiu ao paciente distinguir palavras – prova concreta de que era possível “hackear” o sistema nervoso sensorial e devolver audição a surdos profundos. Nas décadas seguintes, implantes cocleares evoluíram para um tratamento consagrado: cerca de 400 mil pessoas no mundo já receberam essa prótese, muitas das quais crianças que, graças a ela, puderam desenvolver linguagem oral. O sucesso do implante coclear inspirou esforços em outras frentes sensoriais, como próteses de retina para cegueira (implantes epirretinianos já permitem percepção visual rudimentar em casos de retinite pigmentosa) e interfaces táteis para amputados (ESHLAGHI *et al.*, 2012).

No domínio motor, as interfaces cérebro-computador invasivas emergiram como promissoras neurotecnologias para reabilitação de paralisias. O conceito de conectar diretamente o cérebro a dispositivos externos foi proposto academicamente ainda em 1973 pelo cientista da computação Jacques Vidal, da UCLA (Universidade da Califórnia, Los Angeles), que cunhou o termo “*BCI - Brain-Computer Interface*” e realizou os primeiros estudos experimentais utilizando sinais de EEG para mover um cursor na tela (VIDAL, 1973). No entanto, por décadas as BCIs permaneceram restritas a laboratórios, progredindo lentamente de demonstrações com voluntários saudáveis para aplicações em pacientes. Um salto notável

⁸ Graeme Clark (doctor). Disponível em: [https://en.wikipedia.org/wiki/Graeme_Clark_\(doctor\)](https://en.wikipedia.org/wiki/Graeme_Clark_(doctor)). Acesso em: 3 de julho de 2024.

ocorreu no início do século XXI com pesquisas implantando microeletrodos diretamente no córtex motor de indivíduos tetraplégicos. Em 2006, a equipe *BrainGate* (*Brown University*) publicou resultados impressionantes: pacientes com lesão medular alta, portando um sensor de 96 microeletrodos no córtex motor, conseguiram mover um cursor de computador apenas pensando no movimento, chegando a abrir e fechar uma mão mecânica virtual (HOCHBERG *et al.*, 2006). A prova de princípio de que sinais corticais podiam controlar dispositivos externos reacendeu o campo. Nos anos seguintes, avanços em decodificação neural possibilitaram comandos cada vez mais complexos.

Um marco histórico foi atingido em 2012, quando dois participantes tetraplégicos conseguiram usar uma BCI implantada para controlar braços robóticos reais em movimentos 3D úteis. No estudo liderado por Leigh Hochberg (2012), os pacientes mentalizavam mover seus próprios braços; os sinais neuronais captados no córtex motor eram decodificados em tempo real para manipular um braço mecânico ao lado do leito. Com isso, eles puderam realizar tarefas funcionais, como alcançar objetos, apertar a mão de alguém e até beber de uma garrafa usando o braço robótico – algo que não faziam sozinhos havia muitos anos (HOCHBERG *et al.*, 2012). Uma participante de 58 anos, tetraplégica há 14 anos, usou o braço mecânico para levar uma garrafa à boca e tomar café através de um canudo, relatando a alegria de realizar, de forma independente, um ato tão simples (HOCHBERG *et al.*, 2012). Essa demonstração, amplamente divulgada, evidenciou o potencial das *BCIs* motoras de restaurar movimento e aumentar a autonomia de pessoas com paralisia.

O caso acima envolveu um membro robótico externo, mas avanços recentes apontam também para a reconexão do próprio corpo de pacientes paralisados. Em 2023, cientistas suíços implantaram uma interface capaz de retransmitir os sinais do cérebro diretamente para a medula espinhal abaixo de uma lesão, reestabelecendo a comunicação neural interrompida. Um paciente com tetraplegia crônica pôde recuperar controle sobre as pernas e voltou a andar com auxílio, graças a essa “ponte digital” entre cérebro e coluna (LORACH *et al.*, 2023). Neste experimento, eletrodos implantados na região motora cortical captavam a intenção de movimento das pernas; esses sinais eram transmitidos sem fio a um neuroestimulador inserido na medula lombar, que ativava músculos das pernas em tempo real. O resultado, após treinamento, foi que o paciente recuperou a capacidade de levantar-se, dar passos e até subir rampas, controlando voluntariamente os movimentos que antes estavam bloqueados pela lesão medular (LORACH *et al.*, 2023). Trata-se de um feito notável, combinando *BCI*, estimulação epidural e algoritmos inteligentes, que inaugura uma nova era na reabilitação de lesões medulares – integrando cérebro, computador e o próprio corpo para devolver funções motoras.

2.1.2.4 Interfaces Cérebro-Computador Modernas e IA na Neurotecnologia

À medida que as interfaces cérebro-computador evoluíram, a inteligência artificial (IA) tornou-se uma aliada poderosa para decodificar os complexos padrões da atividade neural. Algoritmos de aprendizado de máquina – especialmente redes neurais profundas – têm sido aplicados para interpretar sinais cerebrais com precisão sem precedentes, ampliando o leque de aplicações médicas das neurotecnologias (DURAIVEL *et al.*, 2023). Um dos focos mais nobres é a comunicação para pacientes com síndrome do encarceramento ou outras condições que os impedem de falar ou escrever. Pesquisas recentes demonstraram que é possível traduzir diretamente os sinais do cérebro em palavras e frases, devolvendo voz a quem a perdeu (BARNOVA *et al.*, 2023).

Em 2023, uma colaboração entre a Universidade da Califórnia em São Francisco (UCSF) e a Universidade de Berkeley conseguiu, pela primeira vez, sintetizar fala e expressões faciais a partir de sinais cerebrais de uma paciente com paralisia completa da musculatura da fala. Nessa pesquisa, uma mulher com paralisia de membros e aparato vocal (devido a um AVC de tronco cerebral) recebeu um implante de 253 eletrodos sobre a área da fala no córtex. Utilizando métodos de IA avançados, os padrões neurais da paciente – enquanto ela imaginava falar – foram decodificados em tempo real e convertidos em uma voz sintetizada, reproduzida através de um avatar digital animado que replicava seus movimentos faciais (MARKS; KURTZMAN, 2023). Foi a primeira vez que uma interface cérebro-computador gerou fala articulada completa com entonação e expressão facial, em tempo quase real (cerca de 80 palavras por minuto, aproximando-se da conversação normal). O estudo, publicado na revista *Nature* em agosto de 2023, representa um marco de convergência entre neurotecnologia e IA: redes neurais artificiais foram treinadas para reconhecer os padrões únicos dos sinais cerebrais da paciente correspondentes a fonemas, permitindo reconstruir frases inteiras que refletem o que ela quis dizer. Esse avanço oferece esperança concreta de, num futuro próximo, dispormos de sistemas comerciais aprovados que deem voz a pacientes incapazes de falar – por exemplo, vítimas de AVC, ELA (doença do neurônio motor) ou outras condições neurológicas (MARKS; KURTZMAN, 2023; METZGER *et al.* 2024).

De forma semelhante, aplicações de IA têm aprimorado *BCIs* motoras. Algoritmos de aprendizado profundo permitem decodificar sinais do córtex motor com mais acurácia e velocidade, melhorando o controle de próteses robóticas ou estimuladores musculares. Em 2021, pesquisadores demonstraram que um modelo de rede neural treinado em sinais

intracorticais pôde interpretar até mesmo a tentativa de escrita manual de um paciente tetraplégico, convertendo a atividade cerebral pretendida em texto na tela com velocidade de 90 caracteres por minuto (WILLETT *et al.*, 2021). Esses sistemas baseados em IA capturam padrões sutis nos dados neurais – impossíveis de serem extraídos por algoritmos lineares tradicionais – e assim ampliam a capacidade de um BCI transmitir informações complexas. Com tal abordagem, um indivíduo completamente paralisado conseguiu comunicar frases inteiras simplesmente tentando escrever letra por letra mentalmente, enquanto o algoritmo decodificador traduzia essas intenções em tempo real.

Outra frente emergente é o uso de IA para interpretar sinais cerebrais não-invasivos de modo inovador. Em 2023, uma equipe da Universidade do Texas desenvolveu um “*decodificador semântico*” capaz de transformar registros de fMRI (ressonância magnética funcional) do cérebro em uma narrativa de texto correspondente ao pensamento do indivíduo (TANG *et al.*, 2023). Nesse estudo, voluntários ouviram histórias ou imaginaram contar uma história enquanto tinham seu cérebro escaneado; um modelo de IA do tipo *transformer* (semelhante ao ChatGPT) inferiu, a partir dos padrões de ativação cortical captados, o conteúdo semântico das frases que estavam sendo ouvidas ou pensadas. O decodificador não gerava cada palavra exata, mas captava o sentido geral das sentenças aproximadamente 50% do tempo, um feito notável considerando tratar-se de um método não invasivo (fMRI) e de linguagem contínua (TANG *et al.*, 2023). Embora ainda em estágio inicial, essa linha de pesquisa indica que, com IA e *big data*, poderemos talvez “ler a mente” de forma limitada para fins clínicos – por exemplo, compreender pensamentos de pacientes não responsivos ou monitorar estados mentais internos relevantes (sempre com os devidos cuidados éticos) (BERTONI; IENCA, 2024; MOLNÁR-GÁBOR; IENCA; FINS *et al.*, 2021).

Por fim, vale mencionar que a IA também otimiza a análise de sinais para diagnóstico e monitoramento. Algoritmos de *deep learning* já são empregados para identificar padrões sutis em EEG (como predição de crises epiléticas iminentes), para segmentar automaticamente lesões em imagens de RM, ou para classificar estados de consciência em pacientes com trauma craniano a partir de múltiplos sinais (ROY *et al.*, 2019). Essas aplicações, embora menos vistosas que as *BCIs*, fazem parte do espectro das neurotecnologias médicas contemporâneas impulsionadas por IA – em todos os casos visando melhorar o cuidado de pacientes neurológicos através de diagnósticos mais precoces e precisos ou de novas possibilidades terapêuticas.

Como se observa, há um contínuo avanço tecnológico e uma crescente sofisticação das ferramentas de inteligência artificial e de neurotecnologias aplicadas ao diagnóstico médico.

Desde as primeiras iniciativas baseadas em sistemas especialistas até os avançados algoritmos de aprendizagem profunda atuais, cada etapa representou um progresso significativo que transformou gradualmente a prática diagnóstica. De fato, já na segunda metade do século XX surgiram pesquisas pioneiras em IA médica, enquanto as neurotecnologias passaram de registros eletroencefalográficos rudimentares a sofisticadas interfaces cérebro-computador, ampliando a precisão na detecção de padrões neurais associados a doenças.

Esses avanços progressivos evidenciam não apenas marcos isolados, mas também a maturação consistente dessas áreas de estudo. Essa trajetória histórica ilustra como o conhecimento acumulado serviu de alicerce para as aplicações contemporâneas que serão abordadas a seguir. Em síntese, os desenvolvimentos retratados preparam o terreno para examinarmos, na próxima seção (item 2.2), de que forma IA e neurotecnologias têm sido atualmente empregadas no diagnóstico e tratamento médicos.

2.2 APLICAÇÕES TECNOLÓGICAS EM DIAGNÓSTICOS MÉDICOS: POTENCIALIDADES E RISCOS

A medicina contemporânea – como demonstrado na seção anterior (item 2.1) – encontra-se no epicentro de uma transformação paradigmática, impulsionada pela convergência entre ciência de dados, inteligência artificial (IA) e neurotecnologias. Trata-se de um movimento que ultrapassa o mero aprimoramento incremental de práticas já existentes, projetando uma reconfiguração profunda dos processos diagnósticos, terapêuticos e até mesmo da compreensão conceitual da saúde e da doença (TELLES; ALCÂNTARA, 2024; TOPOL, 2019). A integração dessas tecnologias possibilita o processamento de vastos conjuntos de dados clínicos e neurofisiológicos, identificando padrões e correlações invisíveis à análise humana, e inaugurando uma era de medicina mais preditiva, personalizada e de alta precisão (OLIVEIRA; LIMA, 2025).

O potencial dessa convergência já se mostra concreto em áreas como a oncologia, a cardiologia e a neurologia, nas quais sistemas de IA, combinados à interfaces cérebro-computador (BCIs) e outras neurotecnologias, vêm permitindo diagnósticos mais precoces, delineamento de terapias personalizadas e, em alguns casos, a restauração parcial de funções neurológicas antes consideradas irrecuperáveis (YUSTE *et al.*, 2017; YUSTE, 2023). Entretanto, essa mesma capacidade de decifrar, prever e intervir em sistemas biológicos complexos – especialmente no sistema nervoso central – traz à tona dilemas éticos e jurídicos

de magnitude inédita, cujos impactos extrapolam a esfera estrita do Consentimento Livre e Esclarecido do paciente, foco central desta dissertação.

Nesse sentido, ainda que a análise da validade do Consentimento Livre e Esclarecido diante de tecnologias opacas e enviesadas seja o eixo condutor desta pesquisa, este tópico também se dedicará a examinar outras questões jurídicas conexas, como a responsabilidade civil e regulatória por eventuais falhas de sistemas, a proteção da privacidade e da integridade mental, a promoção da equidade no acesso e nos resultados da aplicação dessas tecnologias etc.

O objetivo, portanto, é realizar uma análise crítica e detalhada de casos concretos de uso da IA e das neurotecnologias no período de 2023 a 2025, abrangendo experiências nacionais e internacionais. A partir de uma seleção expandida de casos emblemáticos em cada domínio, buscar-se-á avaliar não apenas as potencialidades tecnológicas, mas também os riscos jurídicos e éticos associados, com especial atenção às implicações para a autodeterminação, a justiça distributiva e a salvaguarda dos direitos fundamentais na era da medicina algorítmica.

2.2.1 A Inteligência Artificial como Ferramenta de Aumento da Capacidade Diagnóstica e Terapêutica

Como demonstrado, a inteligência artificial tem se consolidado como uma força transformadora na medicina diagnóstica, processando dados complexos de exames de imagem, genômica e registros clínicos com uma velocidade e precisão que superam as capacidades humanas (TELLES; ALCÂNTARA, 2024). O número de dispositivos médicos baseados em IA aprovados por agências reguladoras, nacional e internacionais, tem crescido exponencialmente, especialmente na área de radiologia (ALTMAN; BLACK; CHEN *et al.*, 2025).

A tabela a seguir apresenta um panorama dos casos de uso que serão analisados nesta seção, destacando suas potencialidades e os principais desafios jurídicos e éticos que suscitam.

Tabela 1 – Panorama das Aplicações de IA na Medicina (2023-2025)

| Aplicação (Caso de Uso) | Tecnologia Principal | Potencialidades (Clínica e Operacional) | Principais Riscos Jurídicos e Éticos |
|---|---|---|---|
| Deteção de Câncer de Mama (Estudo PRAIM) | Rede Neural Convolucional (CNN) para análise de mamografias, atuando como “Rede de Segurança” e “Triagem Normal”. | Clínica: Aumento de 17,6% na deteção de câncer sem aumentar falsos positivos. Operacional: Potencial de redução de até 56,7% na carga de | Responsabilidade Civil: Difusa e dependente do modo de uso (ato médico vs. fato do produto). A não adoção da tecnologia pode se tornar negligência. Viés Algorítmico: Risco de menor acurácia em |

| | | | |
|--|--|--|--|
| | | trabalho dos radiologistas. | populações etnicamente diversas (validação em coortes locais é necessária). Consentimento Livre e Esclarecido: Dificultado pela opacidade do algoritmo (“caixa-preta”). |
| Patologia Digital (Foundation Models - ProV-GigaPath) | <i>Vision Transformer</i> com “atenção dilatada” para análise de lâminas de patologia inteiras (WSI) em acesso aberto. | Clínica: Melhor predição de subtipos de câncer e biomarcadores ao analisar o microambiente tumoral. Operacional: Acelera drasticamente a pesquisa e o desenvolvimento de novas ferramentas de diagnóstico. | Responsabilidade Civil: Fragmentada devido ao modelo <i>open-source</i> , deslocando o ônus para o “agente implementador”. Governança de Dados: Desafios no uso secundário de dados (LGPD) e na anonimização de WSI. Consentimento Livre e Esclarecido: Paciente deve ser informado sobre o uso de IA, a opacidade e os riscos de viés. |
| Medicina Preditiva em Cardiologia (IA de Oxford) | IA para análise de inflamação perivascular em angiotomografias, prevendo risco cardiovascular em 10 anos. | Clínica: Identifica pacientes de alto risco não detectados por métodos tradicionais, alterando o plano terapêutico em até 45% dos casos e permitindo prevenção personalizada. Operacional: Otimiza a estratificação de risco. | Autonomia do Paciente: Tensão com o “direito de não saber” informações preditivas. Consentimento Livre e Esclarecido: Exige consentimento granular e específico para a análise preditiva. Discriminação Algorítmica: Risco de uso dos escores por seguradoras e empregadores. Equidade: Acesso limitado aprofunda desigualdades em saúde. |
| Genômica Preditiva (AlphaMissense) | IA baseada na arquitetura do AlphaFold para prever a patogenicidade de 71 milhões de variantes genéticas <i>missense</i> . | Clínica: Reclassifica “variantes de significado incerto” (VUS), permitindo a elegibilidade para terapias-alvo (ex.: inibidores de PARP). Operacional: Acelera a interpretação de laudos genéticos. | Insegurança Jurídica: Risco de “reclassificação” de variantes por atualizações do modelo e o “dever de recontato”. Viés Algorítmico: Menor acurácia em populações não europeias devido a bancos de dados genômicos enviesados. Consentimento Livre e Esclarecido: Exige aconselhamento genético robusto sobre a natureza probabilística, vieses e o “direito de não saber”. |
| Triagem de Retinopatia Diabética | IA autônoma para análise de imagens de retina, muitas vezes via <i>smartphone</i> , para triagem em atenção primária. | Clínica: Alta acurácia na detecção, permitindo encaminhamento rápido de casos graves. Operacional: Democratiza o acesso | Responsabilidade Civil: Deslocada do ato médico para o fato do produto (CDC), com responsabilidade objetiva do desenvolvedor e da instituição. Viés |

| | | | |
|------------------------------------|---|--|---|
| | | ao diagnóstico em áreas carentes, otimiza o fluxo de atendimento e previne a cegueira. | Algorítmico: Risco de menor acurácia em populações etnicamente diversas. Consentimento Livre e Esclarecido: Paciente deve ser informado que a análise é por um algoritmo autônomo, com suas limitações e riscos de viés. |
| Triagem de Câncer de Pele | Rede Neural Convolutiva (CNN) para análise de imagens dermatoscópicas, auxiliando no diagnóstico diferencial de lesões. | Clínica: Aumenta a acurácia diagnóstica, especialmente para não especialistas, reduzindo biópsias desnecessárias. Operacional: Acelera a triagem inicial (via apps) e o encaminhamento para dermatologistas. | Viés Algorítmico: Risco crítico de menor performance em peles negras e outros fototipos devido a bancos de dados enviesados. Responsabilidade Civil: Altera o padrão de cuidado, criando dilemas para o médico (seguir ou não a IA). Consentimento Livre e Esclarecido: Deve incluir a informação explícita sobre o risco de viés racial da tecnologia. |
| Detecção e Manejo da Sepsis | IA para análise contínua de dados de prontuários eletrônicos (EHR), gerando alertas precoces de risco de sepsis. | Clínica: Permite intervenção rápida, com potencial de reduzir drasticamente a mortalidade. Operacional: Otimiza a vigilância em UTIs e a gestão de pacientes críticos. | Responsabilidade Civil: Complexa (negligência da equipe vs. falha do produto em caso de falso negativo). Governança de Dados: Dependência da qualidade dos dados do EHR. Consentimento Livre e Esclarecido: Dilema da “vigilância algorítmica contínua” e a impossibilidade de não consentir sem recusar o cuidado. Fadiga de Alerta: Risco de alertas falsos diminuírem a confiança e a eficácia do sistema. |

Fonte: Elaborada pelo autor (2025).

2.2.1.1 Detecção de Câncer de Mama em Mamografias: O Caso do Estudo PRAIM

Panorama Geral e Potencialidades:

O estudo *PRAIM* (*Nationwide real-world implementation of AI for cancer detection in population-based mammography screening*), publicado na prestigiosa revista *Nature Medicine* em 2025, representa uma das mais robustas implementações de IA em um programa de rastreamento populacional em larga escala. Conduzido na Alemanha, o estudo avaliou dados

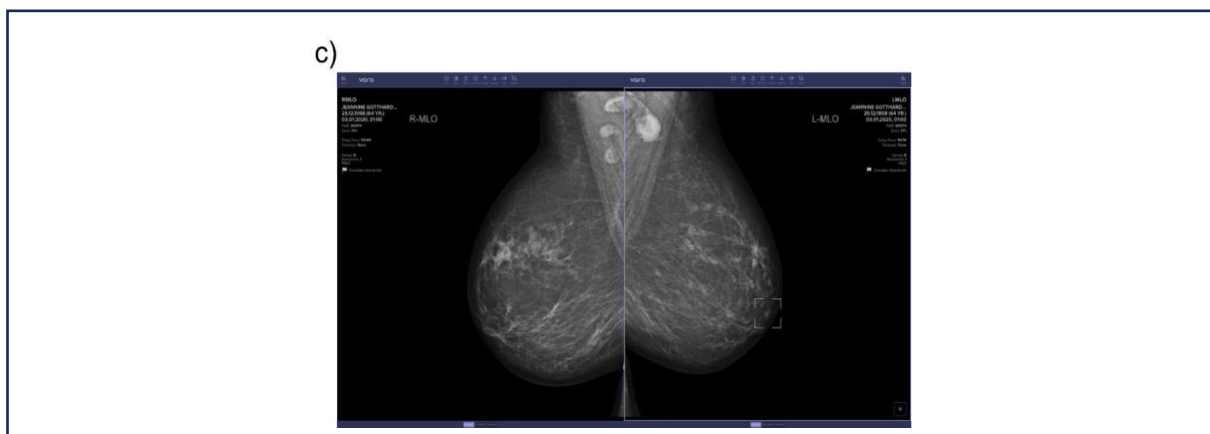
de mais de 460.000 mulheres, comparando o desempenho da leitura de mamografias por dois radiologistas, com o de uma leitura dupla assistida por IA (EISEMANN; BUNK *et al.*, 2025).

A ferramenta de IA funciona como um sofisticado sistema de auxílio à decisão, operando em duas frentes principais. A primeira é a Triagem Normal, na qual o algoritmo pré-classifica exames com altíssima probabilidade de serem normais, otimizando o fluxo de trabalho. A segunda, e talvez mais crucial, é a Rede de Segurança (*Safety Net*), que emite um alerta de segurança quando um radiologista classifica como normal um exame que a IA considera suspeito, destacando a área de interesse (EISEMANN; BUNK *et al.*, 2025).

Os resultados foram notáveis. O uso da IA foi associado a um aumento de 17,6% na taxa de detecção de câncer de mama, identificando 6,7 casos por 1.000 mulheres, em comparação com 5,7 casos no grupo de controle (sem IA). Isso equivale a um caso adicional de câncer detectado para cada mil mulheres examinadas. Fundamentalmente, esse aumento de sensibilidade não veio acompanhado de um aumento na taxa de reconvocação (conhecidos como falsos positivos), que, na verdade, apresentou uma ligeira redução, indicando maior precisão. A funcionalidade de “rede de segurança” foi, sozinha, responsável pela detecção de 204 casos de câncer que, de outra forma, poderiam ter sido perdidos. O potencial de otimização do sistema é igualmente impressionante, com simulações indicando uma possível redução de até 56,7% na carga de trabalho dos radiologistas, se os casos triados como normais pela IA não necessitassem de uma segunda leitura humana (EISEMANN; BUNK *et al.*, 2025).

Figura 1 – Capturas de tela do visualizador com suporte de IA.





Fonte: EISEMANN; BUNK *et al.*, (2025).

Nota: **(a)** A captura de tela mostra uma lista de trabalho com um subconjunto de exames marcados como normais. Os radiologistas também podem optar por ler apenas exames normais ou apenas exames não normais (“potencialmente suspeitos”). Os nomes e datas de nascimento não são reais e foram selecionados aleatoriamente a partir de uma lista de nomes e sobrenomes comuns. **(b)** e **(c)** Quando os radiologistas avaliam um caso como normal (BI-RADS 1 ou 2), mas a rede de segurança é acionada, um alerta é exibido **(b)** e uma região suspeita é destacada no visualizador **(c)**, solicitando que os radiologistas reconsiderem.

Riscos Jurídicos e Éticos:

Apesar do enorme potencial, a implementação clínica desses sistemas de IA suscita complexas questões jurídicas e éticas; isso porque, a sua dupla funcionalidade – ora atuando como “rede de segurança”, ora como ferramentas de “triagem” – cria cenários distintos para a atribuição da responsabilidade civil em caso de erro (FACCHINI NETO; BARBOSA, 2023). No primeiro modo, a IA funciona como consultora, alertando o médico, que mantém autonomia e responsabilidade pela decisão final; se ignorar um alerta pertinente e ocorrer dano, a responsabilidade costuma recair na esfera da culpa profissional (NOGAROLI, 2023). Contudo, no segundo modo – o da “triagem normal” –, o fluxo de trabalho pode ser alterado a ponto de o médico sequer analisar diretamente o exame. Nesse contexto, um diagnóstico falso negativo torna a imputação de responsabilidade muito mais difusa: ela pode recair sobre o hospital, caso implemente um protocolo que confie cegamente na IA; sobre o desenvolvedor do *software*, sob a lógica do Código de Defesa do Consumidor; ou, ainda, sobre o médico, em razão de “culpa na escolha” da ferramenta (*culpa in eligendo*) – deslocando o debate da responsabilidade por ato médico para a responsabilidade pelo fato do produto ou do serviço (NOGAROLI, 2023; KFOURI NETO, 2024).

Em suma, “observa-se que o profissional da Medicina, diante do resultado diagnóstico ou prognóstico trazido pelo algoritmo de IA, estará na complexa posição de justificar: a) por que ele seguiu o diagnóstico ou tratamento sugerido pela IA; ou b) por que – e com base em quais fatores – ele se desviou da recomendação algorítmica. O médico tem soberania, isto é, ele é

livre para escolher seus meios de diagnóstico e propostas de terapia, mas também é responsável por suas escolhas” (NOGAROLI, 2023).

Ademais, a demonstração de superioridade estatística em estudos do mundo real, como o *PRAIM* – um estudo prospectivo sobre uso de IA em rastreamento mamográfico, que elevou significativamente a taxa de detecção de câncer de mama – pode, com o tempo, inverter o risco jurídico e alterar o padrão de cuidado (*standard of care*). Se uma tecnologia comprovar salvar mais vidas, a sua não adoção pela instituição de saúde poderia, no futuro, ser considerada conduta negligente. Isso impõe um novo dever de atualização e investimentos por parte dos gestores de saúde, com implicações especialmente relevantes para o Sistema Único de Saúde (SUS), onde a alocação de recursos é crítica (FACCHINI NETO; BARBOSA, 2023; NOGAROLI, 2023).

Outros dois riscos são proeminentes: o viés algorítmico e a falta de transparência. O estudo *PRAIM* foi conduzido na Alemanha. A aplicação direta de seu modelo no Brasil, um país com imensa diversidade étnica e socioeconômica, é preocupante. Sem uma validação em coortes⁹ locais, o modelo pode ter uma performance inferior em populações sub-representadas nos dados de treinamento originais, como mulheres negras ou indígenas, resultando em iniquidade diagnóstica e perpetuando desigualdades estruturais na saúde (AITH, 2024; CAZANI; ALPINO *et al.*, 2025). Por fim, a opacidade do processo decisório do algoritmo, o chamado efeito “caixa-preta”, dificulta a confiança do médico e a justificativa de um diagnóstico perante o paciente ou, eventualmente, o Judiciário (SIMONASSI; BARROS *et al.*, 2024). A dificuldade em explicar *por que* a IA chegou a uma determinada conclusão afeta diretamente o dever de informação e a validade do Consentimento “Esclarecido” do paciente (TOPOL, 2019; AITH, 2024; CAZANI; ALPINO *et al.*, 2025).

2.2.1.2 Patologia Digital e os Foundation Models: A Revolução do Prov-GigaPath

Panorama Geral e Potencialidades:

A patologia digital está sendo revolucionada pela chegada dos *foundation models* (modelos de fundação), e o *Prov-GigaPath*, apresentado em um artigo na *Nature* em 2024, é um exemplo paradigmático. Desenvolvido por uma colaboração entre a Universidade de

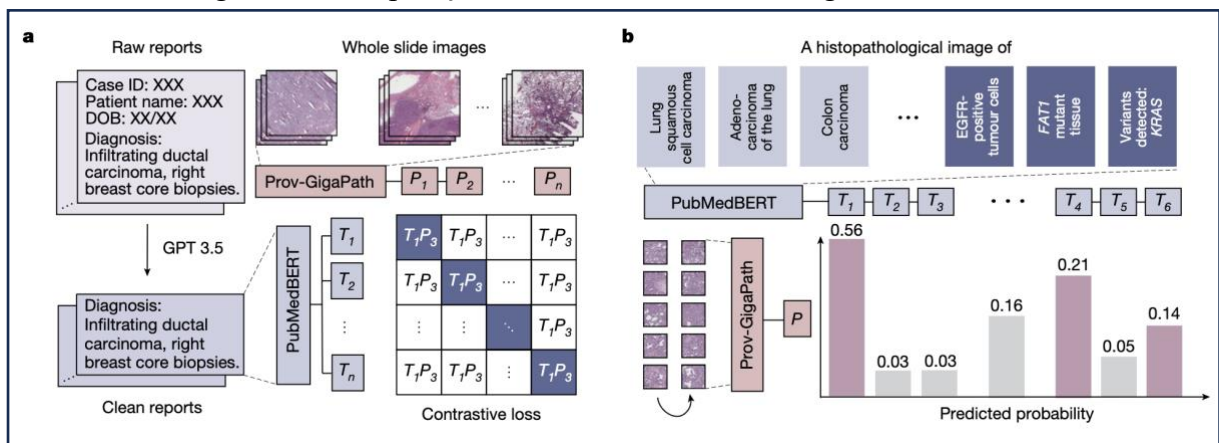
⁹ Por definição, uma coorte é um grupo de indivíduos que tem pelo menos uma característica comum. Num estudo de coorte, esses indivíduos são seguidos ao longo do tempo, para descrever e compreender a evolução da sua saúde e condições de vida. FONTE: Instituto de Saúde Pública da Universidade do Porto (ISPUP). Disponível em: <https://ispup.up.pt/coortes/#:~:text=Por%20definição%2C%20uma%20coorte%20é,saúde%20e%20condições%20de%20vida>. Acesso em: 10 abr. 2024.

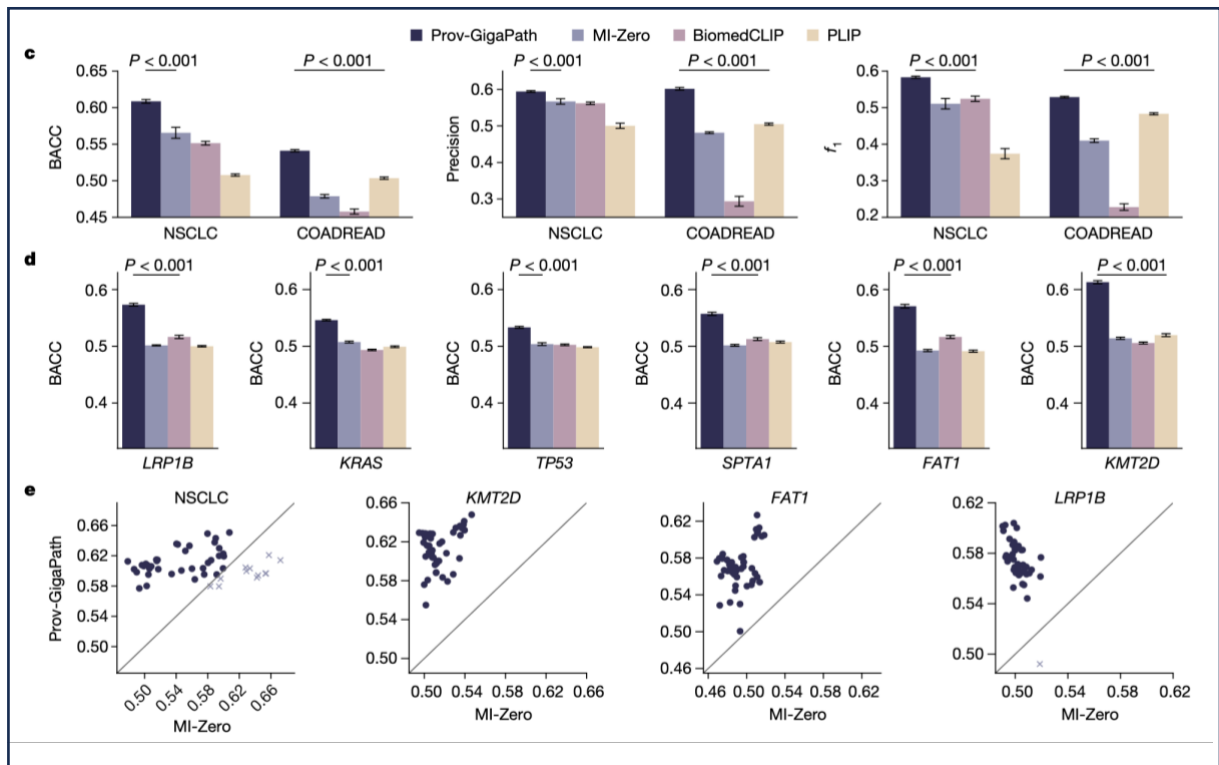
Washington, a *Microsoft Research* e a rede de saúde *Providence*, este modelo de acesso aberto representa um salto quântico em escala e metodologia (XU; USUYAMA *et al.*, 2024).

Diferentemente de modelos anteriores, que eram treinados com porções limitadas de imagens (*patches*) ou em bancos de dados restritos, o *Prov-GigaPath* foi pré-treinado com um conjunto de dados massivo e oriundo do mundo real: 1,3 bilhão de blocos de imagem provenientes de 171.189 lâminas de patologia inteiras (*Whole Slide Images - WSI*), coletadas de mais de 30.000 pacientes com diversos tipos de câncer (XU; USUYAMA *et al.*, 2024; CAMPANELLA *et al.*, 2025).

Sua arquitetura técnica, denominada *GigaPath*, utiliza uma forma de *vision transformer* com um mecanismo de “atenção dilatada” (*dilated attention*), que lhe permite processar a imagem da lâmina inteira em resolução de gigapixel. Essa capacidade é crucial, pois habilita o modelo a capturar não apenas as características de células individuais (morfologia nuclear, mitoses), mas também o contexto do microambiente tumoral – as interações complexas entre células cancerígenas, imunes e estromais – que é fundamental para o prognóstico e a escolha da terapia (XU; USUYAMA *et al.*, 2024). Como resultado, o modelo demonstrou uma performance superior na predição de subtipos de câncer e tem o potencial de acelerar drasticamente o desenvolvimento de novas ferramentas de IA para diagnóstico, prognóstico e identificação de biomarcadores, funcionando como uma base sobre a qual a comunidade científica pode construir aplicações clínicas específicas (BOMMASANI *et al.*, 2022).

Figura 2 – Comparação do alinhamento entre imagens e relatórios





Fonte: XU; USUYAMA *et al.* (2024).

Nota: **(a)** Fluxograma mostrando o ajuste fino do *Prov-GigaPath* usando relatórios patológicos. Relatórios patológicos reais são processados usando o GPT-3.5 da OpenAI para remover informações irrelevantes para o diagnóstico de câncer. Realizamos o aprendizado contrastivo baseado em *CLIP - Contrastive Language-Image Pre-training* (“Pré-treinamento contrastivo de linguagem-imagem”) para alinhar o *Prov-GigaPath* e o *PubMedBERT*. **(b)** O *Prov-GigaPath* ajustado pode então ser usado para realizar a subtipagem de câncer *zero-shot* (= classificação na qual o modelo é solicitado a identificar classes ou categorias para as quais nenhuma imagem de treinamento foi fornecida) e a previsão de mutações. A entrada do *Prov-GigaPath* é uma sequência de blocos segmentados a partir de um *WSI - Whole Slide Image* (“Imagem de Lâmina Total”), e as entradas do codificador de texto *PubMedBERT* são prompts projetados manualmente que representam tipos de câncer e mutações. Com base na saída do *Prov-GigaPath* e do *PubMedBERT*, podemos calcular a probabilidade da entrada *WSI* ser classificada em subtipos específicos de câncer e mutações. **(c)** Gráficos de barras comparando o desempenho da subtipagem *zero-shot* em *NSCLC - Non-Small Cell Lung Cancer* (“Câncer de Pulmão de Não Pequenas Células”) e *COADREAD - Colorectal Adenocarcinoma* (“Adenocarcinoma Colorretal”) em termos de *BACC - Balanced Accuracy* (“Acurácia Balanceada”), precisão e f_1 - *F1 Score* (“Pontuação F1”). **(d)** Gráficos de barras comparando o desempenho na previsão de mutações usando o modelo ajustado para seis genes. **(c), (d)** Os dados são a média \pm s.e.m. em $n = 50$ experimentos. O valor P listado indica a significância do *Prov-GigaPath* superar a melhor abordagem de comparação, com o teste *Wilcoxon* unilateral. **(e)** Gráficos de dispersão comparando o desempenho entre o *Prov-GigaPath* e o *MI-Zero - Multiple Instance Zero-shot Transfer for Histopathology Images* (“Transferência *Zero-Shot* de Múltiplas Instâncias”) em termos de *BACC* na subtipagem de câncer *zero-shot*. Cada ponto indica um ensaio com um conjunto específico de formulações de consulta de texto (XU; USUYAMA *et al.*, 2024).

Riscos Jurídicos e Éticos:

A disponibilização do *Prov-GigaPath* em acesso aberto é um catalisador para a inovação e a democratização da pesquisa. Contudo, essa mesma característica introduz uma complexa

teia de desafios jurídicos e éticos, especialmente no que tange à responsabilidade, à governança de dados e ao consentimento do paciente.

A primeira e mais evidente tensão reside na fragmentação da responsabilidade civil. Se um laboratório hospitalar utiliza o *Prov-GigaPath* como base para desenvolver uma ferramenta comercial de diagnóstico e um erro causa dano ao paciente, a atribuição de culpa torna-se nebulosa. Os desenvolvedores originais (*Microsoft, U. Washington, Providence*), amparados por licenças *open-source* como a MIT ou Apache 2.0, que tipicamente isentam de garantias e responsabilidades, provavelmente se eximirão de culpa (Art. 53, 2, AI Act¹⁰). Essa estrutura pode deslocar todo o ônus para o “agente implementador” – a empresa ou instituição que realizou o ajuste fino (*fine-tuning*) do modelo para uma aplicação clínica específica (XU *et al.*, 2024; HONFUNG POON; USUYAMA, 2024). A propósito, a legislação europeia (AI Act), que serve de inspiração, impõe obrigações de transparência e documentação aos prestadores de modelos de IA de finalidade geral, mas coloca um papel fundamental no responsável pela implantação (o “agente implementador”) para garantir a proteção dos direitos fundamentais ao utilizar sistemas de IA de alto risco (Considerandos nº 93, 100 e 101).

Este cenário desafia o arcabouço regulatório de *SaMD - Software as a Medical Device* (“Software como Dispositivo Médico”)¹¹, que foi concebido para cadeias de desenvolvimento mais lineares e proprietárias, e não para ecossistemas abertos e descentralizados (IMDRF, 2014). A regulação de *SaMD* no Brasil, como a RDC nº 657/2022 da ANVISA (2022a), é um passo importante, mas ainda precisa amadurecer para definir claramente os papéis e as obrigações de cada agente nesse novo ecossistema.

A governança dos dados para o treinamento do modelo é outro ponto de alta complexidade. A utilização de um volume tão grande de dados de saúde para um “uso secundário” – finalidade distinta daquela para a qual foram originalmente coletados (a assistência ao paciente) – exige uma base legal robusta (DONEDA, 2021; REIS; REIS, 2024; UNIÃO EUROPEIA, 2016; BRASIL, 2018). A anonimização nesse tipo de tratamento de dado é uma exigência legal. No entanto, a anonimização de imagens de patologia em gigapixel é um desafio técnico e jurídico. A riqueza de detalhes, quando cruzada com metadados clínicos ou genômicos, aumenta o risco de reidentificação, mesmo que identificadores diretos sejam removidos. Isso exige não apenas técnicas de anonimização cada vez mais sofisticadas, mas também uma governança contratual rigorosa que regule o uso dos dados e proíba tentativas de reidentificação (PRICE II; COHEN, 2019; ZUBOFF, 2021; API4AI, 2025).

¹⁰ Regulamento nº 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024 (UNIÃO EUROPEIA, 2024).

¹¹ FDA, 2025b.

A existência de *foundation models* de acesso aberto também levanta questões de propriedade intelectual: quem detém os direitos sobre um modelo derivado e comercializado a partir de uma base de acesso livre? A regulação deve encontrar um equilíbrio que incentive a inovação sem permitir que o espírito colaborativo do *open source* seja indevidamente apropriado para fins puramente comerciais (BOMMASANI *et al.*, 2022; PINTO, 2025).

Por fim, emerge a questão crucial do Consentimento Livre e Esclarecido do paciente, que não se confunde com o consentimento para tratamento de dados da LGPD. Trata-se do consentimento bioético, pilar da relação paciente-médico, que garante a autonomia do indivíduo em suas decisões de saúde (SOARES, 2025). Quando um diagnóstico é influenciado por uma ferramenta como o *Prov-GigaPath*, o dever de informação do médico se expande. O paciente tem o direito de saber que uma IA foi utilizada, mas também de ser informado sobre as incertezas inerentes a essa tecnologia, como a opacidade de seu funcionamento e o risco de vieses não detectados, que podem levar a um diagnóstico incorreto. A falha em comunicar adequadamente esses riscos pode viciar o consentimento, tornando-o ineficaz e abrindo margem para a responsabilização do profissional e da instituição de saúde, independentemente da responsabilidade do desenvolvedor do software (COHEN, 2019; CALIXTO; SCHULZE, 2025). É sabido, por exemplo, que a responsabilidade civil do profissional pela falha na obtenção do consentimento pode ocorrer mesmo quando a intervenção médica é tecnicamente correta, mas a informação foi inadequada (NOGAROLI, 2023). Isso porque, o médico/hospital está em melhor posição para compreender o contexto de uso e mitigar os riscos para os direitos fundamentais (UNIÃO EUROPEIA, 2024, Considerando nº 93).

2.2.1.3 Medicina Preditiva na Cardiologia: A Ferramenta de IA de Oxford (ORFAN)

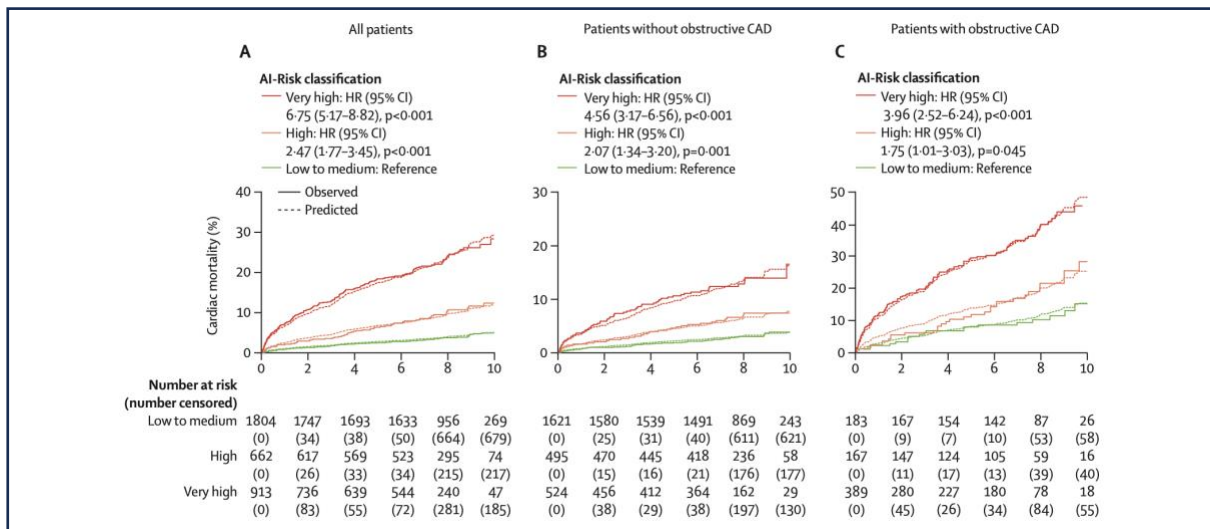
Panorama Geral e Potencialidades:

A cardiologia testemunha uma revolução na medicina preventiva graças a uma ferramenta de inteligência artificial desenvolvida por pesquisadores da Universidade de Oxford. Apresentada em 2023, essa tecnologia analisa angiotomografias computadorizadas cardíacas de rotina, para prever o risco de eventos cardiovasculares graves, como infarto do miocárdio e morte cardíaca, com uma impressionante janela de até 10 (dez) anos de antecedência (CHAN; WAHOME *et al.*, 2024).

A grande inovação da ferramenta reside em sua capacidade de transcender a análise anatômica convencional, focada na detecção de estreitamentos (estenoses) nas artérias coronárias. O algoritmo de IA foi treinado para identificar e quantificar sinais sutis de

inflamação no tecido adiposo perivascular, alterações que são invisíveis ao olho humano, mas que funcionam como um poderoso sinal de alerta precoce para a instabilidade de placas ateroscleróticas. O estudo de validação, que analisou dados de mais de 40.000 pacientes, de oito hospitais do Reino Unido, revelou um fato surpreendente: dois terços dos eventos cardíacos ocorreram em pacientes que não apresentavam obstruções arteriais consideradas significativas pelos métodos tradicionais (CHAN; WAHOME *et al.*, 2024). Esses pacientes, que seriam classificados como de baixo risco e, portanto, subtratados, puderam ser corretamente reestratificados pela IA. O impacto clínico é direto e transformador: em um estudo piloto, a apresentação do escore de risco gerado pela IA aos médicos assistentes levou à alteração do plano terapêutico – como a intensificação de estatinas ou a introdução de novas terapias preventivas – em até 45% dos casos, inaugurando um novo horizonte para a medicina preventiva personalizada (UNIVERSITY OF OXFORD, 2024).

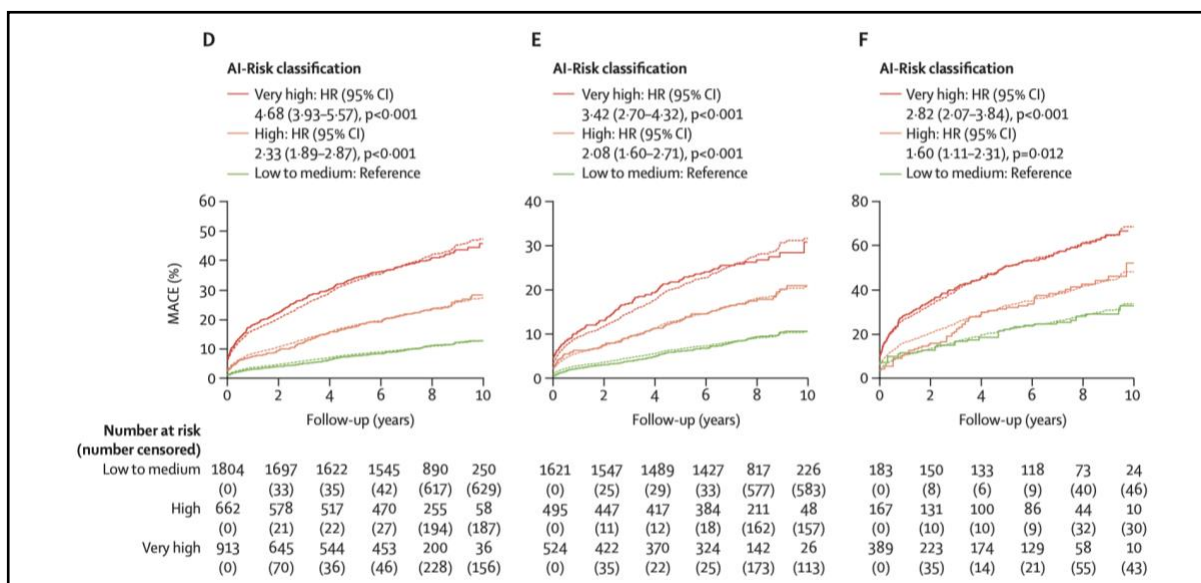
Figura 3 – Classificação AI-Risk e previsão do risco cardiovascular (parte 1)



Fonte: CHAN; WAHOME *et al.*, (2024).

Nota: Curvas de Kaplan-Meier para a capacidade da classificação AI-Risk de prever a mortalidade cardíaca em (A) toda a população, (B) pacientes sem DAC obstrutiva. (C) pacientes com DAC obstrutiva (CHAN; WAHOME *et al.*, 2024).

Figura 4 – Classificação AI-Risk e previsão do risco cardiovascular (parte 2)



Fonte: CHAN; WAHOME *et al.*, (2024).

Nota: As curvas de Kaplan-Meier para a previsão de MACE usando a mesma classificação são apresentadas para **(D)** toda a população, **(E)** pacientes sem DAC obstrutiva e **(F)** pacientes com DAC obstrutiva. IA = inteligência artificial. AI-Risk = o algoritmo assistido por IA previu o risco de mortalidade cardíaca. DAC = doença arterial coronariana. HR = razão de risco. MACE = eventos cardíacos adversos graves (CHAN; WAHOME *et al.*, 2024).

Riscos Jurídicos e Éticos:

A capacidade de antever um risco de doença com uma década de antecedência gera um profundo dilema ético que colide diretamente com a autonomia do paciente: o “direito de não saber” (*right not to know*) – (RÜGER, 2007; RODOTÀ, 2008). O Código de Ética Médica, em seu Capítulo V, art. 31, veda ao médico “desrespeitar o direito do paciente ou de seu representante legal de decidir livremente sobre a execução de práticas diagnósticas ou terapêuticas” (CFM, 2018; VASERINO, 2020). Esse direito de decidir, conforme a doutrina bioética, inclui a prerrogativa de recusar informações sobre a própria saúde (CFM, 2016; SILVA *et al.*, 2021; DADALTO, 2025). Se uma tomografia é realizada para investigar uma dor torácica aguda e, como um subproduto, a IA gera uma predição de risco de longo prazo, o paciente tem o direito de recusar essa informação? A comunicação de uma probabilidade futura, embora não seja um diagnóstico de doença presente, pode gerar profunda ansiedade (“ansiedade preditiva”), impactar decisões de vida e levar a um sentimento de determinismo genético ou algorítmico (ANDORNO, 2004; RODOTÀ, 2008; BEAUCHAMP; CHILDRESS, 2013).

Isso torna o processo de obtenção do Consentimento Livre e Esclarecido um ato eticamente complexo. Não basta obter a autorização para a realização do exame; é necessário um consentimento específico para a análise preditiva por IA. Este consentimento deve ser granular, esclarecendo que a tecnologia pode revelar riscos futuros e, crucialmente, oferecendo ao paciente a opção explícita de *não* receber essa informação. A opacidade do algoritmo

(PASQUALE, 2015; TOPOL, 2019), que impede uma explicação completa sobre *como* o risco foi calculado, e a possibilidade de vieses no treinamento, que podem levar a predições menos acuradas para determinados grupos populacionais (O'NEIL, 2016; AITH, 2024), são informações essenciais que devem ser comunicadas para que o consentimento seja considerado verdadeiramente válido e esclarecido (COHEN, 2019). Se nem médico nem paciente compreendem plenamente os critérios da predição, há um risco concreto de que a decisão clínica se fundamente em informações não totalmente inteligíveis – fragilizando a autonomia do paciente (COHEN, 2019).

A natureza probabilística da predição também tensiona os limites da responsabilidade civil médica. Tradicionalmente, a responsabilidade do médico é uma obrigação de meio, e não de resultado (FACCHINI NETO; BARBOSA, 2023; KFOURI NETO, 2024). No campo da medicina preditiva, a definição do nexos de causalidade torna-se ainda mais nebulosa (NOGAROLI, 2023). Se a IA prevê um alto risco, o médico informa o paciente e ambos optam por uma abordagem conservadora, vindo o paciente a sofrer um infarto, haveria responsabilização? Inversamente, se a IA prevê um baixo risco e um evento ocorre, isso exime o médico de responsabilidade? A ferramenta se torna mais um elemento na complexa equação da decisão clínica, mas não elimina a incerteza inerente à prática médica, exigindo do Judiciário uma análise cuidadosa para evitar o que a doutrina chama de “medicina defensiva” (NOGAROLI, 2023; ROSENVALD; MARTINS, 2024).

Finalmente, emergem os riscos de discriminação e iniquidade. Um escore de risco preditivo, se não houver regulação adequada, poderia ser usado por seguradoras de vida ou de saúde para negar cobertura ou aumentar prêmios, ou até mesmo por empregadores, configurando uma forma de “discriminação algorítmica”. Por exemplo, o Regulamento (UE) 2025/327, relativo ao Espaço Europeu de Dados de Saúde (EEDS), dispõe, expressamente, no Considerando nº 62, que deve:

“ser proibida toda e qualquer tentativa de utilizar dados de saúde eletrônicos para medidas desfavoráveis às pessoas singulares, como o aumento dos prêmios de seguro, o exercício de atividades potencialmente prejudiciais para as pessoas singulares no domínio do emprego, das pensões ou da banca, incluindo as hipotecas de imóveis, a publicidade a produtos ou tratamentos, a automatização da tomada de decisões individuais, a reidentificação de pessoas singulares ou o desenvolvimento de produtos nocivos”.

(UNIÃO EUROPEIA, 2025)

O AI Act também trata desses temas, nos Considerandos nº 57 e 58 (UNIÃO EUROPEIA, 2024). Zuboff (2021) e O'Neil (2016) também demonstraram preocupação com o tema.

Ademais, sendo uma tecnologia de ponta, seu acesso provavelmente será limitado, aprofundando as desigualdades em saúde. Se disponível apenas no sistema privado, criará uma clivagem entre uma parcela da população que pode se beneficiar de uma prevenção cardiovascular de alta precisão e a maioria que depende do Sistema Único de Saúde (SUS), violando o princípio constitucional da equidade (ABUJABER; NASHWAN, 2024; MARTINI, 2023; DANTAS; JEREISSATI, 2024).

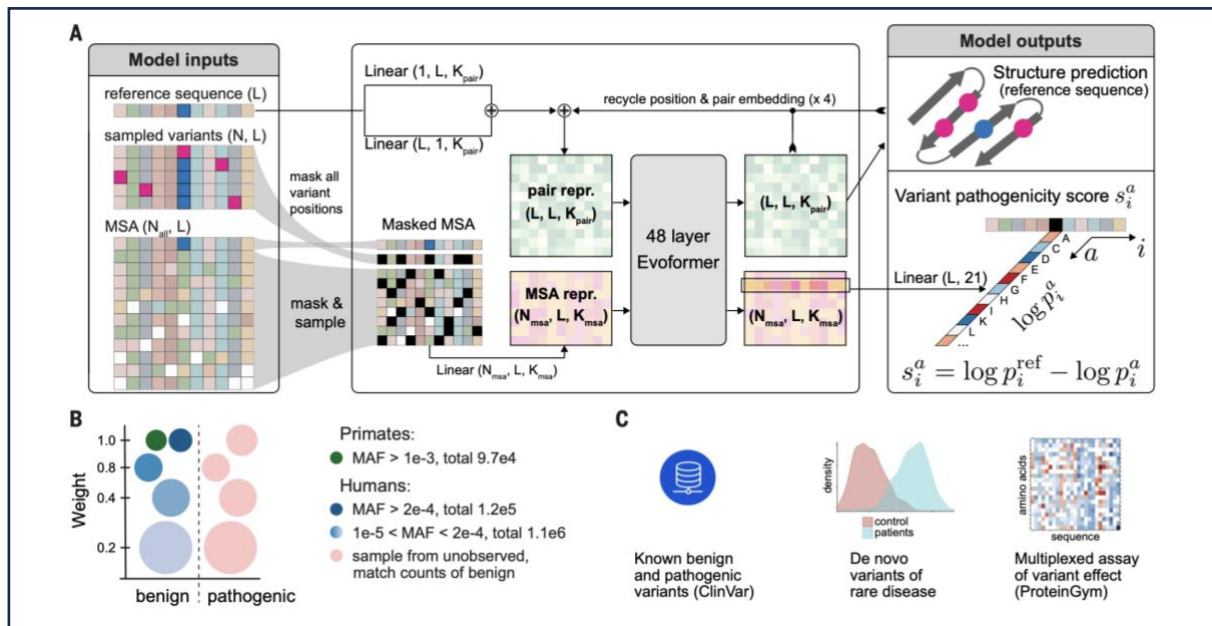
2.2.1.4 Decifrando o Genoma: O AlphaMissense e a Classificação de Variantes Genéticas

Panorama Geral e Potencialidades:

No campo da genômica e da medicina de precisão, a interpretação de variantes genéticas representa um dos maiores desafios diagnósticos. Uma variante de sentido trocado (*missense*) é uma mutação pontual que resulta em um aminoácido diferente na proteína correspondente. Enquanto algumas dessas alterações são inócuas, outras podem ser a causa de doenças graves, como o câncer. Determinar a patogenicidade de uma variante é uma tarefa complexa e demorada. Nesse cenário, o *AlphaMissense*, uma ferramenta de IA desenvolvida pela Google DeepMind e apresentada em 2023, surge como um avanço monumental (CHENG; PAN *et al.*, 2023; NGUYEN; LEE *et al.*, 2025).

Construído sobre a mesma arquitetura do revolucionário *AlphaFold* (que prevê a estrutura 3D de proteínas), o *AlphaMissense* foi treinado para prever a patogenicidade de todas as 71 milhões de variantes *missense* humanas possíveis. O modelo classifica 89% delas como “provavelmente patogênicas” ou “provavelmente benignas”, com uma precisão reportada superior a 90%, quando comparado com bancos de dados de referência curados por humanos, como o *ClinVar* (CHENG; PAN *et al.*, 2023). Seu potencial para a medicina de precisão é vasto. Um estudo publicado em 2025, no *JCO Precision Oncology*, demonstrou que o *AlphaMissense* foi capaz de reclassificar, com sucesso, mutações em genes de reparo de DNA (como BRCA1/2 e ATM) que até então eram consideradas “variantes de significado incerto” (VUS). Essa reclassificação tem implicações clínicas diretas, pois pode ajudar a identificar pacientes elegíveis para terapias-alvo específicas, como os inibidores de PARP (Poli ADP-ribose polimerase)¹², transformando a incerteza diagnóstica em uma conduta terapêutica clara (YAZAKI; PEI *et al.*, 2025; NGUYEN; LEE *et al.*, 2025).

¹² Os inibidores de PARP são medicamentos que interferem no mecanismo de reparo de DNA das células. Normalmente, quando ocorrem danos no DNA, a enzima PARP é ativada para reparar quebras de fita simples. No entanto, em células cancerígenas com mutações nos genes BRCA1 ou BRCA2, o sistema de reparo de DNA já apresenta deficiência, o que as torna vulneráveis a danos adicionais. Ao bloquear a ação da PARP, esses medicamentos impedem que células tumorais “consertem” o DNA danificado, acumulando quebras irreparáveis,

Figura 5 – Visão geral do *AlphaMissense*

Fonte: CHENG; PAN *et al.* (2023).

Nota: (A) Arquitetura do *AlphaMissense*. As entradas do modelo consistem na sequência de proteína de referência [cortada para o comprimento (L) = 256 resíduos], um conjunto de variantes amostradas a partir do conjunto de treinamento para a mesma sequência (até $N = 50$ variantes) e alinhamentos de múltiplas sequências (*MSAs* - *Alinhamentos de Múltiplas Sequências*, até N_{all} - número total de alinhamentos de *MSAs* = 2048). A inferência é realizada para uma variante por vez ($N = 1$). A sequência de referência é repetida na segunda linha do *MSA* com todas as posições variantes amostradas mascaradas. Assim como no *AlphaFold*, o modelo constrói a representação de pares (isto é, codifica informações sobre interações bidirecionais entre resíduos) a partir da sequência de referência (tamanho de *Embedding* K_{pair} = Tamanho do *Embedding* de Pares), e a representação *MSA* a partir do *MSA* mascarado (tamanho de *Embedding* K_{msa} = Tamanho do *Embedding* *MSA*). As representações *MSA* e de pares são processadas por uma pilha de camadas *Evoformer* com reciclagem. Finalmente, o modelo prevê a estrutura da sequência de referência e a pontuação de patogenicidade sai para a variante, que é derivada da *head* de previsão de resíduo mascarado como a diferença de *log-likelihood* entre o resíduo a em relação ao resíduo de referência na posição i . (B) A pontuação de patogenicidade é ajustada (*fine-tuned*) como uma classificação binária de variantes como benignas (variantes *missense* observadas ou frequentes em populações humanas ou de primatas) ou patogênicas (variantes *missense* humanas não observadas). Nós dividimos as variantes benignas em *clusters* por sua frequência alélica menor (*MAF* - *Frequência Alélica Menor*) e introduzimos pesos na função de perda que reduzem a contribuição de variantes raras. Para cada variante observada no conjunto benigno, amostramos uma variante *missense* do conjunto patogênico e atribuímos a ela o mesmo peso de perda que para a variante benigna. (C) Avaliamos o *AlphaMissense* em um conjunto diversificado de *datasets* de *benchmark*, incluindo variantes *missense* anotadas no *ClinVar*, variantes *de novo* de doenças e dados *MAVE* (*Ensaio Multiplexado de Efeito de Variantes*) coletados no *ProteinGym* (CHENG; PAN *et al.*, 2023).

Riscos Jurídicos e Éticos:

resultando na morte celular seletiva das células cancerígenas. Em contrapartida, as células normais, que possuem mecanismos de reparo alternativos, conseguem sobreviver, o que torna a terapia mais específica para células tumorais. FONTE: <https://www.rededorsaoluiz.com.br/onco/oncologador/noticias/artigo/o-que-sao-os-inibidores-de-parp-no-tratamento-do-cancer-de-ovario>. Acesso em: 10 abr. 2024.

A natureza dinâmica e probabilística de ferramentas como o *AlphaMissense* introduz um conjunto de desafios jurídicos e éticos que tensionam a prática da medicina genômica. O primeiro deles é o risco da “reclassificação” e a insegurança jurídica que dela decorre. Um modelo de IA não é estático; ele é continuamente atualizado e aprimorado com novos dados. Uma variante classificada hoje como “benigna” pode ser reclassificada como “patogênica” em uma versão futura. Isso levanta questões complexas: o laboratório que emitiu o laudo genético original tem o dever de monitorar essas atualizações e recontatar o paciente? A doutrina jurídica sobre o “dever de recontato” (*Duty to Recontact*) em genética é controversa e não pacificada, mas a natureza fluida do conhecimento gerado por IA agrava o problema, desafiando os conceitos de finalidade do diagnóstico e de segurança jurídica (PLOEM; GIESBERTZ *et al.*, 2023). Se uma decisão clínica crucial, como a não realização de uma cirurgia profilática, foi tomada com base em uma classificação que posteriormente se provou incorreta, a responsabilização torna-se um campo minado.

A existência de um dever ético de recontato depende “muito das circunstâncias específicas de um caso”, equilibrando valores como respeito pela autonomia, beneficência e o princípio de não causar dano (não-maleficência). Martine C. Ploem, Noor A.A. Giesbertz e outros colegas (2023), desenvolveram, conforme publicado no artigo intitulado: *Duty to recontact in genomic cancer care: A tool helping to assess the professional’s responsibility*, uma ferramenta de suporte à decisão para ajudar os profissionais a avaliar se o recontato é necessário, evidenciando que a doutrina é controversa e não pacificada, requerendo uma análise baseada em critérios específicos (relacionamento profissional, impacto clínico, preferência do indivíduo e viabilidade/custo). O risco de reclassificação de variantes tem implicações clínicas diretas e sérias, afetando decisões sobre medidas preventivas ou tratamento. O artigo cita o seguinte exemplo: uma reclassificação de uma variante BRCA1 (no caso, de classe 4 para classe 2, com implicações negativas) pode influenciar diretamente a decisão de uma paciente sobre uma “salpingo-ooforectomy” bilateral¹³ profilática. Se uma decisão crucial como essa for baseada em uma classificação que posteriormente é revertida pela IA, o ônus da prova de negligência, imprudência ou imperícia, e onexo causal, torna a responsabilização, de fato, um campo minado (PLOEM; GIESBERTZ *et al.*, 2023).

A natureza dinâmica e opaca (*black-box*) da IA agrava a dificuldade de se auditar o processo decisório, tornando a prova do nexo causal um desafio significativo. A

¹³ A salpingo-ooforectomia bilateral é a remoção cirúrgica de ambos os ovários e ambas as trompas de Falópio nov. possibilidade de recorrência. Fonte: <https://www-mercy-com.translate.google.com/health-care-services/cancer-care-oncology/specialties/ovarian-uterine-cervical-cancer/treatments/bilateral-salpingo-oophorectomy? x tr sl=en& x tr tl=pt& x tr hl=pt& x tr pto=tc>. Acesso em: 28 nov. 2024.

responsabilidade pela decisão final é do médico, que deve exercer um dever de vigilância e não pode aceitar acriticamente o resultado do algoritmo, sob pena de incorrer em culpa (como imperícia ou negligência).

É crucial ressaltar, portanto, que as predições *in silico* (computacionais) não substituem a validação clínica (humana). O próprio estudo sobre os genes de reparo de DNA conclui que a acurácia do *AlphaMissense* é dependente do gene em questão e que suas predições “ainda requerem confirmação adicional com validação clínica e funcional” (YAZAKI; PEI *et al.*, 2025). A utilização de uma predição de IA para uma decisão clínica final, sem a devida validação laboratorial e correlação com o quadro do paciente, seria um ato de imprudência. A responsabilidade por um eventual dano recairia pesadamente sobre o profissional e o laboratório que assumiram o risco de agir com base em uma probabilidade computacional (NOGAROLI, 2023; FACCHINI NETO; BARBOSA, 2023). A incorporação dessas ferramentas no arsenal diagnóstico exige a criação de protocolos rigorosos de validação perante agências reguladoras como a ANVISA, que ainda navegam os desafios da regulamentação de *Software as a Medical Device* (SaMD). Sistemas de IA de diagnóstico são classificados como de risco elevado e devem demonstrar conformidade com os requisitos essenciais de segurança e desempenho, o que inclui a verificação e validação do *software* (ANVISA, 2022a, Art. 17; UNIÃO EUROPEIA, 2017, Regulamento (UE) 2017/745).

Contudo, a regulamentação da IA transcende a simples elaboração de uma nova lei, abrangendo um amplo escopo de desafios técnicos e de governança estatal (AITH, 2024). A falta de validação externa rigorosa e a necessidade de dados diversificados para evitar vieses são barreiras importantes que a IA regulatória busca superar antes da implementação clínica generalizada (GUANDALINI *et al.*, 2024; CAZANI; ALPINO *et al.*, 2025).

O viés algorítmico é uma preocupação central na genômica. A vasta maioria dos bancos de dados genômicos usados para treinar e validar modelos como o *AlphaMissense* é composta por dados de populações de ancestralidade europeia¹⁴ (POPEJOY; FULLERTON, 2016). Isso significa que a acurácia da ferramenta pode ser significativamente menor para populações diversas e miscigenadas, como a brasileira, criando um “viés persistente” (segundo POPEJOY; FULLERTON, 2016) ou “viés genômico” (segundo ALSAEDI; OGASAWARA *et al.*, 2025) que pode levar a diagnósticos equivocados e aprofundar as disparidades globais em saúde. Aliás, no contexto brasileiro, onde a população se identifica majoritariamente como negra ou

¹⁴ Uma análise de 2009 revelou que 96% dos participantes de estudos de associação de todo o genoma (GWAS) eram de descendência europeia. Sete anos depois, embora a proporção de indivíduos não descendentes de europeus tenha aumentado para quase 20%, a maior parte desse crescimento se deu em populações asiáticas, e os grupos de ancestralidade africana e latino-americana/hispânicos (que frequentemente são de ancestralidade mista) representam menos de 4% de todas as amostras analisadas (POPEJOY; FULLERTON, 2016).

parda, tais vieses na IA médica são especialmente preocupantes para a equidade em saúde (AITH, 2024). A mensagem atual transmitida pela comunidade científica de genômica é “prejudicial e enganosa”: a de que os genomas dos descendentes de europeus são os mais importantes (POPEJOY; FULLERTON, 2016).

Sakhaa Alsaedi, Michihiro Ogasawara e outros (2025) discutem que as previsões de risco genético devem ser robustas, e ferramentas como o *PolyPred* foram desenvolvidas para lidar com a precisão preditiva em populações diversas, incluindo coortes não europeias, justamente porque a acurácia tende a ser um desafio nessas populações. O *PolyPred*, por exemplo, tem um desempenho elevado em risco preditivo para populações diversas, com $AUC > 0.92$, destacando que a acurácia varia significativamente entre populações. Diferenças na frequência de variantes genéticas entre populações podem fazer com que associações encontradas em europeus sejam falsos positivos em outros grupos. Pacientes de ancestralidade africana e asiática têm maior probabilidade do que os de ancestralidade europeia de receber resultados ambíguos de testes genéticos ou serem informados de que têm variantes de significado desconhecido (VUS). O viés genômico, portanto, aprofunda as disparidades de saúde ao oferecer informações diagnósticas de menor qualidade ou mais incertas para grupos não europeus (ALSAEDI; OGASAWARA *et al.*, 2025).

Por fim, a identificação de uma variante “provavelmente patogênica”, por uma IA, intensifica os desafios do Consentimento Livre e Esclarecido e do aconselhamento genético (DANTAS; NOGAROLI, 2020; NOGAROLI, 2023). O processo de consentimento, neste contexto, deve ser excepcionalmente robusto. O paciente precisa ser claramente informado de que a análise utilizará uma ferramenta de IA, que o resultado é probabilístico (não um diagnóstico definitivo), que existe a possibilidade de vieses e que o conhecimento pode mudar no futuro. A opacidade do modelo impede uma explicação detalhada sobre *como* a conclusão foi alcançada, exigindo que o médico comunique essa incerteza de forma honesta (COHEN, 2019; TOPOL, 2019). Além disso, o aconselhamento genético deve salvaguardar a autonomia do paciente, o que inclui a discussão explícita sobre seu “direito de não saber”, especialmente para condições que não possuem tratamento ou prevenção eficazes (RÜGER, 2007; RODOTÀ, 2008; CFM, 2016). A decisão de conhecer uma predisposição genética com tamanhas implicações para a vida do paciente e de sua família deve ser dele, livre de qualquer coerção, garantindo que sua dignidade seja respeitada (DAVIES, 2020; DADALTO, 2025). Ou, segundo Joaquim Clotet (2009), a informação deve sempre deixar margem para a esperança, ainda que limitada.

2.2.1.5 Democratizando o Diagnóstico: IA na Retinopatia Diabética

Panorama Geral e Potencialidades

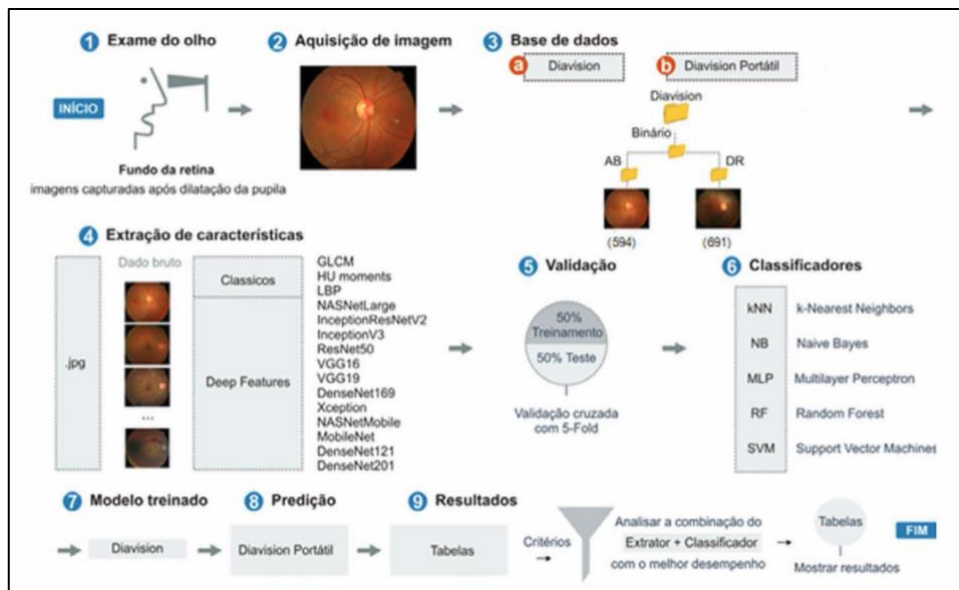
A retinopatia diabética (RD) é a principal causa de perda de visão entre adultos em idade produtiva, e sua prevalência acompanha o crescimento exponencial do diabetes no mundo.¹⁵ O diagnóstico e o tratamento precoces são cruciais para evitar a cegueira, mas o acesso a oftalmologistas, para o exame de fundo de olho anual, representa um gargalo significativo, especialmente em áreas remotas ou carentes (MALERBI; ANDRADE *et al.*, 2021; OLIVEIRA; SILVA *et al.*, 2024). Nesse contexto, a IA surge como uma tecnologia disruptiva com potencial para democratizar o acesso ao diagnóstico. Sistemas de IA autônomos, como o IDx-DR, que recebeu aprovação da agência norte-americana FDA, são capazes de analisar imagens da retina – muitas vezes capturadas por equipamentos portáteis acoplados a smartphones – e detectar a presença de RD com alta acurácia, sem a necessidade de interpretação imediata por um especialista (ABRÀMOFF; LAVIN *et al.*, 2018; OLVERA-BARRIOS; HEEREN *et al.*, 2021). Aliás, o IDx-DR foi o primeiro sistema de diagnóstico de IA autônomo autorizado pela FDA para detectar retinopatia diabética mais que moderada (*mtmDR*) e edema macular diabético, demonstrando a capacidade da IA de levar diagnósticos de nível especializado para ambientes de atenção primária (ABRÀMOFF; LAVIN *et al.*, 2018; COHEN, 2019).

Um ensaio pivotal de um sistema de IA para detecção de RD (*mtmDR*) demonstrou uma sensibilidade de 87,2% (IC 95%, 81,8–91,2%) e especificidade de 90,7% (IC 95%, 88,3–92,7%), excedendo os *endpoints* de superioridade pré-especificados (ABRÀMOFF; LAVIN *et al.*, 2018). Outro estudo retrospectivo com o sistema EyeArt v2.0, envolvendo mais de 100.000 encontros de pacientes, atingiu sensibilidade de 91,3% e especificidade de 91,1% (BHASKARANAND; RAMACHANDRA *et al.*, 2019). O sistema IDx-DR demonstrou uma sensibilidade de 97,6% na identificação de pacientes que requerem avaliação imediata e 99% para retinopatia ameaçadora à visão (OLVERA-BARRIOS; HEEREN *et al.*, 2021).

A principal função desses algoritmos é a triagem: eles não substituem o oftalmologista, mas identificam quais pacientes necessitam de encaminhamento urgente, otimizando o fluxo de atendimento e permitindo que especialistas se concentrem nos casos mais graves. O potencial de ampliação do acesso ao rastreamento é imenso, com a possibilidade de implementação em unidades básicas de saúde ou em pontos de atenção primária, reduzindo as lacunas no cuidado e prevenindo a cegueira em milhares de pessoas (BHASKARANAND; RAMACHANDRA *et al.*, 2019; ABRÀMOFF; LAVIN *et al.*, 2018).

¹⁵ Fonte: <https://diabetesatlas.org/>. Acesso em: 20 ago. 2024.

Figura 6 – Fluxograma da Metodologia Adotada



Fonte: OLIVEIRA; SILVA *et al.* (2024)

Nota: O fluxograma da metodologia adotada teve nove etapas: (1) Aquisição de Imagens utilizando o retinógrafo e o equipamento *Portable Diavision*; (2) exemplo de imagens para criar duas bases de dados: *diavision* e *diavision portátil*; (3) descrição das bases de dados utilizadas; (4) extração de características com três extratores clássicos e 12 arquiteturas com redes neurais (CNN); (5) método de validação utilizado; (6) avaliação com cinco métodos de aprendizagem de máquina; (7) modelos obtidos após o treinamento com as bases de dados *Diavision* e *Messidor*; (8) previsão a partir de imagens da base de dados *Portable Diavision*; (9) avaliação dos resultados em todos os conjuntos de dados.

O fluxograma da metodologia adotada no trabalho de OLIVEIRA; SILVA *et al.* (2024), representada na figura 6, descreve as nove etapas sequenciais do estudo para desenvolver e testar um sistema de visão computacional de baixo custo, baseado em inteligência artificial (IA) e *smartphone*, para o diagnóstico da Retinopatia Diabética (RD).

A metodologia integra a aquisição de imagens com um dispositivo portátil e o treinamento de modelos de *machine learning* (aprendizado de máquina) e *deep learning*.

O módulo de hipótese diagnóstica (*Portable Diavision*), desenvolvido em parceria com a *WDA Technology and Inovation*, é baseado em um diagnóstico binário (ausência ou presença de RD) e emprega a técnica de *deep transfer learning* (“aprendizado profundo por transferência”). Esta abordagem, de baixo custo, demonstrou eficácia satisfatória na triagem de pacientes, com acurácia diagnóstica variando entre 70% e 100% (OLIVEIRA; SILVA *et al.*, 2024).

Riscos Jurídicos e Éticos:

A principal questão jurídica levantada por sistemas autônomos, como o IDx-DR ou o *Portable Diavision*, é a da responsabilidade civil em caso de erro. Se o sistema classifica um exame como normal (falso negativo) e o paciente, confiando no resultado, deixa de procurar um especialista e evolui com perda de visão, a quem responsabilizar? Quando um sistema de IA autônomo fornece um resultado incorreto, como um falso negativo, e o paciente sofre danos (*e.g.*, perda de visão por atraso no tratamento), torna-se difícil ou até impossível descobrir a causa eficiente do dano, ou seja, o defeito do produto e onexo causal com a lesão (NOGAROLI, 2023). A autonomia do sistema dilui a cadeia de responsabilidade tradicional, centrada no ato médico, e a desloca para a esfera da responsabilidade pelo fato do produto, conforme visto anteriormente (UNIÃO EUROPEIA, 2018a). A responsabilidade, nesse caso, tende a ser objetiva e solidária, conforme preconiza o CDC, recaindo sobre o desenvolvedor do *software*, o fabricante do equipamento e a instituição de saúde que implementou o protocolo de triagem (NOGAROLI, 2023). A regulação de *Software as a Medical Device* (SaMD) no Brasil, como a RDC nº 657/2022 da ANVISA (2022a), é um passo importante, mas ainda precisa amadurecer para definir claramente os papéis e as obrigações de cada agente nesse novo ecossistema.

O Consentimento Livre e Esclarecido do paciente adquire aqui contornos específicos. Antes de se submeter à triagem autônoma, o paciente tem o direito de ser informado de que sua retina será avaliada por um algoritmo, e não por um médico (DANTAS; NOGAROLI, 2020; NORAGOLI, 2023). O não cumprimento desse dever de informação leva à obrigação de indenizar (NOGAROLI, 2023). Essa informação deve incluir as limitações da tecnologia, como a opacidade de seu funcionamento (“caixa-preta”), as taxas de acurácia (sensibilidade e especificidade) e, crucialmente, o risco de vieses algorítmicos. Algoritmos treinados predominantemente com dados de uma determinada etnia podem apresentar desempenho inferior em outras, o que é particularmente relevante para um país miscigenado como o Brasil. Um sistema menos acurado para populações negras ou indígenas, por exemplo, perpetuaria e agravaria as desigualdades em saúde (PRICE II; COHEN, 2019; AITH, 2024). A comunicação honesta sobre esses riscos é um pré-requisito para a validade do consentimento e para a legitimidade ética da implementação da tecnologia em larga escala (COHEN; SLOTTJE, 2024; BOSELLI, 2024).

Adicionalmente, limitações técnicas, como a qualidade da imagem, podem impactar o resultado. Condições como catarata ou pupilas pequenas podem impedir a captura de uma imagem adequada para análise, exigindo protocolos claros para o encaminhamento manual desses casos (OLIVEIRA; SILVA *et al.*, 2024). A presença de opacidades do cristalino devido à catarata aumenta significativamente o número de tentativas de imagem necessárias para obter

imagens de qualidade suficiente, bem como a necessidade de dilatação farmacológica. No entanto, em um ensaio pivotal, foi obtida qualidade de imagem suficiente em 96,1% dos participantes. Se um operador não conseguir capturar três imagens de qualidade suficiente, sem dilatação farmacológica, o sistema de IA explicitamente aconselha o uso de gotas de dilatação. Em um ensaio, 4% dos participantes receberam um resultado de qualidade de imagem insuficiente do sistema de IA, mesmo após a conclusão do protocolo do sistema (ABRÀMOFF; LAVIN *et al.*, 2018).

Por fim, a integração dessas tecnologias no Sistema Único de Saúde (SUS), embora promissora, depende de políticas públicas que garantam não apenas a aquisição dos sistemas, mas também a validação em coortes locais, a capacitação de profissionais e a criação de fluxos de referência eficientes para os pacientes triados, a fim de que a promessa de equidade não se transforme em uma nova fonte de disparidade (DANTAS; JEREISSATI, 2024). A falta de treinamento ou de conhecimento aprofundado dos médicos sobre os benefícios e riscos das tecnologias pode se traduzir em piores resultados para os pacientes. A IA deve ser uma ferramenta complementar, e o rastreamento, mesmo com falsos-positivos, deve promover o encaminhamento para o especialista, para um tratamento mais precoce (NOGAROLI, 2023; OLIVEIRA; SILVA *et al.*, 2024).

2.2.1.6 *A Lupa Digital: IA na Triagem de Câncer de Pele*

Panorama Geral e Potencialidades:

O câncer de pele, incluindo sua forma mais letal, o melanoma, figura entre as neoplasias de maior incidência no Brasil e no mundo (GUANDALINI *et al.*, 2024). A dermatoscopia, exame que utiliza uma lente de aumento especial (dermatoscópio), é a ferramenta padrão para a avaliação de lesões cutâneas suspeitas, mas sua acurácia é altamente dependente da experiência do dermatologista (GUANDALINI *et al.*, 2024). Nesse cenário, a inteligência artificial tem demonstrado um potencial extraordinário para aumentar a precisão diagnóstica. Utilizando redes neurais convolucionais (CNNs) treinadas com vastos bancos de imagens, como o *ISIC (International Skin Imaging Collaboration)* e o *HAM10000*, os algoritmos conseguem classificar lesões com uma precisão comparável, e por vezes superior, à de dermatologistas experientes (CAZANI; ALPINO *et al.*, 2025).

Estudos recentes e relatos da Sociedade Brasileira de Dermatologia (SBD) apontam melhorias notáveis: o dermatologista Renato Soriani, membro da SBD, destacou um estudo no *JAMA Dermatology* (WINKLER *et al.*, 2023) mostrando que um sistema de IA elevou a

sensibilidade para distinguir nevos (pintas) de melanomas de 84% para 100%, e a especificidade de 72,1% para 83,7%¹⁶. A tecnologia se mostra especialmente útil como ferramenta de apoio para médicos generalistas ou dermatologistas menos experientes, cuja precisão diagnóstica pode saltar significativamente com o auxílio da IA. Plataformas como a NEVO, desenvolvida no Brasil, utilizam fotos de celular para realizar uma triagem inicial com 93% de acurácia, funcionando como um ponto de acesso rápido ao sistema de saúde e acelerando o encaminhamento para especialistas. O sistema aponta o nível de risco e envia a informação diretamente para o médico responsável (não para o paciente), funcionando como um ponto de triagem antes do dermatologista. O objetivo, portanto, não é substituir o médico, mas atuar como uma poderosa ferramenta de auxílio à decisão, reduzindo biópsias desnecessárias e, crucialmente, salvando vidas através da detecção precoce¹⁷.

Figura 7 – IA em TBP (*Total Body Photography*) – “Fotografia de Corpo Inteiro”



Fonte: BRANCACCIO *et al.* (2024)

Nota: Comparação de duas imagens sequenciais de TBP de um paciente de 36 anos com histórico pessoal de melanoma em estágio inicial. A IA detectou rapidamente a nova lesão (seta preta) que apareceu nas costas após um acompanhamento de seis meses. O sujeito da Figura consentiu com a publicação da imagem. TBP: fotografia de corpo total.

Riscos Jurídicos e Éticos:

O maior risco ético-jurídico da IA na dermatologia é o viés algorítmico relacionado ao tom de pele. A esmagadora maioria dos grandes bancos de dados de imagens dermatoscópicas é composta por lesões em pele clara, de populações de ancestralidade europeia ou norte-americana (ADAMSON; SMITH, 2018; GUANDALINI *et al.*, 2024; CAZANI; ALPINO *et*

¹⁶ FONTE: <https://www.sbd.org.br/no-dia-do-medico-sbd-ressalta-o-impacto-da-inteligencia-artificial-no-diagnostico-dermatologico/>. Acesso em: 5 ago. 2024

¹⁷ IA ajuda dermatologistas a aprimorar diagnósticos de lesões suspeitas. Disponível em: <https://medicinas.com.br/ia-dermatologistas>. Acesso em: 5 ago. 2024.

al., 2025). Isso cria um risco real e documentado de que os algoritmos tenham um desempenho inferior em peles negras e de outros fototipos, o que pode levar a diagnósticos tardios ou incorretos em populações minoritárias, exacerbando as disparidades raciais em saúde (PASQUALE, 2015; FORD; PRICE II, 2016). A criação e validação de bancos de dados que reflitam a diversidade da população brasileira é uma condição *sine qua non* para a implementação ética e segura dessas ferramentas, como se observa dos tópicos anteriores (GUANDALINI *et al.*, 2024; AITH, 2024). Segundo Willian Boelcke, fundador e CEO da AI Pathology, *startup* brasileira que criou a ferramenta NEVO, para o treinamento da IA, foram usados bancos de imagens de pele disponibilizados por instituições como AC Camargo, USP Ribeirão Preto, Hospital das Clínicas, Universidade Federal de Minas Gerais (UFMG), entre outros, que pode mitigar o viés algorítmico¹⁸.

O Consentimento Livre e Esclarecido do paciente, neste contexto, deve ir além da autorização para o exame. O paciente tem o direito de ser informado de que uma ferramenta de IA será usada, que sua eficácia pode variar conforme o tom de pele devido a vieses nos dados de treinamento e que o resultado é um suporte à decisão, não um diagnóstico definitivo (DANTAS; NOGAROLI, 2020; NORAGOLI, 2023). A opacidade do algoritmo (“caixa-preta”) impede que o médico explique *como* a IA chegou à sua recomendação, exigindo uma comunicação transparente sobre essa limitação. Omitir a informação sobre o risco de viés racial pode viciar o consentimento, tornando-o juridicamente questionável (COHEN, 2019; COHEN; SLOTTJE, 2024; BOSELLI, 2024).

Do ponto de vista da responsabilidade civil, a IA permanece como uma ferramenta de apoio, e a decisão final sobre a conduta a ser tomada (biopsiar, acompanhar, tranquilizar) continua sendo um ato médico, sob a responsabilidade do profissional (ARAÚJO, 2024). No entanto, a IA introduz uma nova dinâmica, como já mencionado anteriormente: se um médico discorda de uma IA altamente precisa e um erro ocorre, ele pode ser questionado por não seguir a recomendação da máquina, potencialmente alterando o padrão de cuidado (*standard of care*). Inversamente, se ele segue a IA e há um erro, pode ser questionado por confiar cegamente na tecnologia sem exercer seu julgamento clínico (NOGAROLI, 2023).

Isso cria uma situação jurídica complexa: (i) *Risco I - Discordar da IA altamente precisa (Potencial Mudança no Padrão de Cuidado)* - O médico tem a liberdade de escolher seus meios de diagnóstico e terapia, mas também é responsável por suas escolhas. Se um erro ocorrer após rejeitar a sugestão da máquina, a conduta pode ser questionada. Se a IA é precisa e recomenda

¹⁸ FONTE: <https://www.cnnbrasil.com.br/saude/ia-esta-sendo-usada-para-o-diagnostico-precoce-do-cancer-de-pele>. Publicado em: 9 set. 2025. Acesso em: 5 out. 2025.

o tratamento padrão, mas o profissional ignora o resultado e prescreve algo inadequado, causando lesão, o médico, em princípio, será responsabilizado. Nesse caso, ele terá o ônus de justificar por que desconsiderou a recomendação algorítmica (NOGAROLI, 2023). A falta de diligência do profissional em descartar irrefletidamente um resultado fora do padrão obtido pela IA, poderá constituir um critério para a sua responsabilização. Por exemplo, se a IA propõe um tratamento inovador e mais eficaz baseado em milhões de estudos, o médico tem o dever de ser diligente ao ponto de investigar se não é necessário repensar a forma de tratamento inicialmente correta. Em um evento adverso, o profissional estará na complexa posição de justificar: por que ele seguiu ou por que ele se desviou da recomendação algorítmica (NOGAROLI, 2023); e (ii) *Risco 2 - Seguir cegamente a IA sem julgamento clínico (Viés da Automação)*: o excesso de confiança na tecnologia é reconhecido como um novo risco que pode levar a um diagnóstico culposo. Erros de julgamento decorrentes da influência da IA são conhecidos como viés da automação. A heurística da automação não deve ser utilizada pelos médicos no processo de tomada de decisão diagnóstica. Há um temor concreto de que o médico possa acabar se utilizando da heurística da automação, aceitando acriticamente a sugestão diagnóstica. Se o médico confia cegamente na IA, desconsiderando o padrão de tratamento ou a sua própria experiência clínica, sem proceder à correta anamnese e análise crítica das informações, ele agirá com culpa pela falta de diligência que resultou em danos ao paciente. O médico será civilmente responsável pelo erro de diagnóstico danoso ao qual culposamente der causa. A conduta omissiva (negligência) pode ser verificada a partir da inércia do médico em obter ou analisar informações que eram acessíveis para o correto diagnóstico (NOGAROLI, 2023; FACCHINI NETO; BARBOSA, 2023).

Por fim, a proteção de dados dos pacientes, regida pela LGPD (Lei nº 13.709/2018), é crucial, pois as imagens de lesões são dados de saúde sensíveis e devem ser tratadas com rigorosa segurança e finalidade específica.

2.2.1.7 *Correndo Contra o Tempo: IA na Detecção e Manejo da Sepsis*

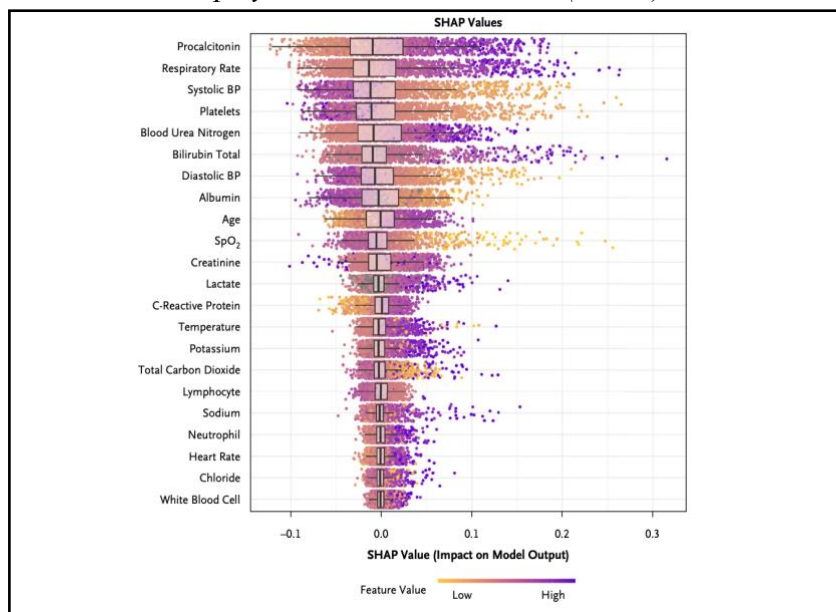
Panorama Geral e Potencialidades:

A sepsis, uma resposta inflamatória desregulada do corpo a uma infecção, é uma emergência médica com altíssimas taxas de mortalidade, onde cada hora no diagnóstico e tratamento é crucial. Sua apresentação clínica heterogênea, no entanto, torna o reconhecimento precoce um desafio constante para as equipes de saúde (BHARGAVA; LOPEZ-ESPINA *et al.*, 2024). A inteligência artificial surge como uma ferramenta vital para superar esse desafio. Por

meio da análise contínua de grandes volumes de dados provenientes de prontuários eletrônicos (*Electronic Health Records - EHR*) – como sinais vitais, resultados de exames laboratoriais e histórico do paciente – algoritmos de aprendizado de máquina podem identificar padrões sutis que precedem a manifestação clínica da sepse (FLEUREN; KLAUSCH *et al.*, 2020).

Esses sistemas funcionam como uma vigilância inteligente, emitindo alertas precoces para a equipe médica e permitindo intervenções mais rápidas, como a administração de antibióticos, que são fundamentais para melhorar os desfechos. Um exemplo concreto é o *Sepsis ImmunoScore*, a primeira ferramenta diagnóstica baseada em IA para sepse a receber autorização da FDA em abril de 2024. Este sistema foi projetado para se integrar aos *EHRs* (“Registros Eletrônicos de Saúde”) e identificar, em até 24 horas, pacientes com alta probabilidade de desenvolver sepse ou de ter uma progressão do quadro, auxiliando na estratificação de risco e na gestão clínica. O potencial para reduzir a mortalidade e otimizar o uso de recursos em unidades de terapia intensiva (UTI) é imenso (BHARGAVA; LOPEZ-ESPINA *et al.*, 2024).

Figura 8 – Gráficos de dispersão dos valores *SHapley Additive exPlanations (SHAP)*



Fonte: BHARGAVA; LOPEZ-ESPINA *et al.* (2024)

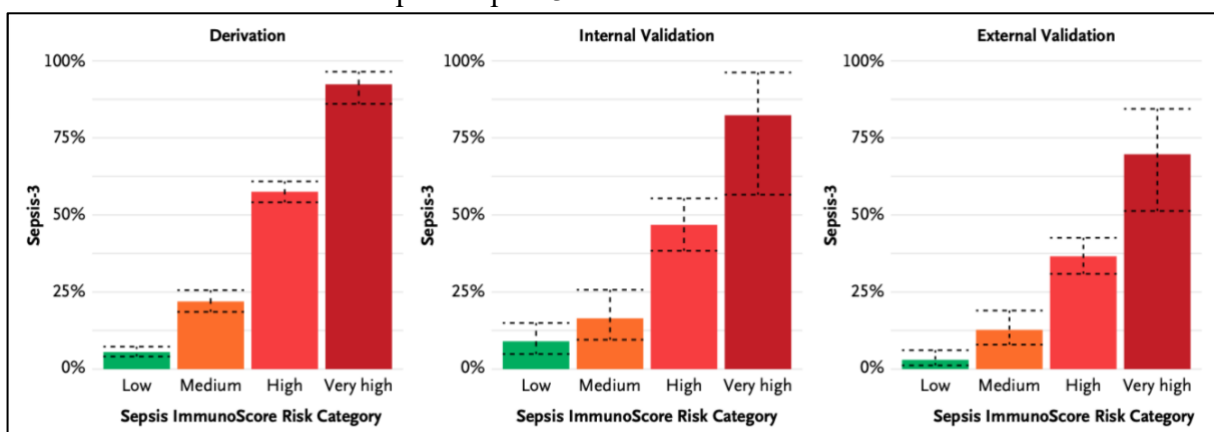
Nota: Gráficos de dispersão dos valores *SHapley Additive exPlanations (SHAP)* intervencionais (uma medida da contribuição das características para o resultado do algoritmo) para cada característica de entrada em todos os pacientes da coorte de derivação. As características estão listadas no eixo y, em ordem decrescente de importância, com as características que mais contribuem no topo da lista, conforme determinado pelo valor *SHAP* intervencional absoluto médio entre os pacientes. Cada ponto de dados é colorido de acordo com seu valor de medição padronizado observado: roxo indica medições elevadas, enquanto

amarelo indica medições mais baixas. Os valores individuais coloridos em roxo oferecem as contribuições mais impactantes *e.g.*, valores elevados de procalcitonina e frequência respiratória e valores baixos de pressão arterial sistólica e plaquetas ofereceram contribuições substanciais). Os pontos de dados cinza representam parâmetros que foram imputados durante a geração do resultado do *Sepsis ImmunoScore*. BP denota pressão arterial; SHAP, Shapley Additive exPlanations; e SpO₂, saturação de oxigênio no sangue (BHARGAVA; LOPEZ-ESPINA *et al.*, 2024).

BHARGAVA; LOPEZ-ESPINA *et al.* (2024) ilustram bem a importância das características do *Sepsis ImmunoScore* para a detecção e o manejo da Sepse na figura acima, que representa como o algoritmo de *machine learning* (ML) toma sua decisão de detecção.

A figura 8 exibe gráficos de dispersão dos valores SHAP (*SHapley Additive exPlanations*), uma métrica que mede a contribuição de cada variável de entrada (característica) para a saída do algoritmo. Mecanismo: o gráfico lista até 22 parâmetros de entrada (demográficos, sinais vitais e testes laboratoriais) em ordem decrescente de importância. Os pontos coloridos indicam como os valores da característica (roxo = elevado, amarelo = baixo) impactam o resultado (score de risco de sepse). Influência Central: a figura revela que os três parâmetros mais influentes para o *Sepsis ImmunoScore* foram a Procalcitonina (PCT), a Taxa Respiratória e a Pressão Arterial Sistólica. Essa figura é crucial por fornecer a interpretabilidade do modelo de IA (BHARGAVA; LOPEZ-ESPINA *et al.*, 2024).

Figura 9 – Estratificação do *Sepsis ImmunoScore* para Sepsis-3 em todas as coortes



Fonte: BHARGAVA; LOPEZ-ESPINA *et al.* (2024)

Nota: Os gráficos de barras mostram os conjuntos de dados das coortes de derivação, validação interna e validação externa para os critérios Sepsis-3, dentro dos valores preditivos de 24 horas para cada categoria de estratificação de risco do *Sepsis ImmunoScore*. As linhas tracejadas indicam os intervalos de confiança de 95% (BHARGAVA; LOPEZ-ESPINA *et al.*, 2024).

Esta figura 9, intitulada pelos mesmos autores como “*Sepsis ImmunoScore Stratification for Sepsis-3 in All Cohorts*” (Estratificação do *Sepsis ImmunoScore* para *Sepsis-3* em Todas as Coortes), é a representação mais direta do desempenho diagnóstico do sistema de IA. Ela utiliza gráficos de barras para mostrar o Valor Preditivo (PV) de cada uma das quatro categorias de risco (Baixo, Médio, Alto e Muito Alto) do *Sepsis ImmunoScore* em relação à ocorrência de sepse (*Sepsis-3*) dentro de 24 horas.

Os resultados são apresentados para as três coortes (Derivação, Validação Interna e Validação Externa), comprovando a capacidade do modelo de IA de identificar e estratificar o risco de sepse de maneira consistente. O aumento gradual e significativo do percentual de pacientes que, de fato, desenvolveram sepse (PV), conforme o risco categorizado pela IA, aumenta (de Baixo para Muito Alto) e demonstra visualmente a alta precisão diagnóstica do *Sepsis ImmunoScore* (AUC de 0.81 na validação externa). Ou seja, a importância da Figura 9 reside no fato de que o risco aumenta de forma consistente e escalonada (relação monotônica crescente) em todos os três grupos de pacientes: (i) *Risco Alto/Muito Alto* - O *Sepsis ImmunoScore* é um forte preditor de que a sepse está iminente, ou seja, tem um Alto Valor Preditivo Positivo. A Razão de Verossimilhança (*Likelihood Ratio - LR*) para o risco Muito Alto foi de 8,3, o que é um indicador muito forte da presença de *Sepsis-3* em 24 horas; e (ii) *Risco Baixo* - O sistema é eficaz para “descartar” a Sepse (LR de 0,1). Isso ajuda os médicos a evitar o *over-triage* (excesso de triagem) e potencialmente a tratar pacientes de baixo risco em regime ambulatorial.

Em suma, a Figura 9 demonstra que a IA não só detecta a Sepse com alta precisão (Área Sob a Curva - AUC de 0,81 na validação externa), mas o mais importante, ela estratifica o risco de forma confiável em bandas de risco que se correlacionam diretamente com a probabilidade real de o paciente desenvolver a condição (BHARGAVA; LOPEZ-ESPINA *et al.*, 2024).

Tabela 2 – Estratificação de risco para sepse segundo o *Sepsis ImmunoScore*

| Nível de Risco (<i>Sepsis ImmunoScore</i>) | Taxa de Pacientes que Tiveram Sepse em 24h (Valor Preditivo - PV) | Razão de Verossimilhança (<i>Likelihood Ratio - LR</i>) | Significado |
|--|---|---|---|
| Low (Baixo) | 3,0% | 0,1 | Ajuda a “excluir” a sepse (<i>rule-out</i>), permitindo considerar o manejo ambulatorial. |
| Medium (Médio) | 12,7% | 0,5 | Risco intermediário de sepse. |
| | | | Alto risco. Serve como um teste “ <i>rule-in</i> ” (confirmação) para |

| | | | |
|-------------------------------|-------|-----|---|
| High (Alto) | 36,6% | 2,1 | suspeita clínica, indicando a necessidade de intervenção imediata. |
| Very High (Muito Alto) | 69,7% | 8,3 | Risco extremamente alto de desenvolver sepse em 24 horas. Requer escalonamento rápido do cuidado e administração imediata de antimicrobianos. |

Fonte: BHARGAVA; LOPEZ-ESPINA *et al.* (2024).

Riscos Jurídicos e Éticos:

A eficácia dos sistemas de IA para sepse depende diretamente da qualidade e integridade dos dados dos prontuários eletrônicos. Dados incompletos, inconsistentes ou incorretamente inseridos podem levar a predições falhas, com consequências graves (DONEDA; SOUZA *et al.*, 2018). A responsabilidade pela curadoria e governança desses dados é um desafio para as instituições de saúde, que se tornam corresponsáveis pela segurança do algoritmo ao fornecerem a matéria-prima para sua operação (FLEUREN; KLAUSCH *et al.*, 2020; NOGAROLI, 2023).

No âmbito da regulamentação de risco, os operadores (responsáveis pela implantação) de sistemas de IA de alto risco devem adotar medidas para garantir o uso correto da tecnologia e, se exercerem controle sobre os dados de entrada, devem assegurar que esses dados sejam pertinentes e suficientemente representativos para a finalidade prevista (Art. 26, nº 4 e 5, AI Act). A correta gestão de riscos no setor de saúde exige que as organizações avaliem continuamente se a coleta e o tratamento de dados pela IA estão ocorrendo de maneira regular, dada a natureza dos dados de saúde como combustíveis para os algoritmos. O princípio da *accountability* (prestação de contas) impõe um dever de cuidado e vigilância, que atua como guia para programadores, fabricantes e usuários da IA, incluindo a curadoria de dados (NOGAROLI, 2023).

Outro risco significativo é a “fadiga de alerta” (*alert fatigue*) – (BIGNAMI *et al.*, 2025; BAYSARI *et al.*, 2025). O fenômeno é reconhecido como um problema que ocorre quando o usuário fica insensibilizado a alarmes, devido à exposição excessiva, o que pode resultar em alertas perdidos (*missed alarms*) (CANADÁ, 2025). Estudos em Unidades de Terapia Intensiva (UTIs) revelaram um alto percentual de falsos positivos (como 88.8% para arritmias), resultando no fato de as equipes clínicas ignorarem, rotineiramente, os alertas, minando a eficácia da ferramenta e criando um risco de omissão em casos reais (HASANZADEH, 2025). A calibração do limiar de alerta do algoritmo é, portanto, uma decisão clínico-gerencial de grande impacto ético e jurídico (GRAAFSMA; MURPHY *et al.*, 2024).

A questão do Consentimento Livre e Esclarecido do paciente é particularmente complexa. Ao ser admitido em um hospital que utiliza essa tecnologia, o paciente (ou seu representante) consente com o tratamento médico geral, mas está ciente de que está sob uma forma de vigilância algorítmica contínua? Esta prática configura a chamada “opacidade pela não revelação”, onde o paciente sequer sabe que a tecnologia está sendo usada para apoiar a decisão clínica, impedindo-o de questionar (NOGAROLI, 2023). De outro lado, a opacidade explicativa impede uma explicação clara sobre *como* a vigilância funciona, e o paciente não tem a opção de “não consentir” com o monitoramento sem, potencialmente, recusar o próprio cuidado. Isso levanta questões sobre a autonomia do paciente em um ambiente onde a tomada de decisão é cada vez mais mediada por sistemas invisíveis (COHEN, 2019; COHEN; SLOTTJE, 2024; DANTAS; JEREISSATI, 2024).

A atribuição de responsabilidade civil é igualmente complexa. Se a IA emite um alerta e a equipe médica, por sobrecarga ou descrença (“fadiga de alerta”), não age a tempo, a responsabilidade recai sobre a equipe por negligência (BIGNAMI *et al.*, 2025; BAYSARI *et al.*, 2025). Mas e se a IA *não* emitir um alerta (falso negativo) e o paciente evoluir para um choque séptico? A responsabilidade seria do desenvolvedor do *software* por um defeito do produto, nos termos do Código de Defesa do Consumidor? Ou da instituição por adotar uma ferramenta falha? A IA atua como um sistema de apoio, e a responsabilidade final permanece com a equipe médica, mas a presença do algoritmo adiciona uma camada de complexidade à análise de um eventual erro (NOGAROLI, 2023; ROSENVALD; MARTINS, 2024).

Por fim, a implementação desses sistemas exige um investimento significativo, levantando questões de equidade no acesso, pois hospitais com mais recursos teriam uma vantagem tecnológica na prevenção de uma condição tão letal, aprofundando as disparidades no sistema de saúde.

2.2.2 As Neurotecnologias e a Fronteira da Mente Humana

Se a IA diagnóstica analisa dados para inferir estados de saúde, as neurotecnologias representam um passo adiante: a interação direta com o sistema nervoso para monitorar, modular ou mesmo substituir funções neurais. Essas tecnologias, que incluem desde interfaces cérebro-computador a sistemas de neuroestimulação, prometem avanços terapêuticos para condições neurológicas e psiquiátricas até então intratáveis (DIAS; ANHOQUE *et al.*, 2024; MUSOLE, 2025). Contudo, ao tocarem no substrato biológico da consciência, do pensamento

e da identidade, elas levantam os mais profundos desafios éticos e jurídicos da nossa era. A tabela a seguir introduz os casos que serão explorados nesta seção.

Tabela 3 – Panorama das Aplicações de Neurotecnologias na Medicina (2023-2025)

| Aplicação (Caso de Uso) | Tecnologia Principal | Potencialidades (Clínica e Operacional) | Principais Riscos Jurídicos e Éticos |
|--|---|---|--|
| Interfaces Cérebro-Computador (BCI) para ELA | BCI intracraniana com IA (Rede Neural Recorrente) para decodificar a “tentativa de fala”. | Clínica: Restaura a comunicação em pacientes com paralisia severa a uma velocidade (62 ppm) próxima da conversação natural. Operacional: Aumenta a autonomia e a conexão social do paciente. | Neurodireitos: Risco à privacidade mental (leitura de intenções). Consentimento Livre e Esclarecido: Desafiado pela opacidade (“dupla caixa-preta”), vulnerabilidade extrema do paciente e invasividade da neurocirurgia. Responsabilidade Civil: Complexa em caso de erro de decodificação. |
| Estimulação Cerebral Profunda (DBS) para Alzheimer | Neuromodulação invasiva (eletrodos cerebrais) para modular circuitos neurais da memória e cognição. | Clínica: Melhora sintomas neuropsiquiátricos (agitação, apatia) e a capacidade de autocuidado em casos severos; busca desacelerar o declínio cognitivo em casos leves. | Consentimento Livre e Esclarecido: Desafio crítico devido à capacidade de consentir comprometida do paciente; exige protocolos rigorosos e diretivas antecipadas. Neurodireitos: Risco à identidade e continuidade psicológica por modulação cerebral. Equidade: Alto custo restringe o acesso. |
| Neuroprótese para Doença de Parkinson | Estimulação elétrica epidural (EES) sobre a medula espinhal para restaurar a ativação dos neurônios motores da perna. | Clínica: Restaura uma marcha fluida e estável, reduzindo o “congelamento” (<i>freezing</i>) e quedas. Operacional: Devolve a autonomia e melhora drasticamente a qualidade de vida. | Consentimento Livre e Esclarecido: Exige diálogo sobre agência híbrida (homem-máquina), dependência tecnológica e a fronteira entre terapia e aprimoramento. Responsabilidade Civil: Fato do produto (CDC) em caso de defeito do dispositivo. Equidade: Alto custo como barreira de acesso. |
| Neuroimagem e IA para Diagnóstico de Alzheimer (DANMLP) | IA multimodal que integra dados de neuroimagem (sMRI), clínicos e | Clínica: Aumenta a acurácia diagnóstica na distinção entre DA, MCI e envelhecimento normal. Operacional: | Privacidade: Risco elevado pela criação de um perfil de dados multimodal e extremamente sensível |

| | | | |
|--|---|---|--|
| | genéticos para diagnóstico precoce. | Potencial para otimizar a triagem e o planejamento de cuidados. | (LGPD). Viés Algorítmico: Risco de associações espúrias e discriminatórias. Consentimento Livre e Esclarecido: Desafiado pela complexidade e opacidade do algoritmo e pela vulnerabilidade da população-alvo. |
| “Ponte Digital” Cérebro-Coluna (BSI) para Lesão Medular | BCI sem fio que decodifica a intenção de movimento no cérebro e a transmite para um neuroestimulador na medula espinhal. | Clínica: Restaura o controle voluntário da marcha em pacientes com paraplegia; promove neuroplasticidade e recuperação neurológica parcial. | Neurodireitos: Risco à privacidade mental e à segurança (vulnerabilidade a <i>brain hacking</i>). Consentimento Livre e Esclarecido: Desafiado pela agência híbrida, vulnerabilidade cibernética e extrema vulnerabilidade do paciente. Responsabilidade Civil: Complexa em caso de erro de decodificação. |
| Neuroestimulação em Malha Fechada para Epilepsia Refratária | Sistema autônomo implantável (RNS) que monitora a atividade cerebral, detecta e interrompe crises epiléticas. | Clínica: Reduz significativamente a frequência das convulsões em pacientes farmacorresistentes. Operacional: Melhora a qualidade de vida e a segurança do paciente. | Consentimento Livre e Esclarecido: Desafiado pela ação autônoma e vigilância contínua do dispositivo e pelos efeitos incertos na subjetividade. Privacidade: Coleta constante de neurodados sensíveis. Responsabilidade Civil: Complexa em caso de falha de detecção (falso negativo) ou estimulação indevida. |
| BCIs Invasivas (Neuralink - Telepathy NI e Blindsight) | <i>Telepathy:</i> BCI intracortical com 1.024 eletrodos para ler a intenção motora. <i>Blindsight:</i> Prótese visual cortical que estimula o cérebro com base em imagens de uma câmera. | Clínica: Restaurar a autonomia digital para pacientes com paralisia (<i>Telepathy</i>) e potencialmente restaurar a visão em cegos (<i>Blindsight</i>). Operacional: Avanço na miniaturização e na implantação robótica de BCIs. | Consentimento Livre e Esclarecido: Criticamente desafiado pela vulnerabilidade do paciente (“desespero terapêutico”) e pela opacidade radical sobre os efeitos de longo prazo na mente e na identidade. Segurança e Continuidade do Cuidado: Risco de falha de <i>hardware</i> e dependência de uma empresa privada. Terapia vs. Aprimoramento: Objetivo explícito de |

| | | | |
|--|--|--|---|
| | | | “destravar o potencial humano” levanta dilemas de equidade. Neurodireitos: Risco à privacidade mental. |
|--|--|--|---|

Fonte: Elaborada pelo autor (2025).

2.2.2.1 Interfaces Cérebro-Computador (BCI): Restaurando a Fala em Pacientes com ELA

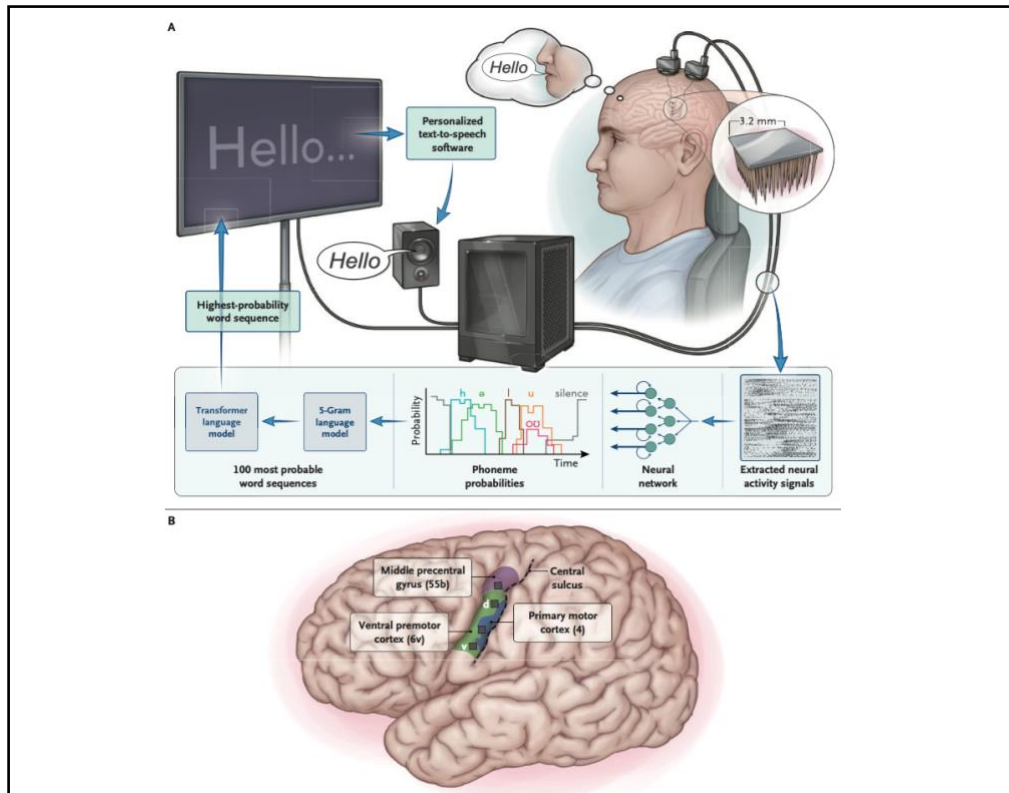
Panorama Geral e Potencialidades:

Em 2023, pesquisadores da Universidade de Stanford e da Universidade da Califórnia, São Francisco, publicaram estudos que representam um marco na história das neurotecnologias. Eles desenvolveram uma interface cérebro-computador (*BCI - Brain-Computer Interface*) capaz de decodificar a atividade cerebral associada à tentativa de fala (*attempted speech*) e convertê-la em texto em uma tela de computador em tempo real (CARD; WAIRAGKAR *et al.*, 2024).

A tecnologia visa restaurar a comunicação para pessoas com paralisia, uma vez que a *BCI* é um tipo promissor de tecnologia de comunicação que pode decodificar diretamente a fala pretendida pelo usuário, a partir de sinais neurais. Um dos estudos mais notáveis envolveu um participante com Esclerose Lateral Amiotrófica (ELA), uma doença neurodegenerativa que leva à perda da capacidade de falar (WILLETT *et al.*, 2023; METZGER *et al.*, 2024).

O sistema utiliza matrizes de microeletrodos, implantados cirurgicamente em regiões do córtex cerebral, responsáveis pela produção da fala. Esses sensores captam os sinais neurais gerados quando o paciente tenta articular palavras. Um sofisticado algoritmo de IA, especificamente uma Rede Neural Recorrente (*RNN*), é treinado para reconhecer os padrões neurais correspondentes a cada fonema. Um modelo de linguagem, que funciona como um sistema de autocorreção avançado, analisa o fluxo de fonemas decodificados e os converte em palavras e frases coerentes. O resultado foi a decodificação da fala a uma velocidade de 62 palavras por minuto, mais de três vezes o recorde anterior, aproximando-se do ritmo da conversação natural e representando uma esperança de restauração da comunicação e da conexão com o mundo para pessoas com paralisia severa (WILLETT *et al.*, 2023; CARD; WAIRAGKAR *et al.*, 2024).

Figura 10 – Localização dos eletrodos e configuração da decodificação da fala



Fonte: CARD; WAIRAGKAR *et al.* (2024)

Nota: O painel **A** mostra a neuroprótese de conversão de fala do cérebro em texto (*Brain-to-Text Speech Neuroprosthesis*). A atividade elétrica é medida com o uso de quatro conjuntos de 64 eletrodos e processada para extrair a atividade neural. Técnicas de aprendizado de máquina, incorporando uma rede neural artificial, são usadas para decodificar os sinais de atividade neural, extraídos em um fonema inglês a cada 80 milissegundos. O painel **B** mostra as localizações aproximadas dos conjuntos de microeletrodos (quadrados cinzentos) sobrepostas numa reconstrução tridimensional do cérebro do participante. As regiões coloridas correspondem às áreas corticais alinhadas ao cérebro do participante com o uso do protocolo de ressonância magnética do Projeto *Human Connectome* antes da implantação. A implantação do dispositivo foi realizada em um centro de pesquisa em neurociência (CARD; WAIRAGKAR *et al.*, 2024).

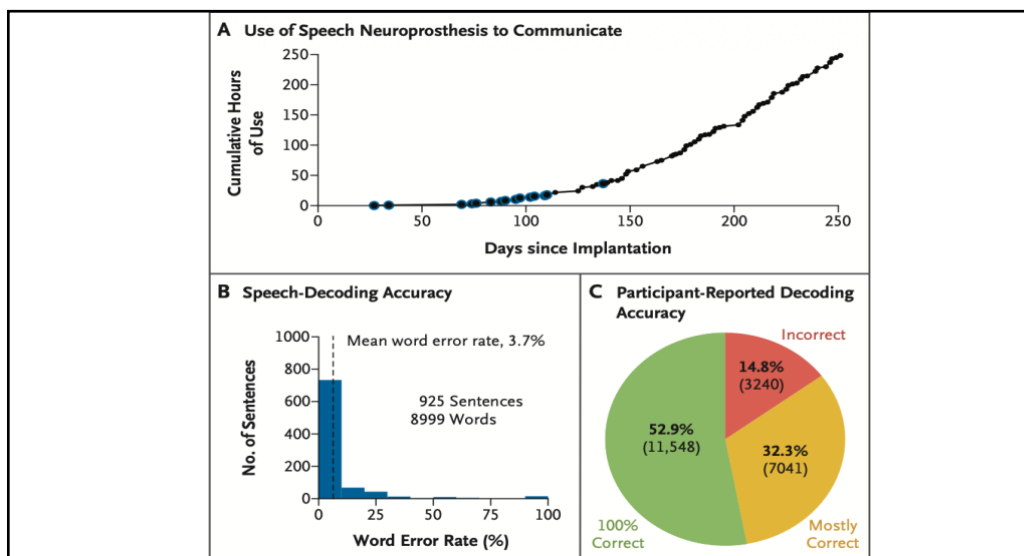
Figura 11 – Interface do usuário no modo de conversação



Fonte: CARD; WAIRAGKAR *et al.* (2024).

Nota: É mostrada uma fotografia do participante usando a neuroprótese de fala no modo de conversação. A neuroprótese detectou quando o participante estava tentando falar apenas com base na atividade neural e concluiu após 6 segundos de inatividade da fala ou após sua ativação opcional de um botão na tela por meio do rastreamento ocular. Após a finalização da frase decodificada, o participante selecionou os botões de confirmação na tela por meio do rastreamento ocular para indicar se a frase decodificada estava correta (CARD; WAIRAGKAR *et al.*, 2024).

Figura 12 – Uso da neuroprótese para fala iniciada pelo participante



Fonte: CARD; WAIRAGKAR *et al.* (2024)

Nota: O painel **A** mostra as horas acumuladas que o participante utilizou a neuroprótese de fala para se comunicar durante sessões de pesquisa estruturadas e uso pessoal. Para as sessões destacadas em azul, a precisão da decodificação no modo de conversação foi quantificada conforme mostrado no painel **B**. O painel **B** mostra um histograma da precisão da decodificação da fala em conversas para as 925 frases com rótulos verdadeiros conhecidos. A taxa média de erros por palavra foi de 3,7% (IC 95%, 3,3 a 4,3). O painel **C** mostra a precisão de decodificação relatada pelo participante para cada frase (que é diferente da taxa de erros por palavra) em todos os dados do modo de conversação (21.829 frases) (CARD; WAIRAGKAR *et al.*, 2024).

Riscos Jurídicos e Éticos:

Este caso de uso transcende a discussão sobre proteção de dados tradicional e torna o debate sobre neurodireitos uma necessidade jurídica concreta e urgente (ANPD, 2025). A tecnologia não está apenas processando dados sobre a saúde de uma pessoa; ela está acessando, lendo e decodificando o substrato neural do pensamento e da intenção de comunicação (BROWN, 2024; DANTAS; JEREISSATI, 2024). Isso implica diretamente a necessidade de proteger o Direito à Privacidade Mental, que resguarda a esfera íntima dos pensamentos contra a vigilância não consentida. Se o sistema pode decodificar a fala intencionada, qual a barreira que o impede de, no futuro, decodificar pensamentos não verbalizados? A LGPD, embora possa

classificar dados neurais como dados pessoais sensíveis (ANPD, 2025; OEA, 2023), pode não possuir a especificidade necessária para tutelar essa nova dimensão da privacidade, o que fundamenta as discussões sobre um arcabouço jurídico específico para os neurodireitos, como a PEC 29/2023 no Brasil (YUSTE; GOERING *et al.*, 2017; BRASIL, 2023; FAZANO; MARTINS, 2024).

A questão do Consentimento Livre e Esclarecido (CLE) do paciente emerge como o epicentro do debate bioético e jurídico. É crucial distinguir o consentimento para tratamento de dados (LGPD) do CLE para o procedimento médico. O primeiro rege o uso da informação – nesse contexto, o consentimento para o uso desses dados deve ser granular e dinâmico, especificando claramente para quais finalidades (clínica, pesquisa, desenvolvimento comercial) eles podem ser utilizados e permitindo sua revogação (BERTONI; IENCA, 2024); o segundo, a legitimidade da intervenção no corpo e na mente do paciente (BROWN, 2024). Para uma tecnologia como a *BCI*, o Consentimento Livre e Esclarecido é desafiado em múltiplos níveis:

Opacidade e Incerteza: a “dupla caixa-preta” (a opacidade do algoritmo somada à inescrutabilidade do cérebro) impede que médicos e pesquisadores expliquem completamente como a *BCI* funciona e quais são seus efeitos a longo prazo sobre a cognição, a personalidade e a própria identidade do paciente (KLEIN, 2016). Como obter um consentimento válido para riscos que são, em grande parte, desconhecidos?

Vulnerabilidade Extrema: pacientes com paralisia severa encontram-se em um estado de extrema vulnerabilidade. A promessa de restaurar a comunicação pode criar uma pressão indireta, afetando a voluntariedade do consentimento e levantando o risco do que se chama de “desespero terapêutico” (BURWELL *et al.*, 2017).

Invasividade: a invasividade dos procedimentos relacionados a interfaces cérebro-computador (*BCI*) acrescenta um nível significativo de complexidade ética e legal ao Consentimento Livre e Esclarecido. Essas tecnologias frequentemente exigem neurocirurgias de alto risco, como implante intracraniano de eletrodos, cujas potenciais complicações – incluindo infecção, hemorragia, formação de tecido cicatricial e rejeição biológica – devem ser exaustivamente detalhadas aos pacientes antes da obtenção do consentimento (GLANNON, 2014). Leuthardt *et al.* (2021) propõem uma classificação semântica clara dos níveis de invasividade: não invasivo, implantado e intracraniano, cada qual acompanhada por diferentes perfis de risco clínico e conseqüente necessidade de explicação do prognóstico aos pacientes. Essa distinção é fundamental, pois a simples menção de “*invasivo*” não transmite adequadamente os riscos que envolvem um procedimento intracraniano, podendo comprometer

a compreensão real do paciente sobre o que está sendo consentido e, conseqüentemente, a validade do Consentimento Livre e Esclarecido.

Como se pode notar, a obtenção de um consentimento, verdadeiramente esclarecido para uma tecnologia tão nova e invasiva, cujos efeitos a longo prazo sobre a cognição, a personalidade e a própria identidade são desconhecidos, representa um desafio bioético monumental (LEITE; REIS *et al.*, 2021).

A titularidade e o controle dos dados neurais também são pontos críticos: quem é o “proprietário” dos sinais cerebrais captados? O paciente, a universidade, a empresa que comercializará a tecnologia? A definição da titularidade é crucial para garantir a autonomia do paciente sobre sua própria atividade mental (IENCA; ANDORNO, 2017). A categoria de neurodireitos, impulsionada por cientistas como Rafael Yuste, estabelece que todos os dados neurais obtidos devem ser mantidos em sigilo, e a venda, transferência e uso comercial devem ser estritamente regulamentados (DANTAS; JEREISSATI, 2024). Um dos princípios centrais é o Direito à Privacidade Mental¹⁹, que exige a proteção da informação cerebral contra o acesso não autorizado ou não desejado (NEGRI *et al.*, 2025).

Finalmente, a responsabilidade civil por dano neural emerge como um novo desafio (OEA, 2023; DANTAS; JEREISSATI, 2024). Sendo um procedimento invasivo, há riscos de danos físicos ao cérebro. Mas e se o dano for psicológico ou social? Por exemplo, se a *BCI* interpretar mal a intenção do paciente e gerar uma fala indesejada ou ofensiva, causando constrangimento, prejuízo financeiro ou rompimento de relações, a quem responsabilizar? A cadeia de responsabilidade é complexa, envolvendo o neurocirurgião, a equipe de pesquisa, o desenvolvedor do *software* e o fabricante do dispositivo (LOPES *et al.*, 2022). O Projeto de Princípios Interamericanos sobre Neurociência, Neurotecnologias e Direitos Humanos (OEA, 2023) inclui o Princípio 10: Acesso à tutela efetiva e acesso a remédios associados ao desenvolvimento e uso das neurotecnologias. Este princípio prevê o acesso à justiça e o estabelecimento de procedimentos legais para reparação em caso de vulneração de direitos. A Declaração da OEA de 2021 aponta a preocupação com o acesso a remédios e proteção judicial em caso de danos. A Carta de Direitos Digitais da Espanha (um marco de referência) destaca a necessidade de criação de procedimentos para a resolução alternativa de conflitos que possam surgir no entorno digital. Isso demonstra que organismos internacionais têm se manifestado sobre a necessidade de garantir a responsabilização e acesso a remédios (LOPES *et al.*, 2022).

¹⁹ NEURORIGHTS FOUNDATION. **Our Approach: The Five Neurorights**. Disponível em: <https://www.neurorightsfoundation.org/mission/impact>. Acesso em: 10 nov. 2024.

2.2.2.2 Estimulação Cerebral Profunda (DBS) na Doença de Alzheimer

Panorama Geral e Potencialidades:

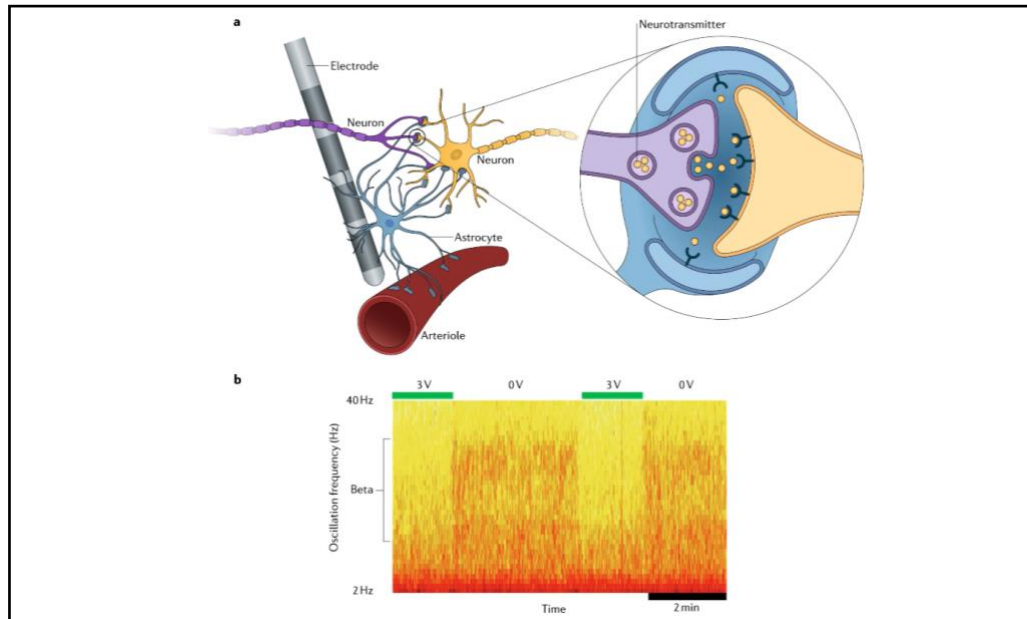
A Estimulação Cerebral Profunda (*DBS - Deep Brain Stimulation*), técnica de neuromodulação já estabelecida para tratar distúrbios do movimento, como a Doença de Parkinson (Síndrome de Tourette, Distonia etc.), consiste no implante cirúrgico de eletrodos em áreas específicas do cérebro, os quais emitem pulsos elétricos para modular circuitos neurais disfuncionais (LOZANO *et al.*, 2019; GUERRA; SATHLER *et al.*, 2025). Recentemente, pesquisadores têm explorado seu potencial para tratar a Doença de Alzheimer (DA) e outras demências.

Um estudo clínico prospectivo e não randomizado, publicado em 2024, investigou o uso de *DBS* bilateral no Núcleo Basal de Meynert (NBM) em 6 pacientes com Doença de Alzheimer em estágio severo (escore MMSE – Mini-Exame do Estado Mental – de 0 a 10, nível CDR – Escala de Avaliação Clínica de Demência – 3). O NBM é um alvo promissor devido à degeneração de neurônios colinérgicos nessa área (crucial para a memória e cognição), que é uma alteração patológica característica da DA e causa importante de comprometimento funcional. Os resultados mostraram que o procedimento foi seguro, sem eventos adversos graves. Embora a estimulação não tenha revertido a progressão da doença, trouxe benefícios significativos: houve melhora nos sintomas neuropsiquiátricos, como agitação e apatia, e aumento na capacidade de autocuidado dos pacientes (XU; LIU; SHANG *et al.*, 2024; 2025).

Outros ensaios, como o ADvance II (fase II)²⁰, avaliaram a *DBS* bilateral direcionada ao fórnix (*DBS-f*), um importante feixe de fibras no circuito da memória do cérebro, em 42 pacientes com DA leve, visando desacelerar o declínio cognitivo. Embora não tenha havido diferenças significativas nos resultados cognitivos (ADAS-Cog 13, CDR-SB) para a coorte total após 12 meses, a análise *post-hoc* sugeriu que pacientes com 65 anos de idade ou mais podem ter se beneficiado com o tratamento *DBS-f ON*, enquanto houve uma possível piora em pacientes com menos de 65 anos. Os achados apoiam a avaliação contínua da *DBS-f* em pacientes idosos com DA e o cálculo do tamanho da amostra para um ensaio clínico de fase III (LOZANO *et al.*, 2024).

Figura 13 – Mecanismos de estimulação cerebral profunda

²⁰ ADvance II: A 12-month Double-blind, Randomized, Controlled Study to Evaluate the Safety and Efficacy of Deep Brain Stimulation of the Fornix (DBS-f) in Patients With Mild Probable Alzheimer's Disease. Disponível em: <https://clinicaltrials.gov/study/NCT03622905>. Acesso em: 12 jul. 2025.

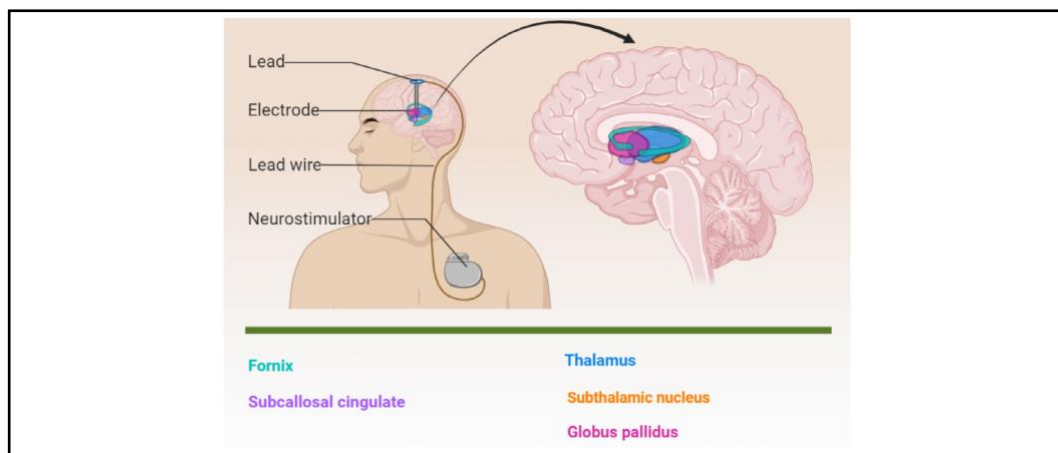


Fonte: LOZANO *et al.* (2019)

Nota: **(a)** Os neurotransmissores (inserção) são liberados em resposta à estimulação, levando a ondas de cálcio e subsequente liberação de gliotransmissores. Essa liberação influencia a plasticidade sináptica, levando à dilatação das arteríolas e ao aumento do fluxo sanguíneo regional. **(b)** Alterações induzidas pela estimulação cerebral profunda (*DBS*) nos potenciais de campo locais dentro do núcleo subtalâmico. A atividade na banda beta é rapidamente reduzida com a *DBS* a 3V e, em seguida, retoma com a estimulação desligada (LOZANO *et al.*, 2019).

A Figura 13, embora se refira à Doença de Parkinson, é essencial para entender os mecanismos de ação da Estimulação Cerebral Profunda (*DBS*). Ao contrário das figuras que analisamos anteriormente, que tratavam de interfaces cérebro-computador para fala ou diagnóstico de doenças, esta figura explica as hipóteses sobre como a *DBS* funciona. As análises anteriores tratavam-se de interfaces cérebro-computador para fala ou diagnóstico de doenças, esta figura explica as hipóteses sobre como a *DBS* funciona no nível celular e de rede, para tratar distúrbios neurológicos e psiquiátricos.

Figura 14 – Diagrama esquemático da ECP (*DBS*) para o tratamento da DA grave



Fonte: XU; LIU; SHANG *et al.* (2025).

A Figura 14, intitulada “*Schematic diagram of DBS for the treatment of severe AD*” (Diagrama esquemático da DBS para o tratamento da Doença de Alzheimer grave) na pesquisa de XU; LIU; SHANG *et al.* (2025), é a representação conceitual que define o foco central do estudo: a comparação entre dois alvos cirúrgicos distintos para a Estimulação Cerebral Profunda (DBS) no tratamento da Doença de Alzheimer (DA) grave. Ela ilustra a integração do *hardware* da Estimulação Cerebral Profunda (DBS) com os alvos anatômicos cerebrais que estão no centro do debate terapêutico para a DA.

Componentes de Hardware do Sistema DBS – Estes elementos representam a tecnologia implantável utilizada para administrar a estimulação elétrica: (i) *Neurostimulator (Neuroestimulador)* - Também conhecido como Gerador de Pulsos Implantável (IPG). É o dispositivo de bateria implantado (geralmente sob a pele do peito ou abdômen) que gera os impulsos elétricos. Sua função é fornecer a energia e administrar o protocolo de estimulação crônica programado para cada paciente; (ii) *Lead (Eletrodo DBS) / Electrode (Contato do Eletrodo)*: o *Lead* é o fio isolado implantado cirurgicamente no tecido cerebral alvo (Fornix ou NBM). O *Electrode* se refere aos contatos (pequenos pontos metálicos na ponta do lead) que liberam a corrente elétrica no alvo. No estudo, *leads* modelo L301C foram implantados para realizar o DBS bilateral; e (iii) *Lead Wire (Fio do Eletrodo)* - É a extensão que conecta o *Lead* implantado no cérebro ao Neurostimulator (IPG).

Alvos Terapêuticos e Estruturas Anatômicas (Contexto da DA) – Estes elementos representam as regiões cerebrais que são manipuladas pela DBS, ou estruturas relacionadas que definem o circuito neural da DA:

A. Alvos Primários do Estudo

1. *Fornix (f-DBS)*: (i) *Localização e Função* - O *Fornix* é um feixe principal de fibras que conecta os nós do circuito de *Papez*, sendo vital para a função de memória e cognição; e (ii) *Papel na DA Grave* – É um dos dois alvos principais investigados no estudo (14 pacientes receberam implantes bilaterais). A estimulação do *fornix* visa modular a atividade da rede neuronal. No entanto, no estudo, o *f-DBS* teve um efeito mais limitado nos sintomas neuropsiquiátricos em comparação com o *NBM-DBS*, pois não regula diretamente o sistema colinérgico.

2. *Núcleo Basal de Meynert (NBM)* (é o alvo principal de comparação e é crucial no contexto do estudo e da Figura 14): (i) *Localização e Função* - É um centro crítico de neurônios colinérgicos (tipo de célula nervosa que utiliza a acetilcolina como neurotransmissor

principal) que fornecem a maior parte da inervação colinérgica ao córtex, essenciais para cognição, atenção e memória. A DA é caracterizada pela atrofia nessa região; e (ii) *Papel na DA Grave* - Seis pacientes receberam implantes bilaterais no NBM. O NBM-DBS visa impulsionar a neurotransmissão colinérgica. O NBM-DBS demonstrou uma vantagem distinta e estatisticamente significativa sobre o f-DBS na melhoria dos sintomas neuropsiquiátricos (NPI).

B. Estruturas Anatômicas de Referência e Outros Alvos de DBS

1. *Thalamus (Tálamo)*: o Tálamo é uma estação de retransmissão central no cérebro. Ele é relevante porque o circuito de *Papez* (que inclui o *Fornix*) se conecta ao tálamo (via trato mamilotalâmico). Sua inclusão no diagrama ajuda a situar o *Fornix* no contexto do circuito de memória que a *DBS* busca modular.

2. *Subcallosal Cingulate (SCC) / Cíngulo Subcaloso*: o Cíngulo Subcaloso (ou Cingulado Subgenual) é uma região associada ao sistema límbico e é um alvo bem estabelecido para *DBS* no tratamento de depressão maior e transtorno obsessivo-compulsivo (TOC). Embora não seja o alvo principal para DA neste estudo, sua inclusão no diagrama (Figura 14) pode servir para: (i) *Ilustrar a ampla aplicação da DBS em distúrbios neuropsiquiátricos*; e (ii) *Reforçar o foco do estudo em sintomas neuropsiquiátricos (NPS) na DA*.

3. *Subthalamic Nucleus (STN) / Globus Pallidus (GPi) (Núcleo Subtalâmico / Globo Pálido)*: estas estruturas são classicamente os principais alvos para a *DBS* no tratamento de distúrbios do movimento, como a Doença de Parkinson (DP). *Papel na DA Grave*: Embora não sejam alvos primários para DA, sua inclusão na Figura 14 serve para mostrar que a metodologia *DBS* é transferida, com sucesso comprovado em distúrbios motores (DP), para a DA. Além disso, o estudo cita pesquisas que demonstraram a eficácia do NBM-DBS na melhoria dos sintomas neuropsiquiátricos em pacientes com DP e Demência de Corpos de Lewy (DLB).

A Figura 14 utiliza, portanto, componentes padronizados de *DBS* para configurar o dilema central da pesquisa em DA grave: se o foco deve ser a modulação do circuito de memória (*Fornix*) ou a regulação do sistema colinérgico/comportamental (NBM), que o estudo conclui ser mais promissor para os sintomas neuropsiquiátricos debilitantes da doença. O achado mais importante do estudo é que o NBM-DBS é significativamente mais eficaz na gestão dos sintomas neuropsiquiátricos do que o f-DBS (XU; LIU; SHANG *et al.*, 2025).

Riscos Jurídicos e Éticos:

A pesquisa com *DBS* em pacientes com DA expõe um paradoxo ético-jurídico fundamental: a tensão entre a busca por tratamentos e o dever de proteger indivíduos cuja capacidade de consentir está, por definição, comprometida.

A Doença de Alzheimer causa deterioração progressiva da memória e das funções cognitivas (ARAÚJO, 2024). Pacientes com DA e Parkinson frequentemente apresentam comorbidades como demência ou outras formas de comprometimento cognitivo (KLEIN, 2016). O estudo de 2025 (XU; LIU; SHANG *et al.*, 2025) sobre *DBS* no NBM e no fórnix, por exemplo, focou em pacientes com DA em estágio severo (escore MMSE de 0 a 10, nível CDR 3). O comprometimento cognitivo e comunicativo pode afetar a capacidade dos voluntários de consentir com pesquisas que envolvam Interfaces Cérebro-Máquina (ICCs/*BCIs*) ou *DBS*. Pacientes com DA têm alta variabilidade na função cognitiva, o que pode afetar sua capacidade de entender os riscos e benefícios da tecnologia *DBS* e o propósito do ensaio (SI; LIN *et al.* 2025). Logo, o Consentimento Livre e Esclarecido do paciente, pilar da ética médica e direito da personalidade, é profundamente desafiado.

O Estatuto da Pessoa com Deficiência (Lei nº 13.146/2015) exige, em seu art. 12, o consentimento prévio, livre e esclarecido para qualquer pesquisa, assegurando a participação da pessoa com deficiência “no maior grau possível”. Para pacientes com DA severa, como vimos acima, a capacidade de compreender os riscos e benefícios de um procedimento experimental e invasivo é questionável (XU; LIU; SHANG *et al.*, 2024; 2025). A opacidade da tecnologia agrava o problema: se os próprios pesquisadores não compreendem plenamente os mecanismos de ação da *DBS* no cérebro com Alzheimer, a informação transmitida ao paciente (ou seu curador) será, por natureza, incompleta. A pesquisa, nesses casos, só se justifica eticamente se houver indícios de benefício direto²¹, conforme o Estatuto. Isso torna o processo de consentimento um ato de extrema delicadeza, que demanda o envolvimento ativo da família e, idealmente, o recurso a diretivas antecipadas de vontade, elaboradas pelo paciente em um estágio anterior da doença (SILVA *et al.*, 2021; DADALTO *et al.*, 2021; DADALTO, 2025).

No estudo de *DBS* para DA severa, os pesquisadores obtiveram o consentimento informado por meio de representantes legais e auxílios de comunicação para respeitar a autonomia dos pacientes com deficiências cognitivas profundas (XU; LIU; SHANG *et al.*, 2024; 2025). A vulnerabilidade dos pacientes com DA, somada à complexidade da tecnologia, aumenta o risco de o consentimento ser viciado, seja por coação, indução a erro, ou até mesmo

²¹ O Regulamento (UE) 2017/745 (relativo aos dispositivos médicos – MDR) também estabelece que, em investigações clínicas envolvendo participantes incapazes, uma condição fundamental é que a pesquisa esteja diretamente relacionada ao quadro clínico do participante e haja motivos científicos para esperar que a participação comporte benefícios diretos que superem os riscos e inconvenientes da investigação (UNIÃO EUROPEIA, 2017a).

por vieses cognitivos que afetam o julgamento e a capacidade de previsão das consequências (SI; LIN *et al.* 2025; NEGRI *et al.*, 2025).

Além do consentimento, a *DBS* toca no cerne do Direito à Continuidade Psicológica, um dos neurodireitos propostos, que visa proteger a identidade e a coerência do “eu” de uma pessoa contra alterações não consentidas por terceiros (IENCA, 2021; LOPES *et al.*, 2022). A modulação direta da atividade cerebral, mesmo com fins terapêuticos, pode induzir efeitos secundários como mudanças de humor, comportamento e até de personalidade, como já observado em alguns pacientes com Parkinson tratados com a técnica (POMES *et al.*, 2025; ANPD, 2025). A ponderação entre o alívio de um sintoma (como a agitação) e a potencial alteração da essência de quem o paciente “é” constitui um dos dilemas éticos mais profundos desta tecnologia, exigindo uma discussão franca com a família sobre os valores e desejos do paciente (YUSTE; GOERING *et al.*, 2017).

Por fim, a questão da equidade no acesso é inevitável (LOPES, 2024). A *DBS* é uma terapia de altíssimo custo, envolvendo neurocirurgia, dispositivos sofisticados e acompanhamento multidisciplinar contínuo. Sua disponibilidade, em um futuro próximo, provavelmente ficará restrita a uma pequena elite econômica, criando um abismo no tratamento da DA e desafiando frontalmente os princípios da universalidade e da equidade que fundamentam o SUS (SANTOS, 2025).

2.2.2.3 Neuroprótese para a Doença de Parkinson: Recuperando a Marcha

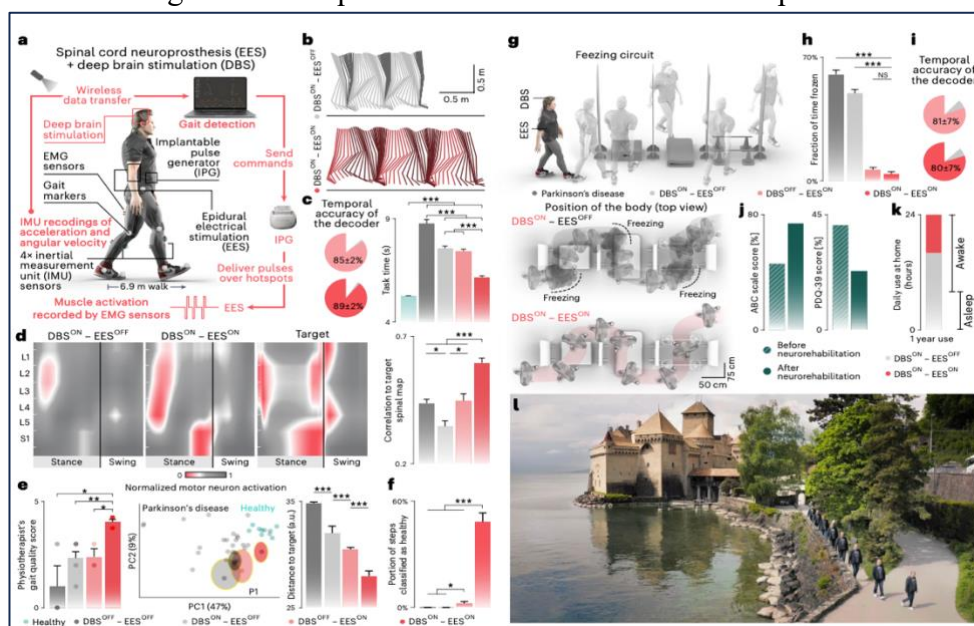
Panorama Geral e Potencialidades:

Os distúrbios da marcha, como a instabilidade e o “congelamento” (*freezing of gait*), estão entre os sintomas mais incapacitantes da Doença de Parkinson (DP) em estágio avançado, levando a quedas frequentes e a uma drástica perda de autonomia. Em 2023, um estudo publicado na *Nature Medicine* apresentou uma neuroprótese inovadora que oferece uma nova esperança para esses pacientes (MILEKOVIC; MORAUD *et al.*, 2023).

A neuroprótese foi desenvolvida com base na Estimulação Elétrica Epidural (*EES*), especificamente mirando as zonas de entrada das raízes dorsais que inervam os segmentos lombo-sacrais da medula espinhal. O objetivo do sistema é modular a atividade dos neurônios motores, reproduzindo o padrão natural de ativação espaço-temporal durante a marcha. O implante consiste em um eletrodo com 16 canais (*Specify 5-6-5*), clinicamente aprovado para dor crônica, que é adaptado para atingir os seis pontos-chave (*hotspots*) de ativação neural necessários para a locomoção. Após ser validada em modelos de primatas não humanos, a

neuroprótese foi implantada em um paciente de 62 anos com DP avançada e severos problemas de locomoção. Os resultados foram transformadores: o paciente, que antes mal conseguia andar e sofria com quedas constantes, recuperou uma marcha fluida e estável. Quando a neuroprótese era ativada (com ou sem *DBS*), os episódios de *freezing of gait* (que o paciente experimentava frequentemente ao girar ou passar por caminhos estreitos) quase desapareciam, resultando em uma melhora substancial em sua qualidade de vida e autonomia (MILEKOVIC; MORAUD *et al.*, 2023).

Figura 15 – Esquema de funcionamento da Neuroprótese



Fonte: MILEKOVIC; MORAUD *et al.* (2023).

Nota: A neuroprótese alivia os problemas de locomoção, melhora o equilíbrio, reduz a frequência de congelamento da marcha, aumenta a qualidade de vida e apoia a mobilidade em ambientes comunitários em pessoas com DP (MILEKOVIC; MORAUD *et al.*, 2023).

A figura apresentada por MILEKOVIC; MORAUD *et al.* (2023) demonstra, por meio de múltiplos painéis, como a Eficácia da Neuroprótese Espinhal (*EES*), combinada com a Estimulação Cerebral Profunda (*DBS*) existente e a terapia dopaminérgica, melhora significativamente a marcha, o equilíbrio, a qualidade de vida e, crucialmente, elimina o *freezing of gait* (congelamento da marcha) no paciente P1. Ela é dividida em painéis que abrangem o mecanismo de controle (*A*), os resultados imediatos de marcha (*B–F*) e o impacto a longo prazo e na comunidade (*G–L*).

Mecanismo e Detecção (Painéis A e C): (i) O Painel A apresenta o esquema da neuroprótese espinhal controlada cinematicamente, combinada com a *DBS* do núcleo subtalâmico (*NST*). O sistema utiliza Unidades de Medição Inercial (*IMUs*) vestíveis, fixadas às pernas, para detectar eventos de marcha (intenções motoras) em tempo real. Essa detecção

aciona rajadas de Estimulação Elétrica Epidural (*EES*) no gerador de pulsos implantável (IPG); e (ii) O Painel **C** confirma a alta precisão temporal do decodificador (*decoder*) baseado em IMU (usando o algoritmo rLDA), que detecta os eventos de ativação dos hotspots da perna (como o *leg lift*). A acurácia foi alta, variando de 85% a 89%, tanto com a DBS ligada ($DBS_{ON} - EES_{ON}$) quanto desligada ($DBS_{OFF} - EES_{ON}$).

Melhoria da Marcha e Normalização (Painéis B, D, E e F) – Estes painéis comparam, principalmente, a condição de tratamento duplo ($DBS_{ON} - EES_{ON}$) versus a condição de DBS e medicação sozinha ($DBS_{ON} - EES_{OFF}$): (i) Painel **B** (*Stick Diagram*) - Ilustra visualmente que a neuroprótese (EES_{ON}) restaura a cinemática da perna, reforçando a atividade dos músculos da perna mais afetada e aumentando a ativação dos hotspots de propulsão; (ii) Painel **D** (*Mapas Espaciais e Correlação*) - Mostra que a *EES* ($DBS_{ON} - EES_{ON}$) restaura a ativação espaciotemporal natural dos neurônios motores da perna durante a marcha. O gráfico de barras confirma que a correlação entre o mapa de ativação real de P1 e o mapa de marcha “alvo” (ideal, simulado biomecanicamente) é significativamente mais alta quando a *EES* está ligada ($DBS_{ON} - EES_{ON}$); (iii) Painel **E** (*Análise de Componentes Principais – PC*) - O *scatter plot* mostra que a marcha de P1 (representada pelo balão) se desloca drasticamente para o *cluster* de pessoas saudáveis quando a *EES* está ativa, distanciando-se do *cluster* de pacientes com DP. O gráfico de barras anexo mostra que a distância Euclidiana (na alta dimensão de parâmetros cinemáticos) para a marcha alvo é minimizada quando a *EES* está ligada; (iv) Painel **F** (*Classificação LDA*) - Apresenta a prova objetiva da normalização da marcha. Quando a neuroprótese foi ligada ($DBS_{ON} - EES_{ON}$), a porção de passos classificados como marcha de pessoas saudáveis aumentou drasticamente (próximo a 60%), em comparação com quase 0% quando a *EES* estava desligada; e (v) Painel **C** (*Bar Plot - Tempo de Tarefa*) - O tempo de tarefa (velocidade) é significativamente reduzido (melhorado) com a neuroprótese ligada, chegando a um desempenho semelhante ou melhor do que a DBS sozinha.

Eliminação do Freezing of Gait (Painéis G, H e I) – O *freezing of gait* (*FoG*) era um dos déficits mais debilitantes de P1, levando a quedas frequentes: (i) O Painel **G** (*Esquema do Circuito*) ilustra o circuito complexo (passagens estreitas e obstáculos) projetado para induzir o *FoG*. As linhas tracejadas no esquema de posição corporal mostram as ocorrências de *FoG* sem a *EES*; (ii) O Painel **H** (*Redução do FoG*) quantifica o tempo que P1 passou congelado. O gráfico de barras demonstra que, com a neuroprótese ligada ($DBS_{OFF/ON} - EES_{ON}$), o *FoG* quase desaparece (a fração de tempo congelada é significativamente reduzida); e (iii) O Painel **I** (*Acurácia no Circuito*) confirma que o decodificador manteve uma alta acurácia (80% a 81%) mesmo ao navegar neste ambiente complexo.

Impacto a Longo Prazo e na Qualidade de Vida (Painéis J, K e L): (i) O Painel J (Resultados de Reabilitação) mostra que a reabilitação de 3 meses, aumentada pela neuroprótese, levou a melhorias significativas: equilíbrio – ganhos na escala ABC (Activities-specific Balance Confidence), que mede a confiança no equilíbrio e qualidade de vida – ganhos relatados no PDQ-39 (Parkinson's Disease Questionnaire); (ii) Painel K (Uso Diário) - Um ano após o implante, P1 utilizava a neuroprótese cerca de 8 horas por dia (durante o tempo que estava acordado). Isso comprova a praticidade e a utilidade do sistema na vida diária; e (iii) Painel L (Mobilidade Comunitária) - Fotografia de P1 desfrutando de caminhadas recreativas na natureza, sugerindo que a neuroprótese apoia a mobilidade em ambientes comunitários.

Em suma, a Figura 15 comprova que a neuroprótese espinhal é uma terapia complementar eficaz para a DP em estágio avançado, superando as limitações da DBS e da terapia dopaminérgica, especialmente em déficits axiais refratários como o *freezing of gait* (MILEKOVIC; MORAUD *et al.*, 2023).

Riscos Jurídicos e Éticos:

Como um dispositivo médico implantável²², a neuroprótese para Parkinson está sujeita ao regime de responsabilidade pelo fato do produto/serviço, previsto no Código de Defesa do Consumidor (BRASIL, 1990, art. 12; art. 14). Em caso de defeito do dispositivo que cause dano ao paciente – como uma falha no gerador de pulsos, um curto-circuito que cause lesão na medula ou uma programação defeituosa que leve a movimentos descontrolados – a responsabilidade recai sobre toda a cadeia de fornecimento, incluindo o fabricante, o importador.

Com relação ao hospital temos que fazer algumas ponderações: sua responsabilidade é objetiva (Art. 14 do CDC) para serviços extramédicos (como a má conservação das neuropróteses ou falha na manutenção) e paramédicos (como intervenção incorreta de enfermeiros). O hospital tem o dever de garantir a manutenção e funcionamento regular dos equipamentos e de avaliar a qualidade e *performance* da tecnologia continuamente. No entanto, para danos decorrentes de atos essencialmente médicos (ex.: imperícia ou negligência do médico no uso da prótese ou no diagnóstico), a responsabilidade do médico é subjetiva (com

²² Art. 4º [...]. XIV - **dispositivo médico implantável**: qualquer dispositivo, incluindo os que são parcial ou totalmente absorvidos, destinado a ser introduzido totalmente no corpo humano; ou a substituir uma superfície epitelial ou a superfície ocular, mediante intervenção clínica, e que se destine a permanecer neste local após a intervenção, ou ainda aquele destinado a ser introduzido parcialmente no corpo humano mediante intervenção clínica e a permanecer neste local após a intervenção por um período de, pelo menos, 30 dias. RESOLUÇÃO – RDC ANVISA Nº 751, DE 15 DE SETEMBRO DE 2022. Disponível em: <https://www.in.gov.br/en/web/dou/-/resolucao-rdc-n-751-de-15-de-setembro-de-2022-430797145>. Acesso em: 10 set. 2024.

necessidade de comprovação de culpa: imperícia, imprudência ou negligência), conforme o Art. 14, § 4º, do CDC. A responsabilidade do hospital, por atos de seu profissional vinculado, é, em regra, *objetiva mitigada*, ou seja, depende da comprovação prévia da culpa do médico (NOGAROLI, 2023; KFOURI NETO, 2024).

Ademais, o titular da regularização do produto perante a ANVISA, em particular, detém responsabilidade administrativa, civil e penal por garantir a segurança e a eficácia do dispositivo²³. De acordo com a Resolução nº 751/2022 (ANVISA, 2022b), o “detentor (de notificação ou de registro)” é a pessoa jurídica (fabricante ou importador) responsável pelo dispositivo médico no Brasil e detém a concessão de comercialização. É responsabilidade do detentor manter o dossiê técnico atualizado e garantir a conformidade do produto. O responsável legal e técnico da empresa solicitante (detentora) é responsável pelas informações e documentos apresentados. Assim, o detentor está sujeito, por exemplo, a sanções administrativas (aplicadas pela ANVISA), em caso de não conformidade com a legislação, e cíveis (responsabilidade objetiva), perante o consumidor (CDC).

Este caso, embora claramente terapêutico, tangencia a fronteira ética entre terapia e aprimoramento (*enhancement*) (DANTAS; JEREISSATI, 2024). A tecnologia foi projetada para restaurar uma função perdida devido a uma doença. Contudo, ela abre a porta para o debate sobre usos não terapêuticos. E se uma tecnologia similar pudesse ser usada para aprimorar a velocidade, a força ou a resistência de uma pessoa saudável? (VILAÇA; MARQUES, 2015). A distinção entre restaurar a saúde e aprimorá-la é um dos debates centrais da bioética e da governança de novas tecnologias (GERARDI; XINARIS, 2025). A principal preocupação ética e jurídica sobre o uso de neurotecnologias para aprimoramento (em pessoas saudáveis) é o potencial de exacerbar as desigualdades sociais e alterar a própria essência humana (UNESCO, 2025). Segundo Vilaça e Marques (2015), o aprimoramento cognitivo, se disponível apenas para poucos, poderia criar duas classes de seres humanos: os “melhorados” e os “naturais”.

Em suma, a neuroprótese para Parkinson, por ser terapêutica, é incentivada. Já a tecnologia similar com potencial de aprimoramento é vista como um risco iminente que requer “cautela especial” na regulação (OEA, 2023) e a garantia explícita, via neurodireitos, de que o

²³ Art. 2º O carregamento de instruções de uso no repositório documental de dispositivos médicos, disponível no portal da Anvisa, é obrigatório e deve ser executado pela **empresa responsável pela notificação ou pelo registro do produto**, a qual atesta que seu conteúdo guarda concordância com a legislação vigente e consistência com o produto regularizado, de acordo com o §4º do art. 33 da Resolução da Diretoria Colegiada - RDC nº 751, de 15 de setembro de 2022 e do §4º do art. 33 da Resolução da Diretoria Colegiada - RDC nº 830, de 06 de dezembro de 2023. RESOLUÇÃO-RE Nº 1.206, DE 27 DE MARÇO DE 2025. Disponível em: https://anvisa.legis.datalegis.net/action/UrlPublicasAction.php?acao=abrirAtoPublico&num_ato=00001206&sgl_tipo=RES&sgl_orgao=RE/GGTPS/ANVISA/MS&vlr_ano=2025&seq_ato=000&cod_modulo=293&cod_menu=8499. Acesso em: 10 ago. 2025.

acesso a esses avanços seja equitativo e não cause novas formas de discriminação (PACHECO, 2025).

Assim como outras terapias avançadas, o acesso justo e equitativo é uma preocupação primordial. O alto custo da neuroprótese e do procedimento cirúrgico associado torna seu acesso um desafio imenso para sistemas de saúde públicos como o SUS. A ausência de políticas que garantam um acesso mais amplo pode fazer com que os benefícios de uma tecnologia tão transformadora fiquem restritos a uma pequena parcela da população, exacerbando as iniquidades em saúde já existentes no país (SANTOS, 2025).

Além dessas questões, vale mencionar que a integração da IA com a neurotecnologia também amplia as ameaças éticas e outros riscos potenciais, incluindo preocupações com a segurança cibernética, a falta de transparência, a possibilidade de viés algorítmico e os riscos à autonomia, à privacidade mental, à identidade pessoal e à manipulação (UNESCO, 2025).

Por fim, vale mencionar que, para além da responsabilidade pós-dano, a legitimidade ética e jurídica de sua implementação repousa sobre a validade do Consentimento Livre e Esclarecido do paciente, que, neste contexto, é multifacetado e excepcionalmente complexo.

O consentimento aqui transcende a autorização para um procedimento cirúrgico (DANTAS; JEREISSATI, 2024). Ele deve ser compreendido como um processo dialógico que abarca, no mínimo, quatro dimensões críticas: (i) *Consentimento para o Risco Físico e de Hardware* - Esta é a camada mais tradicional, envolvendo a comunicação clara sobre os riscos inerentes à neurocirurgia na medula espinhal (infecção, hemorragia, lesão neural); existência ou não de métodos alternativos e à falibilidade do dispositivo implantado (defeito da bateria, quebra de eletrodos, necessidade de cirurgias de revisão). A informação deve ser transparente quanto às estatísticas de sucesso e insucesso, conforme os princípios do Código de Ética Médica (CFM, 2018; ANVISA, 2022a; 2022b); (ii) *Consentimento para a Agência Compartilhada (ou híbrida) e a Incerteza* - O paciente deve compreender que sua capacidade de andar não será mais fruto exclusivo de sua intenção, mas de uma agência híbrida ou compartilhada entre sua vontade e a ação da neuroprótese. Ele precisa consentir com essa “cessão” parcial de controle motor a um dispositivo. Essa dimensão do consentimento deve, obrigatoriamente, incluir o risco de o dispositivo falhar ou interpretar mal os sinais, podendo gerar movimentos involuntários e quedas, o que conecta a discussão da autonomia diretamente à da segurança (IENCA; IGNATIADIS, 2020; LAVAZZA *et al.*, 2025). A opacidade sobre os efeitos de longo prazo da estimulação elétrica contínua na medula espinhal adiciona uma camada de incerteza que deve ser honestamente comunicada, em respeito ao “direito de ser informado sobre o não sabido” (COHEN, 2019; DANTAS; JEREISSATI, 2024); (iii) *Consentimento para a Fronteira*

Terapêutica - A distinção entre restaurar uma função e aprimorá-la (*enhancement*) é crucial (VILAÇA; MARQUES, 2015). O paciente precisa consentir com o objetivo terapêutico de restaurar uma marcha funcional, mas também precisa ser informado sobre as potencialidades da tecnologia (UNESCO, 2025). E se os parâmetros de estimulação puderem ser ajustados para um nível que supere sua capacidade motora pré-doença? Embora o uso atual seja estritamente terapêutico, o consentimento deve abordar essa ambiguidade, discutindo os limites do que se considera “normal” e o que seria “melhor que o normal”. A ausência dessa discussão pode levar a expectativas irrealistas e a um consentimento viciado sobre o verdadeiro escopo da intervenção (GERARDI; XINARIS, 2025; DADALTO, 2025); e (iv) *Consentimento para a Dependência Tecnológica e o Acesso* - O paciente precisa estar ciente de que se tornará dependente de uma tecnologia proprietária para uma função vital. Isso implica dependência de manutenções, atualizações de *software* e da própria continuidade da empresa fabricante (LAVAZZA *et al.*, 2025). O consentimento deve abordar o que acontece se a empresa falir ou descontinuar o suporte ao dispositivo (UGUR, 2024; LAVAZZA *et al.*, 2025). Essa discussão está intrinsecamente ligada à questão do acesso justo e equitativo. O alto custo da tecnologia torna seu acesso um desafio para sistemas públicos como o SUS, e a falta de políticas que garantam a sustentabilidade e a universalização do acesso pode fazer com que os benefícios de uma tecnologia tão transformadora fiquem restritos a poucos, exacerbando as iniquidades em saúde (TORGERSEN *et al.*, 2024; SANTOS, 2025; UNESCO, 2025).

Em suma, o consentimento para a neuroprótese de Parkinson exige uma abordagem holística, que vá muito além da assinatura de um termo, para se tornar um diálogo contínuo sobre riscos conhecidos e desconhecidos, sobre a partilha de controle entre homem e máquina e sobre as implicações sociais de se viver com uma tecnologia que redefine as fronteiras do corpo humano.

2.2.2.4 *Neuroimagem e IA no Diagnóstico Precoce de Alzheimer: O Método DANMLP*

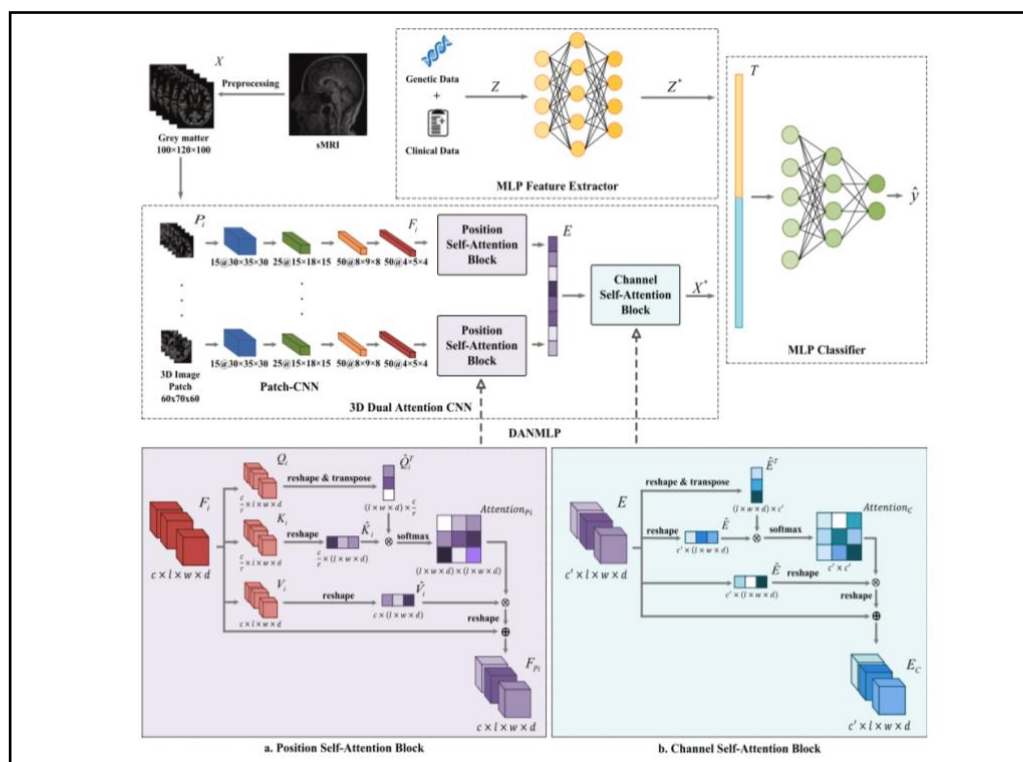
Panorama Geral e Potencialidades:

O diagnóstico precoce da Doença de Alzheimer (DA) e de seu estágio precursor, o Comprometimento Cognitivo Leve (MCI), é fundamental para a intervenção e o planejamento de cuidados. O método *DANMLP* (*Dual Attention Convolutional Neural Network and Multilayer Perceptron*), apresentado em 2023, é um exemplo de como a combinação de neuroimagem e IA pode aprimorar a acurácia diagnóstica (ARAÚJO, 2024).

A principal inovação do DANMLP é sua capacidade de integrar dados multimodais. O sistema utiliza uma arquitetura híbrida que analisa simultaneamente três fontes de informação do paciente: 1) imagens de ressonância magnética estrutural (*sMRI*), que são processadas por uma Rede Neural Convolutiva (*CNN*) equipada com um mecanismo de “dupla atenção” para focar em regiões cerebrais e canais de imagem mais relevantes; 2) dados clínicos, como informações demográficas e escores de testes neuropsicológicos; e 3) dados genéticos, como a presença do alelo ApoE4, um conhecido fator de risco para a DA. Essas duas últimas fontes de dados são analisadas por um Perceptron Multicamadas (*MLP*), e os resultados são fundidos com as características extraídas da imagem para gerar uma classificação final (QIANG; ZHANG *et al.*, 2023).

Essa fusão de dados permite uma acurácia diagnóstica superior à de métodos que se baseiam em uma única modalidade. O estudo que apresentou o *DANMLP* reportou uma acurácia de 93% na tarefa de classificar pacientes com DA *versus* MCI (comprometimento cognitivo leve), e de 82,4% na tarefa mais desafiadora de distinguir pacientes com MCI de controles normais, superando significativamente outros métodos de ponta (QIANG; ZHANG *et al.*, 2023; ARAÚJO, 2024).

Figura 16 – Esquema da estrutura do modelo *DANMLP*



Fonte: QIANG; ZHANG *et al.* (2023).

Nota: O *DANMLP* é composto por três componentes: *CNN* de atenção dupla 3D (3D-DACNN), extrator de características *MLP* e classificador *MLP*. 3D-DACNN é uma rede em nível de *patch* usada para aprender as características das imagens *sMRI*,

que consiste em uma *Patch-CNN* com I ramificações *CNN*, I blocos de autoatenção de posição e um bloco de autoatenção de canal. O extrator de características *MLP* é usado para aprender as características a partir de dados clínicos e genéticos. O classificador *MLP* é adotado para fundir as características aprendidas a partir de imagens, dados clínicos e genéticos para classificação *NC/MCI/DA*. **(a)** A arquitetura do bloco de autoatenção de posição. Este bloco é usado para aumentar as características dependentes da posição dentro do *patch*. **(b)** A arquitetura do bloco de autoatenção de canal. Este bloco é usado para enfatizar mapas de características interdependentes entre *patches* (QIANG; ZHANG *et al.*, 2023).

A Figura 16 ilustra a arquitetura do modelo *DANMLP (Dual Attention CNN and Multilayer Perceptron)* proposto por QIANG; ZHANG *et al.* (2023), um *framework* que combina Redes Neurais Convolucionais (*CNN*) com Atenção Dupla e *Multilayer Perceptron (MLP)* para integrar múltiplos tipos de dados na classificação da DA e Comprometimento Cognitivo Leve (MCI).

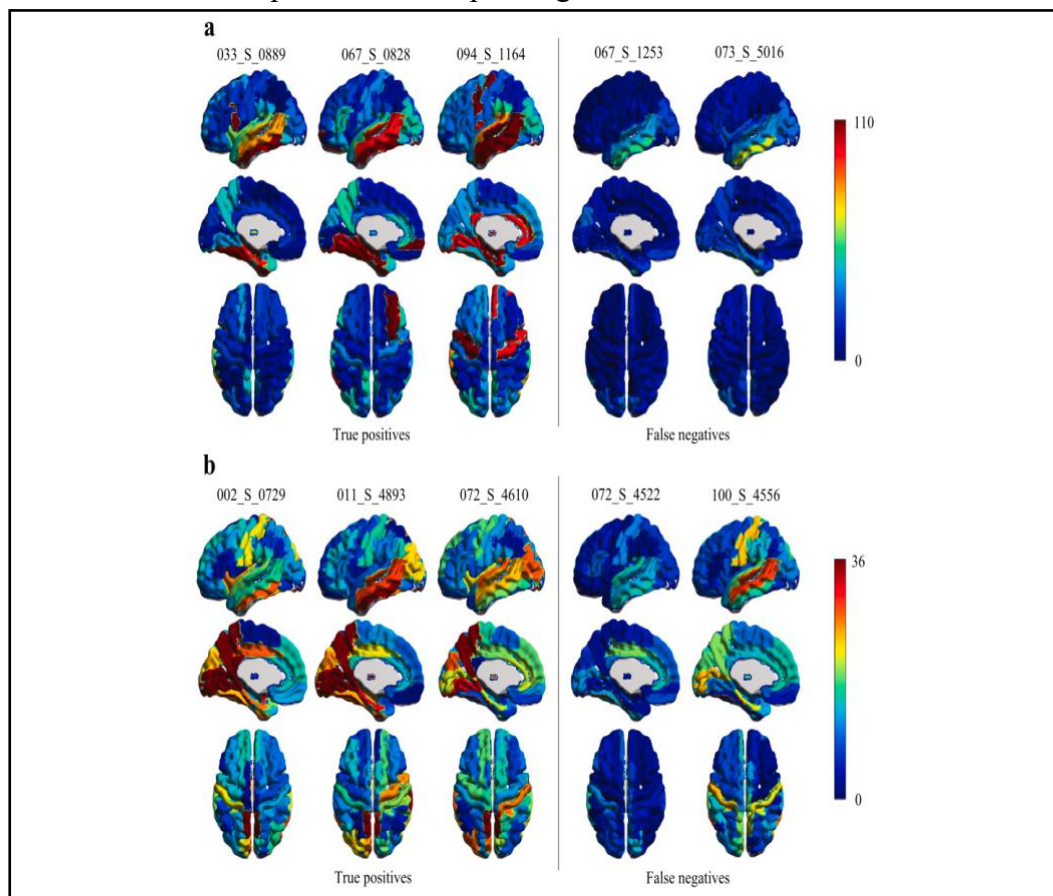
O *DANMLP* foi projetado para superar as limitações dos métodos anteriores, integrando dados multimodais – Imagens de Ressonância Magnética Estrutural (*sMRI*), dados clínicos (demográficos, neuropsicológicos) e dados genéticos *APOE* – e melhorando a interpretabilidade do diagnóstico.

O diagrama da Figura 16 é composto por três componentes principais que trabalham em conjunto: *(i) 3D Dual Attention CNN (3D-DACNN)* - Este componente é responsável por aprender as características discriminativas das imagens *sMRI*, que são cruciais para capturar alterações morfológicas cerebrais (como a atrofia); *(ii) MLP Feature Extractor (Extrator de Características MLP)* - Este componente lida com os dados não-imagem, que são unidimensionais. Usa uma rede *Multilayer Perceptron (MLP)* para aprender as características *embedding (Z*)* desses dados. O uso da *MLP* é adequado para mapear vetores de entrada para vetores de saída *embedding*; e *(iii) MLP Classifier (Classificador MLP)* - Esta é a camada de decisão final do modelo. O Classificador *MLP* recebe e, após processamento, produz o resultado de classificação (Y_n). O modelo realiza classificações binárias, como DA vs. MCI e MCI vs. NC (Controle Normal).

A Figura representa, portanto, um avanço significativo no diagnóstico de DA por IA, em razão de sua: *(i) Integração Multi-modal* - O *framework* reconhece que o diagnóstico de DA é complexo e não pode depender apenas de neuroimagem, incorporando dados genéticos e clínicos que fornecem informações complementares; *(ii) Alta Acurácia* - O *DANMLP* alcançou uma acurácia de 93% na tarefa DA vs. MCI e 82.4% na tarefa mais desafiadora MCI vs. NC, superando outros cinco métodos de última geração, o que confirma a eficácia de sua arquitetura; e *(iii) Interpretabilidade* - O uso de mecanismos de atenção (*PSA* e *CSA*) e a visualização subsequente (Fig. 17) permitem identificar e visualizar as regiões cerebrais discriminativas

afetadas pela DA (como o hipocampo e a amígdala), ajudando os médicos no diagnóstico precoce (QIANG; ZHANG *et al.*, 2023).

Figura 17 – Resultados da visualização das regiões cerebrais discriminatórias identificadas pelo *DANMLP* para alguns indivíduos com DA e DCL



Fonte: QIANG; ZHANG *et al.* (2023).

Nota: **(a)** Resultados da visualização de cinco indivíduos com DA na tarefa de classificação de DA vs. *DCL*. O painel esquerdo (Verdadeiros positivos) exibe três indivíduos com DA classificados corretamente, e o painel direito (Falsos negativos) mostra dois indivíduos com DA classificados incorretamente como indivíduos com *DCL*. **(b)** Resultados da visualização de cinco indivíduos com MCI na tarefa de classificação de *MCI* vs. *NC*. O painel esquerdo (Verdadeiros positivos) exibe três indivíduos com *MCI* classificados corretamente, e o painel direito (Falsos negativos) mostra dois indivíduos com *MCI* classificados incorretamente como indivíduos com *NC*. A barra de cores representa o tamanho dos valores de visualização $\rho_{n,j}^{AD}/\rho_{n,j}^{MCI}$ para cada indivíduo com DA/*DCL*, que pode refletir o grau de alterações patológicas em determinadas regiões do cérebro. Quanto maior o valor de visualização $\rho_{n,j}^{AD}$ (ou $\rho_{n,j}^{MCI}$) de uma região do cérebro, maior a diferença dessa região do cérebro entre DA e *MCI* (ou *MCI* e *NC*) (QIANG; ZHANG *et al.*, 2023).

Riscos Jurídicos e Éticos:

A força do *DANMLP* – a integração de dados multimodais – é também sua maior vulnerabilidade do ponto de vista jurídico e ético. A combinação de dados de neuroimagem, informações clínicas detalhadas e dados genéticos cria um perfil extremamente sensível do

indivíduo. A proteção desse conjunto de dados contra vazamentos, acessos não autorizados e usos indevidos torna-se um desafio crítico de segurança da informação, exigindo a aplicação rigorosa da Lei Geral de Proteção de Dados (LGPD). As hipóteses legais para o tratamento desses dados combinados para fins de diagnóstico e pesquisa devem ser inequívocas, e as medidas técnicas e administrativas de segurança, como criptografia, pseudonimização e controle de acesso, devem ser robustas para mitigar os riscos (BRASIL, 2018; DONEDA, 2021; LEONEL; LEONEL *et al.*, 2025; ANPD, 2025). A literatura ainda alerta que conjuntos genômicos, mesmo pseudonimizados, podem ser reidentificados por ligações externas, ampliando o risco quando associados a imagens cerebrais (BERTONI; IENCA, 2024; GYMREK *et al.*, 2013).

O risco de viés algorítmico também é amplificado neste cenário multimodal. O modelo pode aprender associações espúrias e discriminatórias. Por exemplo, ele poderia correlacionar dados demográficos presentes no prontuário clínico (como etnia, local de residência ou nível socioeconômico) com determinados achados de imagem ou genéticos, levando a diagnósticos sistematicamente menos precisos para grupos populacionais historicamente marginalizados (DONEDA; SOUZA *et al.*, 2018; PRICE II; FORD, 2024; BEZERRA *et al.*, 2024). A interpretabilidade do modelo, que os autores do *DANMLP* citam como um de seus benefícios ao permitir a visualização das áreas cerebrais que mais contribuíram para a decisão (QIANG; ZHANG *et al.*, 2023), torna-se uma ferramenta essencial para auditar e mitigar esses vieses: através de demonstração de desempenho estratificado por subgrupos e de auditorias interpretáveis (mapas de atenção/grad-CAM), sob pena de replicar assimetrias históricas em diagnósticos de demência (MEHRABI *et al.*, 2022; ALJUHANI *et al.*, 2024).

A questão do Consentimento Livre e Esclarecido do paciente é particularmente crítica. Por isso, a aplicação desta ferramenta diagnóstica em populações com potencial comprometimento cognitivo, como é o caso de pacientes com suspeita de DA ou *MCI*, reitera a necessidade de uma atenção redobrada ao processo de consentimento (FERREIRA *et al.*, 2025; DADALTO *et al.*, 2021; DADALTO, 2025), que é desafiado pela: (i) *Complexidade e Opacidade* - Explicar a um paciente, especialmente àquele com suspeita de comprometimento cognitivo, que seus dados de cérebro, vida e genética serão processados por um algoritmo opaco (“caixa-preta”) para gerar uma probabilidade diagnóstica, é uma tarefa hercúlea. A incapacidade de detalhar *como* o algoritmo funciona, em razão da opacidade explicativa, afeta o pilar do “esclarecimento” do consentimento (NOGAROLI, 2023; BROWN, 2024; DANTAS; JEREISSATI, 2024; AITH, 2024); e (ii) *Vulnerabilidade do Paciente* - A aplicação da ferramenta em populações com potencial comprometimento cognitivo exige atenção redobrada.

O Estatuto da Pessoa com Deficiência (Lei nº 13.146/2015), em seu art. 12, garante o direito ao Consentimento Livre e Esclarecido, assegurando a participação da pessoa “no maior grau possível”. Isso impõe o dever de utilizar linguagem acessível e, quando necessário, contar com o apoio de representantes legais, sempre respeitando a autonomia residual do paciente (CFM, 2016; UNESCO, 2025). O diagnóstico preditivo de uma doença incurável também invoca o “direito de não saber”, que deve ser respeitado durante o processo de consentimento (RÜGER, 2007; UNIÃO EUROPEIA, 2025).

O método *DANMLP* ilustra como a convergência entre neuroimagem, dados clínicos e genéticos, potencializada por mecanismos de atenção, pode aprimorar a acurácia diagnóstica em *DA/MCI*. Entretanto, para ser eticamente aceitável, esse ganho técnico precisa vir acompanhado de (i) *Consentimento clínico verdadeiramente esclarecido*, com apoio à decisão em casos de capacidade comprometida; (ii) *Salvaguardas de segurança e governança* para dados altamente sensíveis; e (iii) *Validação e auditoria de vieses* com resultados reportados por subgrupos. Só assim a inovação algorítmica se converte em benefício clínico legítimo, sem sacrificar autonomia e equidade.

2.2.2.5 A “Ponte Digital” cérebro-coluna (BSI) para Lesão Medular

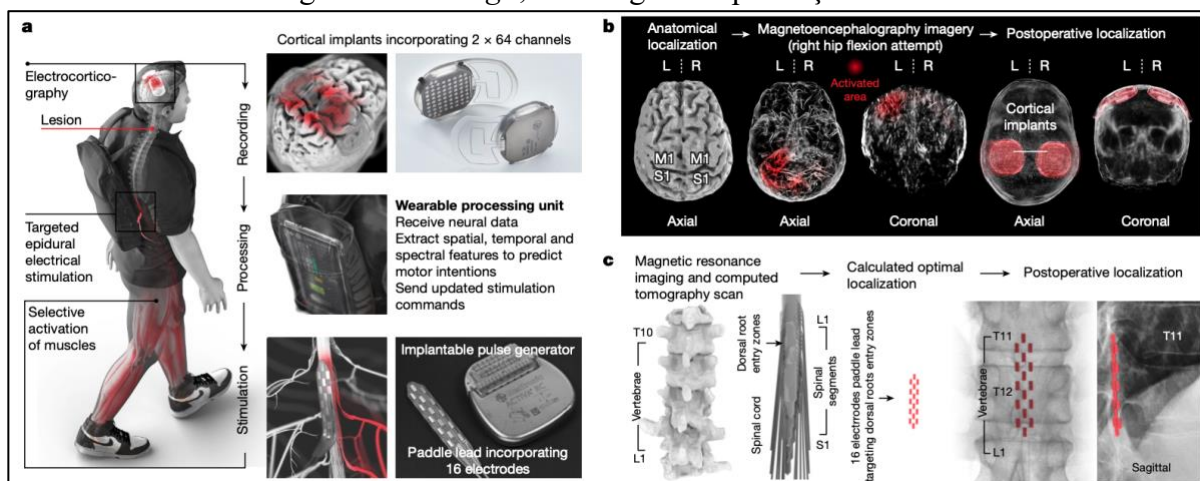
Panorama Geral e Potencialidades:

Um dos avanços mais espetaculares da neurotecnologia recente é a criação de uma “ponte digital” sem fio que reconecta o cérebro à medula espinhal, contornando uma lesão medular (MUIRHEAD *et al.*, 2024). Em um estudo marcante, publicado na *Nature* em 2023, pesquisadores suíços implantaram eletrodos sobre o córtex motor de um paciente paraplégico, na área responsável por controlar as pernas. Esses eletrodos registram a intenção de movimento do paciente. Os sinais são decodificados em tempo real por algoritmos de IA e transmitidos sem fio para um neuroestimulador implantado na região da medula espinhal abaixo da lesão, que por sua vez ativa os músculos das pernas para executar o movimento desejado (LORACH *et al.*, 2023).

O resultado foi a restauração da capacidade de locomoção de um homem de 38 anos, que havia sofrido uma lesão medular cervical incompleta (vértebras C5/C6) e ficado paralisado há mais de uma década. Ele já tinha participado de um ensaio clínico anterior (*STIMO*), que envolvia Estimulação Elétrica Epidural (*EES - Epidural Electrical Stimulation*) da medula espinhal. Esse treinamento inicial permitiu que ele recuperasse a capacidade básica de pisar com o auxílio de um andador com rodas, mas ele não conseguia caminhar de forma independente e tinha grandes dificuldades em terrenos complexos, rampas ou escadas. Após o

implante, ele voltou a ficar em pé, caminhar e até subir escadas com controle natural e voluntário. Mais impressionante ainda, o treinamento com a interface cérebro-espinha ou cérebro-coluna (*Brain-Spine Interface - BSI*) levou a uma recuperação neurológica parcial, com o paciente demonstrando melhora no controle motor e na sensibilidade mesmo quando o dispositivo estava desligado, sugerindo que a tecnologia promove a neuroplasticidade e a formação de novas conexões nervosas (LORACH *et al.*, 2023).

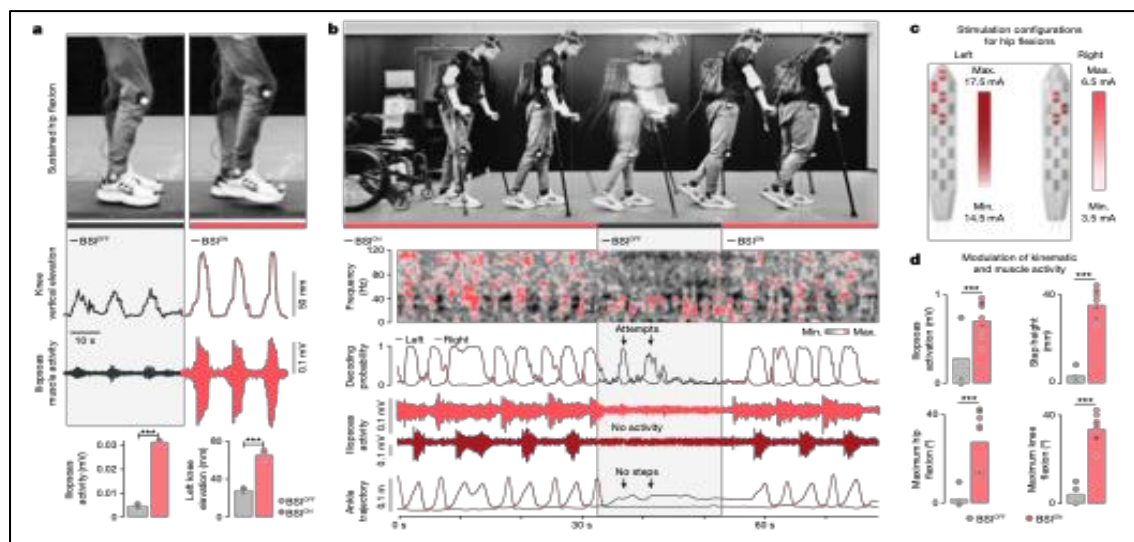
Figura 18 – *Design*, tecnologia e implantação da BSI

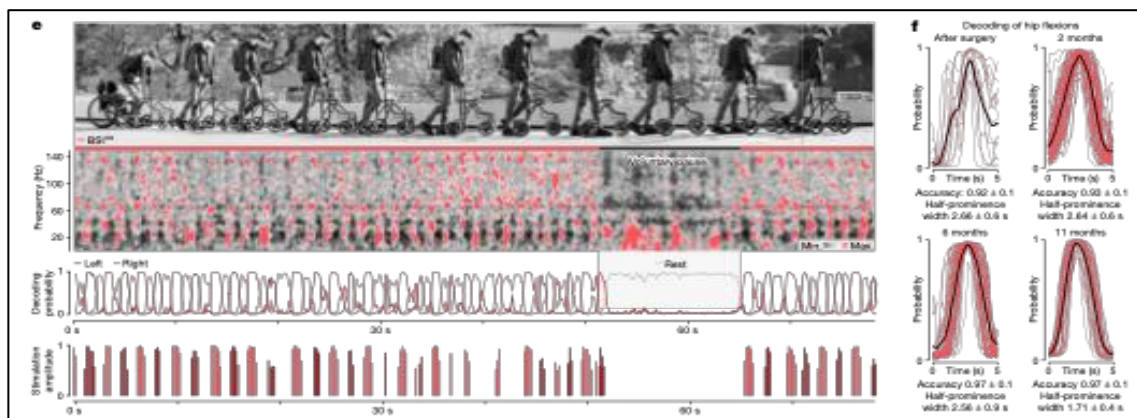


Fonte: LORACH *et al.* (2023).

Nota: **(a)** Dois implantes corticais compostos por 64 eletrodos são posicionados epiduralmente sobre o córtex sensório-motor para coletar sinais ECoG (Eletrocorticografia). Uma unidade de processamento prevê as intenções motoras e traduz essas previsões na modulação de programas de estimulação elétrica epidural direcionados às zonas de entrada da raiz dorsal da medula espinal lombossacral. As estimulações são fornecidas por um gerador de pulsos implantável conectado a um eletrodo de pá com 16 eletrodos. **(b)** Imagens que relatam o planejamento pré-operatório dos locais dos implantes corticais e a confirmação pós-operatória. L, esquerda; R, direita. **(c)** Modelo computacional personalizado que prevê a localização ideal do eletrodo de pá para atingir as zonas de entrada da raiz dorsal associadas aos músculos dos membros inferiores e confirmação pós-operatória (LORACH *et al.*, 2023).

Figura 19 – A BSI que restaura o controle natural da marcha





Fonte: LORACH *et al.* (2023).

Nota: **(a)** Tentativas de realizar flexões voluntárias do quadril sem e com a BSI, incluindo fotografias, elevação vertical do joelho e atividade muscular flexora do quadril. Os gráficos de barras mostram os valores médios dessas medições. ($n = 3$ tentativas por condição, teste t não pareado, $***P < 0,001$.) **(b)** Cronofotografia durante a caminhada com a BSI ligada, desligada e novamente ligada. Observe as duas tentativas decodificadas que não levam à atividade muscular nem à execução de passos. **(c)** Faixa de amplitude de estimulação durante a caminhada. **(d)** Gráficos de barras que apresentam os valores médios dos parâmetros cinemáticos e da atividade muscular durante a caminhada com a BSI desligada e ligada, ($n = 3$ e 8 tentativas para BSI^{OFF} e BSI^{ON} , respectivamente), teste t não pareado de cauda única, $***P < 0,001$, $P(\text{ativação do iliopsoas}) = 3,4 \times 10^{-4}$, $P(\text{altura do passo}) = 5,1 \times 10^{-10}$, $P(\text{ângulo do quadril}) = 2,7 \times 10^{-5}$, $P(\text{ângulo do joelho}) = 1,6 \times 10^{-9}$. **(e)** Cronofotografia de permanência em pé (pausa voluntária) e caminhada com o BSI ao ar livre. O espectrograma, as probabilidades de passos à esquerda e à direita e a modulação das amplitudes de estimulação ilustram a robustez do desempenho e a ausência de detecções falsas positivas durante a pausa voluntária. **(f)** Os gráficos mostram a probabilidade de flexões do quadril direito em passos consecutivos medidos durante a primeira sessão após a implantação neurocirúrgica ($n = 13$ passos, precisão = $0,92 \pm 0,1$ s.d., $w = 2,66 \pm 0,6$ s s.d.), e em 2 ($n = 46$ passos, precisão = $0,93 \pm 0,1$ s.d., $w = 2,64 \pm 0,6$ s s.d.), 6 ($n = 41$ passos, precisão = $0,97 \pm 0,1$ s.d., $w = 2,56 \pm 0,9$ s s.d.) e 11 meses ($n = 29$ etapas, precisão = $0,97 \pm 0,1$ desvio padrão, $w = 1,71 \pm 0,4$ s desvio padrão) após a primeira ativação do BSI usando modelos atualizados (LORACH *et al.*, 2023).

A Figura 19, intitulada pelos autores do estudo como **“The BSI restores natural control of walking”** (“A BSI restaura o controle natural da marcha”), fornece tanto a prova visual imediata, quanto a quantificação estatística de que a BSI, ao traduzir a intenção cortical em comandos de estimulação espinhal, permite que o participante ande com controle natural e adaptativo. Ela é composta por vários painéis que, em conjunto, comprovam a eficácia e a robustez da Interface Cérebro-Coluna (BSI): *(i) Prova da Dependência (Painel b)* – O Painel b (Cronofotografia) oferece o melhor indicador visual da função crítica da BSI como uma ponte digital: a sequência de imagens mostra o participante caminhando com auxílio de muletas quando a BSI está ligada (BSI^{ON}). Em seguida, quando a BSI é desligada (BSI^{OFF}), o participante perde instantaneamente a capacidade de dar qualquer passo, apesar das tentativas detectadas de

caminhar através da modulação da atividade cortical. O caminhar é retomado imediatamente assim que a *BSI* é religada. Este contraste dramático comprova que a *BSI* é o elo direto e necessário (a ponte digital) que restaura o controle do movimento; (ii) *Normalização e Quantificação da Marcha (Painéis a e d)* – Estes painéis fornecem a evidência estatística de que a *BSI* não apenas restaura o movimento, mas o normaliza. O Painel **a** (Tentativas de Flexão do Quadril) demonstra a capacidade de controle voluntário e o Painel **d** (Parâmetros Cinemáticos e de Atividade Muscular) reporta os valores médios de parâmetros cinemáticos durante a caminhada; (iii) *Robustez e Uso Natural (Painel e)* - Este painel, incluindo a cronofotografia e o espectrograma, ilustra a confiabilidade do sistema; e (iv) *Estabilidade a Longo Prazo (Painel f)* - O Painel **f** demonstra que o sistema de decodificação é estável e duradouro, um requisito para a eficácia a longo prazo.

Em suma, a Figura 19 fornece a comprovação mais forte de que a *BSI* traduz a intenção cortical em movimentos funcionais e controlados, restaurando a capacidade de caminhar de forma natural e confiável após a paralisia (*LORACH et al., 2023*).

Em publicação mais recente (7 jul. 2025), cientistas chineses concluíram que esses sistemas assistidos por interface cérebro-coluna fornecem meios para a realização de exercícios de reabilitação e até para a própria movimentação de membros em pacientes com Lesão da Medula Espinhal (LME), incluindo estimulação elétrica e robôs exoesqueletos (LI; GAO *et al.*, 2025).

Importa salientar, por fim, que, embora a neuroprótese espinhal para a Doença de Parkinson (item 2.2.2.3) e a *BSI* partilhem a lógica de empregar Estimulação Elétrica Epidural da medula lombossacral acoplada a algoritmos de decodificação, elas se situam em patamares tecnicamente e normativamente distintos. Na DP, a neuroprótese descrita por MILEKOVIC; MORAUD *et al.* (2023) aciona a estimulação com base em dados cinemáticos captados por unidades de medição inercial e atua sobre circuitos motores ainda estruturalmente íntegros, modulando a marcha de um paciente que conserva vias corticoespinhais funcionais, de modo a configurar, em grande medida, um incremento terapêutico de uma função pré-existente. Já a *BSI* proposta por LORACH *et al.* (2023) reconstrói, por meio de implantes corticais e da decodificação direta de sinais *ECoG*, uma via de comunicação que havia sido fisicamente interrompida por lesão medular crônica, criando uma verdadeira “ponte digital” entre estados neurais e padrões de estimulação medular. Esse deslocamento, do uso de *proxies* corporais de movimento para a manipulação de neurodados propriamente ditos, faz com que o caso da *BSI* represente um salto qualitativo em termos de riscos à privacidade mental, à integridade psicológica e à própria distribuição de agência entre paciente e sistema, aproximando-o, de

forma mais intensa, do núcleo de preocupações que motivam a formulação contemporânea de neurodireitos.

Riscos Jurídicos e Éticos:

A “ponte digital” coloca os neurodireitos, por tanto, no centro do debate jurídico. A tecnologia decodifica a *intenção* de movimento. Isso levanta questões profundas sobre a privacidade mental: que garantias existem de que o sistema não possa, no futuro, decodificar pensamentos ou intenções que o usuário não deseja exteriorizar? (DANTAS; JEREISSATI, 2024). A segurança do sistema é outra preocupação crítica. Sendo uma interface sem fio, ela poderia ser vulnerável a ataques cibernéticos? (HOOPER, 2025; LAVAZZA *et al.*, 2025). A possibilidade de um terceiro mal-intencionado interferir ou controlar os movimentos de uma pessoa é um risco de segurança com implicações éticas e jurídicas sem precedentes (YUSTE; GOERING *et al.*, 2017; CORNEJO, 2024; NEGRI *et al.*, 2025). Dispositivos neuromoduladores invasivos (como implantes utilizados em *BSI*, *BCI* e Estimulação Cerebral Profunda) estão sujeitos também ao risco de “sequestro cerebral” (*brainjacking*) (HOOPER, 2025). Essas possibilidades, portanto, levantam a questão de como determinar a responsabilidade moral e legal por ações realizadas através de uma *BCI* comprometida (LAVAZZA *et al.*, 2025).

A responsabilidade civil por falhas também é outro ponto sensível. Um erro de decodificação que cause uma queda seria responsabilidade do fabricante do *hardware*, do desenvolvedor do *software*, da equipe cirúrgica ou do próprio usuário? A aplicação do Código de Defesa do Consumidor (CDC) ao *hardware* é uma via, mas a prova do nexo causal em um sistema que envolve a intenção do usuário é um desafio. Abaixo, detalhamos a responsabilidade potencial de cada agente, o papel do Código de Defesa do Consumidor (CDC) e os desafios da prova do nexo causal:

1. *Responsabilidade dos Agentes Envolvidos*: o dano resultante de uma falha de decodificação em uma *BCI* (como a queda) insere-se nos chamados “casos de apontar o dedo” (*finger-pointing cases*), onde a causa eficiente do dano é incerta (NOGAROLI, 2023).

A. *Responsabilidade do Fabricante e do Desenvolvedor de Software*: O fabricante do *hardware* (implantes neurais, neuroestimulador) e o desenvolvedor do *software* (algoritmos de decodificação de IA) são sujeitos à responsabilidade civil objetiva perante o paciente (consumidor), conforme o CDC (Art. 12) e o Código Civil (Art. 931) (NOGAROLI, 2023; ALMEIDA JR.; REINAS, 2024): (i) *Natureza do Produto* - A legislação e as propostas de regulamentação europeia (e a interpretação do CDC) expandem a definição de “produto” para

incluir itens intangíveis como *softwares* e sistemas de IA. Portanto, o algoritmo de decodificação em si é considerado parte do produto (NOGAROLI, 2023; HENRIQUES, 2025); (ii) *Defeito do Produto* - Se a falha (a queda) for causada por um *bug* no *software*, um erro de programação inicial, um defeito no *design* do implante, ou um movimento inesperado e incorreto do dispositivo, a responsabilidade recai sobre o fabricante/fornecedor por defeito do produto (NOGAROLI, 2023); e (iii) *Risco do Desenvolvimento* - O fabricante não pode alegar que a falha é um “risco do desenvolvimento” ou um resultado imprevisível da autoaprendizagem (*machine learning*). A doutrina majoritária brasileira entende que o fabricante assume os riscos (previsíveis e imprevisíveis) da atividade, sendo o aprendizado de máquina uma característica intrínseca da tecnologia, e não uma excludente de responsabilidade (NOGAROLI, 2023).

B. Responsabilidade da Equipe Cirúrgica e do Médico: A responsabilidade pessoal do médico (profissional liberal) é subjetiva, dependendo da prova de culpa (negligência, imprudência ou imperícia), conforme o Art. 14, § 4º do CDC (FACCHINI NETO; BARBOSA, 2023; NOGAROLI, 2023): (i) *Uso Indevido da IA* - Se a queda ocorrer devido à negligência médica – por exemplo, se o médico falhar em calibrar o dispositivo corretamente, desconsiderar avisos de erro emitidos pelo sistema de IA, ou usar a tecnologia sem o treinamento adequado – o médico pode ser responsabilizado (NOGAROLI, 2023). O médico tem o dever de exercer a crítica e a vigilância sobre a IA; confiar cegamente no algoritmo pode configurar negligência (FACCHINI NETO; BARBOSA, 2023; NOGAROLI, 2023); (ii) *Dever de Informação* - Se a queda resultar de um risco inerente ao uso da *BSI* (que, sendo invasiva e complexa, possui riscos), e o médico não forneceu um CLE adequado e detalhado sobre esses riscos (incluindo falhas do *software*, *time delay*, ou interrupção do sistema), ele pode ser responsabilizado pela violação do direito à autodeterminação do paciente, independentemente de ter havido culpa no ato técnico (negligência informacional) (DANTAS; NOGAROLI, 2020; NOGAROLI, 2023; SOARES, 2025); e (iii) *Responsabilidade Hospitalar* - O hospital/clínica pode ser responsabilizado objetivamente (Art. 14 do CDC) por falhas em serviços extramédicos, como a ausência de políticas adequadas de treinamento e monitoramento contínuo da IA (NOGAROLI, 2023).

2. *O Desafio da Prova do Nexo Causal e a Intenção do Usuário* – A dificuldade em atribuir responsabilidade se deve a dois fatores principais:

A. A Opacidade da “Caixa Preta” Algorítmica: o sistema de decodificação da *BSI*, que provavelmente se baseia em *deep learning*, opera como uma “caixa-preta” (*black box*). Os algoritmos podem chegar a um resultado (o comando que causou a queda) de forma que é muito

difícil ou até impossível para os humanos (e juristas) explicarem o processo decisório da IA (NOGAROLI, 2023; SOARES, 2025; HENRIQUES, 2025). A complexidade técnica ou científica da IA torna a prova do defeito e/ou do nexo causal excessivamente onerosa para a vítima (o paciente) (NOGAROLI, 2023; HENRIQUES, 2025).

B. A Interferência da Intenção (Agência e Autonomia): a BSI é projetada para captar a intenção de movimento. Contudo, essa tecnologia atua em um circuito fechado (*closed-loop*) e é bidirecional, podendo influenciar o estado mental e a tomada de decisão (ANPD, 2025; LAVAZZA *et al.*, 2025; HOOPER, 2025): (i) *Perda de Controle* - Se a falha for um erro de decodificação, o paciente pode não ter tido a intenção de cair. O sistema pode ter gerado o comando de forma autônoma ou sido influenciado por um mau funcionamento (HOOPER, 2025); (ii) *Manipulação e Brainjacking* - O risco de “sequestro cerebral” (*brainjacking*), onde um atacante obtém o controle não autorizado do dispositivo intracraniano, é uma preocupação real para dispositivos neuromoduladores invasivos. Nesse caso, determinar se a queda foi causada pela intenção original do usuário, por um erro do sistema ou por uma manipulação externa é extremamente difícil, afetando a responsabilidade moral e legal (DANTAS; JEREISSATI, 2024; ANPD, 2025; LAVAZZA *et al.*, 2025); e (iii) *Violação de Neurodireitos* - Essa incerteza sobre a origem da ação (*humana vs. máquina vs. hacker*) mina o direito ao livre-arbítrio e à identidade pessoal (NEGRI *et al.*, 2025; LAVAZZA *et al.*, 2025).

3. Soluções Regulatórias para o CDC e o Nexo Causal – Para enfrentar os desafios probatórios e a opacidade algorítmica, especialmente no contexto do CDC (onde o fabricante responde objetivamente), a doutrina e as propostas regulatórias internacionais sugerem: (i) *Direito à “Caixa Preta” (Disclosure of Evidence)* - O paciente (consumidor) deve ter o direito de solicitar em juízo que o fabricante forneça elementos de prova sobre o sistema de IA suspeito. A recusa em fornecer tais provas pode levar à presunção de defeito do produto em face do fabricante. Essa medida visa facilitar a defesa do consumidor (Art. 6º, VIII do CDC) (NOGAROLI, 2023); (ii) *Presunção de Causalidade* - Em casos de excessiva dificuldade probatória devido à complexidade técnica, pode-se presumir o nexo de causalidade se a vítima demonstrar que o produto contribuiu para o dano e que o dano sofrido é *tipicamente compatível* com o defeito do produto (NOGAROLI, 2023); e (iii) *Responsabilidade Dual* - O sistema pode adotar uma responsabilidade dual: objetiva (externa) para o fabricante (protegendo a vítima-consumidor) e subjetiva (interna) para o médico (preservando o regime profissional) (NOGAROLI, 2023).

A linha tênue entre terapia e aprimoramento também é desafiada. Se a tecnologia pode restaurar a marcha, poderia ser usada para aprimorar a locomoção de pessoas sem deficiência?

As críticas ao aprimoramento apoiam-se frequentemente na noção de normalidade. No entanto, a definição de “pensamento” (e, por extensão, de função neural normal) é insuficiente para proteger contra abusos, pois deve incluir tanto a atividade cerebral quanto a estrutura cerebral. Além disso, a biotecnologia já avançou profundamente, e a neurotecnologia assistiva (como implantes cocleares ou marca-passos) já mudou drasticamente a vida humana, sem ser vista como prejudicial à natureza humana (VILAÇA; MARQUES, 2015; ANPD, 2025). Os bioconservadores questionam até que ponto o melhoramento cognitivo descaracterizaria a natureza humana, fragilizando os direitos humanos, e temem a instauração de uma nova forma de eugenia (LOPES *et al.*, 2022).

Empresas no setor de neurotecnologia, como a Neuralink, têm o objetivo duplo: “restaurar a autonomia de pessoas com necessidades médicas não atendidas hoje e desbloquear o potencial humano amanhã” (LAVAZZA *et al.*, 2025). Esse segundo objetivo visa especificamente a aplicação do dispositivo em indivíduos saudáveis para aprimoramento (LAVAZZA *et al.*, 2025). A regulação deve estabelecer limites claros para evitar o surgimento de desigualdades baseadas em aprimoramentos neurotecnológicos (CABRAL, 2024; GERARDI; XINARIS, 2025).

Por fim, é no Consentimento Livre e Esclarecido do paciente que os dilemas se aprofundam, exigindo um diálogo transparente sobre:

1. *Consentimento para a Agência Híbrida* – o paciente deve compreender que sua capacidade de andar será fruto de uma agência híbrida ou compartilhada entre sua vontade e a ação da BSI (HAAG; IENCA *et al.*, 2025): (i) *As Interfaces Cérebro-Computador (ICCs ou BCIs), como a BSI, estabelecem uma comunicação direta entre o cérebro e dispositivos externos* (DANTAS; JEREISSATI, 2024). O sistema de Lorach *et al.* (2023) converte a atividade cortical (intenção) em modulação analógica para a estimulação espinhal. Esta integração pode influenciar profundamente o senso de controle e personalidade do paciente (DANTAS; JEREISSATI, 2024; UNESCO, 2025). A autonomia e a privacidade se convergem, pois o controle sobre os dados neurais se torna central para a autodeterminação (HAAG; IENCA *et al.*, 2025). O sistema gera uma “agência ambígua” (*ambiguous agency*), onde o controle é compartilhado entre o usuário e a BCI, enfraquecendo a autonomia e levantando questões sobre a responsabilidade pelas ações (SI *et al.*, 2025); e (ii) *O paciente precisa consentir com essa “cessão” parcial de controle a um dispositivo*. Essa dimensão deve incluir o risco de o dispositivo falhar ou interpretar mal os sinais devido à opacidade do algoritmo, podendo gerar movimentos involuntários e quedas, conectando a autonomia diretamente à segurança (SI *et al.*, 2025; HAAG; IENCA *et al.*, 2025). Diretrizes internacionais recomendam que, em tecnologias

neuroimplantáveis, seja adotado um consentimento processual e contínuo (*process/ongoing consent*), com reiterações informativas e reavaliações periódicas de risco/benefício ao longo do acompanhamento (TREMBLAY-MCGAW *et al.*, 2025; JELEMBI, 2025), incluindo planos de saída, deveres pós-ensaio compatíveis com Helsinki e CIOMS (SCHÖNWEITZ *et al.*, 2024) e direito de recusar ou retirar o consentimento a qualquer momento (CFM, 2016; OEA, 2023; UNESCO, 2025).

2. *Consentimento para a Vulnerabilidade Cibernética* – O paciente tem o direito de ser informado sobre os riscos de segurança, incluindo o potencial de *brain hacking* ou acesso não autorizado à interface: (i) *A proteção de seus neurodados e de seu controle motor deve ser uma parte explícita do processo de consentimento, indo além da proteção de dados tradicional da LGPD* (IENCA; ANDORNO, 2017; LAVAZZA *et al.*, 2025). A conectividade sem fio das BCIs introduz riscos de cibersegurança (ARROYO, 2025). Dispositivos neuromoduladores invasivos estão sujeitos ao risco de “sequestro cerebral” (*brainjacking*), como visto acima, que pode gerar manipulação dos dados neurais e potencialmente causar danos ou distúrbios aos usuários (IENCA; ANDORNO, 2017; ANPD, 2025; GUERRA; SATHLER *et al.*, 2025); e (ii) *Os dados coletados pela neurotecnologia, chamados de neurodados ou dados neurais, são considerados altamente sensíveis, pois se referem à estrutura, atividade e função do sistema nervoso, e podem revelar informações íntimas sobre pensamentos e estados emocionais* (UNESCO, 2025; NEGRI *et al.*, 2025; PACHECO, 2025). A privacidade mental e a proteção desses dados são neurodireitos fundamentais (UNESCO, 2025; PACHECO, 2025).

3. *Consentimento em Contexto de Vulnerabilidade Extrema* - Assim como em outras condições neurológicas graves, pacientes com lesão medular encontram-se em um estado de extrema vulnerabilidade (SI *et al.*, 2025; LAVAZZA *et al.*, 2025): (i) *Esses pacientes podem estar tão ansiosos para melhorar sua condição que aceitam riscos elevados sem compreender completamente as implicações*. O processo de consentimento, portanto, deve ser rigorosamente protegido contra o denominado “desespero terapêutico”, garantindo que a decisão de se submeter a um procedimento tão inovador – e com efeitos de longo prazo ainda desconhecidos – seja genuinamente livre e esclarecida (BURWELL *et al.*, 2017; LAVAZZA *et al.*, 2025; UNESCO, 2025); e (ii) *A linha tênue entre terapia e aprimoramento também deve ser discutida, porque pode ser fonte de discriminação, aumentando a vulnerabilidade*: o objetivo é restaurar a “norma” ou superá-la? (CABRAL, 2024). Como analisado anteriormente, a BSI, embora atualmente terapêutica (restaurando a marcha), ilustra a tensão entre o uso médico e o uso não médico da neurotecnologia. O objetivo do desenvolvimento de BCIs é, portanto, duplo: “restaurar a autonomia de pessoas com necessidades médicas não atendidas hoje e desbloquear

o potencial humano amanhã” (aprimoramento) (LAVAZZA *et al.*, 2025). Ocorre que, como já dito anteriormente, o uso para aprimoramento (melhorar a locomoção de pessoas sem deficiência, por exemplo) é um risco explícito que pode exacerbar desigualdades sociais e criar novas formas de discriminação entre os “aprimorados” e os “naturais” (LAVAZZA *et al.*, 2025; PACHECO, 2025; UNESCO, 2025).

Por isso, a regulação deve estabelecer limites claros (UNESCO, 2025). O direito ao acesso equitativo a tecnologias de aprimoramento é um neurodireito proposto para garantir que esse acesso seja pautado por princípios de justiça e equidade (DANTAS; JEREISSATI, 2024; PACHECO, 2025). O processo de consentimento, nesse contexto, deve ser transparente sobre os potenciais riscos e a distinção entre terapia e aprimoramento (UNESCO, 2025; LAVAZZA, 2025).

2.2.2.6 Neuroestimulação em Malha Fechada para Epilepsia Refratária

Panorama Geral e Potencialidades:

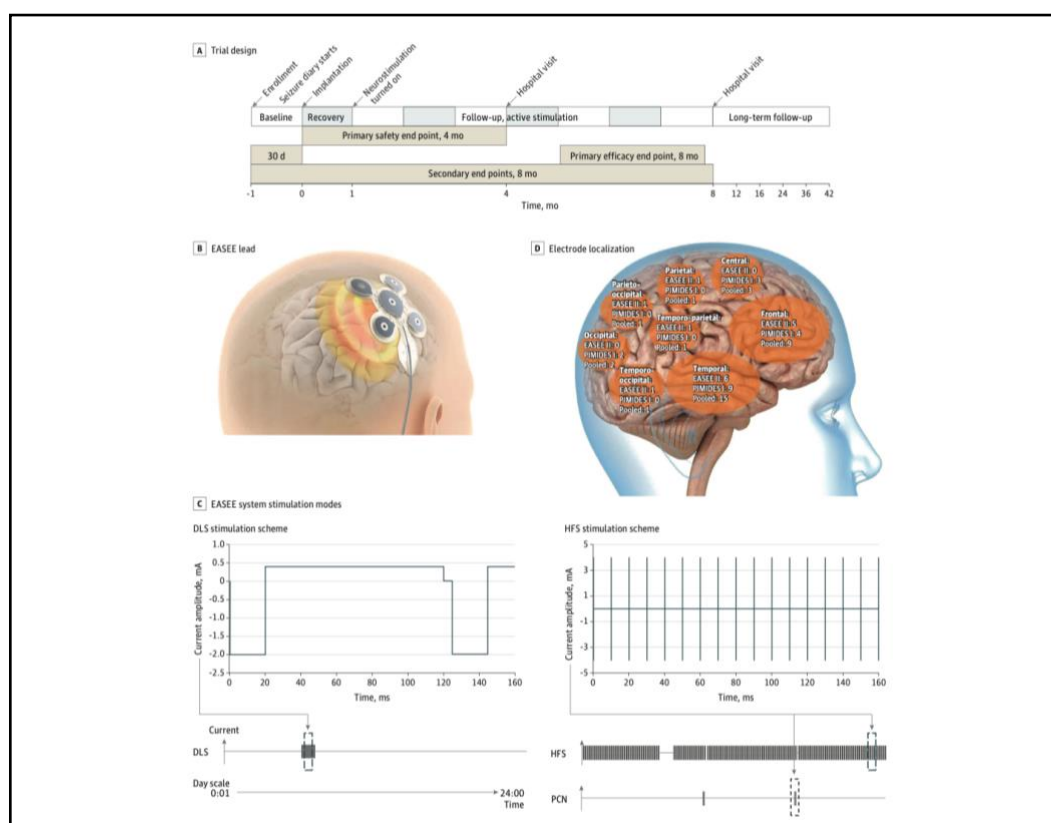
Para muitos pacientes com epilepsia, as crises persistem apesar do uso de múltiplos medicamentos, uma condição conhecida como epilepsia farmacorresistente ou refratária. Para esses casos, a neuroestimulação surge como uma opção terapêutica promissora. Uma das tecnologias mais avançadas nesse campo é a Neuroestimulação Responsiva (*RNS*), também conhecida como sistema de malha fechada (*closed-loop*). Esse dispositivo funciona como um “marca-passo para o cérebro” (SCHULZE-BONHAGE *et al.*, 2023). As Interfaces Cérebro-Computador (*ICCs* ou *BCIs*) de nova geração são tipicamente de malha fechada (*closed-loop*), o que significa que são capazes de ajustar automaticamente seus parâmetros funcionais através de algoritmos de IA (IENCA; IGNATIADIS, 2020; HAAG; IENCA *et al.*, 2025).

A *RNS* atua como um sistema neuroresponsivo. O algoritmo determina o momento em que a crise começa (*seizure onset*), e a estimulação elétrica focal é acionada para interrompê-la antes que a crise se manifeste clinicamente. O sistema de malha fechada detecta a atividade de crise com eletrodos sensores e fornece estimulação elétrica para prevenir a propagação da crise (LOZANO *et al.*, 2019; PATIL *et al.*, 2024).

Na pesquisa de Schulze-Bonhage *et al.* (2023), o implante é feito no crânio do paciente. O sistema de *RNS* monitora continuamente a atividade elétrica do cérebro para detectar os padrões anormais que sinalizam o início de uma convulsão. Ao identificar esses padrões, o dispositivo envia automaticamente pulsos elétricos para a área cerebral específica, interrompendo a crise antes que ela se manifeste clinicamente ou se espalhe. Estudos recentes

sobre sistemas de neuroestimulação, como o *EASEE* (um dispositivo minimamente invasivo), mostraram que mais da metade dos pacientes com epilepsia focal refratária tiveram uma redução de pelo menos 50% na frequência das convulsões após seis meses de tratamento. Os resultados mostraram que não foram observadas alterações significativas no humor, cognição ou na qualidade de vida geral (*overall quality of life*). No entanto, o sistema apresentou um perfil favorável de eventos adversos, com incidências mais baixas de sintomas depressivos e comprometimento da memória em comparação com outras abordagens, como a Estimulação Cerebral Profunda (*DBS*) no núcleo anterior do tálamo (*ANT-DBS*) (SCHULZE-BONHAGE *et al.*, 2023).

Figura 20 – Desenho do estudo, eletrodo do dispositivo de neuroestimulação e modos de estimulação, e localização dos eletrodos

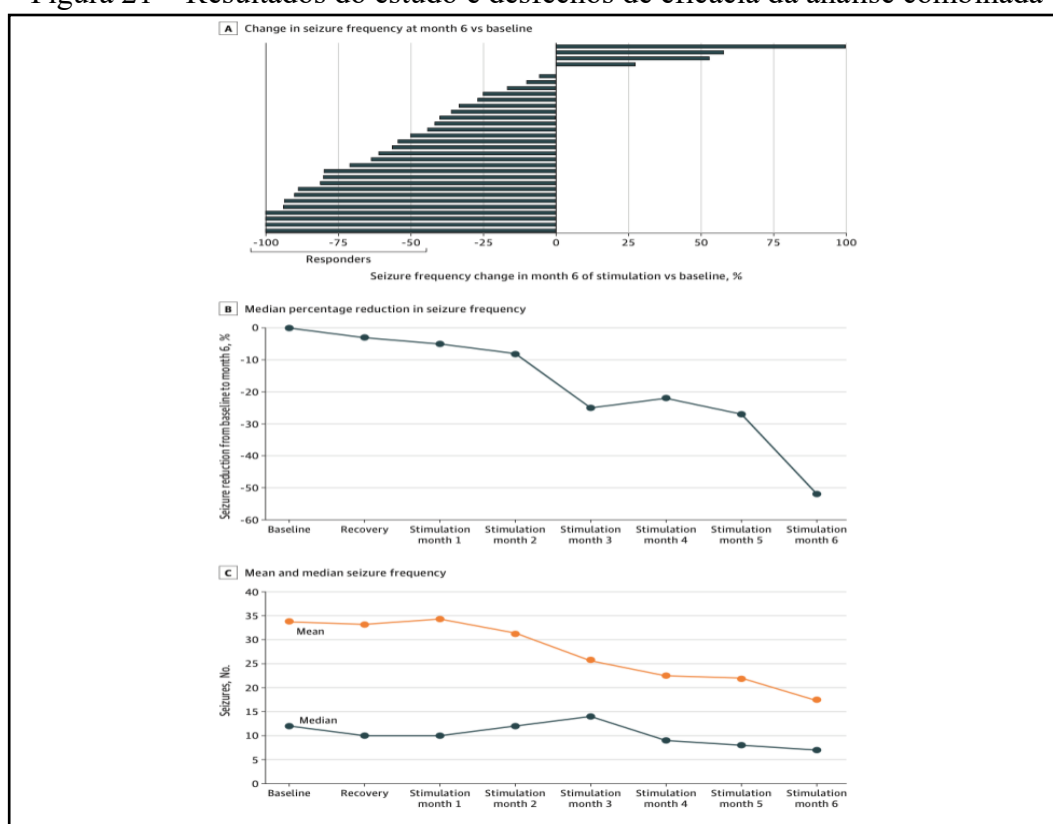


Fonte: SCHULZE-BONHAGE *et al.* (2023).

Nota: **(A)** Cronograma do estudo piloto para avaliar a viabilidade da neuroestimulação com o sistema *EASEE* para tratar epilepsia focal refratária à medicação (*EASEE II*) e estudo piloto para avaliar a viabilidade da neuroestimulação controlada pelo paciente com o sistema *EASEE*, para tratar epilepsia focal refratária à medicação (*PIMIDES I*). **(B)** O dispositivo de neuroestimulação consiste em um eletrodo colocado epicranialmente acima do foco epiléptico respectivo e conectado ao gerador de pulsos no tórax. O eletrodo inclui um eletrodo pseudo-Laplace de 5 canais (um eletrodo central rodeado por 4 eletrodos periféricos; diâmetro total, 77 mm) e um cabo de conexão. **(C)** Os modos de estimulação do dispositivo são estimulação semelhante à corrente contínua (*DLS*; pulsos catódicos de 2 mA, duração de 20 milissegundos, com pulsos de equilíbrio de 100 milissegundos de duração por 20 minutos por dia), estimulação de alta frequência (*HFS*; rajadas de 4 mA, bifásica

simétrica retangular, a 100 Hz, largura de pulso de 160 microssegundos; séries de 500 milissegundos de duração aplicadas a cada 2 minutos) e neuroestimulação controlada pelo milissegundos de duração por 20 minutos por dia), estimulação de alta frequência (*HFS*; rajadas de 4 mA, bifásica simétrica retangular, a 100 Hz, largura de pulso de 160 microssegundos; séries de 500 milissegundos de duração aplicadas a cada neuroestimulação controlada pelo paciente (PCN) sob demanda por 10 a 60 segundos, de acordo com as configurações (disponível para a coorte PIMIDES I). A intensidade da estimulação foi baseada na modelagem do método de elementos finitos e nos dados experimentais de Liu et al (estimulação repetitiva por corrente contínua disponível em Lu et al). **(D)** Implantação do eletrodo do dispositivo foi adaptada a cada paciente e colocada acima dos respectivos focos epilépticos. Aqui é mostrada a distribuição das localizações dos eletrodos na população do estudo, projetada no hemisfério direito (SCHULZE-BONHAGE et al., 2023).

Figura 21 – Resultados do estudo e desfechos de eficácia da análise combinada



Fonte: SCHULZE-BONHAGE et al. (2023).

Nota: **(A)** Resultados individuais da frequência de convulsões no sexto mês de estimulação em comparação com a linha de base. No sexto mês, um total de 17 dos 32 participantes (53%) respondeu ao tratamento e 4 dos 17 (23,5%) ficaram livres de convulsões. **(B)** Mediana da redução percentual na frequência de convulsões desde a linha de base até ao sexto mês de estimulação. **(C)** Média e mediana da frequência de convulsões em um grupo ao longo do tempo (SCHULZE-BONHAGE et al., 2023).

A Figura 20 é um diagrama multifacetado, que descreve o *design* do estudo de SCHULZE-BONHAGE et al. (2023), os componentes do dispositivo e o modo como a estimulação foi administrada aos pacientes.

O **Painel A – Trial Design (Desenho do Estudo)** – é uma linha do tempo que detalha as fases dos dois ensaios clínicos (*EASEE II* e *PIMIDES I*) cujos dados foram agrupados (*pooled analysis*): *Enrollment* - início do estudo, com a seleção dos participantes; *Baseline* (Linha de Base) - um período de 1 mês antes da cirurgia, onde a frequência de crises foi prospectivamente avaliada; *Implantação* - o momento da cirurgia para implantar o sistema *EASEE*; *Recovery* (Recuperação) - um período de 1 mês pós-implantação durante o qual a estimulação foi mantida desligada para separar os possíveis resultados associados à cirurgia dos resultados da estimulação; *Follow-up, Active Stimulation* (Acompanhamento, Estimulação Ativa) - o início da estimulação sem cegamento; *Primary Safety End Point* (“Principal *End Point* de Segurança”) - avaliado 4 meses após a implantação; *Primary Efficacy End Point* (“Principal *End Point* de Eficácia”) - avaliado 6 meses após o início da estimulação (8 meses após a implantação); *Long-term follow-up* - “Acompanhamento de longo prazo”.

O **Painel B – EASEE Lead (Eletrodo EASEE)** – ilustra os componentes físicos do sistema de neuroestimulação. O dispositivo consiste em um eletrodo (*lead*) implantado epicranialmente (acima do crânio) sobre a região do foco epiléptico e conectado a um gerador de pulsos implantado subcutaneamente na região peitoral. O *lead* é um eletrodo *pseudo-Laplace* de 5 canais, composto por um eletrodo central circundado por 4 eletrodos periféricos, com um diâmetro total de 77 mm e um cabo de conexão.

O **Painel C – EASEE system stimulation modes** (“Modos de Estimulação do Sistema *EASEE*) – é crucial por detalhar a combinação de dois modos de estimulação usados no tratamento: (i) *DLS (Direct Current–Like Stimulation)* - *Estimulação Semelhante a Corrente Contínua* - Pulsos catodais de 2 mA, com duração de 20 milissegundos, com pulsos de equilíbrio (100 milissegundos de duração). A aplicação é contínua, por 20 minutos por dia; (ii) *HFS (High-Frequency Stimulation)* - *Estimulação de Alta Frequência*: Rajadas de 4 mA, simétricas bifásicas retangulares, na frequência de 100 Hz, com largura de pulso de 160 microssegundos. *Train* de 500 milissegundos de duração aplicados intermitentemente a cada 2 minutos; e (iii) *PCN (Patient-Controlled Neurostimulation)* - Este modo era adicionalmente disponível para pacientes do estudo *PIMIDES I*. Permitia que os pacientes ativassem pulsos adicionais de *HFS* por 10 a 60 segundos quando percebessem sintomas ictais precoces.

O **Painel D – Electrode localization** (“*Localização do Eletrodo*”) – mostra a distribuição dos locais de implante dos eletrodos em todos os pacientes do estudo (projetados no hemisfério direito). Os eletrodos foram posicionados epicranialmente acima da região do foco epiléptico individual de cada paciente. A maioria dos implantes se concentrou nas regiões Temporais (15 pacientes) e Frontais (9 pacientes). Outras regiões incluíram central, parietal, occipital,

temporo-occipital e temporo-parietal. Isso indica que o estudo incluiu uma gama diversa de focos epilépticos, principalmente nas convexidades dorsolaterais (SCHULZE-BONHAGE *et al.*, 2023).

Riscos Jurídicos e Éticos:

A natureza do sistema de *RNS*, que monitora e intervém continuamente na atividade cerebral, levanta questões éticas e jurídicas profundas. A primeira delas é a privacidade dos dados neurais. A interface cérebro-computador (*BCI*) e a neuroestimulação, ao interagir e potencialmente modificar a estrutura e função neural, perturbam ou violam a dimensão “mental” do indivíduo (LAVAZZA *et al.*, 2025). O dispositivo coleta um fluxo constante de dados cerebrais, que são extremamente sensíveis. Neurodados são definidos como informações qualitativas e quantitativas sobre a estrutura, atividade e função do sistema nervoso (UNESCO, 2025). Eles são os correlatos mais diretos dos estados mentais em nível neurobiológico (MOLNÁR-GÁBOR; IENCA *et al.*, 2021). A doutrina e relatórios internacionais classificam os neurodados como informações altamente sensíveis (LOPES *et al.*, 2022; JELEMBI, 2025; ARROYO, 2025), sendo inerentemente involuntários e intrínsecos ao ser humano (ARROYO, 2025).

A Lei Geral de Proteção de Dados classifica dados de saúde como sensíveis, mas os neurodados contêm informações que vão além, podendo revelar estados cognitivos, afetivos e emocionais (IENCA; ANDORNO, 2017). Neurodados possuem valor proposicional e semântico, contendo informações sobre conceitos abstratos, pensamentos e intenções (BERTONI; IENCA, 2024). Eles fornecem *insights* sobre a dinâmica temporal da atividade cerebral, capturando mudanças em milissegundos que subjazem aos processos cognitivos, percepção e comportamento (BERTONI; IENCA, 2024). A perspectiva de decodificar a atividade neural implica a possibilidade de decodificar ou modificar processos cognitivos e afetivos (ARROYO, 2025).

A governança desses dados – quem os acessa e para que fins são usados – precisa de uma regulamentação específica e rigorosa para proteger a privacidade mental do paciente (YUSTE; GOERING *et al.*, 2017; LOPES *et al.*, 2022; NEGRI *et al.*, 2025). A privacidade mental é um direito fundamental que protege o indivíduo contra a intrusão não consensual de terceiros em suas informações mentais (sejam inferidas de dados neurais ou dados proxy) (BROWN, 2024; DANTAS; JEREISSATI, 2024).

A responsabilidade por falha do dispositivo é outro ponto crítico (NOGAROLI, 2023). Se o sistema não detectar uma crise (falso negativo) e o paciente sofrer um acidente, ou se

estimular o cérebro de forma inadequada (falso positivo) causando um efeito adverso, a quem cabe a responsabilidade? Ao fabricante (fato do produto), ao neurocirurgião ou à equipe que o programou? A complexidade da interação entre *hardware*, *software* e a biologia única de cada paciente torna a atribuição de causalidade um desafio (LIVANIS *et al.*, 2024). Neste ponto, a fim de evitar repetições desnecessárias, remetemos o leitor ao tópico anterior – A “*Ponte Digital*” *cérebro-coluna (BSI) para lesão medular* – onde o tema foi abordado de forma detalhada.

O Consentimento Livre e Esclarecido do paciente é particularmente desafiador devido à autonomia do dispositivo.

Consentimento para a Vigilância e Ação Autônoma: o paciente precisa compreender e consentir com o fato de que um algoritmo estará monitorando seu cérebro 24/7 e agindo sobre ele sem sua intervenção direta. Isso difere de terapias que o paciente controla ativamente. Esses sistemas podem, sob certas circunstâncias, anular a capacidade de tomada de decisão do paciente, delegando parte da agência humana a um sistema autônomo (IENCA; IGNATIADIS, 2020). A opacidade algorítmica (“caixa-preta”) entra em conflito direto com os requisitos padrão do CLE, que exigem divulgação e compreensão da informação. Se o profissional de saúde não consegue explicar por que o sistema tomou uma decisão, o paciente não pode receber a informação adequada para consentir de forma livre e esclarecida. Nesses casos, o paciente pode, no máximo, oferecer assentimento, que não possui o mesmo peso jurídico e procedimental do CLE (DIRECTOR, 2025; BLACKMAN; VEERAPEN, 2025). Portanto, a opacidade do algoritmo, que decide quando disparar o estímulo, impede uma explicação completa, exigindo uma comunicação honesta sobre essa limitação (COHEN, 2019; CILAR BUDLER; STIGLIC, 2025). Em suma, é necessário ser transparente sobre a capacidade e as limitações da tecnologia. O paciente deve ter o direito a uma explicação e justificação, indo além da simples informação (DANTAS; NOGAROLI, 2020; NOGAROLI, 2023).

Consentimento para a Incerteza e Efeitos na Subjetividade: o paciente deve ser informado sobre os riscos de a neuroestimulação contínua afetar não apenas as crises, mas também o humor, a memória ou a personalidade (IENCA, 2021; LOPES *et al.*, 2022). Em geral, intervenções de neurotecnologia, como a Estimulação Cerebral Profunda (*DBS*), demonstram potencial para efeitos neuropsiquiátricos adversos, incluindo apatia, alterações de comportamento (agressividade, impulsividade) e, mais raramente, alucinações (KLEIN, 2016; IENCA; ADORNO, 2017). Tais mudanças podem afetar a continuidade psicológica da pessoa (a sensação de persistir no tempo como a mesma pessoa). Até mesmo nos ensaios clínicos com dispositivos de neuroestimulação, foram relatados sintomas depressivos e comprometimento da

memória em alguns pacientes (IENCA; ADORNO, 2017; HAAG; IENCA *et al.*, 2025). O consentimento é, portanto, para uma intervenção cujos efeitos na subjetividade (identidade e intimidade), a longo prazo, ainda não são totalmente conhecidos, o que demanda um diálogo contínuo, dinâmico e iterativo sobre a identidade pessoal (DANTAS; JEREISSATI, 2024; SI; LIN *et al.* 2025).

Direito de Acesso e Controle sobre os Dados: Embora o sistema seja autônomo, o paciente deve ter o direito de ser informado sobre o funcionamento do dispositivo, de acessar seus próprios dados neurais e de participar das decisões sobre a programação, garantindo que ele mantenha a soberania sobre uma tecnologia que interage diretamente com seu cérebro (IENCA, 2021). Indivíduos devem ter a propriedade sobre seus dados neurais, com o direito de decidir quem pode acessá-los e para quais propósitos, e devem ter a capacidade de transferir esses dados (BERTONI; IENCA, 2024). Os princípios interamericanos da OEA propõem que se garanta o controle de cada pessoa sobre sua própria identidade individual e que o usuário tenha o direito ao controle voluntário sobre o funcionamento de qualquer dispositivo conectado ao seu cérebro (OEA, 2023; LOPES et al, 2022).

2.2.2.7 BCIs Invasivas Neuralink – (*Telepathy NI e Blindsight*)

Panorama Geral e Potencialidades:

A empresa Neuralink, fundada por Elon Musk em 2016 (CORNEJO-PLAZA; CIPPITANI, 2024), posicionou-se como uma das líderes no campo emergente das interfaces cérebro-computador, atraindo intensa atenção midiática e científica (LAVAZZA *et al.*, 2025). Seus avanços sinalizam uma aceleração na transição da neurotecnologia do laboratório para o mercado, com dois projetos principais em desenvolvimento: o *Telepathy NI* e o *Blindsight*.

O implante *Telepathy NI* foi o foco do estudo clínico inicial da empresa, chamado *PRIME* (*Precise Robotically Implanted Brain-Computer Interface*), um estudo de viabilidade precoce para pessoas com tetraplegia ou ELA (CLINICALTRIALS.GOV, 2024; DANTAS; JEREISSATI, 2024).

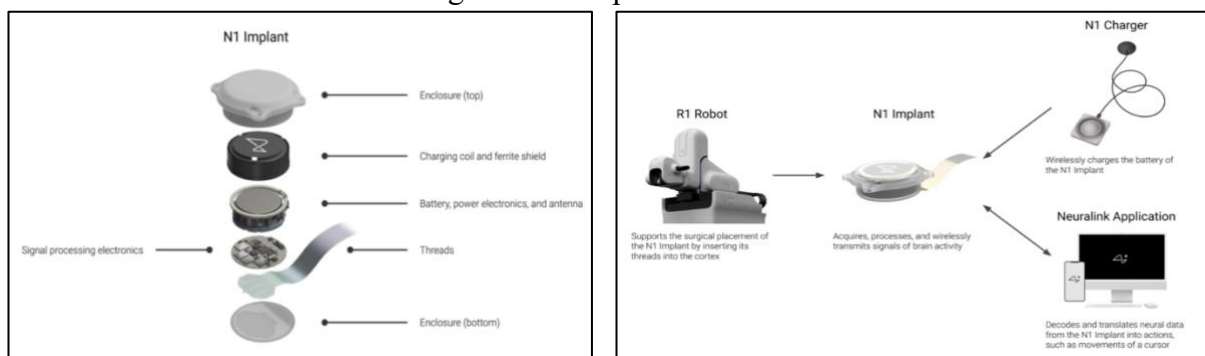
Trata-se de uma *BCI* intracortical projetada para registrar a atividade neural por meio de 1.024 eletrodos distribuídos em 64 fios flexíveis e ultrafinos, inseridos no córtex motor por um robô cirúrgico, com transmissão sem fio (FARAHANY, 2023; MUIRHEAD *et al.*, 2024; PARIKH; VENNIYOOR, 2024; PATIL *et al.*, 2024). O objetivo é permitir que pessoas com paralisia severa (tetraplegia) controlem computadores e outros dispositivos digitais apenas com

a intenção de movimento, restaurando sua autonomia digital (MUIRHEAD *et al.*, 2024; CLINICALTRIALS.GOV, 2024).

O primeiro participante, Noland Arbaugh, implantado em janeiro de 2024, demonstrou controlar um cursor do computador para jogar xadrez, navegar na internet e usar aplicativos, apenas pensando, conforme reportado pela *Nature* (DREW, 2024). Semanas após a cirurgia, ocorreu retração de fios, com queda de desempenho; segundo a empresa, ajustes no algoritmo de decodificação aumentaram a sensibilidade e melhoraram a performance subsequente (REUTERS, 2024; KHADER I. ALKHOURI, 2025).

Posteriormente, a Neuralink mencionou que foram implementadas medidas, como a redução do movimento do cérebro durante a cirurgia e a limitação da folga entre o implante e a superfície do cérebro, para prevenir a retração no segundo paciente (REUTERS, 2024).

Figura 22 – Implante Neuralink N1



Fonte: <https://neuralink.com/updates/prime-study-progress-update>.

Nota: (A) Vista “explodida” do Implante N1. (B) Representação de alto nível de como o Implante N1, o Carregador N1, o Aplicativo Neuralink e o Robô R1 interagem. Disponível em: <https://neuralink.com/updates/prime-study-progress-update>.

Figura 23 – Macaco Pager jogando *pong* com sua mente



Fonte: PATIL *et al.* (2024).

Nota: O vídeo que descreve a figura 23 pode ser acessado através do link: <https://www.youtube.com/watch?v=2rXrGH52aoM>.

A Figura 23, denominada *Pager playing with pong with its mind* (“Pager jogando pong com sua mente”), ilustra a primeira demonstração pública de sucesso em 2021, que estabeleceu o princípio da “*Telepathy*” – o controle de dispositivos digitais diretamente pela atividade neural (PATIL *et al.*, 2024).

A figura mostra o macaco *Pager*, que tinha um implante cerebral, jogando o videogame Pong com sua mente. A eficácia é demonstrada pelo fato de que o macaco inicialmente controlava o jogo com um *joystick*, mas após o *joystick* ser desconectado, a atividade neural capturada pelo chip continuou a ser traduzida em comandos que permitiram a *Pager* jogar o jogo e receber sua recompensa (um *smoothie* de banana).

Este experimento confirmou a viabilidade de usar a Interface Cérebro-Máquina (*BCI*) para decodificar intenções de movimento e traduzi-las em ações de controle digital, o mesmo princípio aplicado ao *Telepathy N1* em humanos. A eficácia do *Telepathy N1* (Estudo *PRIME*) – no contexto do estudo clínico *PRIME* - é descrita da seguinte forma: (i) *Resultados do Primeiro Participante* - O primeiro indivíduo a receber o implante N1 está em uma zona de recuperação total, sem efeitos adversos, e é capaz de mover um *mouse* na tela apenas pensando; e (ii) *Propósito da Eficácia* - O objetivo da *Neuralink* é permitir que pacientes paralisados ou com doenças neuronais controlem telefones ou computadores usando sinais cerebrais. O sucesso na tradução de sinais cerebrais em movimentos de cursor comprova a eficácia primária do sistema.

O segundo grande projeto é o *Blindsight*, um implante de prótese visual cortical que visa restaurar a visão em pessoas cegas, contornando totalmente os olhos e os nervos ópticos e estimulando diretamente o córtex visual (FERNANDEZ; ROBLES, 2024). Em setembro de 2024, o implante experimental *Blindsight*, da *Neuralink*, recebeu o *status* de “dispositivo inovador” (*breakthrough device*) da FDA dos EUA (LAVAZZA *et al.*, 2025).

O mecanismo do *Blindsight* se baseia na estimulação cortical direta para gerar percepções visuais, um processo que depende fortemente da Inteligência Artificial (IA) (WAISBERG *et al.*, 2024).

Em arquiteturas típicas de próteses visuais corticais, uma câmera (geralmente montada em óculos) captura as informações visuais. Uma unidade de processamento externa extrai as características relevantes e as converte em padrões de estimulação elétrica, que são transmitidos sem fio para o sistema implantado (WAISBERG *et al.*, 2024; FERNANDEZ; ROBLES, 2024). A estimulação elétrica intracraniana do córtex visual primário (V1), em pessoas cegas, gera

percepções visuais, historicamente conhecidas como fosfenos (FERNANDEZ; ROBLES, 2024).

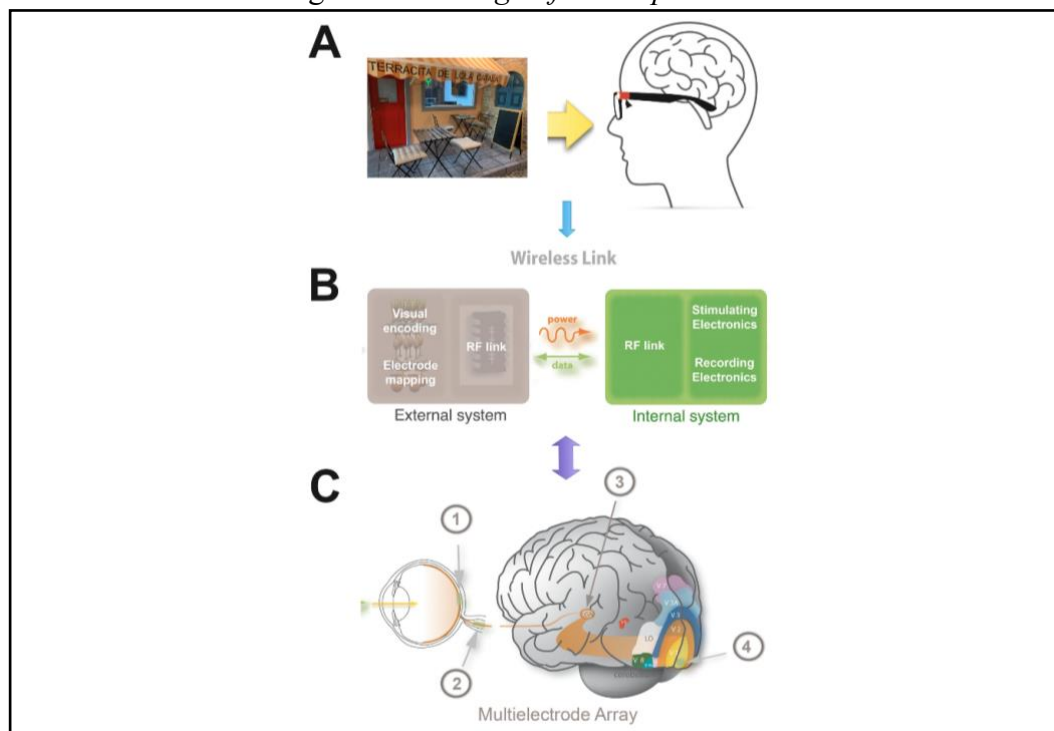
O *Blindsight* usará o chip N1 da Neuralink (ou uma versão dele) que, embora seja uma das inovações na área pela sua miniaturização, capacidade sem fio e grande número de microeletrodos (mais de mil, em comparação com as centenas usuais em outros dispositivos), ainda opera em um sistema que exige decodificação avançada de sinais neurais por algoritmos (LAVAZZA *et al.*, 2025; PARIKH; VENNIYOOR, 2024).

A aplicação de técnicas avançadas de IA e algoritmos de *machine learning* pode ajudar significativamente os usuários a interpretar e usar as informações visuais baseadas em fosfenos (FERNANDEZ; ROBLES, 2024).

Embora a promessa seja de “restaurar a visão”, Fernandez; Robles (2024) e Fine; Boynton (2024) indicam que a visão artificial inicial proporcionada por esses dispositivos é rudimentar e limitada pela própria estrutura do cérebro. O campo das próteses visuais ainda carece das complexas capacidades de processamento do sistema visual humano. O olho humano contém cerca de 125 milhões de fotorreceptores; replicar uma pequena fração dessa funcionalidade exigiria pelo menos 1 milhão de canais paralelos, excedendo as capacidades atuais. As próteses visuais têm fornecido historicamente uma forma limitada de visão, permitindo aos usuários perceber luz, formas e movimento. Simulações com “pacientes virtuais” sugerem que a qualidade perceptual dos dispositivos protéticos corticais será, no futuro previsível, limitada pela organização neurofisiológica do córtex visual primário (V1), e não apenas por restrições de engenharia. Aumentar o número de eletrodos nem sempre resulta em desfechos perceptualmente significativos.

Waisbert *et al.* (2024) e Lavazza *et al.* (2025) afirmam que o desafio reside em transmitir informações biologicamente relevantes para os locais neurais apropriados. Além disso, a manutenção a longo prazo e a segurança são incertas. Em um estudo, a qualidade do sinal e o desempenho diminuíram em implantes crônicos em macacos ao longo de 3 anos.

A visão de longo prazo de Musk, que vai além da terapia, sugere que o *Blindsight* e futuros implantes não se limitarão a restaurar o que foi perdido. A Neurotecnologia (como a Neuralink) está a caminho de um mundo onde será possível decodificar os processos mentais e manipular diretamente os mecanismos cerebrais subjacentes a intenções e emoções (DANTAS; JEREISSATI, 2024). O *Blindsight* pode ser usado para expandir os sentidos para além dos limites biológicos normais, permitindo, por exemplo, a percepção de espectros de luz que os humanos nunca viram ou campos eletromagnéticos (KHADER I. ALKHOURI, 2025).

Figura 24 – *Design of visual prostheses*

Fonte: FERNANDEZ; ROBLES (2024).

Nota: Todas as abordagens geralmente compartilham um conjunto comum de componentes-chave. **(A)** Uma câmera, normalmente montada em óculos padrão, captura o campo visual do sujeito. **(B)** Uma unidade de processamento externa extrai as características mais relevantes do ambiente, convertendo cenas visuais em padrões de estimulação elétrica. Essa unidade transmite energia e dados sem fio por meio de um link de radiofrequência (RF) para o sistema interno implantado, que decodifica os sinais, identifica os eletrodos-alvo e gera as formas de onda de estimulação finais. Em alguns projetos, o sistema interno também suporta registros eletrofisiológicos. **(C)** Por fim, uma matriz multieletródo é implantada em um dos vários locais potenciais ao longo da via visual, incluindo a retina (1), o nervo óptico (2), o núcleo geniculado lateral (3) ou o córtex visual (4) (FERNANDEZ; ROBLES, 2024).

O que distingue o projeto da Neuralink é seu objetivo explicitamente duplo: “restaurar a autonomia de pessoas com necessidades médicas não atendidas hoje e destravar o potencial humano amanhã” (LAVAZZA *et al.*, 2025). Essa segunda meta, focada no aprimoramento (*enhancement*), coloca a empresa no centro dos mais complexos debates éticos e jurídicos da atualidade (WAISBERG *et al.*, 2024).

Riscos Jurídicos e Éticos:

A atuação da Neuralink, por seu alto perfil e por operar na fronteira da ciência, condensa e amplifica os desafios ético-jurídicos das neurotecnologias (LAVAZZA *et al.*, 2025). O mais premente deles é a questão do Consentimento Livre e Esclarecido (CLE) do paciente, que, neste contexto, é multifacetado e de extrema complexidade.

O CLE é o foco das preocupações éticas relacionadas à Neuralink (LAVAZZA *et al.*, 2025). O processo de consentimento é um ponto de apoio fundamental para intervenções médicas eticamente sólidas (PIETRZYKOWSKI; SMILOWSKA, 2021). No entanto, a novidade e a sofisticação técnica das *BCIs* impõem demandas adicionais ao processo de CLE (DIRECTOR, 2025).

A obtenção de um consentimento verdadeiramente livre e esclarecido é complexa, e no campo das neurotecnologias, cria dúvidas sobre a validade do consentimento, pois essas tecnologias podem alterar, imitar, aprimorar ou interferir em diferentes graus no livre-arbítrio e na capacidade mental de fornecer consentimento (DIRECTOR, 2025; ARROYO, 2025).

Primeiramente, a extrema vulnerabilidade dos participantes da pesquisa – pacientes com tetraplegia ou Esclerose Lateral Amiotrófica (ELA) – pode comprometer a voluntariedade do consentimento. A esperança de recuperar funções perdidas pode levar à aceitação de riscos elevados sem a plena compreensão das implicações, um fenômeno conhecido como “desespero terapêutico” (BURWELL *et al.*, 2017; LAVAZZA *et al.*, 2025). Indivíduos que adquirem deficiências em um curto período de tempo podem ter um curto período de decisão para participar da pesquisa, gerando uma “vulnerabilidade decisória” (KLEIN, 2016). Além disso, a complexidade da neurotecnologia e os déficits cognitivos (flutuantes ou persistentes) de pacientes neurológicos e psiquiátricos tornam o processo de CLE ainda mais difícil (SI; LIN *et al.* 2025).

Em segundo lugar, a opacidade radical da tecnologia torna um consentimento “plenamente” esclarecido uma meta quase inatingível (LAVAZZA *et al.*, 2025; ANPD, 2025). A discussão transcende a “dupla caixa-preta” (opacidade do algoritmo e do cérebro). Os próprios pesquisadores admitem a incerteza sobre os efeitos de longo prazo de um implante *BCI* na plasticidade cerebral, na identidade pessoal e o risco de transtornos neurológicos ou de desorientação subjetiva (DANTAS; JEREISSATI, 2024; LAVAZZA *et al.*, 2025; ANPD, 2025). O paciente deve consentir não apenas com os riscos conhecidos da neurocirurgia, mas também com a incerteza fundamental sobre como a tecnologia poderá afetar sua mente e seu “eu” no futuro.

A privacidade mental e a segurança são outros pontos críticos. A natureza bidirecional dos dispositivos da Neuralink, que podem ler e potencialmente escrever no cérebro, levanta o risco de que dados neurais sensíveis sejam extraídos sem autorização (LAVAZZA *et al.*, 2025). Sim, a neurotecnologia pode não apenas registrar e analisar dados do sistema nervoso (leitura), mas também estimular e modular padrões neurais (escrita) (LOPES *et al.*, 2022; ANPD, 2025). Essa capacidade de “ler a mente” (acessar processos mentais) e “escrever na mente” (inserir

informações ou condicionar comportamentos) levanta a possibilidade de manipulação de emoções, sentimentos e decisões, afetando a dignidade e a identidade. A manipulação indevida ou a coerção no uso da neurotecnologia comprometem a autonomia e a liberdade de pensamento (LOPES *et al.*, 2022; DANTAS; JEREISSATI, 2024; UNESCO, 2025).

A falta de um arcabouço legal coeso para neurotecnologias expõe os pacientes a vulnerabilidades, justificando a necessidade de neurodireitos específicos (DANTAS; JEREISSATI, 2024; BROWN, 2024; LAVAZZA *et al.*, 2025). As neurotecnologias, especialmente aquelas voltadas para o consumidor ou que combinam com IA (como a Neuralink), frequentemente caem em uma “zona cinzenta” ou são classificadas sob marcos regulatórios pré-existentes (dispositivos médicos, privacidade de dados) que são inadequados para abordar seus desafios únicos. A tecnologia inova mais rápido do que o sistema regulatório consegue se adaptar (IENCA; ADORNO, 2017; CORNEJO-PLAZA; CIPPITANI, 2024).

Adicionalmente, a continuidade do cuidado por uma empresa privada é uma preocupação jurídica nova: o que acontece se a Neuralink falir ou descontinuar o suporte ao dispositivo, que se tornou parte essencial do paciente? (LAVAZZA *et al.*, 2025). Uma BCI implantada se torna uma parte essencial do indivíduo, e sua manutenção é fundamental para o bem-estar do paciente. Há precedentes de descontinuação de suporte, como no caso da empresa *Second Sight*, que encerrou seu olho biônico *Argus II* devido a problemas financeiros, deixando os clientes sem suporte (ARROYO, 2025). Pensando nisso, a UNIÃO EUROPEIA (2018a) recomenda a criação de entidades independentes de confiança para garantir o acesso contínuo à manutenção, reparos e atualizações de *software* essenciais para dispositivos médicos vitais e avançados, especialmente se o fornecedor original não puder mais prestá-los.

Por fim, como bem advertiu Andrea Lavazza, Marcello Ienca e outros pesquisadores (2025), a prática de a Neuralink comunicar seus avanços significativos primariamente por meio de plataformas de mídia social (como o X/Twitter de Elon Musk) sinaliza um desvio dos padrões estabelecidos de divulgação científica. O processo tradicional de divulgação científica envolve a publicação de resultados em periódicos revisados por pares, conferências especializadas e bancos de dados públicos, o que é crucial para garantir a transparência, a verificabilidade e a integridade na pesquisa científica. A Neuralink, ao optar por comunicar atualizações importantes por meio de plataformas de mídia social, levanta preocupações sobre a transparência e a responsabilidade ética. Outra preocupação dos referidos pesquisadores está relacionada à demora no Registro do Ensaio Clínico. A falha em registrar o primeiro estudo clínico da Neuralink no banco de dados *ClinicalTrials.gov* parecia violar as diretrizes éticas fundamentais para a pesquisa biomédica, como a Declaração de Helsinque. O pré-registro de

estudos clínicos é uma prática essencial na pesquisa biomédica por motivos éticos, que incluem a transparência, a responsabilidade e a proteção dos participantes. A omissão inicial dificultou que a comunidade científica e o público avaliassem a segurança, a eficácia e a ética da pesquisa realizada. A Neuralink submeteu o registro do estudo ao ClinicalTrials.gov em 21 de maio de 2024, antes da implantação de um segundo paciente (LAVAZZA *et al.*, 2025).

2.2.3 Conclusão

Como se pode observar, após a análise aprofundada dos casos concretos explorados neste Capítulo, representativos do estado da arte da Inteligência Artificial e das neurotecnologias no período de 2023 a 2025, estamos vivendo um período de grandes possibilidades terapêuticas e diagnósticas. Da detecção precoce de câncer à restauração da fala, do movimento e da visão, a promessa é a de uma medicina mais precisa, preditiva e personalizada. Contudo, essa mesma fronteira de inovação expõe um conjunto de desafios jurídicos e éticos transversais, de uma complexidade sem precedentes, que demandam uma resposta urgente e estruturada por parte do ordenamento jurídico.

A análise dos casos revela a emergência de temas recorrentes, porém com novas e mais profundas dimensões. A responsabilidade civil é tensionada por cadeias de desenvolvimento fragmentadas, como nos *foundation models* de acesso aberto. O risco de viés algorítmico e discriminação transcende os dados e se manifesta no *design* de *hardware* e em pressupostos neurais, ameaçando aprofundar as desigualdades estruturais em saúde. A privacidade, embora tutelada pela LGPD, é desafiada pela natureza única dos neurodados, que registram o próprio substrato da cognição, tornando o debate sobre neurodireitos – como o direito à privacidade mental e à identidade pessoal – não uma especulação futurista, mas uma necessidade presente.

Acima de tudo, a análise demonstra que o Consentimento Livre e Esclarecido do paciente, pilar da autonomia e da ética médica, deve ser repensado. Ele é desafiado em cada caso de uso: pela opacidade da “dupla caixa-preta” em *BCIs*, *BSIs* e *DBSs*, que impede uma explicação completa dos riscos; pela vulnerabilidade, *e.g.*, de pacientes com Alzheimer ou ELA; pela natureza probabilística de diagnósticos preditivos que invocam o “direito de não saber” (RÜGER, 2007); pela agência híbrida homem-máquina em neuropróteses, que redefine o controle corporal; e pela vigilância algorítmica contínua em sistemas de detecção de sepse. O consentimento, portanto, deixa de ser um ato único para se tornar um processo dialógico, contínuo e rigoroso, essencial para legitimar qualquer intervenção. Mas será que ele ainda é válido, ou seja, realmente livre e esclarecido? Isso nós discutiremos no Capítulo 5.

3 OPACIDADE, INTEGRIDADE E O DIREITO À EXPLICAÇÃO EM SISTEMAS DE SAÚDE INTELIGENTES

Conforme explorado no Capítulo anterior, a convergência entre inteligência artificial (IA), neurotecnologias e a medicina diagnóstica inaugura uma era de potencialidades transformadoras, prometendo diagnósticos mais céleres, precisos e personalizados, capazes de redefinir os paradigmas do cuidado em saúde (TELLES; ALCÂNTARA, 2024). Algoritmos de aprendizado profundo e interfaces cérebro-computador não são mais objetos de ficção científica, mas ferramentas cada vez mais presentes na prática clínica, com o potencial de otimizar desde a leitura de exames radiológicos até a restauração de funções neurológicas complexas. Contudo, essa revolução tecnológica traz consigo um desafio intrínseco e profundo, que pode ser considerado o seu “calcanhar de Aquiles” ético e jurídico: a opacidade.

Muitos dos sistemas de IA mais avançados, especialmente aqueles baseados em redes neurais profundas, operam como “caixas-pretas” (do inglês, *black boxes*), cujos processos decisórios internos permanecem em larga medida indecifráveis, mesmo para seus próprios desenvolvedores (PASQUALE, 2015). Essa característica transcende a mera limitação técnica para se tornar um problema jurídico central, que tensiona e ameaça direitos fundamentais do paciente. A “caixa-preta” não é apenas um desafio para engenheiros e cientistas da computação; é um obstáculo direto à autonomia do paciente, que depende de informação clara para consentir de forma válida; um complicador para a responsabilidade do médico, que precisa justificar suas decisões; e um corrosivo para a própria confiança que sustenta a relação paciente-médico (MITTELSTADT, 2021). Quando essa opacidade algorítmica se soma à já conhecida inescrutabilidade do cérebro humano, no campo das neurotecnologias, o desafio se torna exponencial.

Este Capítulo se propõe a dissecar o fenômeno da opacidade em suas múltiplas dimensões. A análise partirá do problema técnico, aprofundando o conceito de uma “dupla caixa-preta”, que emerge na interseção da IA e das neurotechs com a neurociência, e explorará suas implicações para a confiança, a validação clínica e a atribuição de responsabilidade. Em seguida, examinará a resposta tecnológica a esse desafio, o campo da Inteligência Artificial Explicável (XAI), diferenciando as necessidades de justificação em diagnósticos preditivos da necessidade de compreensão em ações neurotecnológicas interativas. Por fim, o Capítulo culminará na análise das consequências jurídicas diretas da opacidade, argumentando que ela provoca uma reconfiguração do direito à explicação e do Consentimento Livre e Esclarecido, e

justifica a emergência de novas formas de tutela da integridade física e, sobretudo, psicológica do paciente, em um debate que já alcança o nível constitucional no Brasil.

3.1 O PROBLEMA DA “DUPLA CAIXA-PRETA”: OPACIDADE ALGORÍTMICA E A INESCRUTABILIDADE NEURAL

A promessa de uma medicina aprimorada pela inteligência artificial encontra seu mais formidável obstáculo na natureza intrinsecamente opaca de seus sistemas mais poderosos. Essa falta de transparência, longe de ser um mero detalhe técnico, representa uma barreira fundamental que desafia a confiança, a validação e a responsabilidade no ecossistema da saúde. Quando essa complexidade algorítmica é aplicada não a imagens estáticas, mas à dinâmica e misteriosa atividade do cérebro humano, o problema se intensifica, dando origem ao que se pode conceituar como uma “dupla caixa-preta”.

3.1.1 Análise da “Caixa-Preta” Algorítmica em IA Diagnóstica

O termo “caixa-preta”, popularizado na literatura jurídica e tecnológica pela obra seminal de Frank Pasquale (2015), *The Black Box Society*, designa um sistema cujos mecanismos internos são misteriosos ou indecifráveis. Em tal sistema, é possível observar as entradas (os *inputs*, como os dados clínicos e de imagem de um paciente) e as saídas (os *outputs*, como uma recomendação diagnóstica), mas a lógica que transforma um no outro permanece oculta (PASQUALE, 2015). Essa opacidade não é, em geral, um defeito de projeto, mas uma característica inerente aos modelos de aprendizado de máquina mais avançados, notadamente as redes neurais profundas (*deep learning*).

Arquiteturas como as Redes Neurais Convolucionais (*CNNs*), que revolucionaram a radiologia e a patologia digital, são compostas por milhões ou até bilhões de parâmetros (os “pesos” das conexões entre neurônios artificiais) que são ajustados automaticamente durante o processo de treinamento (LITJENS *et al.*, 2017). Esses parâmetros permitem que a rede aprenda a reconhecer padrões hierárquicos e representações de dados extremamente abstratas e não-lineares, de uma forma que não corresponde à lógica humana (PASQUALE, 2015). O sistema não aprende “regras” diagnósticas como um médico faria; ele aprende correlações estatísticas complexas nos dados, resultando em um processo decisório que, embora frequentemente acurado, carece de uma justificativa intuitiva ou verbalizável (MITTELSTADT, 2021).

Os casos concretos explorados no Capítulo anterior ilustram vividamente este desafio. No estudo *PRAIM*, que demonstrou um aumento significativo na detecção de câncer de mama com o auxílio de uma IA, como um radiologista pode justificar perante a paciente – ou, em um cenário litigioso, perante um juiz – a decisão de convocar para uma biópsia baseada em um alerta de uma IA que destaca uma área suspeita com base em padrões texturais sutis que o olho humano treinado não consegue perceber? (EISEMANN; BUNK *et al.*, 2025).

A confiança do profissional no sistema depende de sua performance estatística, mas sua capacidade de explicar a recomendação individual é severamente limitada. Da mesma forma, no caso do *foundation model* Prov-GigaPath, treinado com 1,3 bilhão de imagens de patologia, a vastidão e a complexidade do modelo tornam virtualmente impossível uma auditoria completa de sua lógica interna para garantir que não existam vieses ocultos ou “atalhos” de raciocínio que possam levar a erros sistemáticos em subpopulações específicas de pacientes (XU; USUYAMA *et al.*, 2024). A opacidade, portanto, não apenas dificulta a explicação de decisões corretas, mas também mascara as causas de decisões erradas, comprometendo a segurança do paciente (PRICE II, 2017).

3.1.2 Introdução e Aprofundamento do Conceito de “Dupla Caixa-Preta” em Neurotecnologias

Se a opacidade algorítmica já representa um desafio imenso na IA diagnóstica convencional, sua aplicação às neurotecnologias inaugura um problema de ordem qualitativamente distinta e exponencialmente mais complexo: a “dupla caixa-preta”. Este conceito se refere à sobreposição de duas camadas de inescrutabilidade: a opacidade do algoritmo de IA e a opacidade do próprio cérebro humano.

A primeira camada, a caixa-preta algorítmica, é análoga àquela encontrada na radiologia. O *software* que decodifica os sinais neurais para, por exemplo, controlar uma prótese ou sintetizar a fala, é frequentemente baseado em arquiteturas de redes neurais complexas (como Redes Neurais Recorrentes - *RNNs*), cujos processos internos são igualmente opacos (BARNOVA *et al.*, 2023). O sistema aprende a mapear padrões de atividade neural a comandos ou palavras, mas o “como” exato desse mapeamento permanece indecifrável.

A segunda camada, e a que torna o desafio único, é a caixa-preta neural ou biológica. O próprio cérebro humano é, para a ciência atual, a caixa-preta por excelência. Apesar dos notáveis avanços da neurociência, persiste o que filósofos da mente como Joseph Levine (1983) denominaram de “lacuna explicativa” (*explanatory gap*). Trata-se da profunda dificuldade em

explicar como processos físicos e objetivos – o disparo de neurônios, a liberação de neurotransmissores – dão origem à experiência subjetiva e qualitativa da consciência, da intenção ou do pensamento (os *qualia*) - (LEVINE, 1983; TELLES; ALCÂNTARA, 2024; WAGNER-ALTENDORF, 2024; PASQUALE, 2015). Não compreendemos, em um nível fundamental, como a atividade eletroquímica do cérebro se traduz em uma intenção de mover o braço para a esquerda ou na formulação mental da palavra “olá”.

Um exemplo dessa opacidade neural é citado por Patricia Coutinho *et al.* (2024), que afirmam que os estimuladores cerebrais (como o *deep brain stimulation* para Parkinson) podem causar efeitos inesperados (“iatrogênicos”). No estudo, intitulado *Elevated Mood Induced by Subthalamic Nucleus Deep Brain Stimulation: A Video-Recorded Case Report*, eles descrevem um caso de um paciente com doença de Parkinson que experimentou “humor elevado” (que pode incluir sintomas de hipomania/mania) após ajustes na programação da *DBS* no núcleo subtalâmico (*STN*). Tais efeitos psiquiátricos adversos representam que o objetivo (controle de tremores) teve outro efeito cognitivo indesejado.

Outro exemplo, citado por Russell e Norvig (2021), está no fato de ainda não se compreender inteiramente como se armazena a memória ou como outras áreas do cérebro assumem funções quando uma área é danificada.

A sinergia dessas duas opacidades resulta em um problema que não é meramente aditivo, mas multiplicativo. Na IA radiológica, o algoritmo opaco analisa uma fonte de dados (uma imagem de tomografia) que, embora complexa, é estática e representa uma realidade física com correlatos anatômicos e patológicos conhecidos. A opacidade reside fundamentalmente no *processo* de análise. Nas neurotecnologias, como na Interface Cérebro-Computador (*BCI*) para restaurar a fala, o algoritmo opaco analisa uma fonte de dados (sinais neurais) que é dinâmica, ruidosa e representa uma realidade subjetiva (a intenção de falar), cujo mecanismo de geração (o cérebro) é, ele próprio, uma caixa-preta. Em outras palavras, a opacidade não está apenas no algoritmo que *lê* os dados, mas na *própria fonte* dos dados (IENCA; IGNATIADIS, 2020). O sistema de IA tenta decifrar um código (os padrões neurais) cuja gramática (as regras de funcionamento do cérebro que geram a intenção) não é completamente conhecida pela ciência. A incerteza do modelo computacional se compõe com a incerteza fundamental da neurociência, criando um nível de indecifrabildade muito maior (WILLETT *et al.*, 2023). Um erro no sistema – por exemplo, a *BCI* sintetizar a palavra “não” quando a intenção era “sim” – pode ter se originado de uma falha no *software*, de um padrão de atividade cerebral atípico ou “ruidoso” do paciente, ou de uma interação imprevisível e não-linear entre os dois sistemas opacos (RAMACHANDRAN, 2025).

3.1.3 Implicações da Dupla Opacidade para a Confiança, Validação Clínica e Atribuição de Nexo Causal

Esta indecifrabildade multiplicada tem consequências jurídicas e práticas devastadoras. Primeiramente, ela mina a confiança do usuário e do médico no sistema. A confiança em uma tecnologia não se baseia apenas em sua eficácia geral, mas também na capacidade de diagnosticar e corrigir falhas (LITJENS *et al.*, 2017; PASQUALE, 2015). Em *BCIs*, a não-estacionariedade dos sinais e a opacidade do decodificador tornam a atribuição de erro (do dispositivo *vs.* do usuário) incerta. Exemplifico: se um paciente com uma *BCI* não pode ter certeza se um erro de comunicação foi causado pelo dispositivo ou por sua própria “falha” em gerar um sinal neural “limpo”, sua relação com a tecnologia torna-se frágil e ansiogênica (MITTELSTADT, 2021; EISEMANN; BUNK *et al.*, 2025). A incapacidade de obter um diagnóstico claro da falha impede a construção de uma confiança robusta.

Em segundo lugar, a validação clínica de tais dispositivos por agências reguladoras, como a ANVISA no Brasil ou a FDA nos Estados Unidos, torna-se exponencialmente mais complexa. O paradigma tradicional de validação, baseado em testes com dados retrospectivos e ensaios clínicos controlados, mostra-se inadequado para sistemas que são, por natureza, adaptativos e que interagem com a neuroplasticidade de um cérebro individual, que está em constante mudança (XU; USUYAMA *et al.*, 2024; PRICE II, 2017; BARNOVA *et al.*, 2023). Como garantir a segurança e a eficácia de um algoritmo que aprende e se modifica em tempo real, em um dueto com um cérebro que também aprende e se modifica? A validação não pode ser um evento único, pré-mercado; ela exige um novo modelo de monitoramento contínuo, vigilância pós-mercado e planos de controle de mudanças predeterminados (*Predetermined Change Control Plans - PCCP*) – já objeto de orientação final da FDA (FDA, 2024; FDA, 2025; CANADÁ, 2025) – enquanto no Brasil os *SaMD* são regulados pela RDC 657/2022 e a vigilância pós-mercado integra o arcabouço de dispositivos médicos (ANVISA, 2022a; ANVISA, 2025).

Finalmente, e de forma mais crítica para o Direito, a dupla opacidade cria um verdadeiro “abismo probatório” para a atribuição de nexos de causalidade na responsabilidade civil. O nexo causal é o elo que liga a conduta do agente ao dano sofrido pela vítima, sendo um pilar essencial para qualquer pleito indenizatório (KFOURI NETO, 2024). Se um paciente submetido à Estimulação Cerebral Profunda (*DBS*) para Alzheimer sofre uma alteração de personalidade deletéria, como sua família poderia provar em juízo a causa do dano? A alteração foi causada

por uma estimulação elétrica defeituosa (fato do produto), uma programação negligente dos parâmetros pelo médico (erro profissional), ou foi uma reação idiossincrática e imprevisível do cérebro doente do paciente à estimulação, um risco inerente ao procedimento (iatrogenia)? A dupla opacidade torna a prova pericial, que buscaria isolar a causa do dano, uma tarefa hercúlea, senão impossível (PRICE II; COHEN, 2024; LEVINE, 1983; WAGNER-ALTENDORF, 2024). Isso pode criar “zonas cinzentas” de irresponsabilidade, onde o paciente sofre um dano claro, mas não há como apontar a sua causa real.

3.2 O DESAFIO DA EXPLICABILIDADE (*XAI*): DA JUSTIFICAÇÃO DE PREDIÇÕES À COMPREENSÃO DE AÇÕES NEUROTECNOLÓGICAS

Diante do abismo de opacidade criado pelos sistemas de saúde inteligentes, a ciência da computação tem buscado construir pontes. O campo da Inteligência Artificial Explicável (do inglês, *Explainable Artificial Intelligence - XAI*) emerge como a principal resposta técnica ao problema da “caixa-preta”, com o objetivo de dotar os modelos de IA de transparência, interpretabilidade e, em última instância, de confiabilidade. Contudo, a aplicação da *XAI* no domínio da saúde revela que a “explicação” não é um conceito monolítico; as necessidades de um médico que busca justificar um diagnóstico são fundamentalmente diferentes das de um paciente que tenta compreender a ação de uma neuroprótese (BLAJ, 2025; AZIZ *et al.*, 2025).

3.2.1 Apresentação do Campo da IA Explicável (*XAI*) e suas Técnicas

O propósito central da *XAI* é desenvolver um conjunto de técnicas e métodos que possam elucidar o funcionamento interno de um modelo de IA, tornando seu processo decisório compreensível para um ser humano. A premissa é que, ao entender “por que” um sistema chegou a uma determinada conclusão, os usuários – sejam eles médicos, pacientes ou reguladores – podem avaliar melhor a sua confiabilidade, identificar potenciais vieses e tomar decisões mais informadas (GUNNING, 2019; CANADÁ, 2025; FDA, 2025).

A *XAI* engloba as estratégias e metodologias utilizadas na construção de sistemas de IA que visam permitir aos usuários finais – como médicos, enfermeiros e até mesmo pacientes – compreender e interpretar as saídas e previsões feitas pelos modelos de IA. Em termos mais específicos, a interpretabilidade refere-se à capacidade de fornecer regras compreensíveis que governam o processo de tomada de decisão de um sistema, enquanto a explicabilidade foca na criação de uma interface que torna o funcionamento interno da IA acessível e compreensível

aos humanos. O propósito central da *XAI* é, portanto, desvendar a lógica por trás das decisões da IA, promovendo um “farol de confiança, conformidade, eficácia e robustez” (SADEGHI *et al.*, 2024).

Segundo Zahra Sadeghi *et al.* (2024), a pesquisa em *XAI* pode ser categorizada em seis grupos principais de técnicas, cada um com abordagens distintas para aumentar a transparência dos modelos de IA, como será demonstrado a seguir.

3.2.1.1 Métodos Orientados a Características (*Feature-oriented methods*)

Essas técnicas buscam identificar quais características de entrada (por exemplo, sintomas do paciente, *pixels* em uma imagem médica) mais influenciaram a decisão do modelo.

SHAP (*SHapley Additive exPlanation*): utiliza a teoria dos jogos para calcular a contribuição de cada característica individual para a previsão de um modelo, oferecendo uma compreensão clara de quais fatores foram mais importantes. Exemplos na saúde: o *SHAP* foi empregado para explicar a previsão de valores prototípicos específicos em relação a comunidades microbianas e fenótipos. Também foi usado para a detecção precoce da Doença de Parkinson a partir de imagens *SPECT DaTscan* e para detectar doenças agudas críticas usando dados de prontuários eletrônicos. Outras aplicações incluem o diagnóstico de epilepsia, a detecção de arritmia cardíaca, a identificação de câncer de pulmão e a validação de um modelo de previsão de risco para câncer de próstata. No contexto de Doença de Alzheimer, o *SHAP* foi usado em um modelo de duas camadas para identificar e prever a progressão da doença, fornecendo explicações para o raciocínio do modelo. Adicionalmente, foi aplicado para investigar a relação entre a composição de células imunes tumorais e as taxas de sobrevida no câncer de mama, identificando células significativas como macrófagos M0, células B, CD8+ e células NK T.

CAM e Grad-CAM (*Class Activation Maps*): usados em redes neurais convolucionais (*CNNs*), geram “mapas de calor” visuais que destacam as regiões de uma imagem que mais contribuíram para a decisão do modelo (por exemplo, em diagnósticos por imagem, onde o modelo “olhou” para fazer sua conclusão). Exemplos na saúde: mapas de saliência (como os gerados por *CAM/Grad-CAM*) são usados para interpretar decisões de redes neurais convolucionais no diagnóstico de glaucoma com base em imagens de fundo de olho. O *Grad-CAM* foi aplicado em imagens de fundo de olho para triagem de doenças da retina, Retinopatia Diabética (RD) e avaliação da qualidade de imagens retinianas. Foi também utilizado para a

identificação de pneumonia com alta precisão através de imagens de raios-X, e para detecção de eventos hipoglicêmicos baseados em ECG.

3.2.1.2 Métodos Globais (*Global methods*)

Têm como objetivo explicar as previsões de redes neurais em diversas populações, permitindo a identificação e análise de subpopulações distintas. Eles ajudam a entender as relações intrincadas entre diferentes características em um nível mais amplo.

GAMs (Global Attribution Mappings): criam uma matriz de distância de classificação pareada para capturar relações complexas entre características e, por meio de agrupamento, representam *clusters* de características com “medoides” para construir uma atribuição global, permitindo o exame de características únicas de vários grupos de amostras.

Exemplo na saúde: a utilidade dos *GAMs* está em elucidar previsões de redes neurais em diversas populações, oferecendo vantagens para identificar e analisar distintas subpopulações em vários níveis de granularidade. Isso pode ser aplicado em cenários onde é necessário entender como um modelo de IA se comporta para diferentes grupos de pacientes ou características em um nível mais abrangente.

3.2.1.3 Modelos de Conceito (*Concept Models*)

Permitem representar o estado interno de uma rede neural em um formato compreensível para humanos, utilizando conceitos definidos pelo usuário.

CAVs (Concept Activation Vectors): medem o nível de contribuição de conceitos definidos pelo usuário (por exemplo, a presença de listras no corpo de um animal para classificá-lo como zebra) para a saída de uma rede neural.

Exemplo na saúde: os *CAVs* oferecem uma descrição conceitual amigável para o estado interno de uma rede neural e podem ser utilizados para o diagnóstico prático de Retinopatia Diabética (RD) usando imagens de fundo de olho, uma condição curável que pode causar cegueira.

3.2.1.4 Modelos Substitutos (*Surrogate Models*)

Simplificam a explicação de modelos complexos e opacos. Eles constroem um modelo mais simples e interpretável que se comporta de forma semelhante ao modelo original para uma previsão específica.

LIME (Local Interpretable Model-agnostic Explanation): cria explicações ótimas localmente de modelos de *ML* complexos, gerando um modelo interpretável (o “substituto”) que aproxima o comportamento do modelo “caixa-preta” na vizinhança de uma amostra de entrada específica.

Exemplos na saúde: o método *LIME* foi utilizado para o diagnóstico da Doença de Alzheimer. Também foi empregado para classificar a Doença de Parkinson a partir de exames *DaTscan*, para diagnóstico de glioblastoma e para um sistema de diagnóstico de câncer de pulmão assistido por computador que transforma características críticas em linguagem natural. Outras aplicações incluem a previsão de alta hospitalar pós-AVC, a análise de doenças cardíacas e o diagnóstico de câncer de próstata. Para hepatite, o *LIME* foi parte de uma estrutura de IA explicável para a previsão de risco de deterioração.

3.2.1.5 Métodos Locais Baseados em Pixels (*Local, pixel-based methods*)

Focados em determinar quais partes de uma entrada (como *pixels* em uma imagem) contribuíram para a decisão de uma rede neural de múltiplas camadas.

LRP (Layer-wise Relevance Propagation): produz um mapa de calor que oferece *insights* sobre quais *pixels* da imagem de entrada contribuíram para a previsão do modelo e em que medida, destacando as variáveis que afetam positivamente a escolha da rede.

Exemplo na saúde: o *LRP* pode ser muito útil para ajudar médicos a explicar diagnósticos de redes neurais para a Doença de Alzheimer (e possivelmente outras condições) realizados com base em dados de ressonância magnética estrutural (*MRI*). Também é usado para identificar os estágios da DA.

3.2.1.6 Métodos Centrados no Humano (*Human-centric Methods*)

Diferentemente das abordagens puramente estatísticas ou baseadas em características/*pixels*, estes métodos buscam uma compreensão antropomórfica, focando em como os humanos avaliam informações (por exemplo, por similaridade com protótipos previamente aprendidos). Esses métodos têm uma abordagem fundamentalmente nova para a explicabilidade, baseada na compreensão antropomórfica em vez de estatísticas puras. Porém,

“falharam em oferecer explicações lógicas” em sua utilização, por isso este campo vem buscando preencher essa lacuna, visando uma compreensão mais intuitiva e humana do raciocínio da IA. Isso é crucial em contextos médicos, onde a confiança e a compreensão humana são primordiais.

Em suma, as técnicas de *XAI* buscam fornecer diferentes “janelas” para o funcionamento interno dos modelos de IA, cada uma com sua perspectiva sobre o que torna uma decisão compreensível. Essa diversidade é fundamental para adaptar as explicações às necessidades específicas dos usuários, como os profissionais de saúde, que frequentemente preferem formatos visuais e lógicos em vez de explicações puramente matemáticas (SADEGHI *et al.*, 2024). Embora a *XAI* seja um campo em constante evolução, seu desenvolvimento é vital para garantir que a IA na saúde não apenas seja eficaz, mas também confiável, transparente e ética.

Há quem divida as técnicas de *XAI* em duas grandes categorias. De um lado, há os modelos intrinsecamente interpretáveis (ou *ante-hoc*), concebidos para serem transparentes desde o início – por exemplo, árvores de decisão ou regressão linear. Esses modelos são simples de interpretar, mas em geral têm desempenho inferior aos algoritmos complexos (*deep learning*). Do outro lado, estão os métodos pós-hoc, aplicados *após* o modelo ter sido treinado. Esses últimos explicam “caixas-pretas” já existentes, independentemente de sua arquitetura interna. Em geral, o uso de métodos intrínsecos facilita a explicação, porém pode haver *trade-off* com acurácia; já os métodos *post-hoc* permitem interpretar qualquer modelo complexo, sacrificando por vezes a fidelidade ou precisão da explicação (MCGRATH; JONKER, 2024).

Dentro dos métodos *post-hoc*, que, de acordo com Zahra Sadeghi *et al.* (2024), são os mais comuns para modelos de *deep learning*, destacam-se: (i) *Métodos Agnóstico de Modelo (model agnostic method)* - Estas técnicas podem ser aplicadas a qualquer tipo de modelo de “caixa-preta”. Dois exemplos proeminentes na área da saúde são: o *LIME* e o *SHAP*, já citados acima; e (ii) *Métodos Específicos do Modelo (model-specific method)* - Estas técnicas são desenhadas para arquiteturas específicas, como o *Grad-CAM*, também abordado nos métodos globais acima.

3.2.2 A Persistência da Opacidade e os Limites da Explicabilidade

Embora a Inteligência Artificial Explicável (*XAI*) tenha se estabelecido como uma abordagem crucial para conferir clareza e interpretabilidade aos modelos de IA, especialmente em setores de alto risco como a saúde, a opacidade ainda persiste como um desafio significativo. A demanda por maior transparência e o desejo dos usuários finais (pacientes e médicos) por

interpretabilidade evidenciam uma lacuna entre o desenvolvimento de modelos complexos de “caixa-preta” e a capacidade de fornecer explicações compreensíveis aos usuários. Mesmo com os avanços da *XAI*, diversas barreiras impedem a sua aplicação disseminada e a plena eliminação da opacidade em sistemas de saúde.

Os principais aspectos que demonstram a persistência da opacidade, apesar dos esforços da *XAI*, segundo Zahra Sadeghi *et al.* (2024), são os que passaremos a analisar a seguir.

3.2.2.1 Complexidade Inerente dos Modelos de ML

A análise de modelos de *Machine Learning* (*ML*) complexos, frequentemente exige conhecimento avançado em matemática e estatística, o que se torna uma barreira para usuários finais, como os profissionais de saúde. Modelos de alto desempenho são intrinsecamente complexos, com muitas camadas e interconexões, tornando quase impossível rastrear seu processo de treinamento em milhões de amostras.

3.2.2.2 Limitações das Metodologias XAI

As abordagens de *XAI*, muitas vezes, destacam apenas as regiões ou *features* relevantes para as saídas do modelo de *ML*, sem necessariamente determinar as causas subjacentes ou o “porquê” dessa relevância. Isso significa que, embora se saiba *o que* influenciou a decisão, o *como* e *porquê* profundo podem permanecer obscuros. Além disso, a avaliação adequada das explicações fornecidas ainda é uma questão aberta, e alguns pesquisadores expressam dúvidas sobre a confiabilidade dos métodos *XAI*, considerando-os potencialmente enganosos.

3.2.2.3 Compromisso entre Desempenho e Transparência (Performance vs. Transparency Trade-off)

Existe uma relação inversamente proporcional entre a explicabilidade e o desempenho/acurácia dos sistemas de IA (DIRECTOR, 2025). Aumentar a transparência de um modelo de *ML* pode torná-lo menos eficiente na consecução de seus objetivos. Modelos de *caixa-branca* (como modelos lineares ou árvores de decisão) oferecem alta explicabilidade, mas geralmente com menor desempenho, enquanto modelos de *caixa-preta* (como *Deep Learning* e métodos de *ensemble*) alcançam alto desempenho, mas são inerentemente opacos

(NORORI *et al.*, 2021). Encontrar um equilíbrio entre alta explicabilidade e desempenho aceitável é um desafio fundamental em sistemas de saúde.

3.2.2.4 *Lacuna entre Especialistas e Usuários Finais*

Embora especialistas em *ML* possam se sentir confortáveis com explicações matemáticas, os clínicos, na área médica, geralmente preferem receber explicações em formatos visuais. Essa diferença impõe limitações no formato de saída dos métodos *XAI* e destaca que a compreensibilidade para o usuário final ainda não é plenamente atendida. A colaboração entre especialistas em *ML* e clínicos é essencial para depurar e melhorar modelos, mas pode ser desafiadora devido às agendas dos clínicos.

3.2.2.5 *Necessidade de Supervisão Humana e Confiança Limitada*

Em aplicações de segurança crítica, como a saúde, a tomada de decisões não pode ser deixada inteiramente para sistemas de *ML*, sendo necessária a supervisão de especialistas humanos para evitar catástrofes. Mesmo com a *XAI*, a IA é vista como uma inteligência assistiva, que acelera a análise de dados e a extração de conhecimento, mas ainda requer estruturas *human-in-the-loop*. Isso indica que a confiança completa e, conseqüentemente, a total transparência, ainda não foram plenamente alcançadas para que os sistemas operem autonomamente sem supervisão humana.

3.2.2.6 *Ambiguidade na Definição de Interpretabilidade*

A interpretabilidade é um termo complexo e com múltiplas nuances, sem uma definição única. Essa falta de clareza conceitual dificulta a criação de um sistema de *ML* idealmente interpretável, especialmente na saúde, onde requisitos de solidez, compreensão e escopo (local ou global) precisam ser balanceados.

Em suma, embora a *XAI* seja um avanço promissor para desvendar a “caixa-preta” da IA em saúde, ela ainda enfrenta desafios inerentes à complexidade dos modelos, às suas próprias limitações metodológicas e aos compromissos práticos. A opacidade, portanto, permanece como um obstáculo a ser superado para a plena integração e confiança nos sistemas de IA na prática clínica.

3.2.3 Explicação e Lacunas de Explicabilidade em Neurotecnologias: da Justificação de Predições à Compreensão de Ações (*XAI4BCI*)

A literatura de *XAI* frequentemente trata a “explicação” como um objetivo monolítico. Todavia, o contexto de uso na saúde revela uma taxonomia de necessidades explicativas crucial para o debate jurídico. A natureza da interação entre o humano e o sistema de IA define o tipo de explicação que é necessária e útil em cada caso específico (MITTELSTADT, 2021; RAJPURA *et al.*, 2023). Em outras palavras, a explicação demandada varia conforme o papel da IA – se está apenas fornecendo uma predição ou executando uma ação – e isso impacta diretamente as estratégias de *explainability* requeridas.

3.2.3.1 Justificação de predições (IA diagnóstica “estática”).

Em *pipelines* clássicos de diagnóstico por imagem (*e.g.*, triagem de Retinopatia Diabética - RD), o algoritmo retorna uma conclusão e, para que clínicos e pacientes a avaliem, necessitam de uma justificação retrospectiva: “*por que o sistema chegou a essa conclusão?*”.

O caso do sistema *IDx-DR* – avaliado clinicamente e aprovado pela FDA – ilustra o ponto: modelos autônomos conseguem identificar RD com elevada sensibilidade/especificidade e são usados para triagem, exigindo explicações que tornem *auditável* a base visual da decisão (*e.g.*, mapas de saliência/*Grad-CAM*) (ABRÀMOFF; LAVIN *et al.*, 2018).

Nessa família de tarefas, técnicas *post-hoc* de visualização local (como *Grad-CAM*) estão relativamente maduras para “mostrar” quais regiões da imagem foram determinantes – uma forma de *justificação* adequada ao ato médico, embora, como se observa do item anterior, a opacidade ainda persiste na *XAI*, assim como a explicabilidade também possui limitações. (SADEGHI *et al.*, 2024).

3.2.3.2 Compreensão de ações (neurotecnologias interativas/fechadas).

Em interfaces cérebro-computador (*BCI*) e interfaces cérebro-coluna (*BSI*), a IA decodifica estados neurais e executa ações motoras ou de estimulação em tempo real (HAAG; IENCA *et al.*, 2025; METZGER *et al.*, 2024). Aqui, a pergunta do usuário muda para “por que o sistema fez isto agora?” – isto é, exige-se uma explicação situacional e temporalmente indexada da ação (não apenas da acurácia global).

Dois exemplos concretos evidenciam essa mudança de paradigma: (i) *Controle de braço robótico (BCI motora)* - Estudos com participantes tetraplégicos demonstraram controle 3D de membros robóticos por decodificação de atividade cortical em tempo real, permitindo atos funcionais como alcançar objetos e beber de uma garrafa. A ação do sistema (“mover o braço para a esquerda”) deriva de uma interpretação contínua da intenção motora – contexto em que a necessidade do paciente é entender o *porquê daquela ação naquele instante* (HOCHBERG *et al.*, 2012); e (ii) *“Ponte digital” cérebro-medula (BSI)* - Como explorado no Capítulo anterior, em 2023, foi demonstrada uma interface que reconecta sinais corticais a um neuroestimulador medular, permitindo que um paciente com lesão medular volte a andar, controlando voluntariamente as pernas após treinamento. A IA decodifica intenções de marcha e aciona padrões de estimulação na medula lombar em sincronia com a intenção – um exemplo paradigmático de malha fechada em que a explicação precisa ser operacional, local e temporal para sustentar agência e segurança do usuário (LORACH *et al.*, 2023)

Enquanto a justificção de predições em imagens médicas é hoje servida por um ecossistema robusto de métodos *post-hoc* (*Grad-CAM, LIME/SHAP* etc.), a compreensão de ações em *BCI/BSI* permanece um desafio de ordem superior: é preciso explicar em tempo real decisões tomadas sobre fluxos neurais de alta dimensionalidade, ruidosos, dinâmicos e idiossincráticos (variam entre e dentro de indivíduos). A literatura de *XAI* para *BCI* (*XAI4BCI*) ainda é incipiente e predominantemente orientada a objetivos de engenharia (validação neurofisiológica, *debug* e otimização de controladores), com déficit de explicações inteligíveis ao usuário final (RAJPURA *et al.*, 2023; SADEGHI *et al.*, 2024).

A natureza dos dados neurais, segundo Param Rajpura *et al.* (2023), apresenta complexidades notáveis, tornando a aplicação de técnicas padrão de *XAI* extremamente desafiadora: (i) *Alta dimensionalidade* - Os modelos de *machine learning* empregados em *BCI* frequentemente lidam com dezenas ou centenas de canais de sinais neurais (*EEG, ECoG, LFPs* etc.), gerando milhões de parâmetros. Explorar e explicar quais combinações de características desses sinais levam a determinada decisão é significativamente mais complexo do que em dados estáticos de imagem; (ii) *Baixa razão sinal-ruído* - Os sinais neurais são notoriamente ruidosos. Uma grande parte da atividade registrada pode não ser relacionada ao sinal de interesse. Essa característica torna difícil identificar com clareza quais padrões foram relevantes para a decisão da IA, e a *XAI* pode ser fundamental para melhorar a robustez dos sistemas (por exemplo, entender como o modelo lida com artefatos ou ruídos no *EEG*); (iii) *Temporalidade intrínseca* - Diferentemente de uma imagem estática, os sinais cerebrais têm uma dinâmica temporal intrincada. As interfaces de explicação para *BCI* precisam

considerar padrões que evoluem no tempo, além de aspectos espaciais (quais regiões/canais) e espectrais (frequências). Isso adiciona camadas extras de complexidade à geração de explicações, já que é necessário comunicar quando e em qual sequência certos padrões levaram a uma ação da IA; e (iv) *Enorme variabilidade* - Há uma considerável heterogeneidade nos sinais neurais, tanto entre diferentes indivíduos, quanto no mesmo indivíduo ao longo do tempo. Essa variabilidade dificulta a criação de modelos generalizáveis e confiáveis. Consequentemente, torna-se crucial explicar aos usuários *por que* um determinado *BCI* pode não estar funcionando bem em certo momento ou em certo paciente – em vez de atribuir a falha ao acaso, a explicação deve apontar fatores como variações fisiológicas ou desvios em relação aos dados de treinamento.

Essa distinção é de suma importância, pois as técnicas de *XAI* atuais são muito mais maduras e eficazes para o primeiro cenário (justificação de previsões) do que para o segundo (compreensão de ações). Gerar um mapa de calor em uma imagem estática é um problema computacionalmente tratável. Fornecer uma explicação inteligível, em tempo real, para uma ação baseada na análise de fluxos de dados neurais de alta dimensionalidade, ruidosos e que variam no tempo, é um desafio de uma ordem de magnitude muito maior (SADEGHI *et al.*, 2024). A necessidade de explicação, portanto, se bifurca: a IA diagnóstica demanda uma justificação que apoie a decisão clínica, enquanto a neurotecnologia interativa demanda uma compreensão que garanta a agência e o controle do usuário sobre seu próprio corpo e mente estendidos (HAAG; IENCA *et al.*, 2025).

3.3 O DIREITO À EXPLICAÇÃO EM SAÚDE E SUAS NOVAS DIMENSÕES JURÍDICAS

A opacidade inerente aos sistemas de saúde inteligentes não é apenas um problema técnico a ser resolvido pela engenharia; é um desafio que reverbera no núcleo do ordenamento jurídico, exigindo uma releitura de direitos e garantias fundamentais. A legislação de proteção de dados, tanto no Brasil quanto na Europa, já antecipou a necessidade de salvaguardas contra decisões automatizadas opacas. No entanto, no contexto da saúde, esse “direito à explicação” transcende sua moldura original e se funde com a doutrina secular do Consentimento Livre e Esclarecido, tornando-se uma condição essencial para a autonomia do paciente, especialmente daqueles em situação de vulnerabilidade (HURLEY *et al.*, 2025).

3.3.1 Análise do Direito à Explicação à Luz da LGPD (Art. 20), do GDPR (Art. 22) e do *AI Act* Europeu (Arts. 13 e 14)

O direito à explicação, fundamental para mitigar os riscos da opacidade algorítmica, encontra suas raízes mais diretas nas legislações de proteção de dados, mas sua evolução recente demonstra uma expansão para marcos regulatórios específicos de inteligência artificial. Essa trajetória revela um amadurecimento do arcabouço jurídico, que se move de um direito reativo do titular de dados, para uma obrigação proativa de transparência imposta aos desenvolvedores e implementadores de sistemas de IA, especialmente na área da saúde.

A base fundamental desse direito está consolidada no Regulamento Geral sobre a Proteção de Dados (*GDPR*) da União Europeia (UNIÃO EUROPEIA, 2016) e na Lei Geral de Proteção de Dados (LGPD) do Brasil (BRASIL, 2018). O Artigo 22 do *GDPR* estabelece uma regra geral restritiva: o titular dos dados tem o direito de não ser sujeito a uma decisão baseada *exclusivamente* no tratamento automatizado, que produza efeitos jurídicos ou que o afete significativamente de modo similar. Embora existam exceções – como a necessidade para a execução de um contrato ou o consentimento explícito –, quando aplicáveis, o Regulamento exige a implementação de salvaguardas que incluam, no mínimo, o direito de obter intervenção humana, de manifestar o seu ponto de vista e de contestar a decisão (Art. 22, 3). Implicitamente, para que se possa contestar uma decisão de forma eficaz, é necessária uma explicação sobre como ela foi alcançada. As diretrizes do Comitê Europeu para a Proteção de Dados (*EDPB*) têm reforçado que a explicação fornecida deve ser “significativa”, indo além de uma mera descrição superficial do sistema, enfatizando que a informação deve ser “suficientemente abrangente para que o titular dos dados compreenda as razões da decisão” (UNIÃO EUROPEIA, 2018b).

De forma análoga, a LGPD, em seu Artigo 20, concede ao titular de dados o direito de solicitar a revisão de decisões tomadas *unicamente* com base em tratamento de dados pessoais automatizado que afetem seus interesses. O § 1º deste artigo é ainda mais explícito que o seu correspondente europeu, determinando que o controlador deverá fornecer, sempre que solicitadas, “informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial” (BRASIL, 2018). Esta disposição representa a positivação mais direta do direito à explicação no ordenamento jurídico brasileiro, funcionando como uma ferramenta essencial contra a discriminação e a arbitrariedade algorítmica.

Ambos os diplomas legais, contudo, apresentam uma limitação textual que pode se tornar problemática na prática: a aplicação restrita a decisões tomadas “unicamente” ou “exclusivamente” de forma automatizada. Muitos sistemas de IA em saúde são projetados como

ferramentas de auxílio à decisão, mantendo um “humano no ciclo” (*human-in-the-loop*) que formalmente toma a decisão final. Nesses casos, poder-se-ia argumentar que a decisão não é “unicamente” automatizada e, portanto, as garantias dos Artigos 20 da LGPD e 22 do GDPR não se aplicariam (SOLOVE, 2025). Tal interpretação, contudo, seria excessivamente formalista e ignoraria a realidade da interação humano-máquina. A influência de uma recomendação algorítmica, especialmente quando o sistema demonstra alta acurácia, pode ser tão poderosa a ponto de gerar um “viés de automação”, onde o profissional humano tende a aceitar a sugestão da máquina de forma acrítica. Nesses cenários, a decisão é, *efetivamente*, automatizada, e o espírito da lei deve prevalecer, garantindo ao paciente o direito à revisão e à explicação (FACCHINI NETO; BARBOSA, 2023).

Ou seja, tanto o *GDPR* quanto a *LGPD* apresentam uma limitação conceitual no contexto de sistemas de saúde complexos: seu enfoque é primariamente *ex post facto* (reativo) e centrado no titular dos dados. O direito é exercido após a decisão ter sido tomada. Essa abordagem, embora crucial, é insuficiente para lidar com os riscos sistêmicos de tecnologias como as analisadas nesta dissertação. A “dupla caixa-preta” não é um problema que pode ser resolvido apenas com a revisão de casos individuais; ela exige que a transparência e a explicabilidade sejam incorporadas ao sistema desde sua concepção.

É aqui que a Lei de Inteligência Artificial da União Europeia (*AI Act*) representa uma mudança de paradigma. Em vez de focar apenas no direito do indivíduo, o *AI Act* adota uma abordagem baseada em risco e focada no ciclo de vida da tecnologia, tratando a IA como um produto ou serviço a ser regulado *ex ante* (preventivamente). Os sistemas de IA para diagnóstico médico e a maioria das neurotecnologias podem ser interpretados como sendo de “alto risco” (Anexo III), o que aciona um conjunto rigoroso de obrigações de transparência e supervisão humana.

O Artigo 13 do *AI Act*, sobre “Transparência e fornecimento de informações aos *deployers*” (“implementadores” ou “responsáveis pela implantação”, dentre os quais se encontram os profissionais de saúde), determina que os sistemas de IA de alto risco devem ser concebidos e desenvolvidos de modo a serem “suficientemente transparentes para permitir que os *deployers* interpretem os resultados do sistema e os utilizem de forma adequada” (UNIÃO EUROPEIA, 2024). Isso se materializa na exigência de que as instruções de uso forneçam informações concisas, completas e claras sobre: as características, capacidades e limitações de desempenho do sistema (Artigo 13, nº 3, alínea b); sua finalidade pretendida e os tipos de decisões que ele é projetado para tomar (Artigo 13, nº 3, alínea b, “i” e Considerando 72); sua lógica subjacente, incluindo os principais fatores que influenciam a decisão (Artigo 13, nº 3,

alínea b, “iv” e Considerando 72); e os grupos de pessoas para os quais o sistema foi validado e seus níveis de precisão, bem como os riscos de vieses conhecidos (Artigo 13, nº 3, alínea b, “ii” e “v”; artigo 15 e Considerando 67).

Essa obrigação transfere o ônus da opacidade para o desenvolvedor. Um médico que utiliza um sistema de IA para auxiliar no diagnóstico de um câncer, sob o *AI Act*, tem o direito de receber um manual detalhado que explique como o sistema funciona, quais são suas margens de erro e em quais populações de pacientes ele foi testado. Isso eleva o direito à explicação de uma prerrogativa do paciente para um requisito técnico do produto, capacitando o profissional de saúde a exercer a supervisão adequada.

Complementarmente, o Artigo 14 (“Supervisão Humana”) exige que os sistemas de alto risco sejam projetados para poderem ser eficazmente supervisionados por humanos, que devem ser capazes de compreender o funcionamento do sistema para poder decidir “não utilizar o sistema de IA de alto risco, ignorar, anular ou inverter o resultado” (Art. 14, 4).

A coexistência desses regimes cria um arcabouço jurídico complementar e robusto. O *GDPR* e a *LGPD* garantem o direito do paciente a uma explicação individualizada (*ex post*), enquanto o *AI Act* garante que o sistema utilizado pelo médico seja intrinsecamente explicável por *design* (*ex ante*). Um não substitui o outro; eles se reforçam mutuamente. A capacidade de um médico de cumprir o dever de informar seu paciente e de atender a um pedido de explicação (fundado na *LGPD/GDPR*) é imensamente fortalecida pelas obrigações de transparência que o *AI Act* impõe ao fabricante do dispositivo. Essa abordagem integrada, que ecoa no projeto de lei brasileiro sobre IA (PL 2.338/2023) – (BRASIL, 2023), é o caminho necessário para garantir que o direito à explicação seja não apenas uma formalidade legal, mas uma realidade prática na medicina diagnóstica do século XXI.

Finalmente, vale mencionar que o Artigo 86 do *AI Act*, que consagra o direito à explicação, não se restringe a decisões unicamente automatizadas. Ele se aplica a qualquer pessoa afetada por uma decisão tomada pelo responsável pela implantação (o usuário final, como o médico ou hospital), “com base nos resultados de um sistema de IA de risco elevado”, que produza efeitos jurídicos ou afete significativamente a pessoa (UNIÃO EUROPEIA, 2024; CAMÕES, 2025). Isso abrange diretamente as decisões tomadas pelo médico com base na recomendação do sistema de IA.

O projeto de lei brasileiro para IA de Alto Risco (PL 2.338/2023) exige supervisão humana que busque prevenir ou minimizar riscos. A supervisão deve viabilizar que as pessoas responsáveis possam compreender, interpretar, decidir e intervir nos sistemas de IA. Além disso, o PL estabelece o direito à explicação e à revisão humana para decisões que gerem efeitos

jurídicos relevantes ou que impactem significativamente os interesses do afetado (BRASIL, 2023a).

3.3.2 A Fusão do Direito à Explicação com o Consentimento Livre e Esclarecido no Contexto da Saúde

No âmbito da saúde, o direito à explicação previsto na legislação de proteção de dados adquire uma nova e mais profunda dimensão, fundindo-se com a doutrina do Consentimento Livre e Esclarecido e tornando-se uma condição de sua própria validade, como será devidamente aprofundado no Capítulo 5. A doutrina clássica do Consentimento Livre e Esclarecido, pilar da bioética e do direito médico, exige que, para que o consentimento de um paciente a um procedimento seja válido, ele deve ser livre, esclarecido e precedido de informações completas sobre os riscos, benefícios e alternativas disponíveis. O paciente não tem apenas um direito à informação, mas um direito à compreensão.

A introdução de sistemas de IA opacos na medicina diagnóstica introduz um novo tipo de risco que deve ser comunicado ao paciente: o risco epistêmico, ou seja, o risco de que a decisão seja baseada em uma lógica indecifrável, potencialmente falha ou enviesada, que nem mesmo o médico compreende plenamente (AMANN *et al.*, 2020; WÓJCIK, 2022; BLACKMAN; VEERAPEN, 2025). Se o profissional de saúde, devido à natureza de “caixa-preta” da ferramenta, não consegue explicar a “lógica subjacente” que levou a uma recomendação diagnóstica, ele falha em seu dever de informar o paciente sobre um aspecto crucial do processo decisório. Consequentemente, o paciente, privado dessa informação, não pode ponderar adequadamente todos os riscos envolvidos e, portanto, seu consentimento é viciado (BLACKMAN; VEERAPEN, 2025).

À luz desses instrumentos, o direito à explicação previsto no Art. 20 da LGPD deixa de operar apenas como faculdade de contestar decisões automatizadas *a posteriori* e passa a constituir requisito prévio do Consentimento Livre e Esclarecido em saúde. De forma análoga, o §3º do art. 22 do *GDPR* impõe salvaguardas substanciais sempre que tratamento exclusivamente automatizado produza efeitos significativos, exigindo que o paciente receba informações “significativas sobre a lógica envolvida”. A propósito, um dos objetivos da Proposta de Emenda Constitucional 29/2023 é incluir, no rol de direitos fundamentais, a proteção à transparência algorítmica (ENGELMANN, 2023; BRASIL, 2023c). O *AI Act*, como analisado acima, reforça essa obrigação ao classificar sistemas diagnósticos baseados em IA

como “*high-risk*”, condicionando sua comercialização à prestação de explicações claras, documentação técnica acessível e supervisão humana qualificada.

No Brasil, a RDC ANVISA nº 657/2022, de 24 de março de 2022, dispõe sobre a regularização de Software como Dispositivo Médico (*SaMD*) – (ANVISA, 2022a). Embora a resolução não mencione explicitamente “inteligência artificial” (IA), a definição de *SaMD* é abrangente, aplicando-se a *softwares* destinados a uma ou mais indicações médicas que realizam suas finalidades sem fazer parte de *hardware* de dispositivo médico, o que engloba *softwares* médicos baseados em IA.

Para os *SaMDs* das classes de risco I e II, por exemplo, a resolução impõe a apresentação e manutenção de um dossiê técnico pela empresa detentora da notificação, conforme detalhado no Art. 11.

Esse dossiê deve conter: (i) *Verificação e validação*, que pode consistir em validação analítica ou clínica, dependendo da indicação de uso do *SaMD*. A validação clínica ou utilidade clínica mede a capacidade do *SaMD* em produzir uma saída clinicamente significativa, com impacto positivo na saúde individual ou populacional; (ii) *Avaliação clínica e associação clínica válida*, definida como o conjunto de atividades para avaliar a segurança clínica, eficácia e desempenho de um *SaMD*; (iii) *A descrição dos algoritmos e/ou rotinas utilizadas* para gerar o processamento das sugestões clínicas (prevenção, diagnóstico, tratamento, monitorização fisiológica, reabilitação ou anticoncepção) e suas fundamentações clínicas ou científicas. Além disso, as instruções de uso ou o próprio *SaMD* devem apresentar o princípio de funcionamento, incluindo descrições genéricas dos algoritmos, rotinas e fórmulas utilizadas e suas associações clínicas válidas; e (iv) *As evidências de validação* devem ser suficientes para garantir exatidão, confiabilidade e desempenho pretendido consistentes.

Nos Estados Unidos, o guia da FDA sobre transparência em dispositivos médicos habilitados por *machine-learning* define que informações críticas sobre riscos, desempenho e limitações devem ser comunicadas aos usuários antes da utilização clínica (FDA, 2025).

Do ponto de vista ético-acadêmico, autores como Samuel Director (2025) sustentam que a opacidade algorítmica introduz um risco epistêmico que impossibilita avaliação autônoma dos pacientes, enquanto Hernandez-Boussard *et al.* (2025) demonstram que transparência e participação do paciente são condições para confiança clínica. Assim, a doutrina do Consentimento Livre e Esclarecido absorve o direito à explicação: se o profissional de saúde não puder expor, de modo inteligível, a lógica, a acurácia e as limitações do algoritmo – incluindo sua natureza de “caixa-preta” – o consentimento não será verdadeiramente esclarecido (PRICE, 2023; TORGERSEN *et al.*, 2024).

Director (2025), porém, apesar de reconhecer o risco epistêmico da opacidade, argumenta que, quando a explicabilidade é obtida ao custo de menor precisão, isso tende a produzir piores desfechos para os pacientes; daí concluir que, se a *XAI* (explicabilidade) for menos precisa do que a IA intrinsecamente opaca, o paciente teria razão prudencial para preferir modelos opacos, em nome da beneficência. Nas suas palavras: “Para que a IA explicável resolvesse o problema, ela teria que nos dar tanto a autonomia quanto a beneficência, mas parece que a IA explicável meramente apresenta outra forma de preferir a autonomia em detrimento da beneficência”. Como solução, propõe o “consentimento de ordem superior”, pelo qual o paciente recebe informações de segunda ordem – sobre confiabilidade e desempenho geral da IA —, ainda que não compreenda o raciocínio clínico específico (informação de primeira ordem), preservando-se assim a autonomia enquanto se utiliza a tecnologia mais eficaz. Com a devida vênia, não subscrevemos essa conclusão. Ela parece reeditar o paternalismo na medicina – agora sob a roupagem tecnológica da IA.

À luz dos argumentos desenvolvidos nesta dissertação, julgamos mais adequada a conclusão de Eduardo Dantas e Rafaella Nogaroli (2020): quem utiliza – ou é de algum modo impactado por – sistemas de IA não possui apenas um direito à informação. Como sustenta Nelson Rosenvald, não basta ao criador do algoritmo afirmar “*the algorithm did it*”; exige-se, na medida do possível, a oferta de explicação e justificação (DANTAS; NOGAROLI, 2020; NOGAROLI, 2023). Nesse horizonte, ferramentas de “*Explanatory AI*” (analisada a seguir) constituem meios técnicos para mitigar a opacidade explicativa (MESKE *et al.*, 2025). Porém, como se demonstrará no Capítulo 6, o Consentimento Livre e Esclarecido já não se revela suficiente, por si só, como salvaguarda única e individualista da autonomia do paciente.

3.3.3 O Desafio do Consentimento Livre e Esclarecido em Populações Vulneráveis Diante da “Dupla Caixa-Preta”

O desafio de obter um consentimento verdadeiramente esclarecido para o uso de tecnologias opacas se agrava exponencialmente ao lidar com populações vulneráveis, como pacientes com doenças neurodegenerativas que afetam profundamente a capacidade cognitiva, especificamente a Doença de Alzheimer (DA) grave (XU; LIU; SHANG *et al.*, 2024; XU; LIU; SHANG *et al.*, 2025). O caso da aplicação de Estimulação Cerebral Profunda (*DBS*) em pacientes com DA grave ilustra dramaticamente esse dilema. Os pesquisadores reconhecem as “complexidades e considerações éticas envolvidas na condução de um estudo de *DBS* em

pacientes com DA, com deficiências cognitivas profundas”, priorizando o processo de consentimento com “máximo cuidado e sensibilidade” (XU; LIU; SHANG *et al.*, 2025).

Pacientes com DA grave já enfrentam severas limitações para compreender, reter e ponderar informações sobre procedimentos neurocirúrgicos experimentais, como a DBS direcionada ao fórnix ou ao núcleo basal de Meynert (NBM) - (XU; LIU; SHANG *et al.*, 2025). Essa limitação cognitiva, que se manifesta como grave perda de memória e deficiências intelectuais, foi a base para que os estudos adotassem um processo de consentimento cuidadoso, com a participação de representantes legais, o que implica déficits significativos na capacidade decisória para avaliar riscos e benefícios por parte dos pacientes (XU; LIU; SHANG *et al.*, 2024).

Quando o dispositivo implantado incorpora sistemas de aprendizado de máquina cujos parâmetros permanecem inacessíveis até mesmo aos engenheiros – a chamada “dupla caixa-preta” (cérebro doente + algoritmo opaco) – o médico se vê incapaz de explicar com precisão a lógica subjacente à recomendação terapêutica. Tal cenário coloca em xeque a própria noção de esclarecimento que fundamenta o Consentimento Livre e Esclarecido (DIRECTOR, 2025). O autor ainda afirma que, “sem a possibilidade de o paciente receber uma explicação adequada sobre como a decisão de tratamento foi justificada, seu consentimento não é considerado suficientemente informado e, portanto, não é válido de acordo com os padrões atuais”.

O ordenamento jurídico brasileiro, por meio do Estatuto da Pessoa com Deficiência (Lei nº 13.146/2015), estabelece salvaguardas importantes. O Artigo 12 é taxativo ao afirmar que “o consentimento prévio, livre e esclarecido da pessoa com deficiência é indispensável para a realização de tratamento, procedimento, hospitalização e pesquisa científica” (BRASIL, 2015). Crucialmente, o § 1º do mesmo artigo determina que, mesmo em situação de curatela, “deve ser assegurada sua participação, no maior grau possível, para a obtenção de consentimento” (BRASIL, 2015).

Em um estudo empírico pan-europeu, que investigou as necessidades de conhecimento e treinamento percebidas pelos membros de Comitês de Ética em Pesquisa (CEPs) e especialistas em ética para a integração da ética no desenvolvimento de pesquisas e tecnologias, Bernd Carsten Stahl *et al.* (2025) constataram a necessidade do fortalecimento dos CEPs com treinamento específico sobre IA e neurodispositivos, a fim de melhorar sua capacidade de avaliar pesquisas impulsionadas pela tecnologia. A IA e as neurotecnologias foram identificadas como uma tecnologia em rápido desenvolvimento, muitas das quais consideradas de alto risco e que apresentam várias ameaças, inclusive ao Consentimento Livre e Esclarecido.

A incompreensibilidade tecnológica ameaça transformar o Consentimento em ficção jurídica: se nem médico nem paciente compreendem o mecanismo causal da intervenção, falar em “esclarecimento” torna-se improdutivo. Frente a esse impasse, a literatura neuroética recomenda: (i) *Avaliação de capacidade decisional em múltiplas etapas* (SOARES, 2025); (ii) *Fortalecimento dos comitês de ética em pesquisa* com treinamento específico sobre IA e neurodispositivos (MOLNÁR-GÁBOR; IENCA *et al.*, 2021); e (iii) *Incentivo à elaboração de diretivas antecipadas de vontade* enquanto a capacidade cognitiva ainda permite decisões autônomas (DADALTO, 2025).

Portanto, sem tais salvaguardas, o Consentimento Livre e Esclarecido (que será aprofundado no Capítulo 5) perde sua função protetiva, reduzindo-se a um ritual documental que não protege a vulnerabilidade acrescida pela “caixa-preta” da IA e da Neurotecnologia.

4 VIESES, DISCRIMINAÇÃO E MANIPULAÇÃO NA SAÚDE DIGITAL E NEUROTECNOLOGICA

No Capítulo anterior, a análise se debruçou sobre o desafio procedimental da opacidade, a “caixa-preta” que impede a compreensão de *como* um sistema de saúde inteligente chega a uma conclusão diagnóstica ou terapêutica (PASQUALE, 2015; PRICE II, 2017). A incapacidade de analisar a lógica interna de um algoritmo ou de uma interface cérebro-computador foi apresentada não como uma mera limitação técnica, mas como um obstáculo jurídico fundamental que tensiona o direito à explicação e corrói as bases do Consentimento Livre e Esclarecido. Contudo, a opacidade, por si só, é apenas o véu que encobre um problema ainda mais profundo e substantivo: o viés. Se a opacidade questiona o processo, o viés corrompe o resultado.

Este Capítulo, portanto, atravessa o espelho da “caixa-preta” para confrontar a imagem que ela reflete: a de uma sociedade marcada por desigualdades históricas e estruturais que são silenciosamente absorvidas, replicadas e, não raro, amplificadas por algoritmos revestidos por uma aura de objetividade científica (O'NEIL, 2016; CRAWFORD, 2021). Os vieses não são meras falhas técnicas ou “bugs” a serem corrigidos; são o produto inevitável de sistemas treinados com dados que carregam em si os recortes sociais, culturais e econômicos do mundo que os gerou (AQUINO, 2023). A promessa de que a inteligência artificial (IA) e as neurotecnologias irão revolucionar a medicina, tornando-a mais precisa e personalizada, só poderá ser cumprida se essa fronteira tecnológica for capaz de confrontar, criticamente, as falhas da sociedade que a criou.

Além de mapear as fontes de viés nos dados e modelos, este Capítulo avançará para uma dimensão mais sutil, mas igualmente crítica: o viés inerente à própria cognição humana e como ele pode ser explorado pela arquitetura das tecnologias. A partir da análise de conceitos como heurísticas, *nudges* e *dark patterns*, demonstraremos que a tomada de decisão, tanto do médico quanto do paciente, não é um ato de pura racionalidade, mas um processo sujeito a influências e manipulações que podem comprometer a autonomia e a equidade (NEGRI *et al.*, 2025). Argumenta-se, portanto, que a crise da autodeterminação informativa desafia a suficiência do consentimento como único mecanismo de proteção, exigindo uma busca por equidade que considere não apenas a imparcialidade do algoritmo, mas também a integridade do ambiente decisório em que ele opera.

4.1 FONTES MÚLTIPLAS DE VIÉS NA FRONTEIRA TECNOLÓGICA DA SAÚDE

A ideia de que ferramentas baseadas em IA são inerentemente neutras e objetivas é um mito perigoso. Na prática, sistemas de inteligência artificial e neurotecnologias refletem os preconceitos existentes nos dados e no próprio processo de desenvolvimento. Como observa Ted A. James (2024), algoritmos de saúde aprendem a partir de *datasets* históricos que carregam as escolhas humanas prévias e as desigualdades sociais nelas contidas. Desse modo, erros e limitações presentes nos registros do passado – ou nos próprios profissionais que rotulam os dados – podem ser “espelhados” pela IA, reproduzindo discriminações. Por exemplo, já foi demonstrado que um algoritmo americano priorizou pacientes brancos saudáveis em detrimento de negros com maior necessidade de cuidados, simplesmente porque foi treinado com dados de “custos do tratamento” e não de gravidade clínica (OBERMEYER *et al.*, 2019). Em suma, os vieses podem infiltrar-se em cada etapa do ciclo de vida tecnológico, criando uma verdadeira “arquitetura do preconceito” incorporada às soluções digitais. Na saúde, onde decisões de diagnóstico e tratamento têm impacto crítico, é imperativo identificar e corrigir tais vieses (JAMES, 2024; HOLDSWORTH, 2023).

4.1.1 Viés de Representação e Amostragem de Dados

A fonte mais comum e perniciosa de viés reside nos dados utilizados para treinar os modelos. Se os dados são o combustível da IA, dados enviesados inevitavelmente produzem resultados enviesados. Esse problema se manifesta de diversas formas.

O “viés de representação” ocorre quando os dados usados para treinar modelos de IA não refletem a diversidade demográfica e clínica da população-alvo. Alguns autores (POPEJOY; FULLERTON, 2016) denominam como “viés persistente”. Na genômica, são denominados de “viés genômico” (segundo ALSAEDI; OGASAWARA *et al.*, 2025), como visto no item 2.2.1.4. Modelos treinados em populações homogêneas tendem a falhar para grupos sub-representados (NAZER *et al.*, 2023). Ou seja, o viés de representação surge da forma como as amostras são coletadas de uma população durante o processo de coleta de dados. Amostras não representativas carecem da diversidade da população, apresentando subgrupos ausentes e outras anomalias (MEHRABI *et al.*, 2022). Algoritmos de IA aprendem com dados, e se esses dados de treinamento refletirem vieses (como subdiagnóstico histórico de certos grupos) ou não forem representativos da população, os resultados da IA podem ser distorcidos (AITH, 2024).

Por exemplo, Luana Cazani *et al.* (2025) destacam que “a presença limitada de imagens de pessoas de pele escura em conjuntos de dados dificulta a precisão diagnóstica” de algoritmos

dermatológicos. Na prática, sistemas de detecção de melanoma (como o NEVO), treinados sobretudo em pele clara, podem diagnosticar mal pacientes de pele escura, acarretando atrasos diagnósticos e agravamento dos desfechos clínicos, além de violarem princípios de equidade (NAZER *et al.*, 2023; HASANZADEH, 2025). De modo semelhante, ferramentas de genética (como o *AlphaMissense*, da *Google DeepMind*), calibradas em bases de dados majoritariamente europeias (NGUYEN; LEE *et al.*, 2025), podem errar ao classificar variantes em populações africanas, asiáticas ou latino-americanas, o que poderia levar a decisões terapêuticas equivocadas em oncologia.

No campo da endocrinologia também se observa a presença desse viés. Hispânicos, por exemplo, tendem a ter taxas mais altas de diabetes e complicações relacionadas ao diabetes em comparação com brancos (NAZER *et al.*, 2023). Os autores também demonstram que o mesmo fato clínico pode levar ao “viés de agregação”, se, por exemplo, o uso de algoritmos, baseados em IA para diagnosticar e monitorar o diabetes em populações hispânicas, não forem sensíveis ao fato de que existem diferentes grupos étnicos hispânicos (por exemplo, mexicanos, porto-riquenhos).

Vale citar, por fim, o caso *PRAIM* como exemplo de “viés de representação”. O modelo de IA utilizado no estudo *PRAIM*, que demonstrou notável sucesso na detecção de câncer de mama, foi treinado e validado em uma população alemã (EISEMANN; BUNK *et al.*, 2025). Sua aplicação direta no Brasil, um país de profunda miscigenação, sem uma rigorosa validação local, poderia resultar em menor acurácia para mulheres negras, indígenas ou asiáticas, cujas características “mamográficas” (aquelas presentes no estudo) podem apresentar variações sutis não capturadas pelo modelo original. Isso não seria apenas uma falha técnica, mas a criação de uma grave iniquidade diagnóstica (AITH, 2024).

4.1.2 Viés Histórico e de Medição (*Proxy*)

Além da representação demográfica, o “viés histórico e de medição” é igualmente problemático. Os modelos de IA na saúde treinam-se em dados históricos que carregam práticas e desigualdades passadas (HOLDSWORTH, 2023). Um exemplo emblemático é o uso de *proxies* inadequados: Ziad Obermeyer *et al.* (2019) demonstraram um algoritmo de risco populacional que usava gastos em saúde como *proxy* de gravidade. Como historicamente se gasta menos no tratamento de pacientes negros, o algoritmo passou a subestimar as necessidades desses pacientes, marcando-os como “menos doentes” (quando, de fato, estavam mais enfermos). O estudo descobriu que os pacientes negros tinham 26,3% mais doenças

crônicas do que os pacientes brancos com o mesmo nível de pontuação de risco (4,8 contra 3,8 condições). Em outras palavras, “o viés surge porque o algoritmo prevê custos de saúde em vez de doença, mas o acesso desigual à atenção médica leva a menos gastos com pacientes negros, gerando grandes vieses raciais” (HASANZADEH, 2025). Esse padrão é preocupantemente geral: qualquer indicador de saúde que reflita práticas discriminatórias (como número de consultas, histórico de exames, taxas de readmissão) pode induzir viés semelhante.

Aplicando ao contexto de sepse, analisado no Capítulo 2, algoritmos de previsão (como o *Sepsis ImmunoScore*) que usam dados de prontuário eletrônico podem, por analogia, inferir menor risco em pacientes historicamente desassistidos. Se grupos vulneráveis (negros, pobres, periféricos) receberam menos cuidados intensivos, o modelo aprende um *proxy* enviesado e subestima sua gravidade real – aumentando o risco de omitir intervenções críticas nessas populações (FLEUREN; KLAUSCH *et al.*, 2020).

Em ferramentas diagnósticas para condições que envolvem avaliação da dor, Harini Suresh e John V. Guttag (2021) destacaram que existem disparidades de gênero e raciais. Isso implica que a forma como a dor é percebida, reportada e, conseqüentemente, diagnosticada pode ser diferente para mulheres (em comparação com homens) e pessoas de diferentes raças, levando a um viés nos dados de diagnóstico. No contexto de diagnósticos médicos, afirmam os autores, o viés de medição ocorre quando a condição “diagnosticado com a condição X” é usada como um indicador (*proxy*) para “tem a condição X”, mas a precisão desse diagnóstico varia significativamente entre diferentes grupos de pacientes.

4.1.3 Viés de Anotação e Rotulagem

O “viés de anotação e rotulagem” introduz a subjetividade humana diretamente no cerne do processo de aprendizado de máquina. Em modelos supervisionados, as *etiquetas* de dados – fruto de decisões clínicas humanas – podem incorporar vieses dos próprios especialistas. Na perspectiva de Daniel Kahneman (2012), vieses são erros sistemáticos que se repetem de forma previsível em circunstâncias particulares. Os julgamentos de especialistas são suscetíveis à “ilusão de validade”, na qual se superestima a precisão de suas previsões. Como exemplo, Kahneman cita que clínicos, “absolutamente certos” de um diagnóstico *ante mortem*, erraram em 40% das vezes.

Portanto, rótulos subjetivos ou inconsistentes (por diferenças de formação, experiências ou preconceitos) levam o modelo a aprender premissas enviesadas como se fossem “verdades”. Por exemplo, dados capturados em registros eletrônicos de saúde (*EHRs*) refletem informações

registradas por avaliação de pessoas, o que implica o risco de que o modelo de IA replique erros cognitivos humanos. Quando incorporados ao treinamento, esses vieses subjetivos tornam-se “objetivos” para a IA (NOGAROLI, 2023; SMITH *et al.*, 2024).

4.1.4 Viés Cognitivo e de Automação na Prática Clínica

Mesmo algoritmos bem treinados podem provocar vícios na tomada de decisão humana. Mesmo que os sistemas de IA apresentem bom desempenho isoladamente, eles podem causar consequências prejudiciais devido a fenômenos como o “viés de automação ou de confirmação”, especialmente quando operam em sistemas sociotécnicos complexos com tomadores de decisão humanos (SURESH; GUTTAG, 2021). Profissionais de saúde podem confiar excessivamente nas orientações de um sistema de IA, deixando de buscar evidências confirmatórias.

Farzad Hasanzadeh (2025) relata um estudo do viés de automação em mamografia, no qual os autores descobriram que a precisão dos relatórios dos radiologistas diminuiu com a exposição a sugestões incorretas do *BI-RADS (Breast Imaging Reporting and Data System)* da IA. A precisão caiu de 79,7% para 19,8% para radiologistas inexperientes, de 81,3% para 24,8% para radiologistas moderadamente experientes e de 82,3% para 45,5% para radiologistas muito experientes quando a IA sugeriu uma categoria *BI-RADS* incorreta. Isso demonstrou a suscetibilidade pronunciada ao viés de automação, particularmente para médicos menos experientes.

O fenômeno conhecido como “viés de automação ou de confirmação” ocorre, portanto, quando profissionais tratam a saída de um sistema como infalível, aceitando recomendações sem questionar (o oposto, excesso de confiança). Emely Rosbach *et al.* (2024) quantificaram esse efeito: em experimento com patologistas experientes, cerca de 7% das avaliações corretas foram alteradas para incorretas após recomendação errônea da IA – isto é, uma taxa de viés de automação de aproximadamente 7%, onde 38 de 560 avaliações inicialmente corretas foram revertidas para incorretas por uma orientação incorreta da IA. Ou seja, o único fato de um sistema sugerir um resultado levou médicos treinados a reavaliarem equivocadamente casos que haviam respondido corretamente por conta própria.

Esse viés de automação combina-se a outros “vieses cognitivos” que afetam o clínico sob pressão de tempo ou volume. Fatores ambientais, como a pressão do tempo, podem sobrecarregar os recursos cognitivos e levar ao uso de heurísticas ou à dependência excessiva da automação (ROSBACH *et al.*, 2024).

Os “vieses cognitivos” são definidos como padrões sistemáticos de desvio da racionalidade no julgamento, e a heurística intuitiva está ligada a esses erros de raciocínio. Segundo André Mendonça Machado (2018), eles são erros de raciocínio causados por estratégias mentais de simplificação, geradas no esforço de processamento de informações. Sua ligação ao fenômeno da heurística intuitiva se dá por se tratar de atalhos mentais que simplificam procedimentos complexos que frequentemente acarretam erros de análise ou de raciocínio (NEGRI *et al.*, 2025).

Por exemplo, ao receber uma sugestão automatizada, o médico pode ignorar sinais clínicos *ad hoc* ou confiar cegamente no “diagnóstico” computacional, levando a erros por comissão (seguindo recomendações incorretas) ou por omissão (não percebendo problemas que o sistema não sinaliza) – (ROSBACH *et al.*, 2024).

4.1.5 Viés em Neurotecnologias e Interfaces Cérebro-Computador

As neurotecnologias, por sua vez, introduzem fontes de viés ainda mais complexas e intrínsecas à própria tecnologia. O viés aqui não é apenas um reflexo passivo das desigualdades sociais; ele pode ser ativamente *criado* pela própria arquitetura tecnológica, em um fenômeno conhecido como “viés de covariáveis não controladas” (LA FISCA *et al.*, 2024).

Em sistemas de Interface Cérebro-Computador (*BCI*), por exemplo, pode ocorrer o viés de covariáveis não controladas, onde o algoritmo aprende a classificar os sinais neurais com base em propriedades espúrias do estímulo (como a complexidade visual de uma imagem mostrada ao usuário) em vez de aprender a partir da intenção real do usuário. O algoritmo encontra um “atalho” nos dados (LA FISCA *et al.*, 2024).

Como explicam Luca La Fisca *et al.* (2024), o classificador é esperado para aprender padrões espaço-temporais específicos das categorias, e não padrões relacionados a propriedades não controladas do estímulo. No entanto, se as covariáveis não forem modeladas ou controladas, o algoritmo de *BCI* pode explorar qualquer viés nessas variáveis nos dados. Isso leva a uma maior probabilidade de classificação errada, pois o algoritmo baseia sua decisão em valores de covariáveis em vez do efeito categórico real que deveria estar medindo. Exemplos dessas propriedades espúrias (covariáveis) incluem variáveis psicolinguísticas (como número de fonemas, familiaridade e idade de aquisição de um item) e propriedades da imagem (como contraste, compacidade e homogeneidade). A complexidade visual de uma imagem mostrada ao usuário seria uma “propriedade da imagem”. Em vez de aprender os processos neurais que deveriam ser medidos (a “intenção real do usuário” ou o que difere entre as categorias de

estímulo), a classificação pode se concentrar nessas propriedades não controladas. Outro exemplo: se em um experimento onde o usuário deve pensar em objetos “vivos” *versus* “não vivos”, as imagens de objetos vivos forem consistentemente mais complexas visualmente do que as de objetos não vivos, o algoritmo pode começar a classificar os sinais cerebrais com base na “complexidade visual” em vez da “categoria vivo/não vivo” (LA FISCA *et al.*, 2024).

Consequência do “atalho”: O sistema passa a basear sua decisão em valores de covariáveis em vez do efeito categórico em si, o que aumenta a probabilidade de erros de classificação. Ou seja, o modelo não mede o fenômeno que deveria medir (a intenção do usuário ou a categoria neural), mas sim uma característica incidental do estímulo que coincidiu com a categoria.

Utilizando o exemplo apresentado por Sergio Negri *et al.* (2025), embora se refiram a atalhos cognitivos humanos, essa lógica se estende aos sistemas de IA: na psicologia cognitiva, heurísticas são descritas como “atalhos” mentais ou “regras do polegar” que simplificam julgamentos difíceis. O Sistema 1 (pensamento rápido) – *expressão cunhada por Kahneman (2012)* –, frequentemente “responde a perguntas mais fáceis do que essa que foi feita”, o que é um tipo de atalho. O princípio *WYSIATI (What You See Is All There Is* – “o que você vê é tudo que há”) do Sistema 1 implica que ele foca na evidência existente e ignora a ausente, construindo a melhor história possível a partir de informações limitadas e suprimindo a dúvida. Isso pode levar a que o sistema se baseie em um “atalho” nos dados. No contexto de *BCIs*, a necessidade de separar efeitos de covariáveis do efeito categórico sugere que, sem essa separação, o sistema poderia aprender com propriedades espúrias (“atalhos”) em vez da intenção real do usuário.

Outro fenômeno notável é o “*viés de feedback*”: Barbero e Grosse-Wentrup (2010) mostraram que interfaces cérebro-computador (*BCI*) podem enviesar o aprendizado do usuário dependendo de como exibem o *feedback* de desempenho. O *feedback* é reconhecido como fundamental no aprendizado de como operar uma *BCI*. O estudo investigou a influência de um *feedback* que troca a acurácia do *feedback* por motivação, por exemplo, enviesando artificialmente a crença dos sujeitos sobre seu sucesso no processo de aquisição de habilidade. *Feedback* falso-positivo (o sistema diz “acerto” indevidamente) acelera o treino de novatos, mas prejudica especialistas, enquanto *feedback* falso-negativo atrasa a habilidade dos iniciantes. Em outras palavras, a engenharia de como a *BCI* informa o usuário já gera um viés na amostragem dos sinais neurais, distinto de qualquer viés social ou de dados. Essa constatação tem consequências profundas para a governança, pois indica que estratégias de mitigação

focadas apenas em “limpar os dados” são insuficientes; é preciso auditar a própria arquitetura da interação.

Além disso, outros vieses surgem da própria base científica: a neurociência experimental tradicional é baseada em amostras restritas (estudantes de países desenvolvidos), o que significa que dispositivos neurais podem acabar calibrados para “um tipo de cérebro” específico.

Conforme Shayna La Scala *et al.* (2023), mulheres não-caucasianas, que são ainda mais marginalizadas pela interação entre gênero e raça, estão particularmente ausentes nas pesquisas em neurociência. Uma revisão sistemática de literatura, por exemplo, identificou sub-representação de mulheres e de minorias raciais/étnicas em estudos de *fMRI* (ressonância magnética funcional) voltados a doenças cardiovasculares, apresentando taxas elevadas de doenças cardiovasculares em comparação com seus pares brancos. Essa lacuna é particularmente grave porque certos grupos minoritários, como pessoas negras, exibem maior carga de doença cardiovascular em comparação com seus pares brancos. De modo semelhante, no campo da saúde mental, meninas latinas, também sub-representadas na pesquisa neurocientífica, apresentam níveis superiores de ansiedade não tratada quando comparadas às suas colegas negras e brancas. A exclusão sistemática de minorias racializadas na produção de conhecimento neurocientífico prejudica diretamente essas comunidades e favorece a formulação de estratégias preventivas e interventivas enviesadas.

Muitos dispositivos eletrofisiológicos não foram projetados para lidar com a variabilidade fenotípica, contribuindo para a exclusão de dados de pessoas com tons de pele mais escuros e tipos de cabelo específicos. Isso pode levar a interpretações imprecisas de dados de ressonância magnética (LA SCALA *et al.*, 2023).

Isso, por óbvio, cria riscos de eficácia reduzida ou de segurança para populações não estudadas. A falta de representatividade nos dados de treinamento cria riscos de perpetuar disparidades no atendimento e tratamento de doenças, exacerbando as desigualdades em saúde. No Brasil, onde mais da metade da população se identifica como negra ou parda, tais vieses na IA e Neurotecnologias médicas são especialmente preocupantes para a equidade em saúde (AITH, 2024).

Por fim, após a profunda análise que este tópico realizou, uma certeza fica evidente: a crise do Consentimento Livre e Esclarecido tem se intensificado. Como será desenvolvido no Capítulo 5, no âmbito da IA e das neurotecnologias, o CLE somente se concretiza quando o paciente recebe e compreende uma informação, materialmente adequada, e pode deliberar sem induções derivadas do desenho de interfaces ou do ambiente de cuidado. Na ausência dessas condições, a autonomia é comprometida e o consentimento tende a reduzir-se a um rito formal,

que, na prática, legitima assimetrias tecnológicas (PASQUALE, 2015; CRAWFORD, 2021; PRICE II, 2019).

4.2 ARQUITETURA DA MANIPULAÇÃO: *NUDGES* E *DARK PATTERNS*

A autonomia do paciente pode ser comprometida não apenas pela presença de informações enviesadas, mas também pela forma como tais informações e opções de decisão são apresentadas. O *design* da interface de um *software* de saúde pode explorar vieses cognitivos humanos – como ancoragem, efeito enquadramento e aversão à perda – para induzir escolhas específicas, valendo-se de estratégias conhecidas como *nudges* e *dark patterns* (MATHUR *et al.*, 2019).

O *Nudge*, ou “empurrãozinho”, é definido como qualquer aspecto da arquitetura de escolha que altera o comportamento das pessoas de maneira previsível, sem proibir opções ou alterar incentivos econômicos (THALER; SUNSTEIN, 2008). Ele guia o indivíduo para um resultado que se alinha com seus melhores interesses ou com o bem-estar social (ZUBOFF, 2021). No contexto clínico, *nudge* é qualquer ajuste sutil na arquitetura da decisão – de interfaces, fluxos, *defaults*, alertas ou *feedbacks* – que altera previsivelmente o comportamento de profissionais ou pacientes, sem eliminar opções ou impor penalidades explícitas. O *nudge* atua explorando vieses cognitivos, influenciando a decisão clínica de modo sutil.

Um passo além do *nudge* na escala da manipulação são os *deceptive patterns*, mais conhecidos como *Dark Patterns*. Estes não são meros empurrões, mas “truques usados em sites e aplicativos que levam você a fazer coisas que não pretendia” (BRIGNULL *et al.*, 2023). São estratégias de *design* deliberadamente enganosas. No contexto da saúde digital, os riscos são imensos. Diferem de *nudges* por adicionarem opacidade/assimetria informacional, fricções injustificadas ou ocultação de alternativas.

Alguns exemplos deixarão esses conceitos mais claros. Em Interfaces Cérebro-Computador, o *feedback* enviesado (como falsos positivos ou negativos) atua como um mecanismo que modula o aprendizado e o desempenho do usuário, sendo que sujeitos capazes são impedidos por informações imprecisas, enquanto usuários iniciantes ou com baixo desempenho podem se beneficiar de uma crença incorreta sobre seu controle, aumentando a motivação (BARBERO; GROSSE-WENTRUP, 2010). Esse mecanismo de reforço opera como um “empurrão” (*nudge*). Contudo, a implicação é que, se o sistema prioriza um padrão de *feedback* que induz melhoras aparentes sem garantir a validade científica do sinal neural, isso pode se configurar como um padrão enganoso (*dark pattern*) de desempenho, deliberadamente

concebido para mascarar as limitações do algoritmo classificador e explorar a vulnerabilidade do usuário.

Em *BCIs*, as propriedades não controladas dos estímulos visuais (covariáveis, como a saliência da imagem) podem introduzir um viés na classificação, fazendo com que o algoritmo aprenda atalhos espúrios (focando nos atributos do estímulo, e não nos padrões neurais de intenção). O *design* do experimento, ao usar *feedback* como reforço, funciona como um “empurrão” (*nudge*) atencional, moldando o treinamento do usuário (LA FISCA *et al.*, 2022). Contudo, a implicação crítica é que quando a interface falha em balancear as condições experimentais e oferece uma realimentação convincente e distorcida (*e.g.*, falsos positivos/negativos que aumentam artificialmente a percepção de controle), ela cria uma arquitetura decisória que opera com assimetria e opacidade informacional (BARBERO; GROSSE-WENTRUP, 2010; MITTELSTADT, 2021). Isso se torna um padrão enganoso (*dark pattern*) de desempenho (RAJPURA *et al.*, 2023), destinado a explorar a vulnerabilidade do usuário ao mascarar as verdadeiras limitações ou imprecisões do algoritmo classificador.

Em tecnologias de Estimulação Cerebral Profunda (*DBS*) e neuromodulação, a escolha dos parâmetros de estimulação (incluindo contato, amplitude e frequência) tem a capacidade de modular o humor, a impulsividade e o comportamento; existem casos documentados de hipomania e alterações de personalidade associados a configurações específicas de estimulação (COUTINHO *et al.*, 2024). Implicação: quando as interfaces de programação do *DBS* (ou de neuroestimulação) favorecem valores predefinidos (*presets*) ou padrões de estimulação pouco transparentes, a facilidade e o viés do caminho mais fácil (*default effect*) podem atuar como um “empurrão” (*nudge*) técnico sobre a decisão clínica do médico. Se o *design* da interface, intencionalmente ou não, dificulta os ajustes finos necessários para a personalização do tratamento, ou oculta os *trade-offs* clínicos (como o risco de efeitos adversos emocionais em troca de ganhos motores), o desenho se aproxima de um padrão enganoso (*dark pattern*), explorando a limitação cognitiva humana e a opacidade algorítmica em um contexto de alto risco.

A opacidade explicativa – consistente no risco de o profissional de saúde não informar que utilizou a tecnologia de IA para apoiar uma decisão clínica e/ou não explicar o funcionamento e o papel da tecnologia utilizada de forma compreensível ao paciente –, se utilizada de maneira intencional, para manipular a aceitação do paciente, pode ser considerada um *dark pattern* (NOGAROLI, 2023).

Por fim, vale citar o risco de Ataques de Estímulos Enganosos (*Misleading Stimuli Attacks*), que representa uma grave vulnerabilidade de segurança e privacidade em Interfaces

Cérebro-Computador, particularmente aquelas baseadas em Eletroencefalografia (FARAHANY, 2023). Esses ataques utilizam estímulos subliminares para inferir dados pessoais altamente sensíveis a partir das respostas cerebrais detectadas, uma operação que ocorre sem o conhecimento ou consentimento explícito do indivíduo (MOLNÁR-GÁBOR; IENCA *et al.*, 2021; FARAHANY, 2023). Embora esse ataque seja tecnicamente classificado como “*brain-hacking malicioso*” (IENCA; ADORNO, 2017), ele se alinha diretamente com a definição de um padrão enganoso (*dark pattern*), porque: (i) *Utiliza técnicas para ocultar a coleta e a intenção de extrair dados*, aproveitando a capacidade da neurotecnologia de acessar processos cerebrais abaixo do limiar da percepção humana (ANPD, 2025); e (ii) *Permite a inferência de informações sensíveis* (como endereços, detalhes bancários ou faces) por meio de respostas cerebrais inconscientes, driblando o controle consciente do usuário (FARAHANY, 2023; ANDORNO; MUHL, 2023).

As análises acima demonstram que a IA em saúde e as neurotecnologias constituem um terreno particularmente propício à proliferação de *nudges* e padrões enganosos (*dark patterns*). No Capítulo 5, examinaremos de que modo essa verdadeira “arquitetura da manipulação” aprofunda a crise do Consentimento Livre e Esclarecido e revela sua insuficiência como mecanismo exclusivo de tutela do paciente e impõe uma agenda de equidade que contemple não apenas a imparcialidade do algoritmo, mas também a integridade do ecossistema decisório em que ele é concebido, implementado e utilizado.

4.3 RISCOS DE DISCRIMINAÇÃO: DAS INIQUIDADES DIAGNÓSTICAS À “NEURODISCRIMINAÇÃO”

A análise técnica das fontes de viés, detalhada nas seções anteriores, estabelece o fundamento para a compreensão de suas consequências jurídicas e sociais. Um viés algorítmico não é um problema abstrato confinado ao código; ele se materializa no mundo real como discriminação, gerando danos tangíveis e violando direitos fundamentais. Esta seção fará a transição do “como” o viés surge, para o “porquê” ele importa para o Direito, argumentando que a discriminação algorítmica representa uma afronta direta ao princípio da igualdade e inaugura novas e preocupantes formas de exclusão.

4.3.1 A Discriminação Algorítmica como Violação do Princípio Constitucional da Isonomia

A pedra angular do ordenamento jurídico brasileiro, a Constituição Federal de 1988 (BRASIL, 1988), estabelece em seu Artigo 5º, *caput*, o princípio da isonomia, garantindo que “todos são iguais perante a lei, sem distinção de qualquer natureza”. A discriminação algorítmica – *definida como o resultado discriminatório gerado por uma decisão automatizada que, baseada em informações e dados estatísticos, pode impactar diretamente o acesso de um indivíduo a bens, serviços e ao mercado de trabalho, afetando direitos como autonomia, igualdade e personalidade* – ao resultar, portanto, em tratamento desigual para indivíduos com base em sua pertença a um determinado grupo, mesmo que de forma não intencional, constitui uma violação direta desse preceito fundamental (SILVA, 2022).

Ademais, a Constituição Federal de 1988 (BRASIL, 1988), em seu Artigo 3º, inciso IV, estabelece como objetivo fundamental promover o bem de todos, “sem preconceitos de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação”, e o Artigo 5º, inciso XLI, determina que “a lei punirá qualquer discriminação atentatória dos direitos e liberdades fundamentais”. Portanto, a discriminação algorítmica, ao gerar tratamento desigual, pode constituir uma violação direta desses preceitos fundamentais.

É crucial distinguir entre a discriminação direta e a indireta (SILVA, 2022). A *discriminação direta* ocorreria se um algoritmo utilizasse explicitamente um atributo protegido, como raça ou gênero, em seu processo decisório (MEHRABI *et al.*, 2022). Embora mais fácil de identificar, essa é a forma menos comum de discriminação em sistemas modernos. A forma mais prevalente e insidiosa é a *discriminação indireta*, que acontece quando um critério, aparentemente neutro, como o custo histórico de tratamento, o código postal ou mesmo padrões texturais em uma imagem, tem um impacto desproporcionalmente adverso sobre um grupo protegido (MEHRABI *et al.*, 2022). O algoritmo não “vê” a raça, mas aprende a usar *proxies* que estão fortemente correlacionados a ela, produzindo um resultado discriminatório sob um véu de neutralidade.

A Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709/2018 (BRASIL, 2018), surge como o principal instrumento infraconstitucional para combater essa violação (AITH, 2024). O Artigo 6º, inciso IX, da LGPD²⁴ eleva a não discriminação a princípio basilar do tratamento de dados, proibindo a realização de tratamento para fins discriminatórios ilícitos ou abusivos. Em conjunto com o princípio da transparência e o direito à explicação sobre decisões automatizadas (Art. 20), a LGPD fornece o arcabouço jurídico para questionar e responsabilizar por resultados algorítmicos que violem a isonomia (OLIVEIRA JR., 2025).

²⁴ LGPD. Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios: [...] IX - não discriminação: impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos.

A proposta brasileira de incluir os “neurodireitos” no Código Civil confirma a gravidade desses riscos. Entre as salvaguardas previstas figura expressamente “a proteção contra vieses” (TEFFÉ; CEIA, 2024), reconhecendo que o uso de dados neurais para fins diagnósticos ou de triagem pode reproduzir discriminações estruturais, sobretudo quando combinado a bases de dados historicamente excludentes (ANPD, 2025; NEGRI *et al.*, 2025). Tal reconhecimento reforça que a discriminação algorítmica em saúde viola não apenas o art. 5º, *caput*, da Constituição, mas também um bem jurídico emergente – a integridade mental, livre de manipulações enviesadas – que o legislador pretende alçar à condição de direito fundamental (TEFFÉ; CEIA, 2024).

A Proposta de Emenda à Constituição (PEC) nº 29/2023 (BRASIL, 2023), portanto, propõe a inclusão do inciso LXXX ao Art. 5º da Constituição Federal, visando assegurar a “integridade mental e a transparência algorítmica”. Esta PEC consagra os neurodireitos como “cláusulas pétreas” e “limites éticos e normativos”, reafirmando a Constituição como “viva” e responsiva às demandas sociais. O anteprojeto do Código Civil também reconhece os neurodireitos como “parte indissociável da personalidade e recebem a mesma proteção desta”, não podendo ser transmitidos, renunciados ou limitados. O Chile, por exemplo, já alterou sua Constituição para incluir a proteção da integridade física e mental das pessoas, destacando a atividade cerebral e informações relacionadas (TEFFÉ; CEIA, 2024).

4.3.2 A Materialização do Dano: Iniquidades em Saúde e Diagnósticos Falhos

A violação abstrata do princípio da igualdade se traduz em danos concretos e, por vezes, irreparáveis à saúde e à vida dos pacientes. As iniquidades diagnósticas são a manifestação mais direta da discriminação algorítmica: (i) *um diagnóstico de câncer de pele tardio, para um paciente negro*, porque o algoritmo não foi treinado com imagens de seu fototipo (CAZANI; ALPINO *et al.*, 2025); (ii) *uma subestimação do risco de infarto, em uma mulher*, porque os modelos cardiológicos foram historicamente baseados em dados masculinos (NORORI *et al.*, 2021); ou (iii) *a falha na detecção de retinopatia diabética em uma população indígena*, cujos dados não compuseram o treinamento do sistema (OLIVEIRA; SILVA *et al.*, 2024) – todos esses cenários, derivados dos casos analisados no Capítulo 2, ilustram como o viés deixa de ser uma questão teórica para se tornar uma ameaça real.

Esses erros sistemáticos não são aleatórios. Aliás, nas palavras de Daniel Kahneman (2012), esses erros sistemáticos – conhecidos como vieses – se repetem de forma previsível em circunstâncias particulares. Eles afetam desproporcionalmente grupos que já enfrentam

barreiras históricas no acesso à saúde, com o risco de a tecnologia, em vez de mitigar, aprofundar as desigualdades existentes.

A discriminação algorítmica, no âmbito da saúde, configura ilícito civil reparável em múltiplas dimensões (WÓJCIK, 2022; SMOLENAARS; GONÇALVES, 2021): *(i) Dano material, expresso nos custos adicionais de terapias mais complexas exigidas por um diagnóstico tardio; (ii) Dano moral, resultante do sofrimento psíquico e da lesão à dignidade do paciente; e (iii) Perda autônoma de uma chance de cura ou de obtenção de prognóstico mais favorável, categoria já acolhida pela jurisprudência pátria como passível de indenização*²⁵.

Encerramos este Capítulo demonstrando que a saúde algorítmica e neurotecnológica opera sob um conjunto de vulnerabilidades estruturais – vieses de dados e dinâmicas de interação humano-máquina – que, somadas, reconfiguram o próprio ambiente decisório clínico. À luz desse diagnóstico, retomamos a pergunta central formulada na INTRODUÇÃO dessa dissertação: **de que maneira as características operacionais intrínsecas da IA e das neurotecnologias – em especial a opacidade algorítmica e o viés sistêmico – corroem a autonomia do paciente e precipitam o colapso funcional da doutrina do Consentimento Livre e Esclarecido na prática médica?** É o que pretendemos responder no próximo Capítulo.

²⁵ REsp nº 1.662.338 – SP. Relatora: Ministra Nancy Andrigh, DJe: 12 de dezembro de 2017. Disponível em: https://processo.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencial=1654017&num_registro=201503075580&data=20180202&formato=PDF. Acesso em: 14 mar. 2024.

5 A EROSÃO DA AUTONOMIA E O COLAPSO DO CONSENTIMENTO LIVRE E ESCLARECIDO: REPERCUSSÕES DA OPACIDADE E DO VIÉS NA SAÚDE INTELIGENTE E NEUROTECNOLÓGICA

Os Capítulos anteriores desta dissertação delinearão os desafios fundamentais que a inteligência artificial (IA) e as neurotecnologias impõem ao direito e à ética na saúde. O Capítulo 3 desvendou o desafio procedimental da opacidade algorítmica, a “caixa-preta” que, ao impedir a compreensão da lógica subjacente a um diagnóstico, tensiona o direito à explicação e o dever de informação do médico (PASQUALE, 2015; MITTELSTADT, 2021). Em seguida, o Capítulo 4 atravessou esse véu de indecifrabildade para expor o desafio substantivo do viés, demonstrando como os algoritmos, longe de serem neutros, podem absorver, replicar e amplificar desigualdades sociais, resultando em discriminação, iniquidades diagnósticas e até mesmo em manipulação sutil da tomada de decisão (O'NEIL, 2016; CRAWFORD, 2021).

Este Capítulo se propõe a analisar a confluência desses dois fenômenos, argumentando que a sua interação – a incapacidade de explicar (opacidade) e a propensão a discriminar e manipular (viés) – não apenas desafia, mas provoca um colapso funcional dos pilares que sustentam a relação clínica e o direito médico: a autonomia da vontade e o Consentimento Livre e Esclarecido (CLE). O ponto central a ser defendido é que a opacidade torna o consentimento fundamentalmente *não esclarecido*, pois o paciente é levado a anuir a um processo decisório cujos riscos e mecanismos são indecifráveis. Simultaneamente, o viés e a manipulação tornam o consentimento fundamentalmente *não livre*, pois a escolha do paciente é corrompida por informações de qualidade desigual ou por uma arquitetura de escolha deliberadamente enganosa.

Antes de prosseguir, no entanto, é necessário fazer um *disclaimer*: o consentimento do paciente no contexto clínico, é tratado pela doutrina como Consentimento Informado (DANTAS; NOGAROLI, 2020). Porém, para os fins desta dissertação sobre IA e neurotecnologias, optamos por sua designação completa: Consentimento Livre e Esclarecido. Isso porque a ênfase em ambos os elementos – “livre” e “esclarecido” – é deliberada. Ou seja, embora a doutrina bioética frequentemente utilize a expressão “consentimento informado” para diferenciar a prática clínica do campo da pesquisa em saúde (onde a nomenclatura utilizada é Consentimento Livre e Esclarecido), no cenário das novas tecnologias, a retomada da expressão completa se torna imperativa. Conforme será abordado no presente Capítulo, os vieses algorítmicos e a opacidade dos sistemas de IA (“efeito caixa-preta”) representam ameaças diretas a cada um desses pilares. O pilar “esclarecido” é desafiado pela dificuldade de explicar

o funcionamento de algoritmos complexos, enquanto o pilar “livre” pode ser comprometido por recomendações enviesadas e manipuladoras que se apresentam como objetivas, influenciando a decisão do paciente de forma sutil.

Por delimitação metodológica, circunscrevemos a análise ao CLE prestado por pessoas adultas e plenamente capazes. Não examinaremos, portanto, o assentimento de menores nem as hipóteses de representação ou substituição decisória, isto é, situações em que o consentimento é fornecido por terceiro em razão de incapacidade do titular.

Logo, para desenvolver essa dissertação, o percurso argumentativo se desdobrará em três etapas. Primeiramente, será analisado como a opacidade, especialmente em sua forma agravada de “dupla caixa-preta” nas neurotecnologias, introduz um “risco epistêmico” que vicia o consentimento em sua origem, como descrevem Pasquale (2015) e Price II (2017) ao apontarem a impossibilidade de auditoria e explicação de determinados sistemas decisórios. Em segundo lugar, será demonstrado como o viés algorítmico e as técnicas de manipulação corrompem a liberdade de escolha, tornando a autonomia uma ilusão – dinâmica já observada por Zuboff (2021) ao tratar do uso de arquiteturas de escolha para moldar decisões de forma imperceptível.

Por fim, como terceira etapa (a ser aprofundada no Capítulo 6), à luz da literatura crítica sobre IA e neurodireitos, propõe-se superar o modelo de consentimento individual como mecanismo exclusivo de proteção, defendendo a transição para um paradigma de autodeterminação efetiva e de tutela estrutural. Tal paradigma deve se concentrar na governança da tecnologia e na integridade do ambiente decisório. No âmbito da *neurotech* essa tutela será garantida, como sustentam Ienca e Andorno (2017), com a existência de novos direitos para salvaguardar a integridade mental e a identidade pessoal frente a tecnologias neuroinvasivas. Sergio Negri *et al.* (2025) reforçam que a proteção efetiva exige mecanismos institucionais e regulatórios capazes de atuar sobre a arquitetura e o funcionamento das próprias tecnologias, garantindo que as condições para o exercício da autonomia não sejam degradadas por vieses estruturais, opacidade ou manipulação.

5.1 AUTONOMIA E CONSENTIMENTO: PERSPECTIVAS BIOÉTICAS E JURÍDICAS

A autonomia, entendida como a capacidade de um indivíduo governar a própria vida, segundo suas escolhas e valores, constitui hoje um valor central tanto no campo jurídico quanto na bioética. Negar ao ser humano o poder de decidir como conduzir sua vida privada, frustra sua realização existencial. O *princípio da autonomia* fundamenta-se no reconhecimento da dignidade e da liberdade inerentes a cada pessoa humana (SARMENTO, 2005). No contexto

da bioética, especialmente a partir da segunda metade do século XX, a afirmação da autonomia do paciente emergiu em resposta a práticas médicas historicamente pautadas pelo paternalismo: antes, decisões de tratamento eram impostas pelos profissionais de saúde, ao passo que agora se reconhece o direito do paciente de escolher – de forma livre e esclarecida – o que será feito com seu próprio corpo (CÓDIGO DE NUREMBERG, 1947; CLOTET, 2009).

Portanto, sob a perspectiva bioética, o respeito à autonomia do paciente tornou-se um dos quatro princípios fundamentais que devem guiar a relação paciente-médico, ao lado da beneficência, não maleficência e justiça (BEAUCHAMP; CHILDRESS, 2013). Isso significa que qualquer intervenção médica, de caráter preventivo, diagnóstico ou terapêutico, só deve ser realizada com o consentimento prévio, livre e esclarecido do paciente. O consentimento exige informações claras, pertinentes e suficientes sobre justificativas, objetivos esperados, benefícios, riscos, efeitos colaterais, complicações, duração e cuidados do procedimento (UNESCO, 2005; CFM, 2016; DANTAS; JEREISSATI, 2024). Trata-se de assegurar que o indivíduo tenha condições de tomar decisões conscientes sobre sua saúde, livre de coerção ou influências indevidas. O Consentimento Livre e Esclarecido, portanto, concretiza, na prática, o respeito à pessoa enquanto agente moral autônomo, invertendo a lógica tradicional em que o paciente era visto apenas como objeto dos cuidados médicos; constituindo-se, dessa forma, no “paradigma da autonomia do paciente” (CFM, 2016).

Do ponto de vista jurídico, a chamada *autonomia privada* é um princípio basilar do direito civil, pelo qual se reconhece aos indivíduos a liberdade de autodeterminação nas esferas pessoal e patrimonial. Cada pessoa, dotada de personalidade jurídica, pode livremente celebrar contratos, dispor de seus bens e conduzir os rumos de sua vida privada, dentro dos limites da lei e dos bons costumes (SARMENTO, 2005). Essa autonomia da vontade privada está intimamente ligada ao direito fundamental à liberdade e à ideia de que o Estado deve respeitar as escolhas individuais, como ressalta Sarmento (2005), ao afirmar que a autonomia é expressão concreta da dignidade humana. No ordenamento brasileiro, embora não haja um artigo específico intitulado “direito à autonomia”, diversos dispositivos consagram aspectos desse princípio. A Constituição Federal de 1988 (BRASIL, 1988), por exemplo, erige a dignidade da pessoa humana como fundamento da República (art. 1º, III), demonstrando que o Estado está para a pessoa, e não a pessoa para o Estado (art. 3º), o que implica, portanto, reconhecer a pessoa como um fim em si mesma, capaz de traçar seu próprio projeto de vida (BRASIL, 1988; DANTAS; JEREISSATI, 2024). Ademais, garante-se no art. 5º, II e X, a inviolabilidade da liberdade e da intimidade, assegurando a cada um o poder de fazer ou deixar de fazer algo, salvo imposição legal (BRASIL, 1988).

No campo do direito à saúde, a autonomia do paciente aparece como extensão específica da autonomia privada, incidindo sobre decisões relativas ao próprio corpo e tratamentos médicos (NOGAROLI, 2023). Aliás, o “direito ao próprio corpo” é uma espécie de direito da personalidade, decorrente da autonomia privada, onde a autodeterminação pessoal prepondera (DANTAS; JEREISSATI, 2024). A legislação brasileira reforça essa prerrogativa: o Código Civil, em seu artigo 15, estabelece que ninguém pode ser constrangido a submeter-se a tratamento médico ou intervenção cirúrgica contra a sua vontade, mesmo que se trate de medida necessária para salvar-lhe a vida (BRASIL, 2002). Em complemento, o dever jurídico de obter o Consentimento Livre e Esclarecido do paciente antes de qualquer procedimento invasivo ou de risco está consagrado tanto em normas éticas quanto na jurisprudência.

O Código de Ética Médica vigente (Resolução CFM nº 2.217/2018) proíbe o médico, em seu art. 22, de realizar procedimentos sem o consentimento do paciente ou de seu representante legal (salvo em caso de risco iminente de morte), e determina como imperativo o fornecimento de todas as informações sobre diagnósticos, prognósticos e terapêuticas propostas (Art. 34) – (CFM, 2018). Em situações de urgência e emergência, a Recomendação CFM nº 1/2016 prevê que o médico pode atuar em favor da vida do paciente, sem consentimento prévio, amparado no princípio da beneficência, mas deve registrar e justificar o fato em prontuário (CFM, 2016). Ademais, em casos de transtornos mentais, que impeçam o consentimento, a Resolução CFM nº 2.057/2013 (Art. 29, §2º) admite o tratamento sem consentimento, buscando-o do representante legal (CFM, 2013).

Corroborando esses preceitos, o Superior Tribunal de Justiça (STJ) já decidiu que a falta de esclarecimento adequado e de obtenção do consentimento livre e esclarecido configura violação ao direito de autodeterminação do paciente, acarretando responsabilidade civil ao profissional ou à instituição de saúde envolvida²⁶, vejamos:

[...] 3. O dever de informação é a obrigação que possui o médico de esclarecer o paciente sobre os riscos do tratamento, suas vantagens e desvantagens, as possíveis técnicas a serem empregadas, bem como a revelação quanto aos prognósticos e aos quadros clínico e cirúrgico, salvo quando tal informação possa afetá-lo psicologicamente, ocasião em que a comunicação será feita a seu representante legal. 4. O princípio da autonomia da vontade, ou autodeterminação, com base constitucional e previsão em diversos documentos internacionais, é fonte do dever de informação e do correlato direito ao consentimento livre e informado do paciente e preconiza a valorização do sujeito de direito por trás do paciente, enfatizando a sua capacidade de se autogovernar, de fazer opções e de agir segundo suas próprias deliberações. 5. Haverá efetivo cumprimento do dever de informação quando os esclarecimentos se relacionarem especificamente ao caso do paciente, não se mostrando suficiente a informação genérica. Da mesma forma, para validar a

²⁶ STJ, Recurso Especial n. 1.540.580/DF. Rel. Min. Lázaro Guimarães (Convocado TRF-5ª Região), Rel. p/ acórdão Min. Luis Felipe Salomão. 4ª Turma. Julgado em 02/08/2018, DJe 04/09/2018.

informação prestada, não pode o consentimento do paciente ser genérico (blanket consent), necessitando ser claramente individualizado. 6. O dever de informar é dever de conduta decorrente da boa-fé objetiva e sua simples inobservância caracteriza inadimplemento contratual, fonte de responsabilidade civil per se. A indenização, nesses casos, é devida pela privação sofrida pelo paciente em sua autodeterminação, por lhe ter sido retirada a oportunidade de ponderar os riscos e vantagens de determinado tratamento, que, ao final, lhe causou danos, que poderiam não ter sido causados, caso não fosse realizado o procedimento, por opção do paciente. 7. O ônus da prova quanto ao cumprimento do dever de informar e obter o consentimento informado do paciente é do médico ou do hospital, orientado pelo princípio da colaboração processual, em que cada parte deve contribuir com os elementos probatórios que mais facilmente lhe possam ser exigidos. 8. A responsabilidade subjetiva do médico (CDC, art. 14, §4º) não exclui a possibilidade de inversão do ônus da prova, se presentes os requisitos do art. 6º, VIII, do CDC, devendo o profissional demonstrar ter agido com respeito às orientações técnicas aplicáveis. Precedentes. 9. Inexistente legislação específica para regulamentar o dever de informação, é o Código de Defesa do Consumidor o diploma que desempenha essa função, tornando bastante rigorosos os deveres de informar com clareza, lealdade e exatidão (art. 6º, III, art. 8º, art. 9º).

Em outras palavras, do prisma jurídico positivo, a autonomia do paciente não é apenas um ideal bioético, mas um verdadeiro direito subjetivo exigível, tutelado pelo ordenamento jurídico e cuja violação pode ensejar responsabilidade civil, especialmente nos casos de inexistência de consentimento ou de utilização de *blanket consent* – termo que designa autorizações genéricas e insuficientemente informadas (CALIXTO; SCHULZE, 2025).

Importante salientar que, embora fundamental, a autonomia (seja privada em geral, seja do paciente em particular) não possui caráter absoluto. Tanto a bioética, quanto o direito, impõem limites quando o exercício da vontade colide com outros bens jurídicos ou princípios igualmente relevantes. No direito civil, vigora a regra de que ninguém pode dispor de forma ilimitada sobre a própria integridade física: o artigo 13 do Código Civil veda atos de disposição do corpo que resultem em diminuição permanente da saúde ou que violem os bons costumes (BRASIL, 2002). Isso significa que a liberdade individual encontra fronteiras no respeito à vida, à integridade e à ordem pública. No contexto médico, essa limitação se reflete, por exemplo, na proibição de procedimentos flagrantemente desproporcionais ou danosos, ainda que o paciente os consinta (GUERRA; SCHAEFER, 2009).

Sob a ótica bioética, reconhece-se que o respeito à autonomia deve ser exercido sempre que possível, mas não de maneira irrestrita. O princípio da beneficência (fazer o bem) e da não maleficência (evitar o mal) podem justificar intervenções limitadoras da vontade do paciente em situações extremas – por exemplo, quando o paciente não tem plena capacidade de decisão, ou quando sua escolha pode causar dano grave e irreversível a si mesmo ou a terceiros (PAGE, 2012). Todavia, mesmo nesses cenários, procura-se minimizar as violações à autonomia, buscando alternativas menos invasivas e privilegiando, sempre que cabível, a vontade

previamente manifestada pelo indivíduo (como ocorre no caso das diretivas antecipadas de vontade, em que o paciente pode, antecipadamente, delinear quais tratamentos aceita ou recusa, orientando os médicos caso ele próprio fique impossibilitado de expressar sua vontade) – (SILVA, 2021).

No horizonte da chamada “era neural”, em que tecnologias neurocientíficas avançadas e inteligências artificiais aplicadas à saúde ganham espaço, o debate sobre a autonomia do paciente adquire novos contornos. A capacidade de escolher “com consciência” – isto é, de forma informada e genuinamente livre – enfrenta desafios inéditos trazidos por recursos como interfaces cérebro-máquina, dispositivos de neuromonitoramento e algoritmos de apoio à decisão clínica (DANTAS; NOGAROLI, 2020; DANTAS; JEREISSATI, 2024).

Por um lado, essas inovações prometem empoderar ainda mais os pacientes, fornecendo-lhes dados sobre seu próprio cérebro e saúde mental, e potencialmente ampliando sua autonomia (por exemplo, pacientes paralisados que, graças a interfaces neurais, conseguem comunicar suas decisões e desejos de tratamento). Por outro lado, surgem preocupações quanto à voluntariedade e privacidade mental. As neurotecnologias são capazes de decodificar e manipular com precisão os estados e processos mentais das pessoas, afetando a base de suas intenções, emoções e decisões (YUSTE; GOERING, 2017). A mente humana pode se tornar controlável à distância, subordinada a controle externo ou a interesses comerciais, com o risco de realinhar hábitos e alterar a autonomia da pessoa, já afirmava Rodotà (2008).

Portanto, se aparelhos podem acessar ou mesmo alterar estados mentais, pensamentos e emoções, como garantir que as escolhas do paciente continuem sendo autênticas e não influenciadas indevidamente por tais tecnologias? A tomada de decisão humana pode ser influenciada, como vimos nos Capítulos anteriores, tanto pela opacidade algorítmica quanto por vieses das mais diversas ordens, tornando a mera manifestação de vontade (consentimento) insuficiente para garantir a autodeterminação (NEGRI *et al.*, 2025; DANTAS; JEREISSATI, 2024).

5.2 A OPACIDADE COMO HIPÓTESE DE AUSÊNCIA DE CONSENTIMENTO: O RISCO EPISTÊMICO E O DIREITO À COMPREENSÃO

Como demonstrado no item anterior, a doutrina clássica do Consentimento Livre e Esclarecido, pilar da bioética e do direito médico, está historicamente fundamentada na comunicação de riscos de resultado: os efeitos colaterais de um medicamento, a probabilidade de sucesso de uma cirurgia e as sequelas de um procedimento. Contudo, a opacidade dos

sistemas de IA introduz uma categoria de risco inteiramente nova e para a qual o arcabouço jurídico tradicional se mostra inadequado: o *risco epistêmico*.

Trata-se do risco inerente ao próprio processo de tomada de decisão, o risco de não saber como uma conclusão foi alcançada e, conseqüentemente, de não ser capaz de auditar sua validade, identificar suas falhas ou justificar sua lógica (PASQUALE, 2015; PRICE II, 2017). Quando um paciente consente com um diagnóstico auxiliado por uma “caixa-preta”, ele não está apenas aceitando os riscos clínicos do procedimento subsequente; ele está, primeiramente, consentindo com um método de decisão que é, por definição, inexplicável. Pasquale (2015) argumenta que a opacidade desses sistemas cria assimetrias de informação que minam a responsabilização e a confiança. Price II (2017) complementa que, na chamada *black-box medicine*, a impossibilidade de explicar a lógica decisória torna o dever de informar quase impraticável e o direito do paciente à compreensão, uma promessa ilusória. Essa ocultação de um risco fundamental do processo vicia o consentimento em sua origem, transformando-o em um ato formal sem o conteúdo material necessário para a autonomia genuína, revelando-se, de fato, como um “não-consentimento”.

5.2.1 O Consentimento para o Inexplicável: A “Dupla Caixa-Preta” e a Inexistência da Informação como Hipótese de Nulidade

O problema da opacidade alcança seu ponto mais crítico no conceito de “dupla caixa-preta”, analisado no Capítulo 3, que expressa a sobreposição entre a inescrutabilidade algorítmica e a própria complexidade inerente ao funcionamento do cérebro humano presentes nas neurotecnologias. Nesse cenário, o Consentimento Livre e Esclarecido se aproxima de um ponto de colapso funcional, pois a informação a ser transmitida ao paciente torna-se tão complexa e incerta que perde seu propósito de esclarecimento.

A Resolução do Parlamento Europeu de 2017 (UNIÃO EUROPEIA, 2018a), no parágrafo 12, recomenda que robôs e sistemas de inteligência artificial sejam dotados de um sistema capaz de registrar as operações realizadas e os passos lógicos que conduziram às suas decisões, com o objetivo de assegurar maior explicabilidade e transparência. Tal mecanismo pode ser compreendido como uma medida paliativa para mitigar os efeitos da opacidade algorítmica. Contudo, sua aplicabilidade revela-se insuficiente no campo das neurotecnologias, sobretudo diante da “dupla caixa-preta”, conforme se demonstrará a seguir.

Os casos concretos analisados no Capítulo 2 ilustram essa fragilidade de forma contundente. Considere-se a Interface Cérebro-Computador (*BCI*) para restauração da fala em

uma paciente com Esclerose Lateral Amiotrófica (ELA) – (WILLETT *et al.*, 2023). A paciente consente com um implante neurocirúrgico que promete traduzir suas intenções de fala em texto ou voz sintetizada. No entanto, se o dispositivo comete um erro – gerando uma palavra ofensiva ou uma frase indesejada –, a causa é, em grande medida, indecifrável. Foi uma falha no algoritmo de decodificação? Um “ruído” atípico nos sinais neurais da própria paciente? Ou uma interação imprevisível e não-linear entre os dois sistemas opacos? O paciente consente com uma tecnologia sem que lhe possa ser oferecida uma explicação clara sobre como diagnosticar ou corrigir uma falha que afeta a manifestação mais íntima de seu “eu”: a sua voz e o seu pensamento (YUSTE *et al.*, 2017; IENCA; ADORNO, 2017; ANPD, 2025).

O mesmo dilema se manifesta na aplicação da Estimulação Cerebral Profunda (*Deep Brain Stimulation – DBS*) para a Doença de Alzheimer (XU; LIU; SHANG *et al.*, 2024). Pacientes com capacidade cognitiva já comprometida são submetidos a um procedimento invasivo que modula a atividade de seus cérebros de forma opaca (DANTAS; JEREISSATI, 2024; ANPD, 2025). O consentimento, muitas vezes obtido por meio de um representante legal, é dado para um procedimento cujos mecanismos de ação e, crucialmente, cujos efeitos de longo prazo sobre a personalidade e a identidade permanecem duplamente inexplicáveis. Xu; Liu; Shang *et al.* (2024) destacam que, no caso da *DBS* em Alzheimer, a incerteza quanto à eficácia e aos impactos neurológicos futuros agrava os dilemas éticos. Nesse cenário, a exigência legal de um consentimento “esclarecido” arrisca-se a tornar-se uma ficção jurídica, um mero formalismo que não cumpre sua função de proteger a autonomia da pessoa em situação de vulnerabilidade, como advertem Dadalto *et al.* (2021) e Nogaroli (2023), ao apontarem que a obtenção formal de consentimento em contextos de incapacidade cognitiva demanda salvaguardas adicionais e não pode ser tratada como ato meramente protocolar.

Estudos qualitativos com pacientes submetido a *DBS* e *BCI* já indicam “alterações significativas no senso de identidade e de controle sobre as próprias ações após o uso prolongado da tecnologia”. Neurotecnologias podem “alterar o sentido e conteúdo da autonomia da pessoa” e “perturbar o senso de identidade e o autogoverno da pessoa” (DANTAS; JEREISSATI, 2024). As implicações de longo prazo para a saúde mental carecem de evidências científicas (ANPD, 2025).

Se tais efeitos ocorrem em neurodispositivos relativamente simples, o futuro de plataformas mais avançadas e invasivas (como será visto no próximo item) ainda é incerto e inquieta muitos especialistas (YUSTE *et al.*, 2017; DANTAS; JEREISSATI, 2024).

Portanto, quando a lógica de um sistema de IA clínica é uma “caixa-preta” – e, no caso de neurotecnologias, enfrenta-se uma “dupla opacidade” (algorítmica e neurobiológica) – torna-

se tecnicamente difícil oferecer ao paciente uma explicação significativa sobre funcionamento, limitações, incertezas e riscos do modelo. Em relações de consumo (*e.g.*, serviços médico-hospitalares privados, planos de saúde, oferta de dispositivos e SaMD ao paciente), essa opacidade colide diretamente com o direito básico à informação adequada e clara (CDC, art. 6º, III) e com o dever de segurança/informação sobre riscos (arts. 8º e 9º) (BRASIL, 1990). A insuficiência informacional, no CDC, não é um detalhe: produto é defeituoso também por “informações insuficientes ou inadequadas sobre sua utilização e riscos” (art. 12), e o serviço é defeituoso quando não oferece a segurança que dele legitimamente se espera, à luz dos riscos e do modo de fornecimento (art. 14, caput e § 1º) (BRASIL, 1990). Por isso, “consentimentos” colhidos sem informação materialmente inteligível sobre um sistema opaco – isto é, sem que o consumidor-paciente compreenda o que o sistema faz, seus limites e riscos relevantes – são ineficazes; e cláusulas que pretendam convalidar esse déficit informacional ou limitar direitos do consumidor são nulas de pleno direito (art. 51, I e IV) (BRASIL, 1990).

A jurisprudência do STJ, como demonstrado anteriormente, reforça: o dever de informar é dever básico nas relações de consumo e não se admite “*blanket consent*” (consentimento genérico) que não individualize riscos e alternativas terapêuticas – sua inobservância caracteriza falha do serviço e gera responsabilidade (STJ, 2018).

Em síntese, sob o CDC, a opacidade que impede informação adequada vicia o consentimento obtido e caracteriza defeito do produto/serviço, atraindo responsabilidade objetiva do fornecedor (cadeia de desenvolvedores, importadores, hospitais) independentemente de culpa (arts. 12 e 14) – e eventuais cláusulas exoneratórias são nulas (art. 51) (BRASIL, 1990; CALIXTO; SCHULZE, 2025).

Ademais, à luz de uma leitura analógica da LGPD, é possível sustentar a nulidade do consentimento do paciente, “caso as informações fornecidas ao titular tenham conteúdo enganoso ou abusivo ou não tenham sido apresentadas previamente com *transparência, de forma clara e inequívoca*” (BRASIL, 2018, Art. 9º, §1º).

5.2.2 A Integridade Psicológica e a Limitação da Autonomia em Terapias Moduladoras

O tema introduzido no item anterior (5.2.1) demanda aprofundamento, em razão da elevada opacidade desses sistemas de IA e neurotecnologias aplicados à saúde, bem como das potenciais consequências psicológicas, ainda pouco conhecidas e insuficientemente mapeadas.

Como visto, o consentimento para o uso de neurotecnologias, que modulam ativamente o cérebro, representa, segundo Yuste *et al.* (2017), a fronteira mais desafiadora para a

autonomia, pois envolve autorizar intervenções capazes de afetar elementos centrais da identidade pessoal. Trata-se de um consentimento que implica aceitar o risco de uma potencial alteração dessa identidade – risco que, devido à opacidade dos mecanismos envolvidos, é impossível de ser plenamente dimensionado *a priori*. A autonomia, nesse contexto, é exercida para autorizar um procedimento que pode, paradoxalmente, modificar a própria base da qual essa autonomia emana. Esse cenário se conecta diretamente ao debate sobre a necessidade de um “direito à continuidade psicológica” como nova salvaguarda fundamental, destinado a proteger a integridade mental frente a tecnologias neuroinvasivas (YUSTE *et al.*, 2017).

Para além dos exemplos apresentados no item anterior, merece destaque a chamada “ponte digital” cérebro-coluna (*Brain-Spine Interface*), já examinada no Capítulo 2. Trata-se de uma *BCI* invasiva de nova geração – mais complexa que as anteriores, porque fecha o circuito motor humano, reintegrando sinais cerebrais ao próprio corpo do paciente –, que é capaz de restaurar a marcha em pacientes com lesão medular, representando um exemplo paradigmático do desafio à autonomia na interação homem-máquina. Nesse procedimento, o paciente consente em ter suas intenções de movimento decodificadas e convertidas em ação por um sistema computacional (LORACH *et al.*, 2023). O processo envolve implantes cranianos detectando atividade elétrica no córtex quando o paciente “pensa em andar”, sendo o sinal transmitido, decodificado por um computador (em uma mochila), e os dados enviados a um gerador de pulsos na coluna vertebral, ativando eletrodos (LAVAZZA *et al.*, 2025). A autonomia motora, embora recuperada, o é por meio da cessão de controle a um sistema cuja lógica operacional permanece, em parte, opaca e indecifrável para o próprio usuário. Forma-se, assim, uma agência híbrida e compartilhada entre cérebro e máquina, cujas regras de interação não são totalmente conhecidas.

Essa configuração reforça a necessidade de um modelo ampliado de consentimento, que não se limite à autorização pontual para o uso da tecnologia, mas que inclua mecanismos contínuos de supervisão, revisão e possibilidade de retirada do consentimento, protegendo a autodeterminação em face da opacidade e da imprevisibilidade tecnológica. Estudos recentes sobre ética em neurotecnologia ressaltam o caráter inadequado de abordagens meramente procedimentais – como aprovações éticas isoladas e formulários de consentimento estático – e apontam para a importância de processos reflexivos e contínuos ao longo de todo o ciclo de interação com dispositivos implantáveis (HAAG; IENCA *et al.*, 2025; DANTAS; JEREISSATI, 2024). Além disso, o princípio ético de não abandono em pesquisas com dispositivos neurais reforça que o consentimento deve estender-se ao acompanhamento pós-

implantação, com responsabilidades claras sobre comunicação e segurança ao longo do tempo (CENOLLI, 2024; CALIXTO; SCHULZE, 2025).

Esse modelo ampliado encontra respaldo no debate contemporâneo sobre os neurodireitos, que defendem a incorporação de mecanismos jurídicos capazes de responder a impactos tecnológicos profundos sobre a identidade, a agência e a integridade mental. Em particular, Timo Istace e Milena Costas Trascasas (2024) argumentam que a natureza emergente das neurotecnologias modernas torna o cumprimento pleno do consentimento livre e informado praticamente inviável, sendo necessários protocolos legais dinâmicos, capazes de atualizar a análise de risco e consentimento conforme o progresso científico. Esses fundamentos apontam para uma estrutura jurídica que transcenda o consentimento individual pontual e abarque salvaguardas estruturais contínuas – como revisões técnicas periódicas, comissões de acompanhamento e canais de retratação – garantindo que a escolha do paciente seja efetivamente respeitada ao longo do tempo.

Outro caso paradigmático é o sistema de neuroestimulação responsiva (*RNS*) para epilepsia. Funciona em malha fechada, como também relatado no Capítulo 2, intervindo de forma autônoma na atividade cerebral para prevenir convulsões (SCHULZE-BONHAGE *et al.*, 2023). O paciente consente em ser continuamente monitorado e ter sua atividade neural modulada sem sua intervenção consciente, com base em uma lógica algorítmica que não pode ser explicada em tempo real. A autonomia, aqui, é exercida primordialmente no momento inicial do consentimento, sendo em seguida delegada a um sistema cuja opacidade tecnológica impede o controle ou a compreensão contínua por parte do paciente. Molnár-Gábor *et al.* (2024) complementam que essa delegação, quando não acompanhada de mecanismos claros de supervisão e possibilidade de revisão, pode comprometer direitos fundamentais ligados à liberdade cognitiva e à autodeterminação (AUSTRÁLIA, 2024).

Portanto, a opacidade dos sistemas inteligentes de saúde não ameaça apenas a autonomia informacional do paciente; ela se projeta sobre a própria integridade da pessoa, em suas dimensões física e psicológica (ROY; MALAFA; ADWER *et al.*, 2025). Enquanto os riscos à integridade física, associados a procedimentos invasivos, são mais facilmente reconhecíveis pelo Direito, as neurotecnologias inauguram uma nova categoria de risco, mais sutil e profunda (ZUK; SANCHEZ *et al.*, 2023): a possibilidade de modular a identidade e a personalidade, desafiando o conceito de continuidade psicológica e exigindo uma resposta constitucional à altura (MUÑOZ; MARINARO, 2023).

5.2.2.1 *Análise dos Riscos à Integridade Física*

Os riscos imediatos e já bem documentados das neurotecnologias invasivas – tais como o implante de eletrodos para estimulação cerebral profunda (*DBS*) ou de sensores intracorticais em interfaces cérebro-computador (*BCIs*) – incidem diretamente sobre a integridade física do paciente (LU; LIU; WEI *et al.*, 2025). Mesmo com abordagens neurocirúrgicas altamente sofisticadas, permanecem possíveis complicações como hemorragias intracerebrais, infecções do sítio de implante, reações adversas à anestesia e lesões ao tecido neural adjacente. Para além dessas contingências agudas, subsistem incertezas relativas aos efeitos de longo prazo da presença de um corpo estranho no cérebro e da estimulação elétrica crônica, incluindo processos inflamatórios, neurotoxicidade e plasticidade neural indesejada. Revisões recentes sobre *BCIs* alertam, ainda, para a degradação progressiva dos materiais condutores e isolantes, a qual compromete a eficácia terapêutica e a segurança do dispositivo (CHEN; CHEN; WANG *et al.*, 2025). Ademais, a ausência de políticas claras de manutenção e substituição desses implantes expõe os pacientes a riscos adicionais, pois falhas de *hardware* podem ocorrer anos após a cirurgia (GAO; YAN; ZHOU *et al.*, 2025).

Documentos estratégicos, como o relatório *BRAIN 2025*²⁷, insistem na necessidade de acompanhar sistematicamente esses problemas ao longo de décadas de uso, enquanto autoridades europeias de proteção de dados recomendam que tais zonas de incerteza sejam comunicadas de forma transparente durante o Consentimento Livre e Esclarecido (ARROYO, 2025). Por fim, estudos recentes salientam que a segurança desses dispositivos não se restringe aos aspectos biológicos, mas abrange vulnerabilidades cibernéticas capazes de afetar a integridade física e cognitiva do usuário (SCHRODER; SIRBU; PARK *et al.*, 2025).

5.2.2.2 *Análise dos Riscos à Integridade Psicológica: a Modulação da Personalidade e o Debate sobre o “Direito à Continuidade Psicológica”*

O desafio mais profundo e inovador para o Direito, no entanto, reside na tutela da integridade psicológica. Como analisamos acima, a capacidade das neurotecnologias de modular diretamente a atividade de circuitos neurais que governam o humor, o comportamento e a cognição vai além do dano físico para tocar no núcleo da psique humana. Relatos na literatura médica sobre pacientes com Parkinson tratados com *DBS* que experimentaram efeitos colaterais como hipomania, apatia, aumento da impulsividade ou outras alterações de personalidade não são incomuns (IENCA, 2021; COUTINHO, 2024; ANPD, 2025). Embora

²⁷ Disponível em: https://braininitiative.nih.gov/sites/default/files/documents/brain2025_508c_2.pdf.

muitas vezes esses efeitos sejam reversíveis com o ajuste dos parâmetros de estimulação, eles ilustram uma nova e potente forma de intervenção sobre o “eu” (POMES *et al.*, 2025).

Essa problemática conecta-se a uma reflexão mais ampla sobre como tecnologias emergentes podem atingir dimensões centrais da autonomia humana. Nesse sentido, Shoshana Zuboff (2021) descreve que as operações de “suprimento” de dados podem ser dirigidas à personalidade, às emoções e às vulnerabilidades individuais, muitas vezes contornando a consciência e os direitos de escolha. Para a autora, a ameaça à autonomia inicia-se justamente por um ataque à consciência, “dilacerando nossa capacidade de regular pensamentos, emoções e desejos” – elementos estruturantes do desenvolvimento pessoal, da autenticidade e do controle dos impulsos.

Isso nos força a confrontar um tipo de dano que o Direito tradicional tem dificuldade em categorizar. O direito da responsabilidade civil está bem equipado para lidar com danos físicos (lesões corporais) e danos morais (dor, sofrimento, angústia) – (PORTO; WERNECK, 2021). Contudo, como a modulação neurotecnológica pode alterar traços fundamentais da personalidade, memórias ou o senso de si de uma pessoa, sem necessariamente causar dor ou sofrimento no sentido clássico (POMES *et al.*, 2025), o indivíduo pode simplesmente “tornar-se outro”, para o bem ou para o mal, de uma forma que ele próprio ou seus familiares não o reconheçam mais (BROWNING, 2025).

Essa alteração da identidade fundamental de uma pessoa representa um tipo de dano novo e profundo, que ataca a própria essência do ser. Torna-se necessário, portanto, conceituar juridicamente o “dano à identidade” ou o “dano à continuidade psicológica” como uma categoria autônoma de prejuízo indenizável (ASTOBIZA; BERIAIN, 2025). Essa necessidade impulsiona o debate, liderado por teóricos dos neurodireitos como Rafael Yuste e o *Morningside Group*, sobre a proposição de um novo direito humano a ser tutelado: o direito à identidade pessoal e à continuidade psicológica (YUSTE, 2017). Este direito visa proteger a coerência da narrativa pessoal de um indivíduo e seu senso de ser a mesma pessoa ao longo do tempo, resguardando-o contra alterações tecnológicas não consentidas que fragmentem ou perturbem sua identidade (DI SALVO, 2025; YUSTE *et al.*, 2021).

Nesse contexto, a análise jurídica da nulidade do CLE, fundamentada no Código de Defesa do Consumidor e na jurisprudência do STJ, precisa ser expandida para um patamar mais abrangente. O que poderia soar como um debate futurista já recebeu resposta normativa concreta no Brasil. A Proposta de Emenda à Constituição nº 29/2023 (BRASIL, 2023), atualmente em tramitação no Senado Federal, apresenta-se como uma reação direta e tempestiva aos riscos discutidos ao longo deste Capítulo. A PEC propõe a inclusão do inciso

LXXX ao artigo 5º da Constituição Federal, para garantir que o desenvolvimento científico e tecnológico deverá assegurar a integridade mental e a transparência algorítmica, elevando-os à categoria de direitos e garantias fundamentais.

5.2.2.3 *Conexão com o Debate Constitucional Brasileiro: a PEC 29/2023 e a Positivação da “Proteção à Integridade Mental”*

A análise do texto da proposta revela sua profunda conexão com os desafios da opacidade e da modulação da identidade. A positivação da “proteção à integridade mental” pode ser interpretada como a criação do arcabouço constitucional necessário para abrigar e proteger o bem jurídico da identidade pessoal e da continuidade psicológica contra as novas formas de intervenção neurotecnológica (DANTAS; JEREISSATI, 2024; ARANTES, 2025). Ela eleva a tutela da mente do nível infraconstitucional (como um aspecto da privacidade ou da personalidade no Código Civil) para o mais alto patamar de proteção jurídica.

Aliás, apenas abrindo um parêntese, no anteprojeto de reforma do Código Civil, os neurodireitos são incluídos como parte indissociável da personalidade, não podendo ser transmitidos, renunciados ou limitados, e visam preservar a privacidade mental, a identidade pessoal, o livre arbítrio, o acesso justo à ampliação ou melhoria cerebral, a integridade mental e a proteção contra vieses (TEFFÉ; CEIA, 2024; NEGRI *et al.*, 2025; CAMPOS, 2025).

Simultaneamente, a PEC propõe a inclusão da “transparência no uso de algoritmos” como direito fundamental, em resposta direta ao problema da opacidade e da “caixa-preta” (NOGAROLI, 2023). Ela fornece a base constitucional para o direito à explicação, fortalecendo as garantias já previstas na LGPD e assegurando que o Consentimento Livre e Esclarecido no uso de tecnologias de saúde inteligentes seja efetivamente esclarecido (ENGELMANN, 2023).

Dessa forma, a PEC 29/2023 demonstra que os desafios aqui discutidos não são meras especulações acadêmicas. A opacidade sistêmica, ao ameaçar o direito à explicação, ao viciar o Consentimento Livre e Esclarecido e ao colocar em risco a própria integridade psicológica do paciente, eleva o debate do nível técnico-jurídico para o nível dos fundamentos do Estado de Direito. A necessidade de uma tutela fundamental expressa, como a proposta pela PEC, torna-se não uma medida de precaução excessiva, mas uma salvaguarda essencial para garantir que a medicina do futuro, impulsionada pela tecnologia, permaneça firmemente ancorada na dignidade da pessoa humana (NEVES; BARROS, 2024).

5.2.3 O Risco Epistêmico como Novo Elemento do Dever de Informar

A existência do risco epistêmico impõe uma expansão do dever de informação do profissional de saúde. Não basta mais comunicar os riscos e benefícios clínicos; torna-se ética e juridicamente necessário comunicar a própria incerteza metodológica do diagnóstico ou da terapia baseada em IA. O diálogo com o paciente deve incluir, de forma transparente, o fato de que uma recomendação pode se basear em uma lógica que nem mesmo o médico compreende plenamente (TOPOL, 2019; NOGAROLI, 2023). Nesses casos, o médico deverá “considerar outros fatores, como o histórico médico do paciente e quaisquer exames diagnósticos adicionais, e ignorar previsões com alta incerteza ou triá-las para revisão humana detalhada. Portanto, é importante que o médico compreenda essa incerteza e como ela pode impactar o processo de tomada de decisão” (ZOU; ZHIHAO *et al.*, 2023).

No caso do estudo *PRAIM*, que utiliza IA para aumentar a detecção de câncer de mama em mamografias, a comunicação muda radicalmente (EISEMANN; BUNK *et al.*, 2025). Ao receber um alerta da “rede de segurança” da IA, o radiologista não pode mais se limitar a dizer “encontrei uma área suspeita”. Seu dever de informação, agora, exige uma honestidade epistêmica, do tipo: *Sra. Fula de Tal, a inteligência artificial, um sistema que, segundo estudos, detecta 17,6% mais casos de câncer, identificou uma área que eu, como especialista, não consigo ver claramente ou interpretar com certeza. A recomendação para a biópsia se baseia, portanto, na confiança na performance estatística do algoritmo, cuja lógica exata para o seu caso específico não é explicável.* Essa nova forma de diálogo transforma o consentimento, não retira a opacidade do sistema, mas auxilia na relação paciente-médico.

Da mesma forma, a ferramenta de cardiologia de Oxford, que prevê o risco de infarto em até dez anos, com base em sinais de inflamação invisíveis ao olho humano, coloca o consentimento sob uma nova luz (CHAN; WAHOME *et al.*, 2024). A decisão do paciente de alterar seu plano de tratamento, como ocorreu em 45% dos casos do estudo piloto, é baseada na aceitação de um risco futuro calculado por uma “caixa-preta”. Isso se conecta ao “direito de não saber” (DAVIES, 2020; RÜGER, 2007), mas com uma nuance crucial: o paciente agora tem o direito de recusar uma informação preditiva, que não pode ser plenamente justificada em seus fundamentos lógicos, apenas em sua acurácia estatística (PLOUG; HOLM, 2020).

Portanto, é fundamental destacar que o consentimento do paciente deve ser concebido como um processo contínuo de diálogo e de “escolha esclarecida”, e não como um mero ato formal de assinatura (NOGAROLI, 2023). Contudo, mesmo concebido dessa forma, o consentimento ainda necessita de salvaguardas jurídicas adicionais capazes de mitigar os efeitos da opacidade sistêmica, como será discutido adiante. A dependência exclusiva do

consentimento como fundamento para a utilização de tecnologias de inteligência artificial e neurotecnologia revela limitações estruturais. A crença de que os indivíduos tomam decisões de maneira inteiramente autônoma e livre de influências externas é contestada por estudos sobre [opacidade algorítmica], vieses e heurísticas, que demonstram como fatores subconscientes moldam julgamentos e escolhas (NEGRI *et al.*, 2025). Em conclusão, o debate jurídico precisa evoluir para reconhecer tanto a complexidade da tomada de decisão humana, quanto as limitações do consentimento como único mecanismo de proteção.

Além disso, como se examinará adiante, o consentimento não é impactado apenas pela opacidade algorítmica, mas também pelos vieses estruturais que permeiam esses sistemas.

5.3 VIÉS ALGORÍTMICO E MANIPULAÇÃO: A CORRUPÇÃO DA LIBERDADE DE ESCOLHA E OS LIMITES DA AUTONOMIA

Se a opacidade vicia o consentimento por torná-lo não esclarecido, o viés algorítmico e as técnicas de manipulação o corrompem por torná-lo não livre. A autonomia do paciente não é exercida em um vácuo, mas depende de um ecossistema de informação e de opções que seja justo, equitativo e transparente (DANKWA-MULLAN, 2024). O viés, como explorado no Capítulo 4, envenena esse ecossistema.

A decisão humana, já suscetível a vieses cognitivos e heurísticas (KAHNEMAN, 2012), torna-se um alvo fácil para sistemas que podem explorar deliberadamente essas vulnerabilidades. Zuboff (2021) descreve como mecanismos digitais de persuasão podem ser estruturados para moldar escolhas sem que o indivíduo perceba a extensão da influência exercida. O'Neil (2016) acrescenta que, quando esses sistemas são aplicados de forma massiva, eles não apenas exploram falhas cognitivas individuais, mas reforçam desigualdades já existentes, afetando de maneira desproporcional os grupos mais vulneráveis. A autonomia, nesse contexto, corre o risco de se tornar uma ilusão, especialmente para aqueles cuja capacidade de escolha já é restringida por desigualdades estruturais (O'NEIL, 2016). A proteção jurídica, portanto, não pode se concentrar apenas no ato final do consentimento, mas deve garantir a integridade de todo o processo que o antecede, como defende Solove (2013), incluindo a preservação de um ambiente decisório livre de manipulações indevidas.

5.3.1 Escolha Desigual: Como o Viés Algorítmico Prejudica a Autonomia de Grupos Vulneráveis

Quando uma ferramenta de diagnóstico é comprovadamente enviesada, a qualidade da informação que ela fornece a diferentes grupos populacionais é, por definição, desigual, comprometendo a isonomia epistêmica na tomada de decisão clínica (OBERMAYER *et al.*, 2019). Isso significa que a base sobre a qual pacientes pertencentes a grupos marginalizados/discriminados/diversos exercem sua autonomia é inerentemente mais frágil e arriscada, uma vez que se apoia em dados e modelos estatísticos construídos a partir de representações populacionais desbalanceadas (NOBLE, 2018). O consentimento dado por esses pacientes, embora formalmente válido, é materialmente menos “esclarecido”, pois se fundamenta em uma probabilidade diagnóstica de menor acurácia e, portanto, em informação de qualidade inferior (BAROCAS; SELBST, 2016). A discriminação algorítmica, nesse sentido, configura não apenas uma falha técnica, mas uma violação do direito a uma escolha autônoma e informada em condições de igualdade, o que implica afrontar diretamente princípios fundamentais de justiça e não discriminação na saúde (PASQUALE, 2020).

O caso da inteligência artificial aplicada à dermatologia, como exemplifica a plataforma NEVO, abordada no Capítulo 2, evidencia de forma contundente o impacto do viés algorítmico sobre a autonomia do paciente (CAZANI; ALPINO *et al.*, 2025). Imagine-se um paciente negro que, ao utilizar um aplicativo de triagem de câncer de pele, recebe um resultado com probabilidade de acerto significativamente inferior àquela obtida por um paciente branco, em razão do viés relacionado ao tom de pele presente nos dados de treinamento (BUOLAMWINI; GEBRU, 2018; NORORI *et al.*, 2021). A decisão autônoma desse paciente de “não procurar um médico”, tomada com base em um resultado falsamente negativo, não constitui uma escolha plenamente livre e informada. Trata-se, na verdade, de uma decisão cuja qualidade foi degradada pela discriminação algorítmica embutida no sistema (O’NEIL, 2016), o que implica, na prática, uma violação do direito de exercer a autonomia em condições de igualdade (PASQUALE, 2020).

O mesmo raciocínio se aplica à aplicação da inteligência artificial em genômica. Uma paciente de ancestralidade não europeia que utiliza uma ferramenta como o *AlphaMissense* para interpretar uma variante genética pode receber uma classificação de “significado incerto” (NGUYEN; LEE *et al.*, 2025). Esse resultado não é meramente fruto de incerteza científica inerente, mas, como observam Manrai *et al.* (2016), consequência do viés de representação nos bancos de dados genômicos, nos quais populações de ancestralidade não europeia estão sub-representadas. Essa sub-representação, segundo Popejoy e Fullerton (2016), amplia artificialmente a margem de incerteza e compromete a equidade na interpretação genética. A decisão subsequente da paciente – seja realizar cirurgias profiláticas, seja aderir a terapias-alvo

– é tomada sobre um terreno de informação menos sólido, e essa fragilidade informacional, decorrente de desigualdades estruturais no treinamento dos algoritmos, atinge diretamente o núcleo de sua autonomia decisória.

5.3.2 *Nudges e Dark Patterns na Relação Paciente-Médico*

Como analisado no item 4.2, a autonomia do paciente pode ser comprometida não apenas pela presença de informações enviesadas, mas também pela forma como tais informações e opções de decisão são apresentadas. Vimos que o *design* da interface de um *software* de saúde pode explorar vieses cognitivos humanos para induzir escolhas específicas, valendo-se de estratégias conhecidas como *nudges* e *dark patterns* (MATHUR et al., 2019). Vimos também que a IA em saúde e as neurotecnologias constituem um terreno particularmente propício à proliferação dessas arquiteturas da escolha. Nessas circunstâncias, o Consentimento Livre e Esclarecido corre o risco de se tornar uma formalidade meramente procedimental, carente de deliberação genuína e incapaz de garantir o exercício pleno da autonomia decisória (YEUNG, 2017).

A opacidade explicativa dos sistemas de IA médica, que pode ser utilizada intencionalmente para manipular a escolha do paciente; a complexidade intrínseca dos dados neurais e do funcionamento das neurotecnologias, que dificulta a explicação compreensível ao titular, abrem espaço para o consentimento não ser realmente livre, uma tática também de *dark patterns* (NOGAROLI, 2023; ANPD, 2025).

A UNESCO (2025) aborda “profundas questões éticas e de direitos humanos relacionadas à autonomia, ao consentimento, à privacidade e ao potencial de manipulação levantado pela neurotecnologia, que surgem nos contextos de sistemas de recomendação, *priming* e *nudging*, *marketing* durante o sono e o sonho, *neuromarketing* e ambientes de ciclo fechado”. Destaca, também a necessidade de se “exigir consentimento prévio, livre e informado para qualquer uso desses dados que possam influenciar decisões ou comportamentos, o direito de optar por não participar desses sistemas e transparência e divulgações claras no momento da coleta de dados, com limitações estritas ao uso de dados para fins além daqueles explicitamente divulgados” (UNESCO, 2025).

Essas práticas deliberadamente enganosas configuram exemplo claro do que Zuboff (2021), em uma analogia ao capitalismo de vigilância, descreveria como “captura silenciosa”²⁸,

²⁸ Zuboff (2021) descreve extensivamente em sua obra o processo de “desposseção” e “extração” de dados comportamentais sem consentimento real, o que corresponde diretamente à ideia de uma “captura silenciosa”. O

em que o consentimento é obtido por meio de mecanismos assimétricos e opacos, comprometendo a autonomia informativa. O’Neil (2016) acrescentaria que tais estratégias exploram vulnerabilidades cognitivas para moldar decisões, transformando o consentimento em um ato meramente formal, sem liberdade real de escolha²⁹. Solove (2025) reconhece que o ambiente digital é propenso a distorções e, por isso, em vez de conceber “fluxos de consentimento” ideais, as leis de privacidade devem ir “além do controle individual e da autogestão” e focar em “medidas estruturais que não coloquem o ônus sobre os indivíduos” (pp. 17 e 20). Cohen (2024), por fim, argumenta que o consentimento informado, especialmente no contexto da IA/ML, é “largamente uma ficção” e que a IA “complica ainda mais a equação; torna a ficção mais fantasiosa”.

5.3.3 O Consentimento sob Viés Sistêmico: Informação Materialmente Defeituosa e Invalidez

Mutatis mutandis, o raciocínio jurídico desenvolvido para a opacidade algorítmica (item 5.2) aplica-se aos casos de viés em IA e neurotecnologias na saúde. Em relações de consumo (e.g., prestação de serviços médico-hospitalares privados, planos de saúde, fornecimento de SaMD e dispositivos ao paciente), o CDC assegura o direito básico à informação adequada e clara (art. 6º, III) e impõe deveres de segurança e advertência de riscos (arts. 8º e 9º). Produto/serviço é defeituoso também por “informações insuficientes ou inadequadas sobre sua utilização e riscos” (arts. 12 e 14), e cláusulas que pretendam exonerar ou mitigar esse dever informacional são nulas de pleno direito (art. 51, I e IV). Daí resultam nulidade/ineficácia do “consentimento” colhido sob informação materialmente defeituosa, bem como a responsabilidade objetiva dos fornecedores. (BRASIL, 1990).

Além disso, três razões fundamentam essa afirmação: (i) *Viés como risco previsível e relevante exige informação clara*; (ii) *Transparência na oferta e vedação à informação enganosa/abusiva*; e (iii) *Arquiteturas de decisão e viés de automação*.

nudge, para ela, está dentro de uma abordagem da ação, para modificação do comportamento, denominada “sintonizar” (p. 354).

²⁹ O’Neil (2016) faz uma metáfora ao descrever como os modelos de IA são: “como deuses”, [...] são “opacos, seus mecanismos invisíveis a todos exceto os altos sacerdotes de seus domínios” (p. 6), o que impede a contestação e permite que punam “os pobres e oprimidos da sociedade enquanto enriquecia ainda mais os ricos” (pp. 6 e 7). Isso implica uma exploração de vulnerabilidades dos indivíduos que são impactados por esses sistemas sem transparência. Além disso, O’Neil (2016), define algoritmos como representações abstratas que “pegam o que sabemos e usa isso para prever respostas em situações variadas” e “guiam nossas decisões” (p. 20).

Viés como risco previsível e relevante exige informação clara: vieses documentados no Capítulo 4, e.g., uso de *proxies* que subpriorizam pacientes negros (algoritmo de gestão populacional que, ao usar custo como *proxy* de gravidade, subestimou a necessidade de cuidado em pacientes negros) (OBERMEYER *et al.*, 2019), desempenho desigual por subgrupos em IA dermatológica, que apresentou pior acurácia em peles escuras (CAZANI *et al.*, 2025) e vieses estruturais de *EHR* – Prontuário Eletrônico do Paciente (seleção, medição, *missingness*), que distorcem estimativas clínicas (GIANFRANCESCO *et al.*, 2018), são riscos previsíveis que impactam a decisão clínica e devem ser informados ao consumidor-paciente. O não esclarecimento de desempenho por subgrupo, limites de generalização e incertezas vicia o dever informacional e configura defeito do produto/serviço (arts. 12 e 14), tornando ineficaz o “consentimento”, e cláusulas que pretendam convalidar esse déficit informacional ou limitar direitos do consumidor são nulas de pleno direito (art. 51, I e IV) (BRASIL, 1990).

Transparência na oferta e vedação à informação enganosa/abusiva: o CDC exige que informações publicitárias/contratuais sejam corretas, claras, precisas e ostensivas (art. 31) e veda publicidade enganosa (art. 37). Omitir que o algoritmo não foi validado no contexto local, que apresenta pior desempenho em determinados subgrupos ou que a métrica adotada preserva desigualdades (WACHTER; MITTELSTADT; RUSSELL, 2021), converte o material informativo em informação enganosa/abusiva – hipótese de nulidade e de defeito informacional (arts. 12, 14, 31 e 37). A jurisprudência do STJ reforça que “consentimento genérico” (sem individualização de riscos) não convalida o déficit de informação nas relações de consumo; o dever de informar é basilar e evolutivo (STJ, 2018). (BRASIL, 1990; STJ, 2022).

Arquiteturas de decisão e viés de automação: omissão relevante e prática abusiva – Interfaces que ancoram a decisão clínica por *scores* opacos, *defaults* e *nudges*, num cenário de tendência humana à superconfiança na automação, induzem decisões sob pressupostos enviesados. A literatura descreve o *automation bias* (viés de automação) em ambientes clínicos (ROSBACH *et al.*, 2024). Em neurotecnologias, *feedbacks* enviesados em *BCI* e saliência de estímulos funcionam como “empurrões comportamentais” (*nudges*) que mascaram limitações do classificador (BARBERO; GROSSE-WENTRUP, 2010; LA FISCA *et al.*, 2022). Se tais mecanismos são conhecidos ou previsíveis e não forem previamente informados (com salvaguardas e *override* humano), há falha do dever informacional e potencial prática abusiva (art. 39, IV), atraindo nulidade de cláusulas limitativas (art. 51) e responsabilidade objetiva (arts. 12 e 14) (BRASIL, 1990).

Ademais, à luz de uma leitura analógica da LGPD, é possível sustentar a nulidade do consentimento do paciente, “caso as informações fornecidas ao titular tenham conteúdo enganoso ou abusivo ou não tenham sido apresentadas previamente com *transparência, de forma clara e inequívoca*” (BRASIL, 2018, Art. 9º, §1º).

Em termos práticos, o Consentimento Livre e Esclarecido deve expor desempenho por subgrupos, validade contextual, métricas de equidade adotadas e limites de generalização, sob pena de reduzir-se a formalidade destituída de tutela real (PRICE II, 2019; WACHTER; MITTELSTADT; RUSSELL, 2021).

5.4 NEURODISCRIMINAÇÃO E O CONSENTIMENTO COAGIDO

A ameaça final à autonomia emerge com o conceito de “neurodiscriminação”, o tratamento desigual de indivíduos com base em seus dados neurais (IENCA; IGNATIADIS, 2020). A “neurodiscriminação” foi cunhada para designar a discriminação que ocorre quando indivíduos ou grupos populacionais são injustamente prejudicados por inferências feitas a partir de dados neurais e/ou vieses em sua coleta e processamento (ANPD, 2025). O mero temor de ser estigmatizado ou discriminado com base em inferências sobre a própria atividade mental pode coagir um indivíduo a consentir com termos e condições que, em um cenário de livre escolha, ele prontamente recusaria.

Imagine-se um paciente com epilepsia refratária que utiliza um sistema de estimulação responsiva (*Responsive Neurostimulation – RNS*), destinado ao monitoramento e à estimulação contínuos para controle de crises, tecnologia cujo uso clínico tem demonstrado benefícios significativos na redução da frequência e gravidade dos episódios epiléticos (SCHULZE-BONHAGE *et al.*, 2023). No entanto, para manter o acesso ao pleno funcionamento do dispositivo – incluindo atualizações críticas de *software* e serviços de acompanhamento técnico indispensáveis –, o paciente é instado a fornecer seu CLE para a continuidade do tratamento, mesmo diante de incertezas sobre os efeitos de longo prazo e a possibilidade de modulação de aspectos sensíveis de sua atividade cerebral; uma tecnologia que, embora promissora, envolve novas formas de governança da mente (SIRBU *et al.*, 2025).

Diante do receio de que a recusa possa acarretar deterioração da eficácia terapêutica ou perda de acesso a melhorias essenciais, o paciente pode sentir-se pressionado a consentir, não por uma decisão plenamente autônoma, mas por uma coação indireta fundada na profunda assimetria de poder existente entre ele e o fornecedor da tecnologia (DAVTYAN; MESROPYAN, 2025). Nesse cenário, o CLE, pilar da autonomia na prática clínica, torna-se

hipervulnerável, aproximando-se de um ato formal, desprovido da liberdade decisória que deveria garantir (SIRBU *et al.*, 2025).

Nessas circunstâncias, o chamado “Consentimento Livre e Esclarecido” não é livre, nem esclarecido. O que há é uma coação indireta (condicionamento do pleno funcionamento/atualizações essenciais do dispositivo à aceitação de termos incertos) e informação materialmente inadequada sobre riscos, limites e impactos do tratamento de neurodados/IA.

No regime consumerista, isso configura prática abusiva (aproveitamento da hipervulnerabilidade do paciente e imposição de condição desproporcional) e, conforme o caso, até condicionamento vedado (*tying*) (arts. 39, IV, V e I), além de defeito do produto/serviço por informações insuficientes (arts. 12 e 14) e violação do direito básico à informação clara e ostensiva (art. 6º, III; art. 31) (BRASIL, 1990). Consequentemente, cláusulas que pretendam exonerar responsabilidade, impor renúncia de direitos ou transferir ao paciente o risco tecnológico são nulas de pleno direito (art. 51, I e IV; art. 25, §1º), e o “consentimento” obtido sob esse arranjo é ineficaz para convalidar o vício informacional, tornando-se nulo (como abordado anteriormente) e não afasta a responsabilidade objetiva da cadeia de fornecimento (arts. 12 e 14) (BRASIL, 1990).

Por fim, a tabela a seguir sistematiza como os desafios da opacidade e do viés impactam diretamente os pilares da autonomia e do Consentimento Livre e Esclarecido, utilizando os casos concretos como ilustração.

Tabela 4 – Matriz de Impacto da Opacidade e do Viés sobre a Autonomia e o Consentimento Livre e Esclarecido

| Desafio Sistêmico | Princípio Jurídico Afetado | Descrição do Impacto | Caso Ilustrativo (Capítulo 2) |
|------------------------------|--|--|---|
| Opacidade Algorítmica | Consentimento Esclarecido | O paciente não pode ser verdadeiramente “esclarecido” sobre um processo decisório que é intrinsecamente inexplicável (Risco Epistêmico). | Estudo PRAIM (Mamografia): o consentimento para biópsia se baseia na confiança em um alerta de IA cuja lógica específica é desconhecida. |
| “Dupla Caixa-Preta” | Autonomia e Integridade Pessoal | O consentimento é dado para uma intervenção cujos efeitos sobre a identidade e a mente são imprevisíveis e indecifráveis, minando o controle sobre o próprio “eu”. | DBS para Alzheimer: o paciente consente com uma modulação cerebral opaca que pode alterar sua personalidade de forma não prevista. |

| | | | |
|---|---|--|--|
| Viés de Representação | Autonomia (Igualdade na Escolha) | Pacientes de grupos sub-representados recebem informações de menor qualidade, exercendo sua autonomia a partir de uma base de escolha desigual e mais arriscada. | IA em Dermatologia (NEVO): a decisão de um paciente negro de não procurar um médico pode ser baseada em um resultado falsamente negativo devido ao viés pelo tom de pele. |
| Manipulação (Nudges/Dark Patterns) | Consentimento Livre | A arquitetura da interface explora vieses cognitivos para induzir o paciente/médico a uma escolha que não é genuinamente livre, mas direcionada. | BCI Invasiva Neuralink – Telepathy N1: um <i>dark pattern</i> pode influenciar pensamentos e comportamentos por meio de estímulos subliminares ou despercebidos, que operam fora da consciência humana, com “ <i>priming effects</i> ” e “ <i>framing effects</i> ”, por exemplo. |

Fonte: Elaborada pelo autor (2025).

6 SALVAGUARDAS JURÍDICAS AO CONSENTIMENTO: TRANSPARÊNCIA, EQUIDADE E *ACCOUNTABILITY*

As análises precedentes demonstram que a crise do Consentimento Livre e Esclarecido na saúde inteligente não se resolve com mais textos, avisos ou formulários extensos: trata-se de uma crise de paradigma. Partir da figura de um sujeito plenamente racional, capaz de ponderar dados complexos e decidir livre de pressões, revela-se ficcional diante da opacidade algorítmica, dos vieses sistêmicos e da vulnerabilidade cognitiva inerente às decisões em saúde digital (SOLAIMAN *et al.*, 2025; MELLO *et al.*, 2025).

Nesse sentido, “consertar” o consentimento nos moldes tradicionais é insuficiente: é preciso reconhecer seus limites e complementá-lo com instrumentos de governança/tutela estrutural – auditorias independentes, monitoramento contínuo, padrões institucionais de explicabilidade e mecanismos de responsabilização ao longo de todo o ciclo de vida da IA – de modo a proteger a autonomia do paciente em um nível mais fundamental. Trata-se de deslocar o foco da escolha individual para a configuração e o controle da tecnologia em si (HURLEY *et al.*, 2025; SAENZ *et al.*, 2024; SOLAIMAN *et al.*, 2025). Em outras palavras, a arquitetura dos sistemas digitais deve incorporar salvaguardas ético-jurídicas – raciocínio alinhado à máxima de Lawrence Lessig (2011) de que “o código é lei”, isto é, a própria arquitetura tecnológica funciona como forma de regulação e deve ser desenhada de acordo com valores sociais e jurídicos.

Importa, mais uma vez, distinguir o consentimento aqui abordado daquele previsto na LGPD. O consentimento como base legal na Lei Geral de Proteção de Dados refere-se à autorização do titular para tratamento de *dados pessoais* (BRASIL, 2018). Já o Consentimento Livre e Esclarecido (CLE) na área da saúde – de caráter ético-jurídico – diz respeito à anuência do paciente a intervenções médicas ou ao uso de tecnologias em seu corpo e em sua saúde, constituindo expressão direta de sua autonomia e dignidade. Este último envolve princípios adicionais de beneficência, não maleficência e justiça, tendo por finalidade salvaguardar a integridade físico-mental da pessoa (BEAUCHAMP; CHILDRESS, 2013). Carlos Henrique Félix Dantas e Régis Gurgel do Amaral Jereissati (2024), inclusive, defendem a necessidade de um “duplo consentimento” (para interferência corporal e tratamento de dados neurais) para neurotecnologias. Isso se aplica, *mutatis mutandis*, para as ferramentas que utilizam IA na área médica também. Ou seja, são consentimentos distintos.

É um engano, portanto, presumir que a conformidade à LGPD (por exemplo, obter consentimento para uso de dados de saúde) esgote as preocupações com a autonomia do

paciente diante de IA e neurotecnologias – pelo contrário, o consentimento bioético é mais exigente e abrangente, requerendo garantias substanciais de compreensão, voluntariedade e respeito à vontade do paciente. Confundir regimes degenera garantias: a revogação do consentimento de dados não implica, por si, interrupção segura de terapêutica já iniciada; inversamente, o CLE para um implante não autoriza usos ulteriores de dados neurais para pesquisa sem base legal e ética específica (ZUBOFF, 2021; MOLNÁR-GÁBOR, 2024; ANPD, 2025).

Portanto, faz-se necessária uma abordagem multifacetada para superar os atuais limites do CLE, combinando maior transparência técnica, melhorias no processo de consentimento em si (tornando-o mais equitativo e centrado no paciente) e salvaguardas adicionais de natureza regulatória e institucional (AITH, 2024). Ademais, essa análise focalizará exclusivamente as salvaguardas jurídicas destinadas a complementar o Consentimento Livre e Esclarecido do paciente, no contexto do uso de inteligência artificial e neurotecnologias em diagnósticos e tratamentos médicos. O caráter desse consentimento será destacado em sua dimensão ético-jurídica ligada à autonomia do paciente, evidenciando a necessidade de proteções específicas de transparência, equidade e *accountability* (NOGAROLI, 2023; DANTAS; JEREISSATI, 2024; AITH, 2024; NEGRI *et al.*, 2025).

6.1 TRANSPARÊNCIA E EXPLICABILIDADE COMO PRINCÍPIOS FUNDAMENTAIS PARA APOIAR A TOMADA DE DECISÃO INFORMADA

Para enfrentar a opacidade das tecnologias inteligentes, a transparência – em suas várias dimensões – desponta como princípio ético e legal central. No contexto da IA, transparência, frequentemente, se traduz em explicabilidade: a capacidade de explicar, em termos compreensíveis, como um algoritmo funciona e toma decisões. Regulações internacionais já caminham nesse sentido (MITTELSTADT, 2021; DANTAS; JEREISSATI, 2024).

O recém-aprovado Regulamento Europeu de IA (*AI Act*) coloca a explicabilidade no centro das preocupações para sistemas de alto risco, exigindo que decisões assistidas por IA sejam explicadas de forma clara e acessível aos usuários finais (BOSELLI, 2024; UNIÃO EUROPEIA, 2024). Ou seja, não basta que apenas engenheiros entendam o modelo; é imperativo que operadores e impactados – médicos e pacientes, no caso da saúde – possam apreender a lógica e os critérios relevantes do algoritmo (BOSELLI, 2024; NOGAROLI, 2023).

Essa tendência também se reflete no Brasil: o Projeto de Lei 2.338/2023, em tramitação, elenca a explicabilidade como um dos princípios para uma IA confiável, prevendo o dever de

fornecer informações gerais sobre o funcionamento do modelo, explicitando sua lógica e critérios decisórios, especialmente para sistemas de alto risco na saúde (BRASIL, 2023; BOSELLI, 2024; AITH, 2024).

A busca por transparência, entretanto, não implica expor código-fonte complexo ao leigo, mas sim traduzir o essencial em linguagem acessível (DANTAS; NOGAROLI, 2020). Como observa Gisele Boselli (2024), a explicabilidade deve se estender aos profissionais que operarão a IA e aos destinatários finais da tecnologia, para que compreendam suas potencialidades e limitações. Por exemplo, ao utilizar um algoritmo de suporte diagnóstico, o médico precisa ser capaz de questioná-lo apropriadamente e entender suas restrições. Só assim poderá esclarecer adequadamente o paciente sobre os riscos envolvidos, permitindo-lhe consentir ou recusar o uso da ferramenta de forma realmente informada. Isso abrange, dentre outros pontos, comunicar a possibilidade de viés algorítmico – caso o modelo tenha sido treinado em dados pouco diversos – e seus potenciais impactos em minorias ou em perfis específicos de pacientes. Kate Crawford (2021) adverte que a opacidade dos sistemas de IA muitas vezes encobre assimetrias de poder e impactos desproporcionais sobre grupos vulneráveis; portanto, trazer à luz *como e em que condições* o algoritmo opera é essencial para garantir equidade no cuidado de saúde (CRAWFORD, 2021). Em suma, a transparência técnica deve ser convertida em informação significativa para a tomada de decisão clínica.

Desse imperativo emergem dois desdobramentos importantes: (i) *O reconhecimento de um “direito à explicação” do paciente*; e (ii) *Revisão do processo de Consentimento Livre e Esclarecido*.

Reconhecimento de um “direito à explicação” do paciente: doutrinadores defendem que, com a evolução tecnológica, o clássico direito à informação adequada inclui agora o direito de saber como a nova ferramenta opera e por que ela recomenda determinado resultado (FALEIROS Jr.; NOGAROLI, 2021; NOGAROLI, 2023). O *AI Act* europeu e o próprio PL 2338/2023, prevêm isso expressamente (BRASIL, 2023; UNIÃO EUROPEIA, 2024). Isso envolve explicar os objetivos do algoritmo, suas vantagens esperadas, potenciais custos e riscos adicionais, bem como as alternativas disponíveis. A informação deixa de ser meramente descritiva e passa a ser justificativa – o paciente tem direito não só de ouvir que “um algoritmo recomendou tal terapia”, mas de entender minimamente qual lógica ou evidência levou a essa recomendação. Um exemplo prático dado na doutrina é o médico explicar ao paciente que, embora um tratamento inicial tenha falhado, a IA (como o *Watson for Oncology*) fará um cruzamento de dados para apresentar novas propostas de tratamento, informando o grau de falibilidade e os riscos da ferramenta (NOGAROLI, 2023).

Revisão do processo de Consentimento Livre e Esclarecido: correlato ao anterior, impõe-se revisar o próprio processo de Consentimento Livre e Esclarecido, convertendo-o de um evento pontual (a assinatura de um formulário padrão) em um verdadeiro diálogo esclarecedor entre médico e paciente (CFM, 2016; FALEIROS Jr.; NOGAROLI, 2021). Não basta o criador ou o usuário de um algoritmo dizer “*the algorithm did it*” – isto é, atribuir toda decisão ou erro à “caixa-preta” tecnológica (DANTAS; NOGAROLI, 2020; NOGAROLI, 2023). Somente por meio de uma interação dialógica – em que o paciente pode fazer perguntas e obter justificativas em linguagem simples – é possível minimizar o impacto da opacidade e dos vieses tecnológicos, além de se superar uma outra “caixa-preta comunicativa”, que hoje caracteriza muitos termos de consentimento longos e inacessíveis. A explicabilidade, como princípio, encontra respaldo, portanto, na responsabilização profissional (NALIN; NOGAROLI, 2020; NOGAROLI, 2023).

É importante notar que a transparência não se limita à interface paciente-médico, mas permeia todo o ciclo de vida da tecnologia. Desenvolvedores de IA em saúde são instados a seguir diretrizes éticas desde a concepção dos modelos, garantindo qualidade e diversidade nos dados de treinamento (SOLAIMAN *et al.*, 2025). Embora eliminar completamente vieses possa ser inatingível, boas práticas de ciência de dados podem minimizá-los significativamente, resultando em decisões automatizadas mais justas e equânimes para diferentes grupos (BOSELLI, 2024). A promoção da diversidade – de gênero, origem cultural e étnica – tanto nos dados para programação algorítmica, quanto na equipe de criação da tecnologia, é essencial para desenvolver a IA e entender seu impacto ético, social, jurídico e econômico (NOGAROLI, 2023).

Essa perspectiva reforça a dimensão de equidade como objetivo fundamental: a tecnologia deve ser transparente não apenas para ser compreendida, mas também para permitir identificação e correção de discriminações embutidas (O’NEIL, 2016). A matemática algorítmica, como alerta Cathy O’Neil (2016), pode se tornar uma arma de injustiça em massa se não houver mecanismos de escrutínio; portanto, transparência e explicabilidade servem também para expor e mitigar viés e discriminação, promovendo um cuidado de saúde mais equitativo.

Adicionalmente, a clareza sobre o funcionamento da IA favorece a *accountability* (prestação de contas) (SMOLENAARS; GONÇALVES, 2021). Frank Pasquale (2015) destaca que a sociedade da “caixa-preta” algorítmica subverte a responsabilização, uma vez que decisões automatizadas ininteligíveis dificultam atribuir culpa ou corrigir erros sistêmicos. Ao exigir explicações e documentações

auditáveis, transparência e *accountability* andam de mãos dadas: abre-se caminho para avaliar se o algoritmo respeita padrões de segurança e ética, bem como para responsabilizar atores em caso de violações.

Em síntese, transparência/explicabilidade, equidade e *accountability* formam um tripé interdependente que fortalece o pilar “esclarecido” do consentimento na era da IA. Esse tripé transforma princípios abstratos em obrigações concretas – *e.g.* tornar dados de treinamento e métricas de desempenho disponíveis para escrutínio, reportar vieses descobertos – de modo que a autonomia do paciente não seja socavada por segredos tecnológicos.

Por fim, não é demais reafirmar que a legislação europeia já consolidou a transparência e a explicabilidade como obrigações legais para sistemas de IA de alto risco, que incluem a saúde. O GDPR, por exemplo, estabelece o “princípio da cognoscibilidade”, segundo o qual o titular dos dados tem o direito de saber da existência de decisões automatizadas e de receber informações significativas sobre a lógica utilizada para que possa compreendê-la. O novo *AI Act*, por sua vez, impõe que os sistemas de IA, considerados como de alto risco, devem ser projetados para que sua operação seja suficientemente transparente, permitindo aos usuários interpretar os resultados e usá-los adequadamente (FILIPPIS *et al.*, 2024; BOTTINI *et al.*, 2025).

6.2 FORTALECENDO O PROCESSO DE CONSENTIMENTO: DO FORMULÁRIO AO DIÁLOGO HUMANIZADO

Como mencionado anteriormente, além de aprimorar a informação técnica fornecida, é preciso transformar a dinâmica do consentimento. Historicamente, o CLE muitas vezes se reduz a um formulário padronizado, carregado de jargão, entregue para assinatura pouco antes de um procedimento. Na era das IAs e neurotecnologias, tal modelo se mostra ainda mais inadequado. Fortalecer o processo de consentimento implica evoluir de um ritual burocrático para um processo comunicativo genuíno, no qual o paciente possa realmente compreender e deliberar sobre o uso da tecnologia em seu cuidado.

Pesquisas recentes propõem reformular os materiais e práticas de consentimento com linguagem simples e recursos visuais interativos. Por exemplo, informações em formato de diagramas, fluxogramas, animações, recursos interativos ou vídeos explicativos podem ilustrar de forma acessível o papel da IA no diagnóstico ou tratamento (CHAR *et al.*, 2018; GUNDLACK *et al.*, 2025; SOARES, 2025). Jia-Yue Si *et al.* (2025) noticiam que os estudos sugerem que tecnologias como Realidade Virtual (VR) e Realidade Aumentada (AR) podem

melhorar a compreensão do paciente, permitindo visualizar em 3D os planos de tratamento e simulações cirúrgicas.

Ademais, em vez de o paciente apenas aderir ou recusar um “pacote fechado” de intervenções, modelos de Consentimento Modular/Granular permitiriam que ele autorizasse determinados usos da tecnologia e negasse outros, incrementando seu controle sobre a trajetória terapêutica (CHAR *et al.*, 2018; KAYE *et al.*, 2015; NOGAROLI, 2023; SOARES, 2025; SEGAL; HOD, 2025). Essa abordagem modular se coaduna com a ideia de um paciente mais ativo e engajado, capaz de personalizar seu consentimento conforme suas preferências e valores, estando, inclusive, em conformidade com o CFM (2016).

Portanto, a concepção do consentimento deve migrar de um ato único e estático para se tornar um processo dinâmico e contínuo (CFM, 2016; THE SOCIABLE, 2025). Para intervenções de longa duração ou tecnologias de aprendizado contínuo (como certos algoritmos de *machine learning* ou neurodispositivos adaptativos), o consentimento inicial pode tornar-se obsoleto à luz de novas informações. A UNESCO, em seu projeto de Recomendação sobre Ética da Neurotecnologia, defende que os procedimentos de consentimento sejam iterativos e atualizáveis, permitindo ao indivíduo revisitar e reconfirmar sua decisão, à medida que novos dados sobre riscos e benefícios forem emergindo (UNESCO, 2025). Essa abordagem, frequentemente denominada de Consentimento Dinâmico, reconhece que a autonomia se exerce ao longo do tempo. Não basta obter um “sim” do paciente no início; é preciso garantir um canal aberto para diálogo e revisão conforme a tecnologia evolui ou se revelam efeitos inesperados (THE SOCIABLE, 2025; UNESCO 2025). Em pesquisa biomédica, o Consentimento Dinâmico já é explorado como solução para acompanhar estudos de longa duração, e suas lições podem ser transpostas ao contexto clínico de IA (KAYE *et al.*, 2015).

Um elemento indispensável nesse fortalecimento do consentimento é a comunicação franca sobre o uso da IA/Neurotecnologia. Instituições médicas devem implementar políticas de *disclosure*, segundo as quais o paciente sempre seja informado se uma decisão diagnóstica ou terapêutica envolveu um algoritmo (FALEIROS Jr.; NOGAROLI, 2021). Como visto, ocultar essa informação, ainda que com boa intenção, priva o paciente de seu direito fundamental à autodeterminação decisória (BEAUCHAMP; CHILDRESS, 2013). Logo, o médico deve introduzir o assunto explicando, por exemplo: “*Para auxiliar no seu caso, usamos um programa de IA que analisa seus exames e sugere possibilidades. Gostaria de lhe explicar o que isso significa e ouvir suas dúvidas antes de prosseguirmos*”. Tal postura reconhece o

paciente como parte ativa na decisão e promove confiança, pois ele percebe que nada lhe está sendo escondido (RATTI *et al.*, 2025).

Contudo, informar a presença da IA é apenas o primeiro passo – é preciso explicar em termos compreensíveis como ela atua. Os acadêmicos José Faleiros Jr. e Rafaella Nogaroli (2021) denominam de “opacidade explicativa” a situação em que o médico até revela ter usado um algoritmo, mas não consegue elucidar seu funcionamento de forma acessível. Evitar essa opacidade exige que a explicação seja adaptada ao nível de compreensão de cada paciente. Por exemplo, para um paciente leigo, o médico pode explicar que “*uma inteligência artificial analisou as imagens procurando padrões semelhantes aos de milhares de outros casos já estudados, e com base nisso indicou tal diagnóstico; porém, isso é um apoio ao meu olhar clínico, não uma certeza absoluta*”. Da mesma forma, deve-se esclarecer qual o papel exato da IA e do humano: o paciente precisa saber se o algoritmo apenas sugere uma opção ao médico ou se realmente executa parte do procedimento (SCHIFF; BORENSTEIN, 2019). No caso de uma cirurgia robótica guiada por IA, como ilustrado em discussão bioética, recomenda-se que o cirurgião detalhe *quais etapas estarão sob controle da máquina e quais sob sua supervisão direta*, bem como os potenciais riscos de falhas humanas *vs.* falhas da IA em cada etapa (SCHIFF; BORENSTEIN, 2019). Schiff e Borenstein (2019) afirmam, acertadamente, que tal delineamento de papéis evita mal-entendidos como o do paciente que exclama “*não quero um robô operando; quero você*”, sem saber que o robô é apenas uma ferramenta e que o cirurgião permanece no comando.

Nesse contexto, uma ferramenta emergente é a *Explanatory AI*, que visa resolver a questão da opacidade explicativa levantada acima. Em pesquisa publicada por Christian Meske *et al.* (2025), no dia 8 de agosto de 2025 – intitulada *From Explainable to Explanatory Artificial Intelligence: Toward a New Paradigm for Human-Centered Explanations through Generative AI* (“Da inteligência artificial explicável à explicativa: rumo a um novo paradigma para explicações centradas no ser humano por meio da IA generativa”) –, os autores demonstram como a Inteligência Artificial Explicativa (*Explanatory AI*) emerge como uma salvaguarda crucial para redefinir a compreensão do usuário, indo além da transparência algorítmica tradicional, para priorizar as explicações centradas no ser humano e a geração de sentido no contexto. Enquanto as abordagens tradicionais de IA Explicável (*XAI*), como *SHAP* ou *LIME*, já amplamente debatida, focam em revelar “Como a IA decidiu?” para propósitos diagnósticos e técnicos (como depuração de modelos e conformidade regulatória), a IA Explicativa aborda a questão “Por que isso faz sentido?” para apoiar a compreensão humana e a tomada de decisões em contextos sociotécnicos do mundo real (MESKE *et al.*, 2025).

Para que essas melhorias sejam efetivas, é crucial investir na capacitação dos profissionais de saúde em IA e comunicação. Médicos e equipe devem ser treinados não só no uso técnico das novas ferramentas, mas em como traduzi-las pedagogicamente ao paciente. Conforme Danton S. Char *et al.* (2018) argumentam, médicos usuários de aprendizado de máquina devem se educar sobre a construção e limitações desses sistemas; a ignorância quanto ao funcionamento deles pode levar a resultados eticamente problemáticos. Universidades de vanguarda já incorporam tais conteúdos: a Universidade de Toronto, por exemplo, criou certificações em IA em saúde para estudantes de Medicina, reconhecendo a necessidade de formação interdisciplinar desde a graduação (FALEIROS Jr.; NOGAROLI, 2021). Médicos bem-informados sentir-se-ão mais seguros para abordar o tema com pacientes, o que por sua vez melhora a qualidade do consentimento. Ademais, o desenvolvimento de literacia digital em saúde na população geral também contribui – pacientes familiarizados minimamente com conceitos de IA estarão mais aptos a fazer perguntas pertinentes e entender explicações. Por isso, algumas propostas advogam campanhas educativas e recursos informativos públicos sobre IA médica, para empoderar os pacientes na era digital (GUNDLACK *et al.*, 2025).

Outro aspecto essencial é a garantia de momentos de diálogo verdadeiro entre médico e paciente, nos quais haja espaço para perguntas e respostas em linguagem adaptada ao contexto sociocultural do paciente (O'NEIL, 2016; DANTAS; NOGAROLI, 2020). Para isso, é indispensável investir em programas contínuos de capacitação de médicos comunicadores, a fim de que estejam aptos a explicar não apenas o funcionamento básico da IA ou da neurotecnologia empregada, mas também suas limitações e vieses potenciais (CHAR *et al.*, 2018; NOGAROLI, 2023).

Tais medidas humanizam o consentimento, transformando-o em uma extensão da relação de confiança entre paciente e profissional de saúde (TOPOL, 2019; SOARES, 2025). Assim, mesmo diante de ferramentas artificiais altamente complexas, o consentimento deixa de ser mera formalidade e passa a ser um processo vivo e contínuo, sustentado pela transparência e pela corresponsabilidade.

É importante fazer uma ressalva no tocante às neurotecnologias. Como analisado no Capítulo 5, o uso das *BCIs* e *DBSs* podem comprometer o consentimento, por todos os fundamentos ali demonstrados, especialmente os analisados no item 5.2.2, como limitação da integridade física e psicológica que pode ocorrer no uso dessas *neurotechs*. Nesses casos, o consentimento não pode se limitar à autorização pontual para o uso da tecnologia; deve incluir mecanismos contínuos de supervisão, revisão e possibilidade de retirada/recusa do

consentimento, protegendo a autodeterminação em face da dupla opacidade e da imprevisibilidade tecnológica.

Eric Topol (2019) argumenta que a integração da IA na medicina será verdadeiramente benéfica apenas se vier acompanhada de uma medicina mais humanizada, em que a tecnologia liberte tempo do profissional para conversar com o paciente em vez de afastá-lo. Nesse sentido, o reforço do processo de consentimento como encontro dialógico é coerente com a visão de que a IA deve empoderar e não alienar o paciente (TOPOL, 2019). Stuart Russell e Peter Norvig (2021), ao discutirem a necessidade de um *design* de IA centrado no ser humano, enfatizam que sistemas inteligentes devem se alinhar aos valores e preferências das pessoas. Transposto isso à esfera clínica, significa que a vontade do paciente, baseada em informações transparentes e claras, deve guiar o uso da IA, e não o contrário.

Resumindo, reconstruir o pilar “esclarecido” do consentimento passa tanto pela qualidade da informação fornecida (transparência e explicabilidade), quanto pela forma do engajamento com o paciente (diálogo contínuo, linguagem clara, oportunidade de revisão). Essas medidas atuam em sinergia: um paciente bem-informado e continuamente engajado está muito mais apto a exercer sua autonomia de maneira efetiva. No próximo item, abordaremos o pilar complementar – o consentimento “livre” –, analisando como garantir que a escolha do paciente seja genuinamente voluntária, sem coerção ou surpresas algorítmicas, o que envolve em grande parte assegurar o direito de recusa e a existência de alternativas.

6.3 DIREITO AO CONSENTIMENTO LIVRE E ESCLARECIDO E O DIREITO DE RECUSA

Como destacado acima, a análise do consentimento, na era da saúde inteligente, desdobra-se em duas dimensões críticas: o consentimento para o uso da própria tecnologia de Inteligência Artificial, ou de neurotecnologia, no cuidado e o consentimento para um curso de ação específico cujo processo decisório foi mediado por essa tecnologia (WORLD HEALTH ORGANIZATION, 2021; YUSTE *et al.*, 2017). Contudo, a verdadeira efetivação do CLE só ocorre quando ele é acompanhado por seu corolário indispensável: *o direito à recusa*. Ambos os direitos são, portanto, indissociáveis e fundamentais para a preservação da autodeterminação do paciente (ALBUQUERQUE, 2025).

Vieses, heurísticas e estratégias de manipulação (como *nudges* e *dark patterns*), de certos modelos de IA, associada à complexidade técnica de neurodispositivos implantáveis ou não invasivos, impõe barreiras adicionais ao Consentimento Livre e Esclarecido (MITTELSTADT,

2021), afetando o pilar “livre” do Consentimento. Quando o paciente não compreende plenamente a natureza e as implicações do uso dessas tecnologias, a relação de confiança com o profissional de saúde pode ser comprometida, reduzindo a adesão terapêutica e prejudicando os desfechos clínicos (AMANN *et al.*, 2020). Por esse motivo, a previsão do uso de IA ou neurodispositivos – seja no diagnóstico, no prognóstico ou no plano terapêutico – deve constar expressamente no processo de Consentimento Livre e Esclarecido (WORLD HEALTH ORGANIZATION, 2021; KOLFSCHOOTEN, 2022).

Um pressuposto fundamental da liberdade de consentir é a existência de alternativas viáveis (DANTAS; NOGAROLI, 2020). A Recomendação CFM Nº 1/2016 determina que o esclarecimento deve ser claro e pertinente sobre os métodos alternativos e as consequências da não realização do procedimento. Consentir com o uso de uma IA ou neurotecnologia em seu diagnóstico/tratamento só é uma escolha autônoma se o paciente puder optar por não utilizá-la sem ser prejudicado por isso (UNIÃO EUROPEIA, 2016; UNESCO, 2005). Em outras palavras, deve haver um direito efetivo de recusa: o paciente deve ter a opção de não ser submetido a decisões baseadas exclusivamente em processos automatizados e de solicitar que seu caso seja conduzido por métodos tradicionais (humanos), sem que isso implique perda de acesso a cuidados de qualidade ou aumento de custos (CFM, 2019; SOARES, 2025). Um consentimento “livre” é ilusório, se recusar o algoritmo significar não receber nenhuma alternativa de tratamento ou receber um atendimento inferior (CLOTET, 2009). Assim, a liberdade no consentimento requer tanto o poder de dizer não, quanto a garantia de que o “não” será respeitado sem retaliação ou prejuízos ao paciente.

O direito de recusa, neste contexto, assume uma dimensão negativa – a de não ser submetido a decisões baseadas exclusivamente em processos automatizados ou mediadas por neurodispositivos – e uma dimensão positiva, que impõe ao prestador de serviços de saúde o dever de oferecer alternativas equivalentes sem onerar o paciente, tais como diagnósticos e planos terapêuticos conduzidos diretamente por profissionais de saúde (PLOUG; HOLM, 2020). Embora se possa argumentar que, em certos equipamentos, a IA já seja uma parte integral e indissociável – como em eletrocardiógrafos modernos (PLOUG; HOLM, 2020) –, o princípio deve ser mantido.

O arcabouço regulatório europeu compreende profundamente essa interdependência entre consentimento e alternativas. O Artigo 22 do GDPR (Regulamento Geral de Proteção de Dados da UE) consagra o princípio de que ninguém deve ficar sujeito exclusivamente a uma decisão automatizada que produza efeitos significativos, assegurando ao indivíduo o direito de obter intervenção humana, de expressar seu ponto de vista e de contestar a decisão automatizada

(UNIÃO EUROPEIA, 2016). Trata-se, em essência, do “princípio da não exclusividade da decisão algorítmica”. Já o *AI Act*, em seu Artigo 14, reforça e operacionaliza essa garantia ao impor que sistemas de IA de alto risco sejam projetados para permitir uma supervisão humana eficaz (UNIÃO EUROPEIA, 2024). A combinação desses dispositivos revela uma arquitetura jurídica coesa: a transparência exigida pelo *AI Act* torna a decisão algorítmica compreensível, enquanto a supervisão humana garantida pelo GDPR, e pelo próprio *AI Act*, a torna contestável. Juntas, explicabilidade e intervenção humana formam as duas faces da mesma moeda da autonomia efetiva (FILIPPIS *et al.*, 2024).

O documento da *MDCG (Medical Device Coordination Group)*³⁰ afirma que o *AI Act* “sets a legal obligation on the manufacturer to design and develop AI systems, including high-risk MDAI, with appropriate human oversight mechanisms” (“estabelece uma obrigação legal para o fabricante projetar e desenvolver sistemas de IA, incluindo MDAI de alto risco, com mecanismos adequados de supervisão humana”). Ele também destaca que os requisitos de supervisão humana (Artigo 14, *AI Act*) e transparência (Artigo 13, *AI Act*) “reinforce the need to provide deployers and affected persons with sufficient information to understand the system’s capabilities, limitations, and potential risks” (“reforçam a necessidade de fornecer aos responsáveis pela implantação e às pessoas afetadas informações suficientes para compreender as capacidades, limitações e riscos potenciais do sistema”) (AIB/MDCG, 2025).

A Resolução CREMERS nº 6/2025 é explícita ao afirmar que a Inteligência Artificial “não substitui a decisão do médico” e deve ser aplicada como ferramenta, com a primazia e o protagonismo dos médicos e pacientes, norteadas pela supervisão humana e pelo consentimento informado. O médico deve esclarecer sobre o uso da IA e obter o Termo de Consentimento Livre e Esclarecido (TCLE) quando a IA for “ferramenta determinante” (CREMERS, 2025).

No contexto médico, isso significa, portanto, que qualquer recomendação ou decisão assistida por IA deve ser apresentada de forma que o paciente possa questioná-la e, se desejar, recusá-la em favor de uma alternativa humana (KOLFSCHOOTEN, 2022). O paciente tem o direito de saber, por exemplo, que um determinado diagnóstico foi sugerido por um algoritmo e que ele pode optar por uma segunda opinião de um especialista humano antes de prosseguir (PLOUG; HOLM, 2023). Para que esse direito de recusa seja exercido de forma plena, afirmam esses últimos autores, é indispensável que o paciente receba informações claras sobre: (i) a identificação do sistema de IA ou neurodispositivo utilizado; (ii) seus potenciais vieses

³⁰ AIB/MDCG. Joint Artificial Intelligence Board and Medical Device Coordination Group Document. AIB 2025-1 / MDCG 2025-6 - FAQ on Interplay between the Medical Devices Regulation & In vitro Diagnostic Medical Devices Regulation and the Artificial Intelligence Act (June 2025). Disponível em: https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf. Acesso em: 7 jul. 2025.

conhecidos; (iii) *suas métricas de desempenho* (como acurácia geral e em subgrupos populacionais); e (iv) *a divisão de responsabilidades* entre a tecnologia e o profissional humano, no resultado final (PLOUG; HOLM, 2023). Munido dessas informações, o paciente pode avaliar criticamente o valor agregado (ou os riscos) da ferramenta e decidir se consente com seu uso.

Fernando Aith (2024) afirma que a Lei Geral de Proteção de Dados já concede aos pacientes o direito de revisar decisões automatizadas que os afetam significativamente, exigindo uma revisão humana dessa decisão. No entanto, afirma o autor, a complexidade de algoritmos de “caixa-preta” dificulta a explicabilidade (AITH, 2024). O Projeto de Lei brasileiro (PL 2.338/2023) também segue uma linha similar ao prever que, *verbis*:

“A supervisão humana de sistemas de inteligência artificial de alto risco buscará prevenir ou minimizar os riscos [...] viabilizando que as pessoas responsáveis pela supervisão humana possam [...]” compreender, interpretar, decidir e intervir nos sistemas de IA (BRASIL, 2023, Art. 20, Parágrafo único).

Ademais, a recusa deve ser respeitada e operacionalizada pelos sistemas de saúde. Isso significa, por exemplo, que se um paciente optar por não usar um sistema de IA para triagem, o hospital deve ter protocolos para que ele seja avaliado por método alternativo (um especialista, outro exame etc.) sem atrasos indevidos. Limitar ou desestimular a recusa – seja por indisponibilidade de alternativas ou por pressionar o paciente a aceitar a IA sob pena de “ficar sem atendimento” – seria uma afronta direta ao direito à autonomia. O ônus de providenciar alternativas não pode recair sobre o paciente; cabe à estrutura de saúde oferecê-las. Ploug e Holm (2020) destacam que, em certos dispositivos modernos, a IA já é parte tão integrada que pode parecer impossível evitá-la (por exemplo, marcapassos ou desfibriladores implantáveis com algoritmos internos). Nesses casos, defendem os autores, deve-se ao menos assegurar que a decisão de adotar tal tecnologia seja voluntária e que o paciente entenda que não há opção totalmente “*não-IA*” equivalente (PLOUG; HOLM, 2020). Ainda assim, o princípio geral se mantém: sempre que viável, preservar *uma rota humanizada paralela*.

A manipulação algorítmica é outra preocupação correlata e ela afeta diretamente o pilar “livre” do Consentimento. Os sistemas de IA podem influenciar pensamentos e comportamentos por meio de estímulos subliminares ou despercebidos, que operam fora da consciência humana (KAHNEMAN, 2012). Kahneman explica que o Sistema 1 (automático) opera rapidamente com pouco ou nenhum esforço e pode ser influenciado por “*priming effects*” e “*framing effects*” (efeitos de enquadramento), nos quais a formulação de resultados e

contingências pode influenciar a decisão sem que a pessoa tenha consciência disso. Esses efeitos podem ser explorados deliberadamente para manipular a atratividade das opções.

Portanto, sistemas de IA podem influenciar escolhas de modo sutil, por exemplo, personalizando recomendações ou enfatizando certos riscos de forma persuasiva (o equivalente digital de *nudges* ou até *dark patterns* no contexto médico). Isso pode corroer a liberdade do consentimento sem que o paciente perceba, pois ele pode ser direcionado a aceitar a tecnologia acreditando ser sua vontade, quando, na verdade, sua decisão foi calibrada por *designs* de interface ou pela forma como as opções lhe foram apresentadas. Assim, a equidade no consentimento requer também vigilância contra táticas de persuasão oculta.

O paciente deve receber as informações de modo equilibrado, sem vieses de apresentação que favoreçam a decisão desejada pelo provedor ou desenvolvedor da tecnologia (MITTELSTADT, 2021). Transparência aqui significa, também, honestidade na comunicação: admitir incertezas, mencionar *pros e contras* de se usar a IA, e não apenas seus benefícios. Um consentimento livre pressupõe que o paciente esteja decidindo sem ter sua agência minada por assimetrias informacionais ou técnicas de influência indevida.

Resumindo, para garantir o pilar do consentimento “livre” no ambiente da saúde digital, é preciso assegurar: (i) *O direito de recusa e a disponibilidade de alternativas equivalentes*; (ii) *A supervisão humana e possibilidade de contestação de decisões automatizadas*; (iii) *A neutralidade e equilíbrio na forma de apresentar opções ao paciente (evitando manipulação)*; e (iv) *A proteção de grupos vulneráveis, cujo consentimento pode ser mais facilmente comprometido*. Somente com essas salvaguardas será possível dizer que o “sim” ou “não” do paciente, à utilização de uma IA/neurotecnologia, é, de fato, uma expressão autêntica de sua vontade e não um formalismo vazio. Como bem sintetiza I. Glenn Cohen (2019), a autonomia do paciente em contextos de alta tecnologia depende de escolhas informadas e voluntárias, mas essas escolhas só existem se construirmos um ambiente que as sustente – ambiente este moldado pelo Direito. Assim, tal qual Daniel Solove (2013) denunciou no campo da privacidade o “dilema do consentimento” – em que sobrecarregar indivíduos com escolhas complexas sem apoio estrutural os leva a consentir contra seus próprios interesses –, no campo da saúde digital devemos evitar jogar todo o peso da proteção nos ombros do paciente. Urge, pois, uma abordagem em que a liberdade de consentir seja amparada por estruturas que a tornem exequível e significativa.

6.4 SALVAGUARDAS REGULATÓRIAS E ÉTICAS ADICIONAIS

Apesar de todos os aprimoramentos possíveis no processo de consentimento (informação mais clara, diálogo aprimorado, direito de recusa garantido), não se pode delegar exclusivamente ao paciente a responsabilidade de se proteger dos riscos complexos e muitas vezes ocultos da IA e das neurotecnologias (BEAUCHAMP; CHILDRESS, 2013; COHEN, 2019). Superar o consentimento como único escudo implica implementar mecanismos complementares de proteção, de natureza regulatória, institucional e ética, que atuem antes, durante e após o momento do consentimento (RATTI *et al.*, 2025; SOLAIMAN *et al.*, 2025). Em outras palavras, a proteção efetiva do paciente não pode repousar unicamente em sua capacidade individual de consentir, mas deve emergir de um ecossistema de salvaguardas distribuídas, integradas e verificáveis.

Delinearemos a seguir as principais camadas dessa arquitetura de governança proativa.

6.4.1 Avaliação rigorosa e certificação de algoritmos

Uma solução-chave é garantir que os sistemas de IA e Neurotecnologias, usados em saúde, passem por escrutínio independente quanto à segurança, eficácia e equidade antes de sua adoção clínica. Isso pode tomar a forma de aprovação por agências reguladoras (análoga à aprovação de fármacos) ou de certificações técnico-éticas emitidas por órgãos competentes. Por exemplo, órgãos como a FDA, nos EUA, já avaliam e autorizam dispositivos algorítmicos de diagnóstico com base em evidências de desempenho clínico e segurança, assim como a ANVISA aqui no Brasil, como será demonstrado a seguir.

Entre 2021 e 2025, a FDA emitiu diretrizes e recomendações específicas para dispositivos de IA e Neurotecnologias, consolidando princípios de Boas Práticas de Aprendizado de Máquina (*GMLP*) que orientam o desenvolvimento de algoritmos seguros e de alta qualidade (FDA, 2021; FDA, 2023; FDA, 2025). Essas diretrizes enfatizam a necessidade de dados de treinamento representativos e de qualidade, de validação clínica robusta e de transparência no *design* do modelo (como se observa da *Artificial Intelligence/Machine Learning-Enabled – Software as a Medical Device (SaMD)* e *Artificial Intelligence-Enabled Device Software Functions: Lifecycle 2 Management and Marketing 3 Submission Recommendations – FDA, 2025b; 2025d*), por exemplo, exigindo documentação da lógica ou arquitetura do algoritmo e da proveniência dos dados, de modo a garantir rastreabilidade e possibilitar auditorias técnicas.

Além disso, a FDA propôs a incorporação de Planos de Controle de Mudança Predeterminada, através do *Predetermined Change Control Plans for Medical Devices – PCCPs* (FDA, 2024), para algoritmos aprendizes, permitindo que fabricantes obtenham

aprovação prévia para atualizações futuras de seus modelos, dentro de limites e protocolos estabelecidos. Isso garante que mesmo algoritmos que evoluem ao longo do tempo mantenham um nível aprovado de desempenho e segurança, sem precisar reiniciar todo o processo regulatório a cada modificação, ao mesmo tempo em que os regula sob uma perspectiva de ciclo de vida total do *software*.

De forma análoga, na Europa, o *AI Act* – primeiro marco regulatório abrangente sobre IA – classificará sistemas de IA em categorias de risco e exigirá que desenvolvedores de IA de alto risco (como as aplicadas à saúde) obtenham certificação de conformidade antes de colocar o produto no mercado. Entre outras obrigações, será necessário fornecer documentação técnica completa, realizar avaliação de conformidade por órgãos notificados e implementar gestão de risco algorítmico contínua (FILIPPIS *et al.*, 2024; UNIÃO EUROPEIA, 2024). Esses requisitos forçam os fornecedores a identificar e mitigar vieses, garantir a qualidade dos dados e incorporar controles de explicabilidade, antes que a tecnologia chegue ao paciente.

Importante notar que esse arcabouço europeu complementa normas preexistentes, como o Regulamento Europeu de Dispositivos Médicos (*MDR* 2017/745), que já impõe requisitos de segurança, desempenho e validação clínica para *softwares* médicos, incluindo algoritmos de diagnóstico (UNIÃO EUROPEIA, 2017a; 2025), e o Regulamento Geral de Proteção de Dados (*GDPR*), que consagra limites ao tratamento automatizado de dados pessoais sensíveis e assegura, em seu art. 22, o direito a uma intervenção humana em decisões automatizadas de impacto significativo (FILIPPIS *et al.*, 2024; UNIÃO EUROPEIA, 2016; 2024).

No âmbito das neurotecnologias, organismos internacionais e regionais têm se dedicado ativamente ao estabelecimento de critérios rigorosos para avaliação prévia e regulação (DANTAS; JEREISSATI, 2024; ANPD, 2025). A necessidade de uma abordagem preventiva e regulamentação rigorosa é reforçada por iniciativas como as da Organização dos Estados Americanos (OEA), que, em 2023, aprovou os “Princípios Interamericanos em matéria de Neurociências, Neurotecnologias e Direitos Humanos”, promovendo um enfoque de direitos humanos desde a concepção das neurotecnologias (OEA, 2023). O Chile, em 2021, tornou-se o primeiro país a incorporar a proteção dos neurodireitos em sua Constituição, garantindo que o desenvolvimento científico e tecnológico esteja a serviço das pessoas e salvaguardando a atividade cerebral e as informações dela provenientes. A decisão judicial chilena, no caso *Girardi vs. Emotiv Inc.*, também destaca o papel ativo do Estado na análise preventiva antes da comercialização de neurotecnologias, enfatizando a proteção da integridade mental e privacidade dos usuários (DANTAS; JEREISSATI, 2024).

A UNESCO (2025), por exemplo, no “Projeto de Texto de Recomendação sobre a Ética da Neurotecnologia”, propõe a criação de Metodologias de Avaliação de Impacto Ético específicas para neurotecnologia e IA em saúde, que examinariam riscos à dignidade, autonomia, privacidade e outros direitos humanos ao longo do ciclo de vida desses dispositivos. Essa abordagem preventiva distribui parte da responsabilidade para fabricantes e reguladores – que teriam de demonstrar, antes da comercialização, a aderência a padrões éticos e de direitos humanos – aliviando o ônus sobre o paciente de “adivinhar” riscos ocultos ou entender sozinho as opacidades do algoritmo (UNIÃO EUROPEIA, 2025).

No Brasil, reforçando essa tendência mundial, *softwares* de finalidade médica (inclusive os baseados em IA) são regulados como dispositivos médicos e devem cumprir exigências de registro sanitário conforme a classificação de risco (ANVISA, 2022a; 2022b). As Resoluções da Anvisa – RDC nº 657/2022 e RDC nº 751/2022 – estabeleceram um marco específico para “*Software* como Dispositivo Médico” (*SaMD*), alinhado às práticas internacionais, definindo quais *softwares* se enquadram como dispositivos e disciplinando sua regularização (AITH, 2024).

Ou seja, aqui também, essas normas exigem que o desenvolvedor apresente documentação técnica robusta como condição de segurança e eficácia, incluindo descrição detalhada da finalidade do *software*, arquitetura e histórico de versões do algoritmo, caracterização do conjunto de dados de treinamento e validação, justificativa do método algorítmico utilizado, evidências de validação clínica e avaliação de riscos específicos do software (ANVISA, 2022a; 2022b). Com isso, busca-se assegurar que antes do produto entrar no mercado brasileiro, ele atenda a parâmetros de qualidade, desempenho e transparência técnica similares aos exigidos em outros grandes mercados regulados. Ademais, a ANVISA prevê que modificações significativas em *softwares* já aprovados (por exemplo, atualização de algoritmos que afetem desempenho clínico) sejam submetidas à Agência, para avaliação através de petições de alteração, o que mantém o controle regulatório contínuo sobre a evolução do dispositivo (ANVISA, 2022a).

Em março de 2025, através da Resolução-RE nº 1.206/2025, a ANVISA (2025a) tornou obrigatório o depósito dos instrumentos com as instruções de uso, no repositório da Agência, para novos produtos e alterações de produtos já notificados ou registrados. Isso assegura que o conteúdo esteja consistente com o produto regularizado e a legislação, contribuindo para a “transparência técnica” e o “controle regulatório contínuo sobre a evolução do dispositivo”. Exemplo de Regulação de *SaMD*: No anexo dessa Resolução, há o registro de um “Sistema de Monitoramento Contínuo GS1”, classificado como “*IVD* - Registro de produto com *software*

não embarcado (*SaMD*) de uso dedicado”, demonstrando a aplicação prática da regulamentação de *SaMD* no Brasil. Posteriormente, em julho de 2025, a ANVISA publicou o Ofício Circular Nº 2/2025/SEI/GGTPS/DIRE3/ANVISA, detalhando a implementação do sistema de Identificação Única de Dispositivos Médicos (*UDI*), um padrão internacional que segue as regras do *IMDRF* (*International Medical Device Regulators Forum*) e busca a identificação inequívoca de dispositivos médicos para aumentar a segurança do paciente e a transparência. Isso demonstra o atendimento a “parâmetros de qualidade, desempenho e transparência técnica similares aos exigidos em outros grandes mercados regulados” (ANVISA, 2025b).

Apesar dos avanços, o sistema regulatório brasileiro ainda apresenta desafios. A RDC nº 657/2022, por exemplo, exclui de seu escopo os “*softwares* para bem-estar” ou de estilo de vida que não são explicitamente comercializados como dispositivos médicos. Isso cria uma lacuna, permitindo que aplicativos baseados em IA que podem impactar a saúde (como monitores de frequência cardíaca ou oxigênio no sangue para condicionamento físico) escapem da supervisão da ANVISA, posicionando-se como ferramentas gerais de bem-estar (AITH, 2024).

Finalmente, é oportuno citar um caso envolvendo neurotecnologia, no qual a Corte Suprema do Chile instou a autoridade sanitária a realizar uma análise preventiva antes de autorizar a comercialização e o uso dessas tecnologias (DANTAS; JEREISSATI, 2024). Carlos Henrique Félix Dantas e Régis Gurgel do Amaral Jereissati (2024), comentando o caso, destacam a apreciável ênfase do papel do Estado na introdução de novas tecnologias, sobretudo aquelas correlatas à atividade cerebral, anteriormente de natureza reservada e médica. Nesse contexto, a Corte insta a autoridade competente à realização de uma análise preventiva, ou seja, antes de autorizar a comercialização e o uso dessa modalidade de tecnologia, especialmente quando se apresentam questões até então não examinadas. Este ponto da sentença, destacam os autores, “reflete o *princípio da precaução*, proveniente do direito ambiental, que é exercido diante de um risco potencial pela ação do Estado por meio de medidas de precaução. Isso tem como objetivo prevenir e antecipar seus possíveis efeitos, além de proteger diretamente a integridade humana, que inclui sua privacidade, confidencialidade e os direitos próprios da integridade psíquica e do sujeito de experimentação científica (C. S. Girardi c/ Emotiv Inc., 2023)”.

6.4.2 Monitoramento pós-mercado e auditorias contínuas

Mesmo após a aprovação inicial, os algoritmos de sistemas de IA podem aprender ou sofrer alterações em seus dados de entrada, e os contextos de uso desses sistemas podem

evoluir. O desempenho de sistemas de IA de alto risco, em particular, deve ser monitorado ao longo de todo o seu ciclo de vida (NOGAROLI, 2023).

Essa evolução dos sistemas pode levar ao surgimento de novos problemas ou desvios de desempenho. Por exemplo, um aumento de erro pode ocorrer ao aplicar o algoritmo em um subgrupo populacional diferente daquele usado no treinamento. A má qualidade dos dados de treinamento pode gerar resultados tendenciosos, perpetuando ou ampliando a discriminação (DANTAS; JEREISSATI, 2024). Por isso, é vital implementar sistemas de gestão de risco e monitoramento contínuo do desempenho algorítmico, além de conduzir auditorias periódicas de vieses e erros durante toda a vida útil do sistema (UNIÃO EUROPEIA, 2025).

O *AI Act* europeu (UNIÃO EUROPEIA, 2024), no Considerando nº 65, exige que o sistema de gestão de riscos seja um processo iterativo contínuo, planejado e executado ao longo de todo o ciclo de vida de um sistema de IA de alto risco, com revisão e atualização regulares. A documentação técnica deve ser mantida atualizada durante toda a vida útil do sistema, e os sistemas de IA de alto risco devem permitir o registro automático de eventos (Considerando nº 71). As alterações no algoritmo e desempenho de sistemas de IA, que continuam a aprender após a implementação, não são consideradas modificações substanciais, desde que predeterminadas e examinadas na avaliação inicial de conformidade (Considerando nº 128; Art. 43, nº 4). Os prestadores devem ter um sistema de monitoramento pós-mercado para coletar e analisar dados, identificar e relatar incidentes graves às autoridades competentes (Considerandos nº 24 e 25; Art. 72, nº 2).

De igual modo, a FDA – ao recomendar os *PCCPs*, supracitados – enfatiza uma perspectiva de ciclo de vida total do dispositivo, na qual modelos implantados sejam continuamente acompanhados quanto à performance e gestão de riscos de retreinamento (FDA, 2024). Isso significa que os fabricantes devem não apenas validar o algoritmo inicialmente, mas também estabelecer métricas e procedimentos para inspecionar seu comportamento no mundo real, identificar *drifts* ou quedas de acurácia, e acionar mecanismos de atualização ou correção quando necessário.

Nesse sentido, o Regulamento (UE) 2017/745 (relativo aos dispositivos médicos – *MDR*) foi muito feliz ao exigir que os fabricantes mantenham a conformidade da produção em série e considerem a experiência adquirida pelos usuários, com o uso dos dispositivos, no processo de produção (Considerando nº 32). Isso inclui um sistema de gestão de qualidade e um sistema de monitoramento pós-comercialização, proporcionais à classe de risco e ao tipo de dispositivo em questão. Além disso, o Regulamento prevê que os dados e informações da monitorização pós-comercialização devam ser usados para atualizar a documentação técnica, incluindo a avaliação

de riscos e a avaliação clínica. Alterações que afetem o desempenho, a segurança ou a finalidade prevista do *software* exigem nova avaliação de conformidade (Considerando nº 74; Art. 83), como, aliás, também disciplina o ANEXO VII, do *AI Act* (UNIÃO EUROPEIA, 2017a; 2024).

A OMS, em suas diretrizes recentes, igualmente recomenda que governos e desenvolvedores implementem auditorias independentes periódicas e avaliações de impacto pós-implantação, a fim de verificar se os sistemas de IA em saúde mantêm desempenho confiável, equitativo e alinhado aos valores éticos ao longo do tempo (OMS, 2024).

No Brasil, o PL 2.338/2023, propõe que a avaliação de impacto algorítmico seja um processo iterativo contínuo ao longo de todo o ciclo de vida dos sistemas de IA de alto risco, com atualizações periódicas (em sintonia com o Considerando nº 77, *AI Act* europeu (2024)). Os agentes de inteligência artificial devem usar ferramentas para avaliar a acurácia, robustez e potenciais resultados discriminatórios, implementando medidas de mitigação de riscos. Eles devem documentar testes para confiabilidade e segurança. Também devem comunicar riscos ou impactos inesperados e relevantes às autoridades setoriais e outros agentes da cadeia (BRASIL, 2023, Arts. 25 e 20).

No âmbito da ANVISA (2022b), os produtos regularizados no Brasil estão sujeitos a auditorias, monitoramento de mercado e inspeções contínuas, garantindo que permaneçam em conformidade. Alterações em produtos para solucionar problemas de segurança e desempenho não se enquadram em permissões de esgotamento de estoque, ressaltando o foco nesses aspectos. Além disso, no Ofício Circular Nº 2/2025/SEI/GGTPS/DIRE3/ANVISA, a Agência informa sobre a criação do Sistema de Identificação Única de Dispositivo (SIUD), a base de dados *UDI* da ANVISA, que permitirá a consulta de informações sobre os dispositivos médicos, fortalecendo a vigilância pós-mercado e o controle contínuo (ANVISA, 2025b).

Enfim, todas essas medidas visam manter um sistema de monitoramento pós-mercado, que colete e analise dados sobre o uso real e os resultados da tecnologia, reportando imediatamente quaisquer incidentes graves ou mau funcionamento às autoridades (FILIPPIS *et al.*, 2024). De forma complementar, espera-se que hospitais e desenvolvedores adotem políticas sistemáticas para coletar *feedback* de casos reais, rastrear resultados clínicos e reportar os eventos adversos, que forem associados ao uso de IA (PASQUALE, 2020; TOPOL, 2019).

Solaiman *et al.* (2025) argumentam, inclusive, que esse ciclo contínuo de retroalimentação (“*lifecycle governance*”) possibilita a correção de falhas e a atualização das informações comunicadas aos pacientes, criando um processo de melhoria permanente. Por exemplo, caso se constate que um algoritmo de triagem apresenta maior margem de erro em pacientes idosos, a redação dos Consentimentos futuros deve incluir advertências

específicas sobre esse risco, garantindo que a decisão do paciente seja informada à luz das evidências mais recentes (RATTI *et al.*, 2025). Portanto, o monitoramento pós-mercado tem implicações diretas para o CLE contínuo do paciente, tornando-o um instrumento vivo e recalibrado conforme surgem evidências adicionais (*Dynamic Consent*).

Finalmente, é importante destacar que essa prática de monitoramento se alinha ao conceito de “transparência dinâmica”, segundo o qual pacientes e médicos devem ser mantidos atualizados sobre mudanças no perfil de risco/benefício da tecnologia, à medida que novos dados são gerados (NOGAROLI, 2023). Ao reforçar que o consentimento não é um compromisso fixo ou cego, mas um processo adaptativo, fundamentado em evidências atualizadas, cria-se um ambiente de maior confiança e corresponsabilidade entre paciente, profissional e desenvolvedor (CHAR *et al.*, 2018; BEAUCHAMP; CHILDRESS, 2013).

6.4.3 Guias éticos e comitês de supervisão institucionais

No ambiente hospitalar, a incorporação de IA e neurotecnologias deve vir acompanhada de protocolos éticos claros e de supervisão institucional robusta, a fim de mitigar riscos e assegurar a proteção da autonomia do paciente (PRICE II; COHEN *et al.*, 2024; PASQUALE, 2020). Comissões de ética clínica e comitês de revisão tecnológica podem desempenhar um papel central na avaliação de casos de uso sensíveis, verificando se determinadas aplicações respeitam os princípios da bioética – beneficência, não maleficência, autonomia e justiça –, conforme delineado por Beauchamp e Childress (2013). DONEDA; SOUZA *et al.* (2018) já destacavam a necessidade desses órgãos internos de supervisão.

Esses colegiados multiprofissionais funcionam como um contrapeso institucional, garantindo que, mesmo havendo consentimento do paciente, a utilização da tecnologia esteja alinhada com seu melhor interesse e com os valores fundamentais da prática médica (NOGAROLI, 2023; TOPOL, 2019). Resolução do Parlamento Europeu, nº 2015/2103 (INL), que contém recomendações sobre Robótica, sublinha, na recomendação nº 36, “a importância da criação urgente de comissões de ética de robôs em hospitais e em outras instituições de cuidados de saúde dotadas de pessoal competente e incumbidas de analisar e prestar assistência na resolução de problemas éticos complicados e pouco habituais, que envolvam questões que afetam o cuidado e o tratamento” (UNIÃO EUROPEIA, 2018a).

Outra proposta emergente é a nomeação de “oficiais de ética em IA” – ou, segundo Luciano Floridi *et al.* (2018), “*AI ombudsperson*” –, nos hospitais, encarregados de

supervisionar a implementação responsável dessas tecnologias, realizando auditorias periódicas tanto sobre a conformidade legal, quanto sobre a qualidade e efetividade do processo de Consentimento Livre e Esclarecido associado a elas (FLORIDI *et al.*, 2018). Barry Solaiman *et al.* (2025) propõem uma “*True Lifecycle Approach*” - *TLA*, que “integra princípios fundamentais da lei de saúde – como consentimento informado, responsabilidade e direitos do paciente – durante o desenvolvimento, implantação e uso da IA”. A *TLA* é “fundamentalmente centrada no paciente”, priorizando a segurança, a confiança e a sua autonomia.

Por óbvio, tudo que está sendo proposto demanda treinamento profissional específico sobre os limites, vieses e potencial de erro dos modelos de IA. Médicos, enfermeiros e demais atores da saúde precisam ser capacitados para interpretar criticamente as recomendações de algoritmos, reconhecer quando a saída de uma IA pode estar equivocada ou enviesada, e comunicar essas nuances aos pacientes de forma compreensível (COHEN, 2019; KAYE *et al.*, 2015; AMANN *et al.*, 2020). Hospitais devem garantir políticas de treinamento e qualificação constante dos médicos ao implementar novas tecnologias (NOGAROLI, 2023). Essa educação permanente ajuda a equilibrar a relação homem-máquina na prática clínica, prevenindo uma dependência cega em sistemas automatizados e fortalecendo o papel do profissional como garantidor da qualidade da decisão clínica.

Ao deslocar parte da responsabilidade para estruturas organizacionais (hospitais, comitês, oficiais de ética), reconhece-se que o consentimento isoladamente não basta para lidar com a opacidade e os vieses – é preciso uma governança interna que complemente a autonomia do paciente com vigilância profissional e institucional. Superar o Consentimento Livre e Esclarecido como único escudo não é descartá-lo, mas fortalecê-lo e complementá-lo. O equilíbrio desejado é aquele em que inovação e direitos fundamentais caminhem juntos: que os avanços de IA e neurotecnologias sirvam ao bem-estar humano sem suplantarem autonomia, dignidade e segurança do paciente (OMS, 2024; UNIÃO EUROPEIA, 2024; TOPOL, 2019).

6.4.4 Cultura de responsabilização compartilhada

Superar o CLE como única salvaguarda, no uso de IA e neurotecnologias em saúde, requer migrar de um modelo centrado no “termo assinado” para uma arquitetura de responsabilidades distribuídas entre desenvolvedores, instituições de saúde, profissionais clínicos e reguladores. Em termos normativos, essa mudança dialoga com a bioética principialista – na qual a autonomia deve ser garantida sem eclipsar a beneficência, a não maleficência e a justiça (BEAUCHAMP; CHILDRESS, 2013) – e com diretrizes internacionais

que vinculam *accountability* à transparência substantiva, à avaliação de impacto e à supervisão humana contínua (OMS, 2024; UNIÃO EUROPEIA, 2024; FILIPPIS *et al.*, 2024).

Podemos delinear essa teia de responsabilidades da seguinte forma: (i) *Deveres dos desenvolvedores (fabricantes)*; (ii) *Deveres das instituições de saúde*; (iii) *Deveres dos profissionais*; e (iv) *Deveres dos órgãos reguladores*.

Deveres dos desenvolvedores (fabricantes): para fornecedores de IA clínica e de neurodispositivos, a responsabilidade começa antes da colocação no mercado: exige-se projeto orientado a risco, documentação técnica robusta, registro de dados de treinamento e validação (com análise de representatividade e desempenho por subgrupos), mitigação de vieses e planos de monitoramento pós-mercado (OMS, 2024; NIST, 2023; FILIPPIS *et al.*, 2024). No Brasil, *softwares* com finalidade médica (inclusive os baseados em IA) são dispositivos médicos (*SaMD*) e devem atender às RDC nº 657/2022, nº 751/2022 e RDC nº 777/2023 (ANVISA, 2022a; 2022b; 2023a). O titular da regularização do produto perante a ANVISA, em particular, detém responsabilidade administrativa, civil e penal por garantir a segurança e a eficácia do dispositivo³¹. Na União Europeia, o *MDR* 2017/745, o *IVDR* 2017/746 e o *AI Act* 2024/1689 impõem dupla conformidade: evidência clínica e gestão de risco do dispositivo, além de requisitos específicos para IA de alto risco (governança de dados, documentação, testes, supervisão humana e transparência) (UNIÃO EUROPEIA, 2017a; 2017b; UNIÃO EUROPEIA, 2024; AIB/MDCG, 2025). Essas camadas regulatórias reforçam que explicabilidade prática e desempenho auditável são deveres do fornecedor, e não ônus cognitivo do paciente (PRICE II, 2017; PASQUALE, 2015; FILIPPIS *et al.*, 2024).

Deveres das instituições de saúde: hospitais e clínicas devem incorporar governança algorítmica: comitês técnico-clínicos para avaliação de uso sensível (incluindo neurotecnologias), protocolos de segundo parecer humano para decisões críticas, trilhas de auditoria, e transparência dinâmica com atualização das informações de risco/benefício quando novos dados emergem (OMS, 2024; TOPOL, 2019; FILIPPIS *et al.*, 2024). Também lhes cabe vigilância pós-implementação (registro de incidentes, *drift* de desempenho, alertas por subpopulações), devolvendo esse aprendizado ao consentimento futuro e aos materiais

³¹ Art. 2º O carregamento de instruções de uso no repositório documental de dispositivos médicos, disponível no portal da Anvisa, é obrigatório e deve ser executado pela **empresa responsável pela notificação ou pelo registro do produto**, a qual atesta que seu conteúdo guarda concordância com a legislação vigente e consistência com o produto regularizado, de acordo com o §4º do art. 33 da Resolução da Diretoria Colegiada - RDC nº 751, de 15 de setembro de 2022 e do §4º do art. 33 da Resolução da Diretoria Colegiada - RDC nº 830, de 06 de dezembro de 2023. RESOLUÇÃO-RE Nº 1.206, DE 27 DE MARÇO DE 2025. Disponível em: https://anvisa.legis.datalegis.net/action/UrlPublicasAction.php?acao=abrirAtoPublico&num_ato=00001206&sgl_tipo=RES&sgl_orgao=RE/GGTPS/ANVISA/MS&vlr_ano=2025&seq_ato=000&cod_modulo=293&cod_menu=8499. Acesso em: 10 ago. 2025.

informativos (ANVISA, 2022b; NIST, 2023.). Em neurotecnologias invasivas (*e.g.*, interfaces cérebro-computador ou *DBS*), o acompanhamento longitudinal e o suporte ético-clínico são parte indissociável do dever institucional de cuidado (FDA, 2021; YUSTE *et al.*, 2017).

Deveres dos profissionais: a adoção de IA não diminui a responsabilidade clínica dos médicos, a transforma (PRICE II; GERKE; COHEN, 2024). O padrão de cuidado passa a incluir a competência em IA: conhecer indicações, limites, vieses e incertezas, explicá-los ao paciente em linguagem acessível e justificar por que seguiu ou contrariou a recomendação algorítmica (COHEN, 2019; CHAR *et al.*, 2018; FILIPPIS *et al.*, 2024). Em termos de consentimento, o profissional tem o dever de situar a decisão: explicitar que se trata de decisão assistida por IA/neurodispositivo, indicar alternativas e incertezas (inclusive de natureza epistêmica), acolher dúvidas e registrar a deliberação compartilhada (BEAUCHAMP; CHILDRESS, 2013; OMS, 2024). Em neurotecnologias com potencial de impactar identidade e continuidade psicológica, a prudência informativa deve ser redobrada (YUSTE *et al.*, 2017).

Deveres dos órgãos reguladores: os órgãos reguladores devem atualizar continuamente guias e requisitos à medida que evidências e tecnologias evoluem: no Brasil, a ANVISA já consolidou o enquadramento de *SaMD* e os deveres documentais; na UE, o *AI Act* cria obrigação *ex ante* proporcionada ao risco; nos EUA, a FDA publica guias para IA/ML e orientação específica para *BCIs* implantáveis (ANVISA, 2022; UNIÃO EUROPEIA, 2024; FDA, 2021). Em todos os casos, fiscalização e pós-mercado são elementos estruturantes, complementando (e não substituindo) o consentimento individual (OMS, 2024; FILIPPIS *et al.*, 2024).

Essa defesa em profundidade (múltiplas camadas de proteção) funciona como rede de segurança: se uma barreira falha – por exemplo, um viés não detectado pelo fornecedor – outra deve atuar (auditoria institucional, revisão clínica, *recall* regulatório), evitando que todo o peso recaia sobre a “suposta” escolha informada de um indivíduo vulnerável.

Ao deslocar deveres do paciente para os sujeitos que projetam, mediam e fiscalizam a tecnologia, essa cultura de responsabilização compartilhada torna o consentimento efetivamente livre e esclarecido em sistemas complexos e opacos – inclusive no fronte das neurotecnologias (YUSTE *et al.*, 2017).

6.4.5 Novos direitos para salvaguardar a integridade mental e a identidade pessoal frente a tecnologias neuroinvasivas

No âmbito da *neurotech*, a tutela da autodeterminação e do consentimento somente será garantida, como sustentam Ienca e Andorno (2017), com a existência de novos direitos para salvaguardar a integridade mental e a identidade pessoal, frente a tecnologias neuroinvasivas, que serão aprofundados no próximo Capítulo. Sergio Negri *et al.* (2025) reforçam que a proteção efetiva exige mecanismos institucionais e regulatórios capazes de atuar sobre a arquitetura e o funcionamento das próprias tecnologias, garantindo que as condições para o exercício da autonomia não sejam degradadas por vieses estruturais, opacidade ou manipulação.

Como visto no Capítulo 5, especificamente no item 5.2.2, o consentimento para o uso de neurotecnologias, que modulam ativamente o cérebro, representa, segundo Yuste *et al.* (2017), a fronteira mais desafiadora para a autonomia, pois envolve autorizar intervenções capazes de afetar elementos centrais da identidade pessoal. Trata-se de um consentimento que implica aceitar o risco de uma potencial alteração dessa identidade – risco que, devido à opacidade dos mecanismos envolvidos, é impossível de ser plenamente dimensionado *a priori*. A autonomia, nesse contexto, é exercida para autorizar um procedimento que pode, paradoxalmente, modificar a própria base da qual essa autonomia emana. Esse cenário se conecta diretamente ao debate sobre a necessidade de um “direito à continuidade psicológica”, como nova salvaguarda fundamental, destinado a proteger a integridade mental frente a tecnologias neuroinvasivas (YUSTE *et al.*, 2017).

Estudos recentes sobre ética em neurotecnologia ressaltam o caráter inadequado de abordagens meramente procedimentais – como aprovações éticas isoladas e formulários de consentimento estático – e apontam para a importância de processos reflexivos e contínuos ao longo de todo o ciclo de interação com dispositivos implantáveis (HAAG; IENCA *et al.*, 2025; DANTAS; JEREISSATI, 2024). Além disso, o princípio ético de não abandono em pesquisas com dispositivos neurais reforça que o consentimento deve estender-se ao acompanhamento pós-implantação, com responsabilidades claras sobre comunicação e segurança ao longo do tempo (CENOLLI, 2024; CALIXTO; SCHULZE, 2025).

Dantas e Jereissati (2024) propõem a positivação de um novo direito fundamental – o “direito à convivência harmônica” – como via regulatória mais adequada às especificidades das interfaces cérebro-computador (ICCs ou *BCIs*) combinadas com inteligências artificiais. Os autores sustentam que uma defesa intransigente dos direitos à identidade e à continuidade psicológica em sua formulação “original” pode tornar-se contraproducente, chegando até a inviabilizar aplicações benéficas dessas tecnologias. Em lugar dessa postura, sugerem um enfoque específico ao contexto neurotecnológico, orientado por um direito que proteja não apenas o desenvolvimento funcional das *BCIs* e das IAs, mas também sua consonância com

valores, interesses e bem-estar psicológico humanos – isto é, um arranjo normativo voltado a assegurar o alinhamento ético entre humanos, *BCIs* e IAs, antes e durante o seu uso.

Contudo, os próprios autores reconhecem que essa proposta não pode eliminar as anteriores. Ou seja, essa convivência harmoniosa deve englobar também: (i) a busca pela transparência quanto a propósitos, funcionamento e uso de dados cerebrais dos sistemas *BCIs* e IA; (ii) Auditorias éticas pré e pós-implantação das tecnologias; (iii) Monitoramento contra vieses algorítmicos discriminatórios; e (iv) Protocolos rígidos sobre aquisição e compartilhamento de *inputs* e *outputs* neurais (DANTAS; JEREISSATI, 2024).

Em síntese, as mesmas salvaguardas delineadas para sistemas de IA – tal como discutidas nos itens 6.4.1 a 6.4.4 deste tópico – devem ser aplicadas às neurotecnologias; todavia, impõe-se de modo inequívoco a regulamentação dos neurodireitos, sob pena de a proteção pretendida não se efetivar no plano concreto.

6.5 GOVERNANÇA DUPLA DA IA EM DISPOSITIVOS MÉDICOS: A COMPLEMENTARIDADE ENTRE O *AI ACT* E AS REGULAMENTAÇÕES *MDR/IVDR* NA UNIÃO EUROPEIA

A experiência regulatória da União Europeia oferece um estudo de caso ilustrativo de como estruturar salvaguardas jurídicas, em múltiplos níveis, para proteger pacientes diante de IA em dispositivos médicos. Podemos chamá-la de “governança dupla”, pois combina um marco setorial existente (regulação de dispositivos médicos) com um novo marco (regulação de IA), de forma complementar e integrada, formando um arcabouço regulatório europeu para a Inteligência Artificial em Dispositivos Médicos (*MDAI*), que foi disciplinado pelo AIB/MDCG³² (documento conjunto do Conselho de Inteligência Artificial e do Grupo de Coordenação de Dispositivos Médicos), publicado em junho de 2025.

O Regulamento Europeu de Dispositivos Médicos – *MDR* (Regulamento UE 2017/745) e seu análogo para diagnóstico *in vitro* – *IVDR* (Regulamento UE 2017/746) – já impõem requisitos estritos de segurança, eficácia clínica e gerenciamento de risco para qualquer dispositivo médico comercializado na Europa, incluindo *softwares* médicos (UNIÃO EUROPEIA, 2017a; 2017b). Isso significa que um *software* de IA para diagnóstico, para obter o título de conformidade, precisa demonstrar, por estudos e documentação, que é seguro e

³² AIB/MDCG. Joint Artificial Intelligence Board and Medical Device Coordination Group Document. AIB 2025-1 / MDCG 2025-6 - FAQ on Interplay between the Medical Devices Regulation & In vitro Diagnostic Medical Devices Regulation and the Artificial Intelligence Act (June 2025). Disponível em: https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf. Acesso em: 7 jul. 2025.

cl clinicamente valido para a finalidade proposta³³. Por si so, o *MDR* ja exige, por exemplo, evidencias de desempenho, analise de riscos e um sistema de vigilancia pos-mercado para recolher dados de uso real (UNIAO EUROPEIA, 2017a).

Com a entrada em cena do *AI Act* (Regulamento UE 2024/1689), criou-se uma segunda camada normativa especıfica para IA. Como analisado anteriormente, o *AI Act* classifica sistemas de IA por nıvel de risco e submete os de alto risco (categoria na qual se encaixam a maioria das aplicacoes medicas – FILIPPIS *et al.*, 2024; BOTTINI *et al.*, 2025) a uma serie de obrigacoes: gerenciamento de risco algorıtmico, governanca de dados (para mitigar vieses), requisitos de transparencia e explicabilidade, garantia de supervisao humana, robustez e ciberseguranca, entre outros (UNIAO EUROPEIA, 2024). Crucialmente, o *AI Act* nao substitui nem conflita com o *MDR/IVDR*, mas opera em adicao a estes. Conforme esclarecido em diretrizes conjuntas da Comissao Europeia (AIB/MDCG, 2025), a intencao e que o *MDR* cuide da seguranca e eficacia clınicas do dispositivo, enquanto o *AI Act* agrega salvaguardas direcionadas aos riscos especıficos da IA (como vieses e opacidade). E um exemplo claro de arquitetar salvaguardas multinıvel: nenhuma dessas camadas isoladamente garante a tutela plena do paciente, mas sua combinacao cria uma malha protetiva mais densa.

Em termos praticos, um desenvolvedor que queira lancar um algoritmo de diagnostico na Uniao Europeia deve agora cumprir dois conjuntos de exigencias.

Por exemplo, se desenvolver uma IA para detectar nodulos em imagens de tomografia: pelo *MDR* (UNIAO EUROPEIA, 2017a), as seguintes medidas precisarao ser adotadas: (i) *Acuracia diagnostica* - A confirmacao da conformidade com os requisitos gerais de seguranca e desempenho baseia-se em dados clınicos que fornecam evidencia clınica suficiente (Art. 61, n.o 1 e Anexo XIV, Parte A); (ii) *Gestao de risco* - Os fabricantes devem ter um sistema de gestao de risco que seja proporcional a classe de risco do dispositivo (Considerandos n.os 32 a 34 e Anexo II, n.os 5 e 6); (iii) *Qualidade* - E exigido um sistema de gestao da qualidade para manter a conformidade da producao (Considerando n.o 32; Art. 10, itens 8 e 9; Art. 16, “1” e Art. 17, n.o 1); (iv) *Validacao clınica* - A documentacao tecnica deve incluir verificacao e validacao do *software*, com provas de validacao (incluindo resultados de testes internos e em ambiente real) – (Anexo II, 6.1, “b”); e (v) *Vigilancia pos-mercado* - Os fabricantes devem possuir um sistema de monitorizacao pos-comercializacao para recolher e analisar dados sobre a experiencia de

³³A ANVISA, seguindo os padroes europeus, no Art. 2o da RDC no 657/2022, preve a figura da “Avaliacao Clınica”, como sendo um conjunto de atividades para avaliar a seguranca clınica, eficacia e desempenho de um *Software as a Medical Device* (SaMD), conforme a finalidade pretendida. Conjuntamente, preve a “Associacao Clınica Valida” ou “Validade Cientıfica”, que se refere a extensao em que a saıda do SaMD e clinicamente aceita ou bem fundamentada, com base em evidencias cientıficas (ANVISA, 2022a).

utilização e identificar a necessidade de ações corretivas (Considerando nº 32 e Anexo XIV, Parte B, nº 5).

Pelo *AI Act* – para sistemas de IA de alto risco (UNIÃO EUROPEIA, 2024) – o desenvolvedor deverá realizar: (i) *Classificação de risco* - A maioria dos *softwares* de dispositivos médicos são classificados como sistemas de IA de alto risco sob o *AI Act* (FILIPPIS *et al.*, 2024; BOTTINI *et al.*, 2025); (ii) *Redução de vieses nos dados de treinamento* - Exigem-se conjuntos de dados de treino, validação e testagem de elevada qualidade, pertinentes, representativos, isentos de erros e completos, para atenuar enviesamentos que possam levar à discriminação (Considerando nº 67; Art. 10, nº 2, “f” e nº 3); (iii) *Informações explanatórias aos usuários (transparência e explicabilidade)* - Sistemas de IA de alto risco devem ser transparentes, acompanhados de instruções de uso claras para que os responsáveis pela implantação (usuários) compreendam o funcionamento, funcionalidades, pontos fortes e limitações do sistema (Considerando nº 72; Art. 12, nº 3, “c” e “d”; Art. 13, nº 1 e Considerando nº 163); (iv) *Supervisão humana* - Os sistemas devem ser concebidos para permitir a supervisão humana, garantindo que os operadores tenham as competências, formação e autoridade necessárias para intervir (Considerandos nº 73, 91 e 92; Anexo IV, nº 2, “e” e “f”; Arts. 14 e 26, nº 2 e 3); (v) *Gestão de risco algorítmico* - Prestadores de sistemas de IA de risco elevado devem estabelecer um sistema de gestão de riscos contínuo para identificar e atenuar riscos (Considerandos nº 65, 154 e 155; Art. 9º); e (vi) *Robustez e cibersegurança* - Devem apresentar um nível adequado de exatidão, solidez e cibersegurança (Considerandos nº 74 e 77; Art. 15).

O processo de avaliação de conformidade envolverá tanto autoridades ou organismos notificados da área médica, quanto possivelmente entidades competentes em IA (NOGAROLI, 2023; UNIÃO EUROPEIA, 2025). O Espaço Europeu de Dados de Saúde (EEDS) prevê a cooperação entre autoridades de saúde digital (que podem incluir autoridades responsáveis por dispositivos médicos) e o Comitê Europeu para a Inteligência Artificial (UNIÃO EUROPEIA, 2025). O resultado final é que a IA, em ambientes clínicos, tenha passado por salvaguardas multinível, combinando a segurança e eficácia clínicas com a mitigação dos riscos específicos da IA (*AI Act*, Considerando nº 64).

A propósito, o Regulamento europeu 2025/327, relativo ao Espaço Europeu de Dados de Saúde (EEDS) reitera a integração dos Regulamentos 2017/745 (Dispositivos Médicos), 2017/746 (Dispositivos Médicos para Diagnóstico *In Vitro*), com o Regulamento 2024/1689 (*AI Act*), inclusive, no que diz respeito à segurança desses produtos que interagem com os sistemas de Registos de Saúde Eletrónicos (RSE). Componentes de *software* de sistemas de RSE, que se enquadram na definição de dispositivo médico, dispositivo médico para

diagnóstico *in vitro* ou sistema de IA de alto risco, devem ser certificados em conformidade com os respectivos regulamentos. Se esses fabricantes alegarem interoperabilidade, devem provar a conformidade com os requisitos essenciais de interoperabilidade do Regulamento EEDS (UNIÃO EUROPEIA, 2025, Considerando nº 42; Art. 1º, nº 4 e 5).

Do ponto de vista das salvaguardas jurídicas ao Consentimento, essa abordagem europeia é instrutiva. Ela tira das mãos do paciente e do médico parte do fardo de avaliar confiabilidade e ética da tecnologia, transferindo-o para mecanismos *ex ante* de controle. Ao exigir explicabilidade e mitigação de viés já na fase de projeto (e comprovação disso documentalmente), o regulador europeu está antecipando e prevenindo problemas que, de outro modo, recairiam sobre o paciente na hora do consentimento (por exemplo, o paciente não ficará totalmente no escuro sobre a lógica do sistema, pois o sistema só obteve autorização de mercado tendo mecanismos de explicação). Nesse sentido, a dupla governança exemplifica a transição de uma confiança exclusiva no consentimento individual para uma governança da tecnologia, que cerca o Consentimento de garantias estruturais.

Cabe mencionar que os Estados Unidos adotam estratégia distinta: lá não existe um “*AI Act*” nacional, e a regulação de IA em saúde é mais fragmentada. A FDA atua caso a caso na autorização de dispositivos com IA, mas não há requisitos gerais de explicabilidade ou mitigação de vieses estabelecidos em lei federal – o que há são guias de boas práticas (como o *AI Risk Management Framework* do NIST (2023)) e iniciativas da Casa Branca como o *Blueprint for an AI Bill of Rights* (THE WHITE HOUSE, 2022), de natureza não vinculante. Assim, nos EUA a proteção tende a se dar mais por *accountability ex post* (ações judiciais se algo der errado, responsabilidade civil, pressão de mercado) do que por requisitos prévios uniformes. Essa opção oferece flexibilidade para inovação, porém deixa uma zona de incerteza maior para pacientes: as garantias éticas variam conforme o fabricante e a instituição, não havendo um patamar mínimo tão definido como na UE (PRICE II; GERKE; COHEN, 2024).

Já o Brasil se encontra em momento singular, podendo aprender com ambos os modelos. O Projeto de Lei nº 2.338/2023 (BRASIL, 2023a), inspirado no *AI Act*, procura implementar localmente uma estrutura de classificação de risco e obrigações para IA, enquanto a ANVISA já regula *softwares* médicos e poderia ancorar a avaliação técnica. Além disso, propostas como a PEC 29/2023 (neurodireitos) mostram disposição de inovar também no campo dos direitos fundamentais (BRASIL, 2023c), bem como o Projeto de Lei nº 2174/2023, que estabelece as normas e princípios para proteção dos direitos fundamentais relacionados ao cérebro e ao sistema nervoso humano, objetivando garantir a proteção e promoção dos neurodireitos dos indivíduos (BRASIL, 2023b). Desenha-se, assim, um modelo híbrido brasileiro: baseado em

direitos fundamentais (influência latino-americana), preventivo e orientado a risco (influência europeia), mas atento à necessidade de não sufocar inovação (aprendendo com a flexibilidade americana). Em qualquer caso, o importante é que a lição central seja assimilada: CLE do paciente e governança regulatória não competem entre si, mas se complementam. A segurança e transparência requeridas no nível regulatório reforçam o Consentimento Livre e Esclarecido, ao passo que um consentimento robusto legitima o uso das novas tecnologias, quando estas atendem aos padrões exigidos.

Em suma, a “governança dupla” da UE demonstra na prática a efetivação de duas das principais salvaguardas aqui discutidas – transparência/explicabilidade e *accountability* – via legislação coordenada. É um caminho que outras jurisdições estão observando de perto e adaptando conforme suas realidades. No fim das contas, qualquer modelo regulatório eficaz deverá buscar um equilíbrio similar: regras prévias claras (transparência, avaliação de risco, certificação), aliadas a mecanismos contínuos de responsabilidade e aprendizado, tudo isso orbitando em torno do objetivo maior de proteger a pessoa humana em sua dignidade e autonomia, conforme preconizado pelos marcos bioéticos e jurídicos.

6.5.1 A Contribuição Complementar do *AI Act*: Transparência e Supervisão Humana

O AIB/MDCG demonstra, portanto, que o *AI Act*, ao complementar o *MDR/IVDR*, introduz uma série de requisitos que são diretamente instrumentais para fortalecer o Consentimento Livre e Esclarecido do paciente, mesmo que seu foco principal esteja na governança dos sistemas de IA em geral e nos direitos fundamentais.

6.5.1.1 Transparência Ampliada (Questões nº 11, 14 e 15 do AIB/MDCG):

Questão 11. Como o MDR, o IVDR e o AI Act abordam os dados de treinamento, validação e teste usados para MDAI de alto risco?

Questão 14. Quais são os requisitos impostos pelo AI Act e pela MDR/IVDR em matéria de transparência e MDAI?

Questão 15. Como o MDR/IVDR e o AI Act abordam os requisitos de transparência, explicabilidade e processamento de dados para MDAI de alto risco?

A transparência é um requisito central para o *MDAI* de alto risco. O *AI Act* impõe aos fabricantes a obrigação legal de projetar e desenvolver o *MDAI* de forma que sua operação seja suficientemente transparente para permitir que os *deployers* (“responsáveis pela implantação”)

interpretem as saídas corretamente e usem o sistema de maneira apropriada, apoiados por instruções de uso claras e compreensíveis (UNIÃO EUROPEIA, 2024, Artigo 13).

Crucialmente, o *AI Act* exige que os usuários (neste contexto, os pacientes, quando diretamente interagindo) sejam informados de que estão interagindo com um sistema de IA, a menos que isso seja óbvio pelas circunstâncias. Esta é uma provisão fundamental para o consentimento direto do paciente, pois reconhece a necessidade de conscientização sobre a presença de IA no processo de cuidado.

Requisitos de transparência se estendem ao processamento de dados, exigindo qualidade, gerenciamento e documentação que permitam uma compreensão da forma como os dados são utilizados, treinados e validados (Art. 10, *AI Act*). A garantia de que os conjuntos de dados são representativos, livres de erros e vieses é essencial para a confiança nos resultados da IA, sobre os quais as decisões médicas são tomadas e o consentimento é baseado (Art. 10, *AI Act*).

Em essência, a informação fornecida pelos fabricantes aos *deployers* deve ser clara, completa e acionável, especialmente quando as decisões puderem impactar a saúde ou os direitos fundamentais. Isso significa que a transparência e a explicabilidade não são características opcionais, mas sim obrigações vinculantes, garantindo que os *deployers* possam compreender a lógica, as limitações e o comportamento dos componentes de IA dentro de um dispositivo médico.

6.5.1.2 *Supervisão Humana (Questões nº 18, 19 e 36 do AIB/MDCG):*

Questão 18. Que requisitos de supervisão humana estão incluídos no MDR, IVDR e AI Act para MDAI de alto risco?

Questão 19. A supervisão humana para dispositivos médicos pode ser entendida como parte das medidas de gestão de risco existentes, tais como uma medida de design (por exemplo, botão “parar”)?

Questão 36. Os fabricantes de MDAI são obrigados a definir alguma formação mínima em IA, por exemplo, para especialistas em física médica e radiologistas, para que possam compreender os riscos potenciais das ferramentas de apoio ao diagnóstico?

O *AI Act* enfatiza a importância e estabelece uma obrigação legal para o fabricante projetar e desenvolver sistemas de IA com mecanismos apropriados de supervisão humana. Isso significa que o *MDAI* deve ser concebido com restrições operacionais integradas que não podem ser anuladas pelo próprio sistema e deve ser responsivo ao operador humano (Artigo 14, *AI Act*).

A supervisão humana permite a intervenção humana em processos críticos de tomada de decisão (*MDR* e *IVDR*, Anexo I). Esta é uma salvaguarda direta para o paciente, pois assegura que, mesmo diante da autonomia da IA, a decisão final e a responsabilidade recaem sobre um profissional de saúde, que pode exercer seu julgamento clínico.

A supervisão humana é considerada um fator de mitigação de risco, visando prevenir ou minimizar riscos à saúde, segurança e direitos fundamentais quando um sistema de IA de alto risco é usado (Art. 14, *AI Act*). A capacidade do profissional de saúde de compreender as capacidades e limitações do *MDAI* é crucial para essa supervisão efetiva (*MDR* e *IVDR*, Anexo I). Para que um paciente consinta com o uso de uma IA, a certeza de que haverá uma “camada” de supervisão humana é essencial.

6.5.1.3 Responsabilidade (Accountability) - (Questões nº 16 e 17 do AIB/MDCG):

Questão 16. Como o MDR/IVDR e o AI Act abordam a responsabilidade da MDAI?

Questão 17. Como o MDR, o IVDR e o AI Act abordam a “engenharia de usabilidade” para MDAI?

Embora o *MDR/IVDR* já exija documentação e avaliação clínica de desempenho, o *AI Act* introduz obrigações explícitas de transparência que contribuem para a explicabilidade, o que, por sua vez, facilita a responsabilidade (*MDR* e *IVDR*, Anexo I; Art. 13, *AI Act*).

A capacidade de desenvolvedores e *deployers* de demonstrar e comunicar como as decisões baseadas em *MDAI* são tomadas é fundamental para construir confiança e garantir a prestação de contas. Essa rastreabilidade robusta de mudanças e o controle pós-mercado reforçam a implantação segura e confiável do *MDAI*, aprimorando a responsabilidade ao longo do ciclo de vida do produto. A responsabilidade clara é um componente implícito para que o paciente se sinta seguro ao consentir.

O *MDR* e o *IVDR* exigem que os fabricantes apliquem princípios de “engenharia de usabilidade” aos *MDAIs* no projeto e desenvolvimento, para garantir seu uso seguro e eficaz pelos usuários pretendidos. Os fabricantes devem eliminar ou reduzir, na medida do possível, os riscos relacionados a erros de uso, levando em consideração o conhecimento do usuário e a necessidade de treinamento. Recomenda-se que os sistemas de IA, especialmente aqueles na área da saúde, sejam projetados com princípios centrados no usuário, para facilitar uma interação segura e eficaz. A documentação dos processos e resultados da engenharia de usabilidade é exigida pelas três regulamentações.

A tabela abaixo sintetiza muito bem esta relação de complementaridade.

Tabela 5 – Matriz de Requisitos Complementares:
MDR/IVDR vs. AI Act para MDAI de Alto Risco

| Domínio do Requisito | Foco Principal do MDR/IVDR | Complemento Estrutural do AI Act | Referência em AIB/MDCG |
|-------------------------------|---|--|-------------------------------|
| Gestão de Risco | Segurança e desempenho clínico do dispositivo. | Adiciona a obrigação de avaliar e mitigar riscos aos direitos fundamentais, vieses discriminatórios e robustez do sistema de IA. | Questão nº 7 |
| Governança de Dados | Dados clínicos devem ser robustos e confiáveis para a avaliação clínica/de desempenho. | Adiciona requisitos detalhados sobre a qualidade, representatividade e ausência de vieses nos dados de treinamento, validação e teste. | Questões nº 8 e 11 |
| Transparência | Fornecer informações claras sobre o propósito, operação e limitações do dispositivo ao usuário. | Adiciona a obrigação de projetar o sistema para que suas saídas sejam interpretáveis pelo implementador e de informar quando se interage com IA. | Questões nº 14 e 15 |
| Supervisão Humana | Requisitos de usabilidade para reduzir erros de uso pelo profissional. | Adiciona a obrigação de projetar o sistema com mecanismos que permitam a intervenção, supervisão e controle por um humano. | Questões nº 18 e 19 |
| Vigilância Pós-Mercado | Coleta sistemática de dados sobre o desempenho e a segurança do dispositivo após a comercialização. | Adiciona o monitoramento contínuo da conformidade com os requisitos da IA e a detecção de interações com outros sistemas. | Questões nº 32, 33 e 34 |

Fonte: Elaborada pelo autor (2025).

6.5.2 A Confluência para o Consentimento Livre e Esclarecido do Paciente

A harmonização desses requisitos pelo guia conjunto AIB/MDCG demonstra que o consentimento do paciente, no uso de MDAI, é concebido como um processo multifacetado e dinâmico. Não se trata apenas da assinatura de um documento, mas de um ciclo contínuo de confiança e informação.

6.5.2.1 Pré-condições Técnicas (Questões nº 22 e 23 do AIB/MDCG):

Questão 22. Quais medidas de segurança cibernética são exigidas pelo AI Act e pelo MDR/IVDR?

Questão 23. Quais critérios são especificados no AI Act para avaliar o desempenho do MDAI?

O *AI Act* especifica critérios para garantir a segurança, confiabilidade e eficácia de *MDAI* de alto risco. Embora não use explicitamente os termos “avaliação clínica” ou “avaliação de desempenho”, ele exige requisitos como precisão, robustez e cibersegurança, tanto na fase pré-comercialização como na fase pós-comercialização, garantindo que o sistema subjacente seja confiável, reduzindo os riscos inerentes à tecnologia (Artigo 61, *AI Act* e Anexo XIV *MDR*; Artigo 56, *AI Act* e Anexo XIV *IVDR*; Artigo 15 e Considerando 74 do *AI Act*). Isso fundamenta a confiança que o paciente pode depositar no sistema.

O Artigo 9º afirma especificamente que os testes de *MDAI* de alto risco devem ser realizados, conforme apropriado, a qualquer momento durante o processo de desenvolvimento.

6.5.2.2 Informação e Compreensão (Questões nº 14 e 15 do AIB/MDCG):

Questão 14. Quais são os requisitos impostos pelo AI Act e pelo MDR/IVDR em relação à transparência e ao MDAI?

Questão 15. Como o MDR/IVDR e o AI Act abordam a transparência, a explicabilidade e os requisitos de processamento de dados para MDAI de alto risco?

A transparência imposta pelo *AI Act* assegura que o *deployer* (profissional de saúde e, por extensão, o paciente) tenha acesso a informações abrangentes e compreensíveis sobre o *MDAI*, suas capacidades, limitações e riscos potenciais (Artigos 12 e 13 do *AI Act*, tal como esclarecido nos Considerandos 71 e 72 do *AI Act*).

A capacidade de explicar as decisões baseadas em IA é crucial para que o paciente entenda as recomendações ou ações propostas. O *AI Act* estabelece que a transparência, incluindo as instruções de uso que a acompanham, deve auxiliar os implementadores no uso do sistema e apoiar a tomada de decisões informadas por parte deles.

6.5.2.3 Controle e Autonomia (Questões nº 18, 19 e 20 do AIB/MDCG):

Questão 18. Que requisitos de supervisão humana estão incluídos no MDR, IVDR e AI Act para MDAI de alto risco?

Questão 19. A supervisão humana para dispositivos médicos pode ser entendida como parte das medidas de gestão de risco existentes, tais como uma medida de design (por exemplo, botão “parar”)?

Questão 20. Como o MDR/IVDR e o AI Act abordam o consentimento informado no contexto do MDAI?

A supervisão humana garante que o paciente não esteja à mercê de um algoritmo autônomo.

O *MDR/IVDR* e o *AI Act* incluem disposições destinadas a proteger os direitos individuais através da transparência e da interação informada com o *MDAI*. Nos termos do *MDR* e do *IVDR*, o consentimento informado é explicitamente exigido no contexto de investigações clínicas e estudos de desempenho (Art. 69, *MDR*; Art. 59, *IVDR*; Anexo I, *MDR* e *IVDR*). Estas disposições garantem que os indivíduos que participam nesses estudos sejam adequadamente informados sobre os riscos, benefícios e objetivos do *MDAI*.

O *AI Act* complementa esses requisitos introduzindo obrigações adicionais de transparência que se estendem à implantação geral de sistemas de IA de alto risco. Além disso, os requisitos de supervisão humana e transparência reforçam a necessidade de fornecer aos responsáveis pela implantação e às pessoas afetadas informações suficientes para compreender as capacidades, limitações e riscos potenciais do sistema (Art. 14, *AI Act*). Isso respeita profundamente a autonomia do paciente sobre seu próprio corpo e saúde.

Em síntese, a governança dupla da IA em dispositivos médicos na União Europeia, através da aplicação simultânea e complementar do *AI Act* e do *MDR/IVDR*, estabelece um paradigma regulatório que, embora complexo, é projetado para criar um ambiente onde o consentimento do paciente não é uma mera formalidade. Ele é reforçado por um compromisso intrínseco com a segurança, desempenho, transparência, explicabilidade e supervisão humana. Essas camadas de proteção e informação visam garantir que, ao interagir com o *MDAI*, o paciente possa tomar decisões verdadeiramente livres e esclarecidas, alinhadas com seus valores e sua compreensão do tratamento proposto.

6.6 SÍNTESE PARA UMA GOVERNANÇA PROATIVA E MULTISSETORIAL

À luz da opacidade e dos vieses que permeiam sistemas de IA e neurotecnologias, este Capítulo evidenciou que o Consentimento Livre e Esclarecido do paciente não pode operar como salvaguarda isolada: é preciso estruturá-lo sobre o tripé transparência/explicabilidade, equidade e *accountability*, traduzido em obrigações verificáveis ao longo de todo o ciclo de vida da tecnologia (SOLAIMAN *et al.*, 2025; MELLO *et al.*, 2025).

Tal arranjo desloca parte do ônus cognitivo do paciente para uma arquitetura de responsabilidades distribuídas – desenvolvedores, instituições de saúde, profissionais e reguladores – que combine interação dialógica com explicações compreensíveis ao paciente, qualificação contínua das equipes e procedimentos internos de governança (COHEN, 2019;

NOGAROLI, 2023; BEAUCHAMP; CHILDRESS, 2013; MESKE *et al.*, 2025). Em consonância com diretrizes internacionais, essa governança exige gestão iterativa de riscos, documentação técnica auditável, monitoramento pós-mercado e auditorias periódicas de viés, articulando transparência substantiva com supervisão humana contínua (OMS, 2024; UNIÃO EUROPEIA, 2024; NIST, 2023; SAENZ *et al.*, 2024; HURLEY *et al.*, 2025).

Nesse sentido, a abordagem regulatória europeia – reforçada pelo *AI Act* e pela integração setorial em saúde – ilustra como medidas *ex ante* aliviam o peso decisório no momento do consentimento e protegem a autonomia em contextos de alta complexidade técnico-epistêmica.

Assim, em apertada síntese, para se alcançar uma Governança Proativa e Multissetorial é necessário, em primeiro lugar, elevar os padrões de transparência e explicabilidade, de modo que médicos e pacientes não permaneçam “no escuro” quanto à finalidade, limites, fontes de dados, desempenho por subgrupos e incertezas dos sistemas de IA (OMS, 2024; NIST, 2023; UNIÃO EUROPEIA, 2024). No Brasil, *softwares* com finalidade médica (inclusive aqueles baseados em IA) já são regulados como dispositivos médicos pela ANVISA e devem trazer documentação técnica robusta – descrição do conjunto de dados de treino/validação, justificativa do método algorítmico, histórico de versões do modelo – como condição de segurança e eficácia (ANVISA, 2022). Na União Europeia, a dupla conformidade entre o *MDR* 2017/745 (dispositivos médicos) e o *AI Act* traduz essa exigência em práticas de governança de dados, testes de desempenho, mitigação de vieses, documentação extensa e previsão de supervisão humana (UNIÃO EUROPEIA, 2017a; UNIÃO EUROPEIA, 2024; AIB/MDCG, 2025). Nos Estados Unidos, a FDA vem detalhando requisitos para IA médica, como planos de modificação algorítmica previamente acordados (*PCCP*) e critérios para testes de segurança em *neurotech* implantável, tornando explicáveis e auditáveis os critérios de eficácia antes do uso em pacientes (FDA, 2021; YUSTE *et al.*, 2017). Em síntese, transparência e explicabilidade deixaram de ser meros ideais abstratos para se converterem em obrigações técnico-legais nos principais cenários regulatórios – um desenvolvimento crucial para sustentar o pilar “esclarecido” do consentimento.

Em segundo lugar, o próprio processo de obtenção do consentimento precisa ser qualificado em consonância com a arquitetura de transparência/explicabilidade, equidade e *accountability* delineada neste Capítulo: materiais e formulários em linguagem clara e acessível, desenho em camadas (compreensão básica e informação aprofundada), estratégias de verificação de entendimento, utilização de recursos visuais e ferramentas de IA explicativa (MESKE *et al.*, 2025), bem como um diálogo clínico genuíno, com tempo e espaço para perguntas e para a expressão de dúvidas e preferências. Trata-se de substituir a burocracia por

deliberação compartilhada e educação em saúde digital, instituindo um consentimento contínuo e adaptativo sempre que houver novas evidências, alteração de riscos ou surgimento de alternativas terapêuticas (CHAR *et al.*, 2018; COHEN, 2019; KAYE *et al.*, 2015; OMS, 2024; UNESCO, 2025). No plano operativo, isso inclui a divulgação explícita de que a decisão clínica foi assistida por IA, a explicitação de incertezas epistêmicas e limitações do sistema, a apresentação de cursos de ação alternativos – inclusive a possibilidade de segunda opinião e de recusa sem prejuízo do cuidado – e o registro no prontuário de tais informações, com trilhas de auditoria para fins de supervisão e melhoria contínua (OMS, 2024; UNIÃO EUROPEIA, 2024). Nesse sentido, o CLE deixa de ser um evento pontual para tornar-se um processo vivo que acompanha a jornada do paciente, fortalecendo a confiança na relação paciente-médico e prevenindo a erosão silenciosa da autonomia por falta de compreensão ou por desatualização.

Em terceiro lugar, são necessárias salvaguardas adicionais para além do ato de consentir: marcos regulatórios que imponham avaliação de risco, mitigação de viés, explicabilidade operacional e monitoramento pós-mercado; comitês clínico-éticos para casos sensíveis (especialmente no campo da *neurotech*, onde os riscos à identidade requerem escrutínio especial); treinamento profissional sobre os limites, vieses e responsabilidades no uso de modelos algorítmicos; e, quando pertinente, novos direitos ligados à proteção da mente (por exemplo, liberdade cognitiva, direito à continuidade psicológica) – medidas já em discussão ou implementação (UNIÃO EUROPEIA, 2024; ANVISA, 2022; NIST, 2023; OMS, 2024; YUSTE *et al.*, 2017; MORALES; YUSTE, 2022), como será demonstrado no próximo Capítulo. Tais medidas deslocam a proteção do indivíduo isolado para um ecossistema de responsabilidade compartilhada (UNIÃO EUROPEIA, 2024; ANVISA, 2022; FDA, 2021).

Superar o Consentimento Livre e Esclarecido como único escudo não é descartá-lo, mas fortalecê-lo e complementá-lo. O equilíbrio desejado é aquele em que inovação e direitos fundamentais caminhem juntos: que os avanços da IA e das neurociências sirvam ao bem-estar humano sem suplantarem a autonomia, a dignidade e a segurança do paciente (OMS, 2024; UNIÃO EUROPEIA, 2024; TOPOL, 2019). Pacientes adequadamente informados – não só sobre riscos clínicos, mas também sobre as nuances tecnológicas e suas incertezas – tornam-se coparticipantes das decisões, e não sujeitos passivos de uma autoridade algorítmica opaca (COHEN, 2019; CHAR *et al.*, 2018; SOLAIMAN *et al.*, 2025; MELLO *et al.*, 2025). Conquistam-se, assim, dois objetivos interligados: fomentar a confiança pública nas novas tecnologias e assegurar que tal confiança não seja traída por falta de transparência ou de prestação de contas (UNESCO, 2025; THE WHITE HOUSE, 2022; NIST, 2023).

Em última instância, a solução delineada é multifacetada: tecnologia mais explicável, profissionais mais preparados, pacientes mais engajados e normas mais protetivas. Somente com essa defesa em profundidade será possível colher os frutos da medicina do futuro, minimizando seus riscos e respeitando integralmente os direitos e valores que alicerçam a prática médica ética (PASQUALE, 2020; OMS, 2024; UNIÃO EUROPEIA, 2024). Trata-se de um verdadeiro *update* do pacto bioético: da confiança ingênua no consentimento individual para a confiança construída coletivamente através de transparência, equidade e *accountability*. Assim, a autonomia do paciente – entendida não apenas como liberdade abstrata, mas como autodeterminação efetiva – poderá florescer mesmo em meio a algoritmos complexos e interfaces neuronais, pois estará resguardada por todas as instâncias do sistema: pela engenharia responsável, pela regulação inteligente, pela ética institucional e, claro, pelo contínuo respeito à vontade informada de cada pessoa. Somente com essa aliança multidisciplinar e multinível conseguiremos garantir que as tecnologias do amanhã trabalhem a favor, e nunca contra, da dignidade humana.

Portanto, é precisamente essa passagem – da governança do consentimento à governança da tecnologia – que orienta o próximo Capítulo, dedicado a examinar, em perspectiva comparada, como Brasil (e América Latina), União Europeia e Estados Unidos têm desenhado respostas regulatórias para IA e neurotecnologias em saúde: da dupla camada europeia (*MDR/IVDR + AI Act*) e de seus deveres de transparência e supervisão, às propostas brasileiras que combinam um marco geral de IA e iniciativas de neurodireitos, passando pela abordagem setorial e *ex post* norte-americana, com o protagonismo de agências como ANVISA e FDA. Ao avançarmos, investigaremos como esses três modelos – principiológico, preventivo por risco e setorial reativo – podem conferir densidade jurídica às salvaguardas aqui delineadas e, assim, tornar o consentimento verdadeiramente exequível na saúde algorítmica.

7 REGULAÇÃO DE INTELIGÊNCIA ARTIFICIAL E NEUROTECNOLOGIAS NA SAÚDE DIGITAL: UM PANORAMA COMPARATIVO

Este Capítulo examina a emergência de quadros regulatórios para inteligência artificial (IA) e neurotecnologias, aprofundando e contrastando as abordagens filosóficas e práticas adotadas pelo Brasil, União Europeia e Estados Unidos. Argumenta-se que, enquanto a inovação tecnológica avança em escala global, as respostas regulatórias divergem fundamentalmente, refletindo distintas tradições jurídicas e prioridades sociopolíticas. O Capítulo demonstrará a existência de três modelos predominantes: uma abordagem baseada em novos direitos fundamentais, proeminente na América Latina; um modelo preventivo e baseado em risco, característico da Europa; e uma abordagem setorial e reativa, dominante nos Estados Unidos.

A estrutura do Capítulo começa com a gênese conceitual dos neurodireitos e termina com as implicações práticas dos diferentes modelos regulatórios. Inicia-se com a trajetória do debate acadêmico que culminou na positivação jurídica dos neurodireitos, passando para uma análise detalhada das iniciativas legislativas brasileiras. Posteriormente, a análise volta-se para a governança técnica exercida por agências sanitárias, como a Agência Nacional de Vigilância Sanitária (ANVISA) no Brasil e a *Food and Drug Administration* (FDA) nos Estados Unidos. Por fim, realiza-se uma análise comparativa detalhada com os modelos europeu e norte-americano, culminando em considerações sobre os caminhos futuros para a regulação no Brasil, visando a um equilíbrio entre a promoção da inovação e a proteção intransigente da dignidade humana.

7.1 A GÊNESE DOS NEURODIREITOS: DO DEBATE ACADÊMICO À POSITIVAÇÃO JURÍDICA

A discussão sobre a necessidade de novos direitos voltados à proteção da esfera mental não é um fenômeno isolado ou repentino, mas sim o resultado de um processo de maturação conceitual que, inicialmente formulado no campo acadêmico, migrou para a arena política e regulatória. Os estudos filosófico-jurídicos sobre neurociência (neuroética e neurodireito) têm dado crescente proeminência à análise normativa dos desafios ético-legais nas ciências da mente e do cérebro em termos de direitos e liberdades (IENCA, 2021). A base para os neurodireitos foi lançada por uma série de estudos acadêmicos que defendiam a liberdade

cognitiva como um direito fundamental. O termo “neurodireito” foi usado pela primeira vez em 1991 por Taylor, Harp e Elliot (DANTAS; JEREISSATI, 2024).

O conceito de “neurodireitos” surgiu como uma resposta ética e jurídica à crescente capacidade das neurotecnologias de captar, interpretar e manipular dados neurais, gerando implicações diretas para a autonomia, a privacidade mental e a integridade psicológica dos indivíduos (YUSTE *et al.*, 2017; PRICE II; COHEN, 2019), demonstrando sua evolução, de uma preocupação ética formulada por cientistas, para uma pauta legislativa concreta em diversas jurisdições, ilustrando um fascinante processo de difusão normativa transnacional (DANTAS; JEREISSATI, 2024).

O neurobiólogo Rafael Yuste e sua *NeuroRights Initiative*, vinculada à Columbia University, também tiveram um papel fundamental na formalização e divulgação desses conceitos globalmente, como será abordado a seguir.

7.1.1 O Chamado Acadêmico por Novas Garantias

A convergência entre a inteligência artificial e as neurotecnologias, em especial as interfaces cérebro-computador (*Brain-Computer Interfaces – BCIs*), tem despertado crescente preocupação na comunidade científica internacional, em virtude de seu potencial de impacto direto sobre a autonomia, a privacidade mental e os direitos fundamentais dos indivíduos (YUSTE *et al.*, 2017). Inclusive, no Capítulo 5 foi demonstrado como os *BCIs* e *DBSs* podem impactar a integridade física e psicológica do paciente, limitar a sua própria autonomia (item 5.2.2). O debate sobre essas questões se intensificou a partir da publicação do artigo “*Four ethical priorities for neurotechnologies and AI*”, na revista *Nature* em 2017, por um grupo multidisciplinar de vinte e cinco especialistas – conhecido como *Morningside Group* – liderado pelo neurocientista Rafael Yuste.

O artigo soou um alarme sobre os riscos sem precedentes que essas tecnologias representavam para a humanidade. Yuste *et al.* (2017) argumentaram que, embora promissoras para o tratamento de doenças como Parkinson e Alzheimer, as neurotecnologias poderiam também ser usadas para decodificar processos mentais, manipular o comportamento e alterar o senso de identidade de uma pessoa. Eles identificaram quatro áreas que demandavam atenção ética e regulatória urgente: (1) *privacidade e consentimento*, defendendo que a venda de dados neurais deveria ser estritamente proibida e que os indivíduos deveriam ter o direito de optar por não compartilhar seus dados cerebrais; (2) *agência e identidade*, alertando para a possibilidade de que as interfaces cérebro-máquina pudessem borrar a linha entre a consciência humana e os

comandos tecnológicos, afetando a agência e o senso de si; (3) *vieses*, destacando o perigo de que algoritmos treinados com dados enviesados pudessem perpetuar ou amplificar preconceitos existentes; e (4) *aprimoramento humano*, questionando as implicações de equidade e justiça social caso as neurotecnologias fossem usadas para aumentar as capacidades cognitivas de alguns, criando novas formas de desigualdade.

A conclusão do grupo foi contundente: os marcos éticos existentes, como a Declaração de Helsinque, eram insuficientes. Era imperativo que a comunidade internacional agisse para adicionar novas cláusulas aos tratados de direitos humanos, propondo a necessidade de proteção das características humanas essenciais: a vida mental privada, a agência individual e a compreensão dos indivíduos como entidades limitadas por seus corpos (YUSTE *et al.*, 2017), isto é, a criação de “neurodireitos”. Essa formulação acadêmica estabeleceu a base conceitual sobre a qual as primeiras iniciativas legislativas seriam construídas.

Como analisado no Capítulo 5, de maneira aprofundada, segundo Dantas e Jereissati, estudos qualitativos com pacientes submetido a *DBS* e *BCI* já indicam “alterações significativas no senso de identidade e de controle sobre as próprias ações após o uso prolongado da tecnologia”. Neurotecnologias podem “alterar o sentido e conteúdo da autonomia da pessoa” e “perturbar o senso de identidade e o autogoverno da pessoa” (DANTAS; JEREISSATI, 2024), e as implicações de longo prazo para a saúde mental carecem de evidências científicas (ANPD, 2025), confirmando, assim, as previsões daqueles autores.

Só a título de exemplo, o impacto disso para o CLE, objeto central dessa dissertação, é imenso! Conforme desenvolvido também no Capítulo 5, o consentimento para o uso dessas neurotecnologias implica aceitar o risco de uma potencial alteração dessa identidade – risco que, devido à opacidade dos mecanismos envolvidos (a “dupla caixa-preta”, como denominamos), é impossível de ser plenamente dimensionado *a priori*. A autonomia, nesse contexto, é exercida para autorizar um procedimento que pode, paradoxalmente, modificar a própria base da qual essa autonomia emana. Esse cenário se conecta diretamente ao debate sobre a necessidade de um “direito à continuidade psicológica” como nova salvaguarda fundamental, destinado a proteger a integridade mental frente a tecnologias neuroinvasivas (YUSTE *et al.*, 2017).

7.1.2 O Pioneirismo do Chile: A Constitucionalização da Proteção Cerebral

O Chile destacou-se como a primeira nação a traduzir essa preocupação acadêmica em norma jurídica de mais alta hierarquia. Em um movimento pioneiro, o Congresso chileno

aprovou e, em 25 de outubro de 2021, foi publicada a Lei nº 21.383, que alterou o artigo 19 de sua Constituição Política (CHILE, 2021). A emenda, promulgada em 14 de outubro do mesmo ano, o nº 1º daquele artigo, que trata do direito à vida e à integridade física e psíquica, com a seguinte redação:

“O desenvolvimento científico e tecnológico estará ao serviço das pessoas e será realizado com respeito pela vida e pela integridade física e psíquica. A lei regulamentará os requisitos, condições e restrições para a sua utilização em pessoas, devendo proteger especialmente a atividade cerebral, bem como a informação proveniente dela” (CHILE, 2021).

Esta alteração constitucional representa a primeira positividade explícita da proteção da atividade e dos dados cerebrais em nível fundamental no mundo (DANTAS; JEREISSATI, 2024). Ao fazê-lo, o Chile não apenas respondeu ao chamado de Yuste e seus colegas, mas também estabeleceu um precedente poderoso, servindo de inspiração direta para propostas subsequentes em toda a América Latina, incluindo as que hoje tramitam no Brasil, como a PEC nº 29/2023, apresentada pelo Senador Randolfe Rodrigues, que busca alterar o Artigo 5º da Constituição Federal de 1988 para incluir a proteção à integridade mental e à transparência algorítmica como direitos e garantias fundamentais; o PL nº 522/2022, que visa modificar a Lei Geral de Proteção de Dados (LGPD) para conceituar e regulamentar o dado neural, considerando-o dado pessoal sensível; o anteprojeto de reforma do Código Civil brasileiro também acolhe a categoria de neurodireitos, alinhando-se a essa discussão e, por fim, um alteração regional, que ocorreu no estado do Rio Grande do Sul, já incorporando a proteção da identidade mental em sua Constituição Estadual (Emenda Constitucional nº 85/2023), sendo a primeira constituição brasileira a regular a temática desde uma perspectiva de direitos humanos (DANTAS; JEREISSATI, 2024).

7.1.3 A Abordagem Principiológica da Espanha

A Espanha adotou uma abordagem distinta, porém igualmente influente. Em vez de uma emenda constitucional, o governo espanhol publicou, em julho de 2021, a Carta de Direitos Digitais (ESPANHA, 2021). Este documento, embora não possua força de lei vinculante, sendo, portanto, desprovido de caráter cogente (DANTAS; JEREISSATI, 2024) foi concebido como um guia para futuros projetos legislativos e para orientar a implementação de políticas públicas, funcionando como um instrumento de *soft law* (ESPANHA, 2021).

A Carta é notável por sua Seção XXIV, intitulada “Direitos digitais no emprego das neurotecnologias” (ESPANHA, 2021). Este trecho reconhece explicitamente a necessidade de

uma regulação específica para a área, estabelecendo que as leis deverão garantir o controle do indivíduo sobre o uso de seus dados neurais e assegurar que as decisões e processos baseados em neurotecnologias não sejam condicionados por fornecedores de tecnologia. De forma crucial, essa Seção afirma que será regulada especialmente a utilização de interfaces pessoa-máquina que possam afetar a integridade física ou psíquica, garantindo-se o direito à autodeterminação e a decisão livre e informada da pessoa, como se observa a seguir, *verbis*:

XXIV

Direitos digitais no uso de neurotecnologias

1. As condições, limites e garantias de implantação e uso de neurotecnologias em pessoas serão regulamentados por lei com o objetivo de:
 - a. Preservar a identidade individual como consciência da pessoa sobre si mesma.
 - b. Garantir a autodeterminação individual, soberania e liberdade na tomada de decisões.
 - c. Assegurar a confidencialidade e a segurança dos dados obtidos ou relativos aos seus processos cerebrais e o pleno domínio e disposição sobre os mesmos.
 - d. Ordenar o uso de interfaces pessoa-máquina suscetíveis de afetar a integridade física ou psíquica.
 - e. Assegurar que as decisões e processos baseados em neurotecnologias não sejam condicionados pelo fornecimento de dados, programas ou informações incompletas, indesejadas, desconhecidas ou tendenciosas, ou por intromissão em conexões neuronais.
2. Para garantir a dignidade da pessoa, a igualdade e a não discriminação, e de acordo com os tratados e convenções internacionais, a lei regulará os casos e condições de emprego das neurotecnologias que, além de sua aplicação terapêutica, visem o aumento cognitivo ou a estimulação ou potencialização das capacidades das pessoas. (ESPANHA, 2021).

A trajetória dos neurodireitos, portanto, exhibe um padrão claro de difusão normativa que se move do âmbito epistêmico para o legal. O conceito, formulado por uma comunidade de especialistas que identificou um novo risco tecnológico, foi adotado por um “empreendedor normativo” – o Chile – que criou um modelo legal vinculante. Subsequentemente, outras jurisdições, como o Brasil, começaram a debater propostas inspiradas nesse modelo, enquanto outras, como a Espanha, adaptaram a ideia ao seu próprio contexto jurídico e político, optando por uma declaração de princípios. Este padrão demonstra que a regulação de tecnologias emergentes é um fenômeno jurídico transnacional, uma dimensão essencial para qualquer análise comparativa.

Porém, a Europa não possui Regulamento específico. A proteção à privacidade mental e à autonomia cognitiva é assegurada de forma indireta, por meio de um ecossistema jurídico interconectado (CORNEJO-PLAZA; CIPPITANI, 2024). O Regulamento Geral sobre a Proteção de Dados (*General Data Protection Regulation – GDPR*), aplicável desde 2018, já classifica os dados de saúde e os dados biométricos como “categorias especiais de dados

peçoais” (ou dados pessoais sensíveis), exigindo consentimento explícito e a existência de uma base legal sólida para seu tratamento (UNIÃO EUROPEIA, 2016, art. 9º). Embora a norma não mencione expressamente os dados neurais, a literatura especializada aponta que estes, por sua natureza altamente sensível e potencial para revelar informações íntimas sobre estados mentais e processos cognitivos, se enquadrariam no conceito legal de dados biométricos e de saúde (SALLES *et al.*, 2024; YUSTE *et al.*, 2017).

Vimo no item 6.5, acima, o sistema de “*Governança Dupla da IA em Dispositivos Médicos: A Complementaridade entre o AI Act e as Regulamentações MDR/IVDR na União Europeia*”), que também são aplicados subsidiariamente às neurotecnologias. Dentro dessa estrutura de governança dupla, o consentimento livre e esclarecido do paciente emerge como um pilar fundamental, sendo intrinsecamente moldado e reforçado pelas disposições combinadas do Regulamento de Dispositivos Médicos (*MDR*), do Regulamento de Dispositivos Médicos para Diagnóstico *in vitro* (*IVDR*)³⁴ e da Lei de Inteligência Artificial europeia (*AI Act*), que foram disciplinadas pelo AIB/MDCG. Em junho de 2025, o Conselho de Inteligência Artificial e o Grupo de Coordenação de Dispositivos Médicos elaboraram um documento conjunto para integrar o arcabouço regulatório europeu para a Inteligência Artificial em Dispositivos Médicos (*MDAI*)³⁵, que também aponta para a abordagem sistêmica e integrada que permeia a proteção do paciente.

Adicionalmente, a Carta dos Direitos Fundamentais da União Europeia garante, nos artigos 7º, 8º e 3º, respectivamente, o direito ao respeito pela vida privada e familiar e o direito à integridade física e mental, dispositivos que podem servir de base para contestar intervenções neurotecnológicas abusivas (UNIÃO EUROPEIA, 2012). Assim, embora não exista um “marco europeu dos neurodireitos” formalizado, a combinação entre o *GDPR* e a Carta fornece instrumentos jurídicos potentes para salvaguardar a esfera mental e limitar práticas intrusivas decorrentes de tecnologias neuroassistidas.

Porém, como afirmam Andrea Lavazza *et al.* (2025), essa multiplicidade de diferentes fontes legais é, na verdade, um desafio notável para a Europa. O *GDPR*, *AI Act* e *MDR*, que se aplicam às *BCIs*, podem criar conflitos de leis, dificuldade de interpretação e incerteza para cidadãos e empresários. Os autores desse estudo, intitulado “*Neuralink’s brain-computer*

³⁴ UE - Regulamento (UE) 2017/745 e Regulamento (UE) 2017/746.

³⁵ AIB/MDCG. Joint Artificial Intelligence Board and Medical Device Coordination Group Document. AIB 2025-1 / MDCG 2025-6 - FAQ on Interplay between the Medical Devices Regulation & In vitro Diagnostic Medical Devices Regulation and the Artificial Intelligence Act (June 2025). Disponível em: https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf. Acesso em: 7 jul. 2025.

interfaces: medical innovations and ethical challenges”, dentre os quais está Marcelo Ienca, afirmam que, *verbis*:

“[...] os legisladores de ambos os lados do Atlântico parecem estar principalmente preocupados com os riscos para a segurança e a saúde associados à sua utilização, negligenciando o impacto na integridade mental, na autonomia, na liberdade de pensamento e nos valores a eles relacionados. À luz do exposto, desde 2019, organizações internacionais como o Conselho da Europa (Relatório, COE, 2021), a OCDE (9 princípios, OCDE, 2019) e a UNESCO (relatório de 2021 e recomendação futura; UNESCO, 2021) têm respondido a esta lacuna de governança adotando declarações que abordam os aspectos éticos e de direitos humanos das neurotecnologias. Na mesma linha, o Conselho de Direitos Humanos da ONU adotou recentemente um relatório elaborado pelo Comitê Consultivo de Direitos Humanos sobre o impacto, as oportunidades e os desafios da neurotecnologia na promoção e proteção de todos os direitos humanos (Conselho de Direitos Humanos da ONU, 2024)” (LAVAZZA *et al.*, 2025).

A UNESCO, em maio de 2025, publicou o Projeto de texto da Recomendação sobre a Ética da Neurotecnologia (prevista para adoção em novembro de 2025), onde defende uma abordagem baseada em direitos humanos e centrada no ser humano. O documento exige a proteção da autonomia e liberdade de pensamento e a proteção da integridade física e mental contra qualquer interferência indesejada e prejudicial (UNESCO, 2025).

Por fim, vale destacar que a OCDE (Organização para a Cooperação e Desenvolvimento Econômico) tem exercido um papel importante também na luta pelo reconhecimento dos neurodireitos. Em 2019, ela publicou a *“Recommendation on Responsible Innovation in Neurotechnology”*, enfatizando a autonomia, liberdade cognitiva e a proteção de dados cerebrais (OCDE, 2019). Em julho de 2025, a OCDE publicou um guia abrangente, destinado a auxiliar formuladores de políticas na implementação da Recomendação sobre Inovação Responsável em Neurotecnologia. Este documento fundamental, o primeiro padrão internacional na área, define neurotecnologia e estabelece nove princípios essenciais para governar seu desenvolvimento, como a promoção da inovação responsável, a priorização da avaliação de segurança, a salvaguarda de dados cerebrais pessoais e a antecipação de usos indevidos. O *toolkit* (“kit de ferramentas”) é organizado em 13 objetivos temáticos, agrupados em cinco elementos – Valores Orientadores, Inteligência Estratégica e Supervisão, Engajamento das Partes Interessadas, Regulamentação Ágil e Cooperação Internacional –, que visam aproveitar as neurotecnologias para o benefício social, antecipar desafios de governança e construir capacidades regulatórias de longo prazo. Através de exemplos práticos e referências a estudos e eventos, o documento não só detalha a história e a implementação da

Recomendação, mas também oferece ações concretas para integrar considerações éticas, legais e sociais em todo o ciclo de vida da neurotecnologia, promovendo uma cultura de confiança e responsabilidade globalmente (OCDE, 2025).

7.1.4 Principais Categorias e Princípios dos Neurodireitos

Há, ao menos, três correntes sobre como regular essas tecnologias emergentes, considerando os marcos normativos vigentes, seus potenciais impactos e os desafios que impõem aos direitos humanos e fundamentais. A primeira sustenta a necessidade de criar novos direitos específicos para a tutela do cérebro e da mente – os neurodireitos – por entender que o repertório atual de direitos e liberdades não basta para prover proteção jurídica adequada frente a usos indevidos das neurotecnologias. Nessa chave, despontam como centrais o direito à liberdade cognitiva, o direito à privacidade mental, o direito à integridade mental e o direito à continuidade psicológica (IENCA; ANDORNO, 2017; YUSTE *et al.*, 2017).

A segunda corrente defende que novos direitos não são necessários: bastaria reinterpretar e atualizar os direitos já existentes à luz do novo contexto sociotécnico, ajustando sua aplicação às transformações trazidas pelas neurotecnologias. Seus proponentes dispensam a criação de neurodireitos e advogam uma hermenêutica evolutiva dos sistemas de direitos humanos e fundamentais, adequada ao cenário de desenvolvimento e adoção progressiva dessas tecnologias (RAINEY *et al.*, 2020).

Por fim, a terceira corrente entende ser desnecessária qualquer reforma legislativa ou mudança interpretativa: o direito vigente, tal como já interpretado, seria suficiente para enfrentar os desafios colocados pelas neurotecnologias – posição hoje minoritária, mas ainda presente no debate (LIGTHART *et al.*, 2023).

A proteção contra alterações do encéfalo, que possam mudar o conceito que uma pessoa tem de si mesma, é um dos pilares dos neurodireitos. Por isso, com base na primeira corrente, as propostas de neurodireitos variam ligeiramente em sua formulação, mas geralmente convergem em torno de categorias-chave, conforme tabela abaixo:

Com base nas informações fornecidas, aqui está uma tabela que sintetiza as principais categorias e princípios dos neurodireitos:

Tabela 6 – Principais Categorias e Princípios dos Neurodireitos

| Categoria/Princípio | Ienca e Andorno (2017) | Yuste <i>et al.</i> (2017) <i>NeuroRights Initiative</i> | Princípios Interamericanos da OEA (2023) |
|----------------------------|-------------------------------|---|---|
|----------------------------|-------------------------------|---|---|

| | | | |
|--|--|--|--|
| <p>Autonomia e Agência</p> | <p>Direito à liberdade cognitiva: Protege a capacidade de escolha e controle sobre os próprios processos cognitivos, vedando o uso coercitivo de neurotecnologias.</p> | <p>Direito ao livre-arbítrio (agência): Garante que as pessoas possam controlar suas próprias decisões, sem manipulações externas.</p> | <p>Autonomia da atividade neuronal: Garante o direito de desfrutar de uma vida digna, juntamente com os benefícios do progresso científico e tecnológico, preservando os direitos relativos à identidade, autonomia e livre desenvolvimento da personalidade.</p> <p>Integridade neurocognitiva: Assegura a integridade do pensamento e da consciência contra usos ilegítimos.</p> |
| <p>Privacidade Mental</p> | <p>Direito à privacidade mental: Assegura que dados neurais, pensamentos e informações mentais não sejam acessados ou utilizados sem consentimento explícito.</p> | <p>Direito à privacidade mental: Enfatiza a confidencialidade dos dados neurais e o direito de serem excluídos.</p> | <p>Privacidade da atividade neuronal: (v. acima).</p> <p>Dados neurais como dados pessoais sensíveis: Exige medidas reforçadas de privacidade e segurança.</p> |
| <p>Identidade e Integridade</p> | <p>Direito à integridade mental: Protege contra a manipulação do comportamento, pensamento ou consciência.</p> <p>Direito à continuidade psicológica: Protege contra alterações abusivas na dimensão mental e na identidade.</p> | <p>Direito à identidade pessoal: Protege contra alterações cerebrais que possam modificar o conceito de si mesmo de uma pessoa.</p> | <p>Identidade da atividade neuronal: (v. acima).</p> |
| <p>Equidade e Não Discriminação</p> | <p>(Não abordado diretamente)</p> | <p>Direito ao acesso equitativo ao aumento mental/cognitivo: Garante que os benefícios de aprimoramento sejam distribuídos de forma justa.</p> <p>Direito à proteção contra vieses algorítmicos:</p> | <p>Igualdade, não discriminação e acesso equitativo: Garante que o desenvolvimento e uso sejam acessíveis a todos, sem ampliar desigualdades.</p> |

| | | | |
|-------------------------------|----------------------------|--|---|
| | | Resguarda contra a “neurodiscriminação”. | |
| Governança e Regulação | (Não abordado diretamente) | (Não abordado diretamente) | <ul style="list-style-type: none"> - Proteção dos Direitos Humanos desde o <i>design</i>. - Consentimento expresso e informado para dados neurais. - Aplicação terapêutica exclusiva (com limites para aumento de capacidades). - Governança transparente. - Supervisão e fiscalização pelo Estado. - Acesso à tutela efetiva e acesso a recursos associados ao desenvolvimento e uso das neurotecnologias. |

Fonte: Elaborada pelo autor (2025).

Em linhas gerais, portanto, as propostas de neurodireitos convergem para um mesmo núcleo: proteção da esfera mental contra interferências indevidas; garantia de autonomia / agência na tomada de decisão; privacidade dos dados neurais com consentimento expresso e revogável; e salvaguarda da integridade e identidade pessoais – base sobre a qual se assentam também deveres de transparência, não discriminação e acesso equitativo.

Algumas diferenças, porém, aparecem no enfoque dado por cada matriz. Em Ienca e Andorno (2017), o foco recai em direitos “negativos” de tutela contra intervenções: liberdade cognitiva; privacidade mental; integridade mental e continuidade psicológica, isto é, um escudo para impedir manipulações, acessos ou alterações não consentidas da vida mental. Yuste e a *NeuroRights Initiative*, em 2019, preserva esses pilares, mas acrescenta um vetor “positivo” e distributivo: além de identidade pessoal, livre-arbítrio e privacidade mental, defende acesso equitativo a eventuais aumentos cognitivos e a proteção contra vieses algorítmicos (combatendo a “neurodiscriminação”), alinhando direitos mentais à justiça algorítmica. Já os Princípios Interamericanos da OEA (2023), além de manter as ideias e conceitos dos autores anteriores, operam como um quadro de governança: qualificam dados neurais como sensíveis; exigem proteção dos direitos humanos *by design*, consentimento expresso e informado para coleta/uso; igualdade e não discriminação; limites ao *enhancement* (“aperfeiçoamento”), com primazia terapêutica; asseguram integridade neurocognitiva e impõem transparência, supervisão e fiscalização estatais. Em suma, enquanto Ienca e Andorno estruturam um núcleo protetivo de

direitos mentais, Yuste integra equidade e anti-vieses ao catálogo, e a OEA traduz o conjunto em regras de implementação e *enforcement*³⁶.

7.1.4.1 Justificativa e Desafios para a Implementação dos Neurodireitos

Em apertada síntese, a necessidade de neurodireitos é impulsionada pela capacidade das neurotecnologias de: (i) *Acessar e manipular a mente* - Podem decodificar emoções, manipular pensamentos e até prever comportamentos, representando riscos para a autonomia, privacidade e a própria natureza humana (DANTAS; JEREISSATI, 2024); (ii) *Coleta inconsciente de dados neurais* - Muitas operações de tratamento de dados neurais podem ocorrer sem a consciência do titular, desafiando o conceito de consentimento livre e informado (ANPD, 2025); (iii) *Vieses algorítmicos e discriminação* - A combinação com IA e *machine learning* pode introduzir ou amplificar vieses, levando a “neurodiscriminação” em áreas como seguros, emprego e educação (DANTAS; JEREISSATI, 2024; ANPD, 2025); (iv) *Lacunas regulatórias* - A legislação existente, como a Lei Geral de Proteção de Dados (LGPD) no Brasil, abrange dados pessoais em geral, mas há debate sobre a suficiência da proteção para os dados neurais, que são considerados hipersensíveis (DANTAS; JEREISSATI, 2024; LAVAZZA *et al.*, 2025); e (v) *“Transhumanismo” e aprimoramento* - A busca por aprimoramento humano levanta questões éticas profundas sobre o que significa ser humano e a equidade no acesso a essas tecnologias (DANTAS; JEREISSATI, 2024; UNESCO, 2025).

Ou seja, a defesa dos neurodireitos decorre do poder inédito das neurotecnologias de acessar, inferir e até influenciar estados mentais – decodificando emoções, modulando pensamentos e antecipando comportamentos – com impactos diretos sobre autonomia, privacidade e a própria identidade pessoal. Além disso, dispositivos neuromoduladores invasivos (como implantes utilizados em *BSI*, *BCI* e Estimulação Cerebral Profunda - *DBS*) estão sujeitos ao risco de “sequestro cerebral” (*brainjacking*) (HOOPER, 2025).

O *brainjacking* ocorre quando um “atacante” malicioso obtém o controle não autorizado do dispositivo intracraniano (ANPD, 2025). Uma vez no controle, o invasor pode realizar uma variedade de ataques, causando diferentes tipos de danos e violações de direitos, como à privacidade e à saúde. A manipulação externa de *BCIs* ou neurotecnologias pode influenciar a tomada de decisão ou o controle do indivíduo, violando o direito à integridade mental e ao livre

³⁶ O conceito de *enforcement*, no contexto dos Dispositivos Médicos de Inteligência Artificial (*MDAI*) de alto risco, refere-se aos mecanismos regulatórios e procedimentais exigidos para garantir que esses sistemas **cumpram integralmente** os requisitos do Regulamento de Dispositivos Médicos (*MDR/IVDR*) e do Regulamento da IA (*AI Act*) ao longo de todo o seu ciclo de vida (AIB/MDCG, 2025).

arbitrio (DANTAS; JEREISSATI, 2024; NEGRI *et al.*, 2025; LAVAZZA *et al.*, 2025; HOOPER, 2025). Essa possibilidade levanta a questão de como determinar a responsabilidade moral e legal por ações realizadas através de uma *BCI* comprometida (LAVAZZA *et al.*, 2025).

Esse cenário é agravado por duas frentes: de um lado, a coleta e o tratamento inconscientes de sinais neurais, que esvaziam a ideia de consentimento livre e esclarecido tal como hoje praticado; de outro, a acoplagem com IA e *machine learning*, capaz de introduzir ou amplificar vieses e produzir “neurodiscriminação” em esferas sensíveis (emprego, seguros, educação). Some-se a isso as lacunas regulatórias: embora marcos gerais de proteção de dados – como a LGPD (aqui no Brasil), *GDPR* e *AI Act* (na Europa) – ofereçam salvaguardas importantes, persiste a dúvida sobre sua suficiência diante de neurodados hipersensíveis, cuja captação, inferência e reuso exigem controles mais finos. Por fim, a agenda de aprimoramento humano (transumanismo) desloca o debate para questões de justiça distributiva e de limites éticos: quem terá acesso aos incrementos cognitivos? Em que condições? Com que efeitos sobre igualdade e sobre o próprio significado do “ser humano”?

A linha tênue entre terapia e aprimoramento também deve ser discutida, porque pode ser fonte de discriminação, aumentando a vulnerabilidade: o objetivo é restaurar a “norma” ou superá-la? (CABRAL, 2024). Como analisado anteriormente, a *BSI*, embora atualmente terapêutica (restaurando a marcha, por exemplo), ilustra a tensão entre o uso médico e o uso não médico da neurotecnologia. O objetivo do desenvolvimento de *BCIs* é, portanto, duplo: “restaurar a autonomia de pessoas com necessidades médicas não atendidas hoje e desbloquear o potencial humano amanhã” (aprimoramento) (LAVAZZA *et al.*, 2025). Ocorre que o uso para aprimoramento (melhorar a locomoção de pessoas sem deficiência, por exemplo) é um risco explícito que pode exacerbar desigualdades sociais e criar novas formas de discriminação entre os “aprimorados” e os “naturais” (LAVAZZA *et al.*, 2025; PACHECO, 2025; UNESCO, 2025).

Por isso, a regulação deve estabelecer limites claros (UNESCO, 2025). O direito ao acesso equitativo a tecnologias de aprimoramento é um neurodireito proposto para garantir que esse acesso seja pautado por princípios de justiça e equidade (DANTAS; JEREISSATI, 2024; PACHECO, 2025). O processo de consentimento, nesse contexto, deve ser transparente sobre os potenciais riscos e a distinção entre terapia e aprimoramento (UNESCO, 2025; LAVAZZA, 2025).

Como se observa, convergem justificativas e desafios: proteger a mente contra intrusão e manipulação, tornar o consentimento efetivo num ambiente de processamento opaco e contínuo, prevenir discriminações algorítmicas, suprir lacunas normativas para neurodados e enfrentar, com prudência e equidade, as promessas e riscos do *enhancement* (aprimoramento).

7.2 O MOSAICO REGULATÓRIO BRASILEIRO

Influenciado por esse debate global sobre os riscos e oportunidades decorrentes da inteligência artificial e das neurotecnologias, o Brasil vem respondendo por meio de um conjunto multifacetado e ambicioso de iniciativas legislativas (DONEDA; SOUZA *et al.*, 2018; DONEDA; MENDES *et al.*, 2023; ANPD, 2025). Essas propostas operam em diferentes níveis do ordenamento jurídico – constitucional, legal e infralegal – e revelam uma estratégia normativa complexa, que busca tanto estabelecer princípios gerais e direitos fundamentais quanto criar regras técnicas específicas de implementação.

Esta seção pretende trazer uma visão panorâmica das principais propostas em tramitação, evidenciando essa tensão entre avanços científicos e tecnológicos / riscos a direitos fundamentais, característica de áreas tecnologicamente dinâmicas e com alto potencial disruptivo, nas quais a antecipação desses riscos precisa caminhar o mais próximo possível da inovação.

7.2.1 A Proposta de Blindagem Constitucional: PEC 29/2023

No topo da hierarquia normativa, a Proposta de Emenda à Constituição (PEC) nº 29, de 2023, representa a tentativa de conferir o mais alto grau de proteção aos novos direitos (BRASIL, 2023c). Proposta no Senado Federal por um grupo de senadores liderado por Randolfe Rodrigues, a PEC visa alterar o artigo 5º da Constituição Federal para incluir, no rol de direitos e garantias fundamentais, a “proteção à integridade mental” e a “transparência no uso de algoritmos por sistemas de inteligência artificial” (BRASIL, 2023c).

A justificação da proposta é explícita em seu alinhamento com o movimento internacional, mencionando diretamente a iniciativa da *NeuroRights Foundation* de Rafael Yuste e o exemplo pioneiro do Chile (BRASIL, 2023c). Esta é uma estratégia de “blindagem” principiológica: ao inserir esses direitos no núcleo duro da Constituição, busca-se criar um mandado de otimização que deverá guiar toda a legislação futura e a interpretação judicial (ALEXY, 2024), garantindo que o desenvolvimento tecnológico no país se dê sob a égide da proteção da mente humana e da preservação da autonomia cognitiva (MOLNÁR-GÁBOR, 2024). Desde 26 de junho de 2023, a proposta aguarda a designação de um relator na Comissão de Constituição, Justiça e Cidadania (CCJ) do Senado (BRASIL, 2023c)³⁷.

³⁷ Fonte: <https://www25.senado.leg.br/web/atividade/materias/-/materia/158095>. Acesso em: 20 jul. 2025.

7.2.2 A Construção de um Microsistema Legal para Neurodireitos: PL 2.174/2023

Descendo ao nível da legislação ordinária, o Projeto de Lei (PL) nº 2.174, de 2023, atualmente em tramitação na Câmara dos Deputados, busca instituir um microsistema jurídico específico para a tutela dos neurodireitos no Brasil (BRASIL, 2023b). Apresentada em abril de 2023, a proposição legislativa tem como ementa “estabelecer as normas e princípios para proteção dos direitos fundamentais relacionados ao cérebro e ao sistema nervoso humano, objetivando garantir a proteção e promoção dos neurodireitos dos indivíduos” (BRASIL, 2023b).

O texto normativo pretende operacionalizar, no plano infraconstitucional, os princípios abstratos consagrados na Proposta de Emenda à Constituição sobre neurodireitos, atuando como ponte entre garantias constitucionais e regras técnicas previstas em legislações correlatas. Nesse sentido, propõe a definição de conceitos centrais como “neurodireitos”, alinhando-se ao debate internacional sobre a proteção da esfera mental (YUSTE *et al.*, 2017; PASQUALE, 2020; COHEN, 2019; MOLNÁR-GÁBOR, 2024). Além disso, estabelece salvaguardas concretas para assegurar direitos como privacidade mental (no texto: “privacidade cerebral e neurológica”), integridade psicológica (no texto: “integridade cerebral e neurológica”), autonomia e livre-arbítrio frente às neurotecnologias – inovações como interfaces cérebro-computador e técnicas de neuromodulação (BRASIL, 2023b).

Atualmente, a tramitação do PL encontra-se sob análise da Comissão de Constituição e Justiça e de Cidadania (CCJC) da Câmara dos Deputados (BRASIL, 2023b), momento crucial para o refinamento jurídico-técnico do texto e para a consolidação de um marco regulatório nacional alinhado a tendências globais na regulação das neurotecnologias.

7.2.3 A Tentativa de Adaptação da LGPD: PL nº 1.229/2021 e PL nº 522/2022

Uma abordagem alternativa, de natureza mais técnica e incremental, foi tentada por meio de projetos de lei que buscaram alterar a Lei Geral de Proteção de Dados (LGPD, Lei nº 13.709/2018) para abarcar as especificidades dos neurodireitos.

Uma dessas iniciativas foi o Projeto de Lei nº 1.229, de 2021, que propunha conceituar “dado neural” e classificá-lo expressamente como um dado pessoal sensível (BRASIL, 2021). Essa estratégia visava aproveitar a estrutura de governança e fiscalização já existente da LGPD e da Agência Nacional de Proteção de Dados (ANPD). Contudo, a ficha de tramitação

do projeto revela um desfecho importante para a compreensão do cenário político-legislativo: o projeto foi retirado pelo seu autor em 15 de março de 2022³⁸.

Posteriormente, o Projeto de Lei nº 522, de 2022, seguiu uma linha similar, propondo uma alteração mais detalhada na LGPD. A proposta visa incluir os “dados neurais” como uma subcategoria explícita de dados pessoais sensíveis, definindo-os como “toda informação obtida, direta ou indiretamente, da atividade do sistema nervoso central e periférico, que informe sobre a identidade ou sobre características físicas, fisiológicas, genéticas, mentais, econômicas, culturais ou sociais de uma pessoa natural”. Além disso, o projeto propõe a criação de uma nova seção (“Seção II-A. Do Tratamento de Dados Neurais”) na LGPD, dedicado inteiramente à proteção dos neurodireitos, garantindo a privacidade mental, a identidade pessoal, o livre-arbítrio, o acesso equitativo a tecnologias de aprimoramento cerebral e a proteção contra vieses discriminatórios (BRASIL, 2022). Atualmente, o PL 522/2022 aguarda parecer da Relatora na Comissão de Saúde (CSAUDE) da Câmara dos Deputados³⁹.

A existência de múltiplas propostas com o mesmo objetivo, bem como a retirada do primeiro projeto, pode indicar a preferência do legislador por debates mais aprofundados e pela adoção de um modelo regulatório próprio e especializado – a exemplo do PL 2.174/2023 –, em vez de promover apenas ajustes pontuais (“remendos”) na legislação já existente. Embora uma abordagem dedicada tenha o mérito de conferir maior visibilidade e especificidade ao tema, ela também traz consigo o risco de fragmentação normativa, com potenciais sobreposições e conflitos entre diferentes regimes regulatórios.

7.2.4 Anteprojeto de Reforma do Código Civil

Outra frente relevante para a positivação dos neurodireitos no ordenamento jurídico brasileiro é o Anteprojeto de Reforma do Código Civil, fruto dos trabalhos da Comissão de Juristas instituída pelo Senado Federal em 2023. Diferente das propostas que visam criar leis específicas ou alterar a LGPD, esta abordagem inovadora busca inserir a proteção da mente no núcleo do direito privado, por meio da criação de um livro (Livro VI) inteiramente novo dedicado ao “Direito Civil Digital” (PACHECO, 2025).

Dentro deste novo livro, no Capítulo II que trata “Da Pessoa no Ambiente Digital”, o anteprojeto propõe um artigo específico para conceituar e proteger os neurodireitos, alinhando

³⁸ Fonte: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2276604&fichaAmigavel=nao>. Acesso em: 20 jul. 2025.

³⁹ Fonte: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2317524>. Acesso em: 20 jul. 2025.

o Brasil às mais avançadas discussões internacionais sobre o tema. A proposta busca tratar os neurodireitos como uma dimensão intrínseca da personalidade, conferindo-lhes a mais alta proteção no âmbito das relações privadas (PACHECO, 2025).

O artigo proposto estabelece que os “neurodireitos são parte indissociável da personalidade e recebem a mesma proteção desta, não podendo ser transmitidos, renunciados ou limitados” (PACHECO, 2025, p. 402). O texto detalha um rol de garantias fundamentais, que incluem: (i) *Direito à liberdade cognitiva* - Vedando o uso de neurotecnologias de forma coercitiva ou sem consentimento; (ii) *Direito à privacidade mental* - Para proteger contra o acesso não autorizado a dados cerebrais e vedar sua comercialização; (iii) *Direito à integridade mental* - Protegendo contra a manipulação da atividade mental e a alteração do controle sobre o próprio comportamento; (iv) *Direito de continuidade da identidade pessoal* - Para resguardar contra alterações não autorizadas no cérebro ou em suas atividades; (v) *Direito ao acesso equitativo* - Assegurando que tecnologias de aprimoramento cognitivo sejam distribuídas com justiça; e (vi) *Direito à proteção contra práticas discriminatórias* - Para combater vieses a partir de dados cerebrais.

Dessa forma, o Anteprojeto de Reforma do Código Civil não apenas reconhece a urgência de regular o tema, mas o faz de maneira estrutural, integrando os neurodireitos à teoria dos direitos da personalidade e estabelecendo um sólido alicerce para a proteção da dignidade humana na nova era digital.

O documento mais recente da Agência Nacional de Proteção de Dados (ANPD), publicado em junho de 2025, integra o quarto volume da série Radar Tecnológico, com foco específico nas neurotecnologias (ANPD, 2025). Essa publicação destaca as principais inovações no campo das neurotecnologias, avaliando suas aplicações e usos potenciais, e chama atenção para os desafios éticos e jurídicos que emergem quando essas tecnologias lidam com dados neurais – capazes de revelar aspectos íntimos da mente humana, como estados mentais e emocionais (ANPD, 2025).

7.2.5 A Regulação da IA: PL 2.338/2023 (Marco Legal da IA)

Paralelamente aos esforços focados em neurotecnologias, o Brasil avança na construção de uma legislação geral para a Inteligência Artificial. O Projeto de Lei nº 2.338, de 2023, conhecido como o Marco Legal da IA, é a iniciativa mais robusta nesse sentido (BRASIL, 2023a). Originado de um anteprojeto elaborado por uma comissão de juristas e apresentado pelo Presidente do Senado, o texto foi aprovado na casa em dezembro de 2024 e seguiu para

revisão na Câmara dos Deputados, encontrando-se, atualmente, na Comissão Especial sobre Inteligência Artificial para proferir parecer ao Projeto⁴⁰.

Inspirado no *AI Act* europeu, o projeto adota uma abordagem baseada em risco, classificando os sistemas de IA conforme seu potencial de dano e estabelecendo requisitos de governança proporcionais (COALIZÃO DIREITOS NA REDE, 2025). O Art. 14 do PL 2338/2023 (na versão que tramita na Câmara dos Deputado e consultada no dia 20 jul. 2025) define expressamente os sistemas de IA considerados de alto risco. Dentro dessa lista, o inciso VIII inclui: “aplicações na área da saúde para auxiliar diagnósticos e procedimentos médicos, quando houver risco relevante à integridade física e mental das pessoas” (BRASIL, 2023a).

O PL brasileiro segue a lógica da legislação europeia, onde a maioria dos *softwares* de dispositivos médicos que utilizam IA diagnóstica/tratamento é classificada como um sistema de IA de alto risco (com poucas exceções, por exemplo, *MDR* Classe I não estéril, não medidor, não cirúrgico reutilizável e *IVDR* Classe A não estéril), como se observa do próprio AIB/MDCG (2025), analisado no Capítulo 6. No mesmo sentido é o trabalho publicado na Itália, por Filippis *et al.* (2024), sob o título de *The Risk Perspective of AI in Healthcare: GDPR and GELSI Framework (Governance, Ethical, Legal and Social Implications) and the New European AI Act*.

O AIB/MDCG (2025) estabelece que “sistemas de IA utilizados para fins médicos são referidos como Inteligência Artificial para Dispositivos Médicos (*MDAI*)”. Isso inclui dispositivos médicos, acessórios para dispositivos médicos, dispositivos médicos para diagnóstico *in vitro* e seus acessórios, além de produtos do Anexo XVI do *MDR*. Um *MDAI* é considerado um sistema de IA de alto risco sob o Artigo 6º, nº1 do *AI Act* se atender a duas condições: 1) o *MDAI* é um componente de segurança ou o próprio sistema de IA é um dispositivo médico, e 2) o *MDAI* está sujeito a uma avaliação de conformidade por terceiros, realizada por um organismo notificado, de acordo com o Regulamento de Dispositivos Médicos (*MDR*) ou o Regulamento de Dispositivos Médicos para Diagnóstico *In Vitro* (*IVDR*). Portanto, como a grande maioria desses dispositivos médicos é submetida à avaliação de conformidade por um organismo notificado (avaliação de terceiros), devem ser classificados como um sistema de IA de alto risco (AIB/MDCG, 2025; BOTTINI *et al.*, 2025).

Isso significa que, se o PL for aprovado conforme a sua redação atual, os sistemas de IA e neurotecnologias também deverão adotar o tripé transparência/explicabilidade, equidade e

⁴⁰ Fonte: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262>. Acesso em: 20 jul. 2025.

accountability como norma e seguir todas as salvaguardas necessárias a proteger o CLE e a dignidade humana.

O conjunto dessas iniciativas, portanto, revela a complexidade da estratégia brasileira. Há uma clara complementação entre a abordagem da PEC nº 29/2023, que busca fixar princípios constitucionais, e a do PL nº 2.338/2023 e das regulações da ANVISA (analisadas a seguir), que focam em regras técnicas e de conformidade. O sucesso do arcabouço regulatório brasileiro dependerá, crucialmente, da capacidade de articular de forma coerente esses diferentes níveis. Sem essa articulação, corre-se o risco de ter uma Constituição com direitos simbólicos que não se traduzem em práticas efetivas, ou, inversamente, regras técnicas que operam em um vácuo ético e de direitos fundamentais.

7.3 O CONSENTIMENTO LIVRE E ESCLARECIDO NA FRONTEIRA DA IA E DA NEUROTECNOLOGIA

Embora o CLE tenha sido amplamente debatido no Capítulo 5, é necessário revisitá-lo nesse ponto do texto, para fazer algumas considerações ligadas à legislação.

Vimos que a aplicação de inteligência artificial e neurotecnologias na saúde lança desafios profundos a um dos pilares mais sagrados da bioética e do direito médico: o Consentimento Livre e Esclarecido (Capítulo 5). A discussão, porém, transcende a lógica do consentimento para tratamento de dados pessoais, como previsto na LGPD, e exige uma revisão de seus fundamentos éticos para salvaguardar a autonomia do paciente em um cenário de crescente complexidade tecnológica e opacidade algorítmica.

7.3.1 A Distinção Fundamental: Consentimento para Dados vs. Consentimento para o Ato Médico

É crucial estabelecer uma distinção clara entre os dois regimes de consentimento que coexistem no contexto da saúde digital. O primeiro é o consentimento para o tratamento de dados pessoais (gerais e sensíveis), uma das bases legais previstas nos artigos 7º e 11 da LGPD. Seu foco é a autodeterminação informativa: o direito do indivíduo de controlar o fluxo de suas informações. Esse consentimento deve ser uma manifestação livre, informada e inequívoca, para uma finalidade determinada, e é, por natureza, revogável.

O segundo, e mais fundamental no contexto clínico, é o Consentimento Informado, que, para os fins desta dissertação sobre IA e neurotecnologias, está sendo tratado por sua designação

completa: Consentimento Livre e Esclarecido. A ênfase em ambos os elementos – “livre” e “esclarecido” – é deliberada. Embora a doutrina bioética frequentemente utilize a expressão “consentimento informado” para diferenciar a prática clínica do Termo de Consentimento Livre e Esclarecido (TCLE) de pesquisas, no cenário das novas tecnologias, a retomada da expressão completa se torna imperativa. Conforme demonstrado no Capítulo 5, os vieses algorítmicos e a opacidade dos sistemas de IA (“efeito caixa-preta”) representam ameaças diretas a cada um desses pilares. O pilar “esclarecido” é desafiado pela dificuldade de explicar o funcionamento de algoritmos complexos, enquanto o pilar “livre” pode ser comprometido por recomendações enviesadas e manipuladoras que se apresentam como objetivas, influenciando a decisão do paciente de forma sutil.

Portanto, este consentimento não é meramente uma base legal para o uso de dados, mas um processo ético-jurídico que valida a própria intervenção no corpo e na saúde de uma pessoa. Fundamentado no princípio da autonomia, o Consentimento Livre e Esclarecido é a pedra angular da relação paciente-médico (DANTAS; NOGAROLI, 2020). Sua finalidade não é autorizar o processamento de dados, mas legitimar um ato médico, seja ele diagnóstico, terapêutico ou experimental

O Consentimento Livre e Esclarecido, portanto, pressupõe:

“[...] anuência, livre de vícios, do paciente, após explicação completa e pormenorizada sobre a intervenção médica, incluindo sua natureza, objetivos, métodos, duração, justificativa, possíveis males, riscos e benefícios, métodos alternativos existentes e nível de confidencialidade dos dados, assim como de sua liberdade total para recusar ou interromper o procedimento em qualquer momento; tendo o profissional a obrigação de informá-lo em linguagem adequada (não técnica) para que ele compreenda” (DANTAS; JEREISSATI, 2024).

Logo, a aplicação de neurotecnologias e IA na saúde exige uma dupla camada de consentimento: um para o tratamento dos dados pessoais e neurais (sob a égide da LGPD) e outro, muito mais denso e dialógico, para diagnósticos/tratamentos de saúde com intervenção tecnológica com IA ou *neurotech* em si (sob a égide da bioética e das normas do Conselho Federal de Medicina).

7.3.2 A Governança Ética na Prática Clínica e na Pesquisa no Brasil

No Brasil, a governança do CLE em face dessas novas tecnologias é exercida por duas frentes principais: a ética em pesquisa e a regulação profissional.

Qualquer pesquisa clínica que envolva o uso de IA e neurotecnologias experimentais em seres humanos, como os implantes de *BCIs*, está sujeita à avaliação rigorosa do Sistema

CEP/CONEP, composto pelos Comitês de Ética em Pesquisa – CEPs e pela Comissão Nacional de Ética em Pesquisa – CONEP (NOGAROLI, 2023). A CONEP, que está diretamente ligada ao Conselho Nacional de Saúde – CNS, tem, como instância máxima, a atribuição de avaliar os aspectos éticos de pesquisas de alta complexidade, garantindo que os protocolos respeitem a dignidade e a autonomia dos participantes. A Lei nº 14.874/2024 estabelece a necessidade de obtenção do Consentimento Livre e Esclarecido do participante em pesquisas com neurociência e neurotecnologias emergentes em seres humanos, incluindo dispositivos médicos experimentais (BRASIL, 2024; ANPD, 2025).

Esse processo inclui uma verificação minuciosa dos procedimentos de obtenção do Termo de Consentimento Livre e Esclarecido (TCLE), exigindo que os voluntários compreendam plenamente a natureza e a finalidade da pesquisa, bem como os riscos envolvidos – que não se limitam a danos físicos, mas abrangem também impactos psicológicos e sociais, especialmente aqueles relacionados à manipulação ou à exposição de dados neurais sensíveis (YUSTE *et al.*, 2017; O’NEIL, 2016; CFM 2016; 2018). Em um cenário em que as neurotecnologias podem acessar e processar informações íntimas da mente humana, a supervisão ética robusta torna-se imprescindível para prevenir abusos e proteger direitos fundamentais.

No âmbito da prática médica assistencial, o Conselho Federal de Medicina (CFM) e os Conselhos Regionais (CRMs) são responsáveis por normatizar a conduta ética dos profissionais. O CFM, tanto na Recomendação nº 1/2016 (CFM, 2016), quanto no Código de Ética Médica (CFM, 2018), disciplina o CLE. O Termo de Consentimento deve seguir os seguintes parâmetros:

“a) O esclarecimento claro, pertinente e suficiente sobre justificativas, objetivos esperados, benefícios, riscos, efeitos colaterais, complicações, duração, cuidados e outros aspectos específicos inerentes à execução tem o objetivo de obter o consentimento livre e a decisão segura do paciente para a realização de procedimentos médicos. Portanto, não se enquadra na prática da denominada medicina defensiva.

b) A forma verbal é a normalmente utilizada para obtenção de consentimento para a maioria dos procedimentos realizados. Deve-se considerar fazer a anotação em prontuário, quando pertinente. Contudo, recomenda-se a elaboração escrita (Termo de Consentimento Livre e Esclarecido) nas seguintes condições: em procedimentos invasivos a critério médico, em determinados procedimentos que causem desconforto ou quando a complexidade e quantidade de efeitos dos procedimentos não são suficientes para o entendimento por meio de consentimento verbal.

c) A redação do documento deve ser feita em linguagem clara, que permita ao paciente entender o procedimento e suas consequências, na medida de sua compreensão. Os termos científicos, quando necessários, precisam ser acompanhados de seu significado, em linguagem acessível”.

[...]

(CFM, 2016, Anexo I).

É evidente que esse modelo de consentimento não evoluiu com as novas tecnologias, principalmente em razão da opacidade e dos vieses característicos desses sistemas. Ciente desses desafios impostos pela IA, o CFM criou um Departamento de Inteligência Artificial e está em processo de elaboração de uma Resolução nacional para orientar o uso ético da tecnologia (CFM, 2025; FUTURO DA SAÚDE, 2025). Antecipando-se a essa norma, o Conselho Regional de Medicina do Estado do Rio Grande do Sul (CREMERS) publicou, de forma pioneira, a Resolução nº 6/2025, que, em consonância com as preocupações do CFM, estabelece que o médico deve esclarecer ao paciente em que medida a IA será utilizada no diagnóstico, prevenção e tratamento, obtendo o TCLE quando a IA for uma ferramenta determinante no ato médico (CREMERS, 2025).

Portanto, diante da “caixa-preta” e dos vieses presentes nos sistemas de IA e *neurotechs*, aprofundados nos Capítulos anteriores, há uma crise no modelo tradicional de consentimento, que é essencialmente transacional e baseado na divulgação de informações. As neurotecnologias, que interagem diretamente com o cérebro e o sistema nervoso, levantam questões ainda mais profundas sobre o consentimento, pois podem operar de forma a coletar e, potencialmente, manipular dados neurais sem a consciência do titular.

A capacidade das interfaces cérebro-computador (*BCIs*) e sistemas de IA de infiltrar e alterar funções cerebrais coloca em xeque noções básicas de privacidade, identidade e livre-arbítrio. A manipulação de pensamentos e emoções, já demonstrada em estudos com animais, representa um aumento da vulnerabilidade e lesão ao direito da personalidade e à autodeterminação humana (DANTAS; JEREISSATI, 2024). A proteção contra “alterações do encéfalo que possam mudar o conceito que uma pessoa tem de si mesma” é um dos pilares dos neurodireitos.

Se a informação completa sobre o processo decisório não está disponível nem para o médico, a premissa de que o paciente pode ser plenamente “esclarecido” se rompe. O paciente passa a consentir não apenas com os riscos conhecidos de um resultado, mas com a incerteza fundamental do processo. A solução, portanto, não pode se limitar a exigir “mais transparência”, pois a explicabilidade total pode ser tecnicamente inviável. A evolução necessária aponta para um paradigma de consentimento relacional e contínuo, no qual o consentimento deixa de ser um ato único para se tornar um diálogo permanente entre médico e paciente. Nesse modelo, o foco se desloca da mera divulgação de fatos para a gestão conjunta da incerteza, construída sobre a confiança e o alinhamento de valores e objetivos terapêuticos, reconhecendo os limites tanto do conhecimento humano, quanto dos algoritmos, fazendo-se

necessária, portanto, a adoção das salvaguardas jurídicas complementares ao consentimento (devidamente analisadas no Capítulo 6).

7.4 A GOVERNANÇA TÉCNICA EM SAÚDE: A ATUAÇÃO DAS AGÊNCIAS REGULADORAS

Para além dos debates legislativos e éticos, a governança prática de IA e neurotecnologias na saúde depende fundamentalmente da atuação de agências reguladoras técnicas. São elas que detêm o poder de autorizar ou barrar a entrada desses produtos inovadores no mercado, funcionando como um filtro crucial para garantir a segurança e a eficácia das tecnologias que chegam aos pacientes. No Brasil e nos Estados Unidos, a ANVISA e a FDA, respectivamente, desempenham esse papel central, embora com abordagens e níveis de maturidade distintos.

7.4.1 A ANVISA e a Regulação de *Software as a Medical Device (SaMD)*

No Brasil, *softwares* com finalidade médica (inclusive os baseados em IA) são dispositivos médicos (*SaMD - Software as a Medical Device*) e devem atender às RDC n° 657/2022, RDC n° 751/2022, RDC n. 777/2023 e RDC n. 830/2023 (ANVISA, 2022a; 2022b; ANVISA, 2023a; 2023b). O titular da regularização do produto perante a ANVISA, em particular, detém responsabilidade administrativa, civil e penal por garantir a segurança e a eficácia do dispositivo⁴¹.

Embora estas Resoluções da Diretoria Colegiada (RDCs) da ANVISA não utilizem explicitamente o termo “Inteligência Artificial”, elas efetivamente enquadram qualquer software com finalidade médica – seja para diagnóstico, prevenção, tratamento ou monitoramento –, incluindo aqueles baseados em IA, na definição de *Software* como Dispositivo Médico (*SaMD*) e exigem que atendam aos seus requisitos (PASSARINI, 2022; AITH, 2024). Como será abordado adiante, a ANVISA também se alinha a padrões internacionais

⁴¹ Art. 2º O carregamento de instruções de uso no repositório documental de dispositivos médicos, disponível no portal da Anvisa, é obrigatório e deve ser executado pela **empresa responsável pela notificação ou pelo registro do produto**, a qual atesta que seu conteúdo guarda concordância com a legislação vigente e consistência com o produto regularizado, de acordo com o §4º do art. 33 da Resolução da Diretoria Colegiada - RDC n° 751, de 15 de setembro de 2022 e do §4º do art. 33 da Resolução da Diretoria Colegiada - RDC n° 830, de 06 de dezembro de 2023. RESOLUÇÃO-RE N° 1.206, DE 27 de março de 2025. Disponível em: https://anvisa.legis.datalegis.net/action/UrlPublicasAction.php?acao=abrirAtoPublico&num_ato=00001206&sgl_tipo=RES&sgl_orgao=RE/GGTPS/ANVISA/MS&vlr_ano=2025&seq_ato=000&cod_modulo=293&cod_menu=8499. Acesso em: 10 ago. 2025.

como o *IMDRF*, que tem diretrizes para a avaliação clínica de *SaMDs* e *ML-enabled Medical Devices* (NOGAROLI, 2023).

A Resolução de Diretoria Colegiada (RDC) nº 657, de 24 de março de 2022, define *SaMD* como “*Software* que atende à definição de dispositivo médico, podendo ser de diagnóstico *in vitro* (*IVD*) ou não, sendo destinado a uma ou mais indicações médicas, e que realizam essas finalidades sem fazer parte de hardware de dispositivo médico” (ANVISA, 2022a). Explicitamente, a RDC 657/2022 menciona a necessidade de descrever “os algoritmos e/ou rotinas utilizadas para gerar o processamento das sugestões de prevenção, diagnóstico, tratamento, monitorização fisiológica, reabilitação ou anticoncepção e suas fundamentações clínicas ou científicas” (ANVISA, 2022a). A inteligência artificial opera fundamentalmente por meio de algoritmos, o que significa que *softwares* baseados em IA com finalidade médica são abrangidos por essa regulamentação (ANPD, 2025).

A RDC nº 751, de 15 de setembro de 2022, complementa a regulamentação ao dispor sobre a classificação de risco, regimes de notificação e registro, e os requisitos de rotulagem e instruções de uso de dispositivos médicos, incluindo o *SaMD* (ANVISA, 2022b; ANPD, 2025). Ela estabelece que os dispositivos médicos são enquadrados em Classes I (baixo risco), II (médio risco), III (alto risco) ou IV (máximo risco). Dispositivos de Classes I e II são sujeitos a notificação, enquanto Classes III e IV são sujeitos a registro (ANVISA, 2022b, Artigos 5º ao 8º).

Por exemplo, o Anexo I da RDC nº 751/2022 dispõe que:

Regra 8: Todos os dispositivos implantáveis e os dispositivos cirurgicamente invasivos destinados a uso de longo prazo são classificados na classe III [...] (em seguida traz as exceções);

Regra 11: “O *software* destinado a prestar informações utilizadas para a tomada de decisões com fins terapêuticos ou de diagnóstico é classificado na classe II, exceto se tais decisões tiverem um impacto que possa causar:

a) a morte ou uma deterioração irreversível do estado de saúde de uma pessoa, caso em que é classificado na classe IV; ou

b) uma deterioração grave do estado de saúde de uma pessoa ou uma intervenção cirúrgica, caso em que é classificado na classe III”.

(ANVISA, 2022b, Anexo I).

Fernando Aith (2024) corrobora que a ANVISA, *verbis*:

[...] “deu os primeiros passos para regular *software* na saúde – notadamente, em 2022, a ANVISA emitiu a Resolução RDC 657/2022 para regulamentar o ‘*Software* como Dispositivo Médico’ (*SaMD*) e, mais recentemente, a RDC

751/2022, que atualiza a estrutura geral para dispositivos médicos, reconhecendo e abordando o *SaMD*”.

O mesmo autor afirma, ainda, que a inteligência artificial “está sendo cada vez mais implantada no setor de saúde do Brasil, desde algoritmos de diagnóstico até plataformas de telemedicina”, e que esses *softwares* de saúde autônomos com fins médicos (por exemplo, um aplicativo de diagnóstico de IA) exigem aprovação e certificação semelhantes a outros dispositivos médicos (AITH, 2024).

Em 2023, a ANVISA publicou a Resolução RDC nº 830, de 06 de dezembro de 2023, que trata especificamente sobre os dispositivos médicos para diagnóstico *in vitro* (*IVD*), incluindo seus instrumentos, dispendo sobre a classificação de risco, os regimes de notificação e de registro e os requisitos de rotulagem e instruções de uso de dispositivos médicos para diagnóstico *in vitro*, inclusive seus instrumentos (ANVISA, 2023b).

A Orientação de Serviço nº 131, de 29 de maio de 2024 (ANVISA, 2024), de natureza administrativa e procedimental da própria ANVISA, altera o fluxo interno para a elaboração e deliberação de instrumentos regulatórios na agência. Não abordam diretamente os *SaMDs* como produtos regulados, mas sim o processo pelo qual as regulamentações (como as de *SaMD*) são criadas e alteradas pela ANVISA.

Em 2025, a ANVISA, através da Resolução-RE nº 1.206, de 27 de março de 2025 (ANVISA, 2025a), deferiu petições de registro de produtos e, no art. 2º, estabeleceu a obrigatoriedade do carregamento das instruções de uso no repositório documental de dispositivos médicos da Anvisa para *SaMDs*, conforme estabelecido pelas RDC nº 751/2022 e RDC nº 830/2023. Essa ação deve ser executada pela empresa responsável pela notificação ou pelo registro do produto. Ou seja, a empresa deve atestar que o conteúdo das instruções de uso guarda concordância com a legislação vigente e consistência com o produto regularizado.

No mesmo ano, a ANVISA publicou o Ofício Circular nº 2/2025/SEI/GGTPS/DIRE3 (ANVISA, 2025c), tratando das orientações sobre a implementação do sistema de Identificação Única de Dispositivos Médicos (*UDI*) no Brasil. O *UDI* é um padrão internacional que segue as regras do *International Medical Device Regulators Forum (IMDRF)*. Seu principal objetivo é viabilizar a identificação inequívoca de um determinado dispositivo médico no mercado, ampliando a segurança dos pacientes. A RDC nº 591/2021 determina a obrigatoriedade de apor a marcação *UDI* nos dispositivos médicos e em suas embalagens. O Ofício destaca que essa obrigatoriedade se torna efetiva para os dispositivos médicos da classe de risco IV, fabricados a partir de 10 de julho de 2025 (ANVISA, 2025c).

Um aspecto crucial da abordagem da ANVISA, ainda que não formalizado em um guia específico para IA, é a exigência, durante o processo de submissão, de informações detalhadas sobre as bases de dados utilizadas para o treinamento, teste e validação do algoritmo, como demonstrado acima. Os fabricantes devem justificar a metodologia de IA empregada e apresentar evidências que demonstrem a segurança e a eficácia do *software*. Essa exigência funciona como um mecanismo para que a agência possa avaliar indiretamente a robustez do sistema e o risco de vieses, podendo negar o registro caso as informações sejam insuficientes. A ANVISA também demonstra um alinhamento com padrões internacionais, participando ativamente do *International Medical Device Regulators Forum (IMDRF)*, um fórum global que busca harmonizar diretrizes para dispositivos médicos, incluindo aqueles baseados em IA (KASZNAR LEONARDOS, 2022; ANVISA, 2022a; 2022b; 2025).

Para neurotecnologias mais complexas, como dispositivos implantáveis de estimulação cerebral profunda ou interfaces cérebro-computador, ainda não há uma regulamentação específica. Aplica-se o quadro geral de dispositivos médicos da RDC 751/2022. Dada a sua natureza invasiva e inovadora, tais produtos são invariavelmente enquadrados nas classes de risco mais elevadas (III ou IV), o que exige um processo de aprovação rigoroso, incluindo a apresentação de robustos estudos clínicos pré-mercado (MATTOS FILHO, 2023).

Em maio de 2025, por exemplo, a ANVISA concedeu aprovação para o sistema de estimulação da medula espinhal “*Eterna™ SCS*”, da empresa Abbott, destinado ao tratamento de dor crônica intratável (ABBOTT BRASIL, 2025). Trata-se de um neuroestimulador implantável de última geração (classe de risco IV) que libera pulsos elétricos na medula espinhal para modular a sensação de dor. O *Eterna™ SCS* é um dispositivo recarregável implantado sob a pele do paciente, conectado a eletrodos posicionados próximo à medula – configuração típica de sistemas de neuromodulação. Projetado com base em pesquisas com pacientes, médicos e cuidadores, é o menor estimulador medular implantável e recarregável já desenvolvido pela Abbott. O sistema utiliza a tecnologia de estimulação *BurstDR™*, que simula os padrões naturais de disparo do cérebro. Segundo dados da empresa, essa modalidade de terapia é preferida por 87% dos pacientes em comparação à estimulação tradicional do tipo “formigamento”⁴².

Antes da aprovação no Brasil, o produto já havia obtido certificação no exterior (inclusive aprovação FDA nos EUA em 2022). A Abbott submeteu à ANVISA um extenso dossiê com os dados de projeto e desempenho do aparelho (por exemplo, destacando que o *Eterna™ SCS* é o

⁴² Fonte: <https://saudedigitalnews.com.br/26/06/2025/abbott-recebe-aval-da-anvisa-para-sistema-de-estimulacao-medular-voltado-a-dor-cronica/>. Acesso em: 20 ago. 2025.

menor estimulador medular recarregável do mercado e que sua bateria necessita de apenas 5 recargas ao ano), bem como evidências clínicas de sua eficácia em aliviar dores crônicas de difícil tratamento (suportadas por estudos internacionais publicados). A ANVISA avaliou criteriosamente esses aspectos e aprovou o registro sanitário do *Eterna™ SCS* para comercialização no Brasil. Com isso, o dispositivo foi incorporado ao portfólio terapêutico disponível no país, ampliando as opções para pacientes que sofrem de dores refratárias. O caso *Eterna™* ilustra o caminho regulatório de um dispositivo neurotecnológico inovador: a empresa precisou atender a todos os requisitos de qualidade, fornecer comprovação científica robusta e adaptar seu produto às normas nacionais (inclusive rotulagem em português, instruções de uso com informações de segurança), culminando na autorização pela autoridade sanitária brasileira (ABBOTT BRASIL, 2025).

Observa-se, portanto, que a ANVISA (2022a; 2022b) assumiu, de fato, papel de protagonismo na regulação de dispositivos médicos baseados em IA e neurotecnologias no Brasil, especialmente por meio da RDC n.º 657/2022 (que disciplina *Software as a Medical Device*) e da RDC n.º 751/2022 (classificação de risco e requisitos de regularização). Esse movimento tem atenuado a lacuna normativa federal específica ao exigir documentação técnica robusta, evidências de desempenho clínico, gestão de riscos e vigilância pós-comercialização, criando um patamar mínimo de segurança e transparência para soluções algorítmicas em saúde. Não obstante, como adverte Fernando Aith (2024), o país ainda enfrenta desafios institucionais e de capacidade regulatória para acompanhar a velocidade, a opacidade e os riscos sistêmicos dessas tecnologias – o que demanda, tanto do legislador quanto dos órgãos reguladores, um salto rápido em infraestrutura técnica, recursos humanos especializados e coordenação interinstitucional (*e.g.*, com ANPD e Conselhos Profissionais), sob pena de a tutela regulatória permanecer aquém da complexidade do campo.

7.4.2 A FDA e a Regulação de Dispositivos com IA/ML e Interfaces Cérebro-Computador (BCI)

Nos EUA, atualmente, não há leis ou padrões harmonizados que regulem especificamente o uso de aprendizado de máquina em dispositivos médicos. Diversos projetos de lei, como o *Algorithmic Accountability Act of 2023*, foram introduzidos no Congresso, mas permanecem em tramitação, sem aprovação (EUA, 2023a; 2023b; FDA, 2025). A Califórnia também tem um Projeto de Lei do Senado (*Senate Bill n.º 1223*), que visa alterar o Código Civil para definir “informações pessoais confidenciais”, para os fins da *CCPA*, de modo a incluir adicionalmente

os dados neurais do consumidor. Esses “dados neurais” seriam definidos como informações geradas pela medição da atividade do sistema nervoso central ou periférico do consumidor, e que não são inferidas a partir de informações não neurais (CALIFÓRNIA, 2024).

A governança da IA, portanto, é guiada, principalmente, por dois pilares principais: o primeiro, são *frameworks* voluntários (não obrigatórios) desenvolvidos por agências governamentais. O *NIST AI Risk Management Framework (AI RMF 1.0)*, publicado pelo *National Institute of Standards and Technology*, em janeiro de 2023, oferece um guia detalhado para que as organizações possam mapear, medir e gerenciar os riscos associados à IA de forma voluntária (TABASSI, 2023). Da mesma forma, o *Blueprint for an AI Bill of Rights*, lançado pela Casa Branca, em outubro de 2022, estabelece cinco princípios éticos – sistemas seguros e eficazes; proteção contra discriminação algorítmica; privacidade de dados; notificação e explicação; e alternativas humanas – que funcionam como uma declaração de valores nacionais, sem força vinculante (THE WHITE HOUSE, 2022).

O segundo pilar é a regulação setorial por meio de agências e leis existentes. Como examinaremos a seguir, a FDA regula dispositivos médicos com IA. Em outros contextos, a proteção de dados neurais depende da lei aplicável. Em um ambiente clínico, o *Health Insurance Portability and Accountability Act (HIPAA)* impõe regras estritas de privacidade para informações de saúde pessoal (*PHI*) tratados por entidades cobertas (como hospitais e planos de saúde) (LAVAZZA *et al.*, 2025). Fora desse contexto, como no uso de *wearables* de consumo que coletam dados de EEG, a proteção é muito mais incerta e depende de leis de proteção ao consumidor e de movimentos legislativos em nível estadual. A Califórnia, por exemplo, tem sido pioneira ao alterar sua *California Privacy Rights Act (CPRA)* para incluir explicitamente “*neural data*” como uma categoria de “informação pessoal sensível”, conferindo aos consumidores direitos específicos sobre esses dados (CALIFORNIA, 2024; LI *et al.*, 2024).

Essa abordagem, porém, reflete o “dilema do regulador” e gera efeitos de segunda ordem distintos dos observados na Europa. Ao priorizar a flexibilidade e a inovação com uma regulação *ex post* (responsabilização após o dano), os EUA permitem um desenvolvimento tecnológico mais rápido e com menores barreiras de entrada. Contudo, o efeito colateral é um risco maior de que danos ocorram antes que a regulação possa reagir. Além disso, a fragmentação legal – uma colcha de retalhos de leis estaduais e federais setoriais – cria um cenário de conformidade complexo e incerto para empresas que operam nacionalmente, o que pode, paradoxalmente, aumentar os custos legais e de adaptação a longo prazo.

Mas é, de fato a *Food and Drug Administration (FDA)* que vem atuando na regulação de dispositivos médicos com IA e aprendizado de máquina (*Machine Learning, ML*). Desde 2019, a agência tem publicado uma série de documentos de orientação (*guidances*) e planos de ação para lidar com as especificidades dessas tecnologias (KETRYX, 2025).

Em abril de 2019 houve a publicação da “*Proposed Regulatory Framework for Modifications to AI/ML-Based Software as a Medical Device (SaMD)*”⁴³, que é a Estrutura Regulatória Proposta para Modificações em Software como Dispositivo Médico Baseado em IA/ML. Este documento é considerado uma leitura obrigatória e o marco inicial do esforço regulatório detalhado (PRIEWASSER, 2025).

O Plano de Ação para IA/ML em Dispositivos Médicos, de 2021, delineou a intenção da agência de desenvolver guias sobre boas práticas de aprendizado de máquina (*Good Machine Learning Practices - GMLP*)⁴⁴ e de estabelecer um arcabouço para algoritmos adaptativos – aqueles que podem aprender e se modificar continuamente após serem comercializados (NIXON LAW GROUP, 2025; KETRYX, 2025). Essa proatividade culminou na publicação de guias mais recentes, como o de dezembro de 2024 sobre Recomendações de Submissão para um Plano de Controle de Mudanças Pré-determinados (*PCCPs*) para dispositivos médicos. Esse documento descreve o que os fabricantes devem incluir em um *PCCP* – especificamente, uma descrição detalhada das modificações planejadas, um protocolo de modificação para desenvolver e validar essas mudanças, e uma avaliação de impacto para garantir a segurança e eficácia contínuas do dispositivo. O principal objetivo é permitir que os fabricantes implementem certas modificações predefinidas em seus dispositivos sem a necessidade de uma nova submissão de *marketing* para cada alteração, agilizando o processo e sendo menos oneroso, ao mesmo tempo em que mantém a supervisão regulatória. A orientação também detalha quais tipos de modificações são geralmente apropriados (e quais não são) para inclusão em um *PCCP*, enfatizando que essas modificações devem manter o uso pretendido e a segurança/eficácia do dispositivo (FDA, 2024). No ano seguinte a FDA publicou recomendações sobre como as empresas devem apresentar planos de controle de mudanças predeterminados (*PCCPs*) para dispositivos médicos que utilizam IA, especificamente Funções de *Software* de Dispositivo Habilitadas para IA (*AI-DSFs*) (FDA, 2025a).

⁴³ Disponível em: <https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf>. Acesso em: 10 abr. 2024.

⁴⁴ Disponível em: <https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles>. Acesso em: 10 abr. 2024.

Em 2025, a agência publicou o rascunho de guia sobre a Gestão do Ciclo de Vida de Funções de *Software* com IA (INFECTION CONTROL TODAY, 2025; KETRYX, 2025). Trata-se de um guia preliminar que oferece recomendações não-obrigatórias para fabricantes de dispositivos médicos que incorporam funções de software habilitadas para Inteligência Artificial. O objetivo principal é delinear as expectativas da FDA para a gestão do ciclo de vida e submissão de *marketing* desses dispositivos, assegurando sua segurança e eficácia. O guia enfatiza a importância da transparência, avaliação de risco e gerenciamento de dados, incluindo a minimização de vieses, em todas as etapas de desenvolvimento e uso do dispositivo, além de abordar aspectos cruciais como a descrição do dispositivo e do modelo, a interface do usuário, a rotulagem, a validação de desempenho e a cibersegurança. Ele serve como um recurso para auxiliar na preparação de documentação detalhada que apoie a análise regulatória, com exemplos como os “*model cards*” para comunicação clara de informações (FDA, 2025d). Conjuntamente, ela lançou outras recomendações, também não vinculativas, para o gerenciamento do ciclo de vida e a submissão de pedidos de comercialização de dispositivos médicos habilitados para Inteligência Artificial. O texto enfatiza a importância de uma abordagem de ciclo de vida total (*TPLC*), que aborda a segurança e eficácia do dispositivo desde o desenvolvimento até o monitoramento pós-comercialização (FDA, 2025b). Um tema recorrente em ambas é a transparência, incluindo a recomendação de “*model cards*” para comunicar informações essenciais sobre o dispositivo ao público e aos usuários (KETRYX, 2025).

Em seguida, neste mesmo ano, a agência publicou recomendações sobre o uso de inteligência artificial para apoiar decisões regulatórias relativas à segurança, eficácia e qualidade de medicamentos e produtos biológicos. O foco principal é um roteiro de avaliação de credibilidade baseado em risco, composto por sete etapas, para determinar a confiabilidade de modelos de IA em contextos de uso específicos (FDA, 2025c).

Geeta Joshi e outros colaboradores, em artigo publicado na revista internacional *Electronics*, apresentaram uma análise abrangente dos dispositivos médicos habilitados para inteligência artificial e aprendizado de máquina (IA/ML), aprovados pela FDA até 19 de outubro de 2023. Os autores detalham tendências de aprovação, indicando um aumento significativo desde 2018, com a radiologia dominando a aplicação de ferramentas de ML devido à vasta disponibilidade de dados. O estudo também destaca a predominância da via de liberação 510(k)⁴⁵, que frequentemente não exige novos ensaios clínicos (PRICE II; COHEN *et al.*, 2024),

⁴⁵ A via 510(k) (Premarket Notification) é um dos principais caminhos regulatórios nos Estados Unidos, gerenciado pela *Food and Drug Administration* (FDA), para que dispositivos médicos sejam autorizados a entrar no mercado (JOSHI *et al.*, 2024).

e observa a sub-representação de dispositivos e ensaios pediátricos, bem como a limitação geográfica dos ensaios clínicos ao território dos EUA. Até 19 de outubro de 2023, 96,7% dos dispositivos AI/ML aprovados foram liberados por meio do caminho 510(k). Apenas 2,9% foram aprovados via *De Novo* e cerca de 0,4% via *Premarket Approval (PMA)* (JOSHI *et al.*, 2024). Essa, inclusive, foi a principal crítica feita no estudo. Embora a FDA tenha aprovado um grande volume de dispositivos AI/ML (691 até outubro de 2023), o processo regulatório prioriza a velocidade de entrada no mercado através da equivalência (510(k)), em detrimento da exigência de validação clínica abrangente e da diversidade de dados necessários para garantir a segurança e a eficácia em populações variadas.

No campo específico das neurotecnologias, a FDA tem demonstrado atenção especial. Em 2021, a agência emitiu uma diretriz final para submissões pré-mercado de Dispositivos de Interface Cérebro-Computador (BCI) implantáveis, especificamente aqueles que visam restaurar funções motoras ou sensoriais em pacientes com paralisia ou amputação. O principal objetivo é assegurar a segurança e eficácia desses dispositivos inovadores, detalhando os requisitos para testes não clínicos e considerações clínicas. Ele introduz um novo “Nível de Documentação” para *software*, exigindo um “Nível de Documentação Aprimorado” para BCIs implantados, dada a sua natureza de alto risco, e abrange uma vasta gama de aspectos técnicos, desde biocompatibilidade e esterilidade até segurança elétrica, testes de software e ensaios em animais, culminando em diretrizes para o *design* de estudos clínicos e a proteção do paciente (FDA, 2021).

Embora nenhum BCI invasivo de longo prazo tenha recebido autorização plena de mercado, a FDA tem autorizado estudos clínicos experimentais caso a caso. Um marco recente foi a autorização, em 2023, para o primeiro ensaio clínico em humanos do chip cerebral da Neuralink, após uma recusa inicial por preocupações de segurança. Mais recentemente, em abril de 2025, a FDA concedeu uma autorização de mercado (510(k) *clearance* ou *Premarket Notification*)) para o dispositivo *Layer 7 Cortical Interface*, da *Precision Neuroscience*, para uso temporário (até 30 dias) em aplicações de mapeamento cerebral durante neurocirurgias. Construído sobre um filme fino e flexível, o dispositivo e seus 1.024 eletrodos podem ser alocados através de uma incisão submilimétrica e colocados em quase qualquer lugar na superfície do cérebro em um procedimento reversível. É capaz de registrar informações, bem como estimular a atividade neural, e vários implantes foram usados em um único paciente. (HALE, 2025; PRACTICAL NEUROLOGY, 2025).

A Agência está utilizando suas ferramentas regulatórias – como guias, planos de ação e autorizações condicionais de ensaios – para ativamente moldar e direcionar o caminho da

inovação tecnológica. Ao publicar guias sobre temas como algoritmos adaptativos, a FDA sinaliza para a indústria quais são as “regras do jogo”, incentivando o desenvolvimento de tecnologias seguras desde a concepção. Esse processo de diálogo e ajuste, visível na autorização dos estudos da *Neuralink* e da *Precision Neuroscience*, demonstra uma forma de governança adaptativa e experimentalista. Em vez de depender de uma lei rígida e estática, a agência utiliza instrumentos mais flexíveis para aprender sobre a tecnologia em conjunto com os desenvolvedores, ajustando as exigências à medida que o conhecimento avança. Mas, ao mesmo tempo, aqui reside um grande problema: a FDA contrasta com a Europa por sua dependência de vias como a 510(k) ou o processo *De Novo*, que foram originalmente projetados para dispositivos médicos tradicionais, e não para dispositivos baseados em IA (SOLAIMAN *et al.*, 2025; FDA, 2025a). Historicamente, a FDA tem tratado neurotecnologias como tecnologias vestíveis (*wearable technologies*). Neurotecnologias direcionadas ao consumidor nos EUA podem evitar a classificação como dispositivos médicos, alegando uma finalidade de bem-estar (*wellness purpose*). Se o risco associado ao uso for baixo, o dispositivo é considerado um produto de consumo, sujeito a regulamentações de segurança e publicidade de produtos de consumo (FARAHANY, 2023; LAVAZZA *et al.*, 2025).

Em suma, a análise da atuação da ANVISA e da FDA, apesar de terem focos diferentes (e apesar das fortes críticas ao modelo americano), revela o papel crucial que as agências reguladoras assumiram como protagonistas na governança da inteligência artificial e das neurotecnologias na saúde. Um fato é inconteste: na ausência de leis federais específicas, são elas que, por meio de seus instrumentos técnicos e administrativos, estabelecem a primeira e mais importante barreira de proteção ao paciente, adaptando arcabouços existentes para enquadrar inovações disruptivas.

Contudo, essa governança reativa, embora indispensável, evidencia, é verdade, uma lacuna perigosa e insustentável. A complexidade de algoritmos de “caixa-preta” e a profundidade ética das interfaces cérebro-computador levantam questões de transparência, viés, responsabilidade e autonomia que transcendem a avaliação de segurança e eficácia de um dispositivo. Portanto, a atuação das agências, por mais robusta que seja, não substitui, mas sim clama pela urgência de uma regulamentação federal dedicada, que estabeleça um alicerce legal claro tanto para a IA médica, garantindo sua explicabilidade e auditabilidade, quanto para as neurotecnologias em saúde, endereçando seus dilemas éticos singulares. Sem esse passo legislativo, a regulação permanecerá um passo atrás da inovação, deixando a sociedade e os pacientes em uma posição de vulnerabilidade crônica diante das fronteiras tecnológicas mais desafiadoras do nosso tempo.

7.5 ANÁLISE COMPARATIVA E CONSIDERAÇÕES FINAIS

A análise das abordagens regulatórias para inteligência artificial e neurotecnologias no Brasil, na União Europeia e nos Estados Unidos revela a existência de três modelos filosóficos e práticos distintos, cada um com seus próprios méritos, desafios e implicações para o futuro da inovação e da proteção de direitos. A tabela a seguir, sintetiza as principais características de cada jurisdição, servindo como base para uma reflexão conclusiva.

Tabela 7. Quadro Comparativo das Abordagens Regulatórias para IA e Neurotecnologias

| Característica Regulatória | Brasil (Propostas) | União Europeia | Estados Unidos |
|---|---|--|--|
| Filosofia Principal | Baseada em Direitos Fundamentais | Preventiva, Baseada em Risco | Setorial, Reativa, Baseada em Mercado |
| Instrumento Legal Principal (IA) | PL 2.338/2023 (em tramitação) | Regulamento (UE) 2024/1689 (<i>AI Act</i>) | Inexistente (múltiplas leis setoriais) |
| Reconhecimento de Neurodireitos | Explícito (PEC 29/2023, PL 2.174/2023) | Implícito (via <i>GDPR</i> e Direitos Fundamentais) | Inexistente em nível federal; emergente em nível estadual (e.g., Califórnia) |
| Regulação de Dispositivos Médicos com IA | ANVISA (RDC 751/2022, RDC 657/2022) | Dupla camada: <i>MDR</i> + <i>AI Act</i> | FDA (guias específicos, autorizações caso a caso) |
| Proteção de Dados Neurais | Proposta de lei específica; LGPD (dados de saúde) | <i>GDPR</i> (dados de saúde/biométricos) | Fragmentada: <i>HIPAA</i> (saúde) + Leis Estaduais (e.g., <i>CPRA</i>) |
| Papel de Normas Voluntárias | Secundário, foco em legislação | Complementar (padrões harmonizados CEN-CENELEC ⁴⁶) | Central (<i>NIST AI RMF</i> , <i>White House Blueprint</i>) |

Fonte: Elaborada pelo autor (2025).

A União Europeia estabelece a vanguarda na regulação de IA ao implementar, por meio do Regulamento da IA (*AI Act*), uma abordagem baseada no risco (Artigo 43, nº 8), que foca em requisitos *ex ante* (preventivos) para sistemas de alto risco. Este arcabouço legal visa garantir que os sistemas de IA, que entram no mercado europeu, sejam seguros, confiáveis e respeitem os direitos fundamentais e os valores da União, como a dignidade humana, a transparência, e a não-discriminação (BALCERZAK; KAPELAŃSKA-PRĘGOWSKA, 2024).

⁴⁶ O CEN (Comitê Europeu de Normalização) e o CENELEC (Comitê Europeu de Normalização Eletrotécnica) trabalham para criar normas que harmonizam as regras técnicas nos Estados-Membros. Desempenham um papel crucial no cumprimento dos requisitos do Regulamento da IA (*AI Act*) e dos Regulamentos de Dispositivos Médicos (*MDR/IVDR*) na União Europeia (BALCERZAK; KAPELAŃSKA-PRĘGOWSKA, 2024).

No setor da saúde, a legislação impõe um regime de dupla conformidade e supervisão rigorosa, exigida pelo *MDR* e pelo *AI Act*. Em contrapartida, a UE avança de forma mais cautelosa no campo dos neurodireitos, preferindo proteger a esfera mental por meio da aplicação robusta de seu já denso arcabouço de proteção de dados e direitos humanos, sem, contudo, criar um *nomen juris* novo para esses direitos (CORNEJO-PLAZA; CIPPITANI, 2024).

Os Estados Unidos, por outro lado, seguem uma abordagem mais liberal, setorial e baseada na responsabilização *ex post*, fortemente influenciada pelo poder do mercado, ficando com os Tribunais a função de corrigir excessos. A FDA, o órgão regulador dos EUA, adota uma abordagem mais pragmática e flexível, adaptando vias regulatórias existentes (como 510(k) e *De Novo*), em vez de criar uma legislação específica, como a do *AI Act* da UE e das propostas brasileiras (LAVAZZA *et al.*, 2025; MELLO *et al.*, 2025). A vantagem dessa abordagem é a maior flexibilidade e velocidade para a inovação. A desvantagem é o risco de que lacunas regulatórias só se tornem evidentes após a ocorrência de danos. No tema dos neurodireitos, os EUA apoiam-se em seu arcabouço constitucional tradicional de liberdades civis, deixando, no entanto, uma margem de incerteza jurídica, especialmente para neurotecnologias de consumo que operam fora do perímetro de leis de saúde como o *HIPAA*.

O Brasil encontra-se em um momento proativo e singular de formulação normativa, buscando construir um modelo híbrido. Influenciado tanto pelo movimento de direitos humanos digitais da América Latina quanto pela abordagem baseada em risco da Europa, o país articula estratégias em múltiplos níveis: (i) a inserção de princípios de proteção da integridade mental diretamente na Constituição, conferindo-lhes *status* de direito fundamental; (ii) a criação de leis específicas para definir e proteger neurodireitos e dados neurais; e (iii) o estabelecimento de um marco legal abrangente para IA. Essa abordagem é ambiciosa e, em tese, busca um equilíbrio virtuoso entre o incentivo à inovação e a salvaguarda de valores humanistas.

Contudo, o sucesso desse modelo híbrido dependerá criticamente da sua implementação e da articulação coerente entre as diferentes normas e os órgãos responsáveis por sua aplicação. Será crucial coordenar a atuação do Congresso Nacional, do Poder Executivo (incluindo uma futura Autoridade/Agência Nacional de IA) e das agências setoriais, como a ANVISA e a ANPD, para evitar sobreposições, conflitos normativos e a criação de direitos que permaneçam apenas no papel.

No campo da saúde, observa-se um consenso global emergente, ético e político, quanto à necessidade de controle rigoroso sobre dispositivos médicos inteligentes e neurotecnologias, dada a especial vulnerabilidade dos pacientes e a assimetria de informação entre usuário e fornecedor (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2021; NOGAROLI, 2023; NEGRI *et*

al., 2025). Organismos internacionais, como a Organização Mundial da Saúde (OMS), reconhecem o potencial transformador da IA para ampliar o acesso, melhorar diagnósticos e otimizar planos terapêuticos, mas alertam que tais benefícios só podem ser alcançados com a adoção de salvaguardas éticas robustas, contemplando segurança, explicabilidade, supervisão humana, equidade e responsabilidade jurídica (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2021; AMANN *et al.*, 2020; UNESCO, 2025).

A principal divergência entre jurisdições reside no desenho regulatório adotado para implementar esse controle. Em alguns casos, a proteção ocorre por meio de preceitos constitucionais – como no Chile, que introduziu na sua Constituição a proteção da integridade física e psíquica, conferindo base normativa para limitar intervenções neurotecnológicas abusivas (RUIZ *et al.*, 2024). Outros optam por uma regulação *ex ante* detalhada, como a União Europeia, cujo *Artificial Intelligence Act* (Regulamento UE 2024/1689) complementa o arcabouço já existente para dispositivos médicos no *Medical Devices Regulation* (Regulamento UE 2017/745) e no *In Vitro Diagnostic Medical Devices Regulation* (Regulamento UE 2017/746), impondo requisitos de gestão de risco, governança de dados e transparência antes da colocação do produto no mercado (UNIÃO EUROPEIA, 2024; 2017a; 2017b). Já os Estados Unidos mantêm uma abordagem mais flexível e setorial, centrada em responsabilização *ex post* e monitoramento posterior (*post-market surveillance*), com regulação setorial conduzida por agências como a *Food and Drug Administration* (FDA), orientada por *guidances* voluntários e mecanismos de mercado e judiciais para correção de excessos (FDA, 2025; PASQUALE, 2020).

Assim, embora haja alinhamento internacional quanto à necessidade de regular dispositivos médicos inteligentes e neurotecnologias, o formato e o rigor desse controle variam significativamente, oscilando entre abordagens constitucionais de proteção de direitos, modelos regulatórios *ex ante* de alto detalhamento e regimes flexíveis de supervisão posterior.

Em síntese, Brasil, UE e EUA ilustram três caminhos regulatórios distintos. A abordagem de direitos explícitos, como a proposta no Brasil e já implementada no Chile, representa um reconhecimento enfático da centralidade da mente humana na era digital e pode orientar políticas e decisões judiciais de forma mais clara. O modelo europeu, por sua vez, estabelece um padrão elevado de segurança que pode influenciar mercados globais. Já o modelo americano prioriza a agilidade, confiando na capacidade do sistema de se autocorrigir. Provavelmente, uma convergência ocorrerá a longo prazo, à medida que o diálogo internacional e a disseminação de boas práticas se intensificarem. Para o Brasil, a participação ativa nesses fóruns globais e a atualização constante de suas normas técnicas, como a ANVISA vem

fazendo, serão vitais para manter o país alinhado às melhores práticas mundiais e, ao mesmo tempo, inovador na proteção dos direitos humanos diante das mais disruptivas tecnologias do século XXI.

8 CONCLUSÃO

Ao término desta jornada investigativa, que se propôs a navegar as complexas e, por vezes, turbulentas águas da intersecção entre inteligência artificial, neurotecnologias, medicina e direito, emerge uma conclusão central e inevitável: estamos diante de uma crise paradigmática. A promessa de uma medicina mais precisa, preditiva e personalizada, impulsionada por algoritmos e interfaces cérebro-computador, é real e promissora. Contudo, as ferramentas que forjam esse futuro trazem consigo características intrínsecas – notadamente a opacidade e o viés – que abalam os alicerces de um dos mais caros institutos da ética e do direito médico: o Consentimento Livre e Esclarecido.

A pergunta que orientou este trabalho não foi meramente sobre como gerir os “riscos” da tecnologia, mas sobre como o próprio modo de operação dessa tecnologia provoca o colapso funcional do principal mecanismo de tutela da autonomia do paciente. A trajetória percorrida ao longo dos capítulos buscou, portanto, não apenas descrever um fenômeno, mas diagnosticar uma crise e, a partir dela, vislumbrar os caminhos para a reconstrução da proteção da pessoa humana em um novo cenário tecnológico.

A investigação partiu, no Capítulo 2, do mapeamento do terreno tecnológico, traçando a evolução da IA e das neurotecnologias e apresentando casos concretos que ilustram tanto seu imenso potencial quanto seus riscos latentes. Demonstrou-se que a sofisticação crescente dessas ferramentas caminhou em paralelo com sua crescente indecifrável, culminando no que conceituamos como a “dupla caixa-preta” (complexidade do cérebro + algoritmo opaco): o algoritmo opaco que analisa a fonte de dados mais inescrutável de todas, o cérebro humano. Essa contextualização foi fundamental para estabelecer que a crise do consentimento não é um acidente, mas uma consequência direta da trajetória tecnológica.

O Capítulo 3 ergueu o primeiro pilar da dissertação, ao aprofundar a análise sobre a opacidade. Argumentou-se que a “caixa-preta” algorítmica, ao tornar o processo decisório indecifrável, ataca diretamente o pressuposto “esclarecido” do consentimento. Um paciente que anui a um diagnóstico ou a uma terapia baseada em uma lógica que nem mesmo o médico pode explicar plenamente não está, em verdade, consentindo de forma esclarecida. Ele consente com o inexplicável, um ato que se assemelha mais a um cheque em branco do que a uma deliberação autônoma. A opacidade, portanto, introduz um risco epistêmico – o risco de não saber como uma decisão foi tomada – que anula o consentimento em sua raiz e desafia o direito fundamental à explicação, cuja tutela já se vislumbra em marcos como a Lei Geral de Proteção de Dados (BRASIL, 2018) e o *AI Act* europeu (UNIÃO EUROPEIA, 2024), e que clama por

reconhecimento constitucional no Brasil, como propõe o PL 2.338/2023, o PL 2.174/2023 e a PEC 29/2023 (BRASIL, 2023a; 2023b; 2023c).

Sobre o segundo pilar, erguido no Capítulo 4, demonstrou-se que o viés algorítmico e o potencial manipulatório das tecnologias digitais atacam o pressuposto “livre” do consentimento. Ao absorver e amplificar as desigualdades estruturais presentes nos dados de treinamento, os algoritmos deixam de ser ferramentas neutras para se tornarem, nas palavras de Cathy O’Neil (2016), “armas de destruição matemática” que podem perpetuar a discriminação sob um manto de objetividade técnica. Uma escolha baseada em informação de qualidade desigual – como um diagnóstico menos acurado para uma paciente negra por vieses no treinamento do modelo – não é uma escolha genuinamente livre. A liberdade é corrompida pela iniquidade. Ademais, a análise de *nudges* e *dark patterns* revelou como a própria arquitetura da tecnologia pode explorar vieses cognitivos para manipular sutilmente as decisões de médicos e pacientes, afastando-os de uma deliberação verdadeiramente autônoma.

A confluência desses dois pilares, analisada no coração desta dissertação, o Capítulo 5, permitiu a articulação do argumento central da dissertação: a sinergia entre a opacidade (que torna o consentimento não esclarecido) e o viés (que o torna não livre) leva ao colapso funcional da doutrina do Consentimento Livre e Esclarecido. O modelo tradicional, centrado na divulgação de informações e na escolha individual, mostra-se dramaticamente insuficiente para proteger o paciente em um ambiente tecnológico opaco, discriminatório e potencialmente manipulador. A crise, portanto, não é de ajuste, mas de paradigma.

É a partir do diagnóstico dessa crise que a principal contribuição desta dissertação se manifesta no Capítulo 6: a proposição de uma transição para um paradigma de autodeterminação efetiva e de “tutela estrutural”. Superar o consentimento como salvaguarda única e individualista significa reconhecer que a proteção da autonomia do paciente não pode mais repousar exclusivamente sobre seus ombros, sobrecarregando-o com o ônus de decifrar tecnologias complexas e de se precaver contra vieses sistêmicos. A autodeterminação efetiva exige mais do que a mera formalidade de uma escolha; ela demanda a garantia de um ambiente decisório íntegro, justo e transparente.

Essa tutela estrutural desloca o foco da proteção da *escolha* para a governança da *tecnologia*. Em vez de perguntar apenas “o paciente foi informado?”, devemos passar a exigir: “o algoritmo foi auditado?”, “os dados de treinamento são representativos?”, “existem mecanismos de supervisão humana eficazes?”, “a interface foi projetada para empoderar ou para manipular?”. A proteção da pessoa passa a depender menos de um ato de consentimento isolado e mais de um ecossistema de responsabilidade compartilhada, que envolve

desenvolvedores, instituições de saúde, profissionais e reguladores em um compromisso contínuo com a segurança, transparência/explicabilidade, equidade e *accountability*.

O panorama comparativo de modelos regulatórios, apresentado no Capítulo 7, demonstrou que essa transição paradigmática não é uma mera abstração teórica, mas uma tendência que já se materializa, com diferentes matizes, nas respostas jurídicas que emergem globalmente. O modelo europeu, com seu *AI Act*, é talvez o exemplo mais claro de uma tutela estrutural, ao impor obrigações *ex ante* baseadas em risco, que visam garantir a confiabilidade da tecnologia antes que ela chegue ao paciente. A abordagem latino-americana, com o pioneirismo do Chile (CHILE, 2021) e as propostas no Brasil, busca blindar a proteção da mente em nível constitucional, reconhecendo que a integridade mental e a liberdade cognitiva são pré-condições para qualquer exercício de autonomia. Mesmo a abordagem setorial norte-americana, por meio da FDA, avança na direção de uma governança mais adaptativa e focada no ciclo de vida da tecnologia (FDA, 2025).

As implicações desta pesquisa para o ordenamento jurídico brasileiro são profundas. A dissertação aqui defendida sugere que a Lei Geral de Proteção de Dados, embora fundamental, é insuficiente para abarcar a complexidade dos desafios neurotecnológicos e da IA em saúde. A proteção da autonomia do paciente exige mais do que a tutela da autodeterminação informativa (DONEDA, 2021); demanda a proteção da própria integridade do processo de deliberação. Nesse sentido, iniciativas como o Projeto de Lei 2.338/2023 (Marco Legal da IA) e, sobretudo, a PEC 29/2023, são passos na direção correta, ao reconhecerem a necessidade de uma governança específica e de uma proteção fundamental para a mente humana.

Reconhecemos, contudo, as limitações deste trabalho. A velocidade estonteante da inovação tecnológica faz com que qualquer análise corra o risco de se tornar rapidamente datada. Ademais, nosso foco foi eminentemente teórico-jurídico, carecendo de estudos empíricos que possam aferir, na prática, como pacientes e médicos percebem e lidam com os riscos da opacidade e do viés.

Essas limitações, no entanto, abrem uma vasta e fértil agenda para pesquisas futuras. Torna-se imperativo o desenvolvimento de estudos interdisciplinares que unam o Direito, a Ciência da Computação, a Medicina e as Ciências Sociais para: (i) *Criar e validar metodologias de auditoria algorítmica* adaptadas ao contexto da saúde e à diversidade da população brasileira; (ii) *Investigar empiricamente a eficácia de diferentes modelos de comunicação e de consentimento* para tecnologias de IA, como o consentimento dinâmico; (iii) *Aprofundar a reflexão sobre novos regimes de responsabilidade civil* para danos causados por sistemas

autônomos; e (iv) *Analisar os desafios de implementação e fiscalização* do futuro marco regulatório brasileiro para a IA e as *Neurotechs*.

Em suma, esta dissertação buscou demonstrar que a revolução tecnológica na saúde nos confronta com uma escolha fundamental. Podemos abraçar a inovação de forma acrítica, aceitando a erosão da autonomia como um custo inevitável do progresso, ou podemos, de forma consciente e deliberada, redesenhar nossos marcos éticos e jurídicos para garantir que a tecnologia sirva à humanidade, e não o contrário. A transição de um paradigma de consentimento individual para um de tutela estrutural é o caminho para essa segunda via. É o reconhecimento de que a proteção da dignidade humana, na era da inteligência artificial e das neurotecnologias, exige não apenas que se respeite a escolha do indivíduo, mas que se construa um mundo em que essa escolha possa ser, verdadeiramente, livre e esclarecida.

REFERÊNCIAS

ABBOTT BRASIL. **Abbott recebe aprovação da Anvisa para o sistema de estimulação da medula espinhal Eterna™.** São Paulo, 20 maio 2025. Disponível em: <https://www.abbottbrasil.com.br/corpnewsroom/noticias/press-releases/abbott-recebe-aprovacao-da-anvisa-para-o-sistema-de-estimulacao-da-medula-espinhal-eterna.html>. Acesso em: 15 jul. 2025.

ABRÀMOFF, Michael D.; LAVIN, Phillip T. *et al.* **Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices.** NPJ Digital Medicine, 2018. Disponível em: <https://doi.org/10.1038/s41746-018-0040-6>. Acesso em: 25 de maio de 2024.

ABUJABER, Ahmad A.; NASHWAN, Abdulqadir J. **Ethical framework for artificial intelligence in healthcare research: A path to integrity.** World journal of methodology, 2024. Disponível em: <https://doi.org/10.5662/wjm.v14.i3.94071>. Acesso em: 20 ago. 2024.

ADAMSON, Adewole; SMITH, Avery. **Machine Learning and Health Care Disparities in Dermatology.** JAMA Dermatology, 2018. Disponível em: <https://doi.org/10.1001/jamadermatol.2018.2348>. Acesso em: 12 abr. 2024.

AIB/MDCG. **Joint Artificial Intelligence Board and Medical Device Coordination Group Document.** AIB 2025-1 / MDCG 2025-6 - FAQ on Interplay between the Medical Devices Regulation & In vitro Diagnostic Medical Devices Regulation and the Artificial Intelligence Act (June 2025). Disponível em: https://health.ec.europa.eu/document/download/b78a17d7-e3cd-4943-851d-e02a2f22bbb4_en?filename=mdcg_2025-6_en.pdf. Acesso em: 7 jul. 2025.

AITH, Fernando. **Desafios regulatórios da inteligência artificial em saúde.** Jota.info, 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/coluna-fernando-aith/desafios-para-a-regulacao-da-ia-em-saude-no-brasil>. Acesso em: 19 mar. 2025.

ALAVI, Abass; REIVICH, Martin. **Pioneers reflect on history, future of FDG-PET.** Clinical News | Molecular Imaging, 2002. Disponível em: <https://www.auntminnie.com/clinical-news/molecular-imaging/article/15559948/pioneers-reflect-on-history-future-of-fdg-pet>. Acesso em: 2 de julho de 2024.

ALBUQUERQUE, Aline. **Inteligência artificial e direito do paciente: desafios éticos e jurídicos.** Revista Peruana de Derecho Constitucional, [S. l.], n. 16, p. 137–157, 2025. Disponível em: <https://revista.tc.gob.pe/index.php/revista/article/view/439>. Acesso em: 14 ago. 2025.

ALEXY, Robert. **Teoria dos direitos fundamentais.** 3. ed. trad. Virgílio Afonso da Silva. São Paulo: Malheiros, 2024.

ALJUHANI, Manal *et al.* **Use of Artificial Intelligence in Imaging Dementia.** Cells, 2024. Disponível em: <https://doi.org/10.3390/cells13231965>. Acesso em: 15 mar. 2025.

ALSAEDI, Sakhaa; OGASAWARA, Michihiro *et al.* **AI-powered precision medicine: utilizing genetic risk factor optimization to revolutionize healthcare.** NAR Genomics and

Bioinformatics, Volume 7, Issue 2, 2025. Disponível em: <https://doi.org/10.1093/nargab/lqaf038>. Acesso em: 4 jul. 2025.

ALTMAN, Russ; BLACK, Kameron; CHEN, Jonathan *et al.* **Chapter 5: Science and Medicine.** In: *AI Index and Stanford HAI*, Annual Report by Stanford University, 2025. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_AI-Index-Report-2024_Chapter5.pdf. Acesso em: 4 jul. 2025.

AMANN, Julia; BLASIMME, Alessandro *et al.* **Explainability for artificial intelligence in healthcare: a multidisciplinary perspective.** BMC Medical Informatics and Decision Making, v. 20, n. 310, p. 1-9, 2020. Disponível em: <https://doi.org/10.1186/s12911-020-01332-6>. Acesso em: 14 nov. 2024.

ANDORNO, Roberto. **The right not to know: An autonomy based approach.** Journal of Medical Ethics, 2004. Disponível em: <https://doi.org/10.1136/jme.2002.001578>. Acesso em: 17 set. 2024.

ANDORNO, Roberto; MUHL, Ekaterina. **Neurosurveillance in the workplace: do employers have the right to monitor employees' minds?** Front. Hum. Dyn, 2023. Disponível em: <https://doi.org/10.3389/fhumd.2023.1245619>. Acesso em: 16 ago. 2024.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. **Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks.** *ProPublica*, 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 24 jun. 2024.

ANPD – AGÊNCIA NACIONAL DE PROTEÇÃO DE DADOS. **Radar tecnológico: neurotecnologias.** Brasília, DF: ANPD, 2025. (n. 4). Disponível em: <https://www.gov.br/anpd/pt-br/centrais-de-conteudo/documentos-tecnicos-orientativos/radar-tecnologico-4-neurotecnologias.pdf>. Acesso em: 13 ago. 2025.

ANVISA. **Resolução de Diretoria Colegiada - RDC N° 657, de 24 de março de 2022.** Dispõe sobre a regularização de software como dispositivo médico (Software as a Medical Device - SaMD). Brasília, DF: ANVISA, 2022a. Disponível em: <https://in.gov.br/en/web/dou/-/resolucao-de-diretoria-colegiada-rdc-n-657-de-24-de-marco-de-2022-389603457>. Acesso em: 12 ago. 2025.

ANVISA. **Resolução de Diretoria Colegiada - RDC N° 751, de 15 de setembro de 2022.** Dispõe sobre a classificação de risco, os regimes de notificação e de registro, e os requisitos de rotulagem e instruções de uso de dispositivos médicos. Brasília, DF: ANVISA, 2022b. Disponível em: <https://www.in.gov.br/en/web/dou/-/resolucao-rdc-n-751-de-15-de-setembro-de-2022-430797145>. Acesso em: 12 ago. 2025.

ANVISA. **Resolução RDC n° 777, de 01 de março de 2023.** Altera a Resolução de Diretoria Colegiada - RDC n° 751, de 15 de setembro de 2022, que dispõe sobre a classificação de risco, os regimes de notificação e de registro, e os requisitos de rotulagem e instruções de uso de dispositivos médicos. Brasília, DF: ANVISA, 2023a. Disponível em: <https://anvisalegis.datalegis.net/action/ActionDatalegis.php?acao=abrirTextoAto&tipo=RDC&numeroAto=00000777&seqAto=000&valorAno=2023&orgao=RDC/DC/ANVISA/MS&co>

[dTipo=&desItem=&desItemFim=&cod_menu=9434&cod_modulo=310&pesquisa=true](#). Acesso em: 11 ago. 2024.

ANVISA. **Resolução RDC nº 830, de 06 de dezembro de 2023**. Dispõe sobre a classificação de risco, os regimes de notificação e de registro e os requisitos de rotulagem e instruções de uso de dispositivos médicos para diagnóstico in vitro, inclusive seus instrumentos. Brasília, DF: ANVISA, 2023b. Disponível em: https://antigo.anvisa.gov.br/documents/10181/6763806/OS_131_2024.pdf/34eeb5d1-f5ae-4d24-baf1-2a737c72c167. Acesso em: 11 ago. 2024.

ANVISA. **Orientação de Serviço nº 131/ANVISA, de 29 de maio de 2024**. Dispõe sobre o fluxo para a elaboração e a deliberação de instrumentos regulatórios. Brasília, DF: ANVISA, 2024. Disponível em: https://antigo.anvisa.gov.br/documents/10181/6763806/OS_131_2024.pdf/34eeb5d1-f5ae-4d24-baf1-2a737c72c167. Acesso em: 11 ago. 2024.

ANVISA. **Resolução-RE nº 1.206, de 27 de Março de 2025**. Defere as petições relacionadas à Gerência-Geral de Tecnologia de Produtos para Saúde e dá outras providências. Brasília, DF: ANVISA, 2025a. Disponível em: https://anvisalegis.datalegis.net/action/UrlPublicasAction.php?acao=abrirAtoPublico&num_ato=00001206&sgl_tipo=RES&sgl_orgao=RE/GGTPS/ANVISA/MS&vlr_ano=2025&seq_ato=000&cod_modulo=293&cod_menu=8499. Acesso em: 7 jun. 2025.

ANVISA. **Identificação Única de Dispositivos Médicos: próximas etapas de implementação**. Brasília, 10 jul. 2025b. Disponível em: <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2025/identificacao-unica-de-dispositivos-medicos-proximas-etapas-de-implementacao>. Acesso em: 25 ago. 2025.

ANVISA. **OFÍCIO CIRCULAR Nº 2/2025/SEI/GGTPS/DIRE3/ANVISA**. Orientações sobre as determinações da RDC nº 591/2021 que se tornam obrigatórias a partir de 10 de julho de 2025. Brasília, 2025c. Disponível em: <https://www.gov.br/anvisa/pt-br/assuntos/noticias-anvisa/2025/identificacao-unica-de-dispositivos-medicos-proximas-etapas-de-implementacao>. Acesso em: 25 ago. 2025.

API4AI. **Why Image Anonymization is Vital for GDPR Compliance**. Medium, 2025. Disponível em: <https://medium.com/@API4AI/why-image-anonymization-is-vital-for-gdpr-compliance-11a714418522>. Acesso em: 4 jul. 2025.

AQUINO, Yves Saint James. **Making decisions: Bias in artificial intelligence and data-driven diagnostic tools**. AJGP, 2023. Disponível em: <https://www.doi.org/10.31128/AJGP-12-22-6630>. Acesso em: 10 ago. 2024.

ARANTES, Maria Clara Rodrigues. **Da Proteção Jurídica à Integridade Mental: Uma Análise sobre os Neurodireitos e sua Regulamentação**. Revista Âmbito Jurídico, 2025. Disponível em: <https://ambitojuridico.com.br/da-protecao-juridica-a-integridade-mental-uma-analise-sobre-os-neurodireitos-e-sua-regulamentacao>. Acesso em: 22 jul. 2025.

ARAÚJO, Amanda de. **Um Estudo Sobre a Utilização de Inteligência Artificial na Detecção de Câncer de Pele**. Instituto Federal Goiano, 2024. Disponível em: <https://repositorio.ifgoiano.edu.br/handle/prefix/5013>. Acesso em: 7 ago. 2024.

ARAÚJO, Maria Carolina Gomes De *et al.* **Neuroimagem Avançada e Identificação Precoce da Doença de Alzheimer: Uma Revisão Narrativa.** In: *Anais do Congresso Nacional de Neurologia Multidisciplinar*. Sete Lagoas (MG) Online, 2024. Disponível em: <https://www.even3.com.br/anais/1-congresso-nacional-de-neurologia-multidisciplinar-412958/785230-neuroimagem-avancada-e-identificacao-precoce-da-doenca-de-alzheimer--uma-revisao-narrativa>. Acesso em: 23 abr. 2025.

ARDILA, Diego *et al.* **End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography.** *Nature Medicine*, v. 25, p. 954–961, 2019. Disponível em: <https://doi.org/10.1038/s41591-019-0447-x>. Acesso em: 23 abr. 2024.

ARROYO, Verónica. **Emerging Uses of Neurotechnology.** Literature review prepared for the Office of the Information and Privacy Commissioner of Ontario, 2025. Disponível em: <https://www.ipc.on.ca/en/media/5349/download?attachment>. Acesso em: 20 jul. 2025.

ASTOBIZA, Anibal M.; BERIAIN; Íñigo de Miguel. **From neurorights to neuroduties: the case of personal identity** [version 2; peer review: 2 approved, 1 approved with reservations]. *Bioethics Open Res*, 2024. Disponível em: <https://doi.org/10.12688/bioethopenres.17501.2>. Acesso em: 10 mai. 2025.

AUSTRÁLIA. Australian Human Rights Commission. **Protecting Cognition: Background Paper on Human Rights and Neurotechnology.** Sydney: Australian Human Rights Commission, 2024. Disponível em: <https://humanrights.gov.au/our-work/technology-and-human-rights/publications/protecting-cognition-background-paper>. Acesso em: 11 mar. 2025.

AZIZ, Noor A.; MANZOOR, Awais *et al.* **Unveiling Explainable AI in Healthcare: Current Trends, Challenges, and Future Directions.** *medRxiv*, 2025. Disponível em: <https://doi.org/10.1101/2024.08.10.24311735>. Acesso em: 25 jul. 2025.

BALCERZAK, Michał; KAPELAŃSKA-PRĘGOWSKA, Julia. **Artificial Intelligence and International Human Rights Law Developing Standards for a Changing World.** Edward Elgar Publishing, 2024. Disponível em: <https://doi.org/10.4337/9781035337934>. Acesso em: 25 jul. 2025.

BALICER, Ran D.; COHEN-STAVI, Chandra. **Advancing Healthcare Through Data-Driven Medicine.** In: *Healthcare and Artificial Intelligence*. Springer Cham, 2020. Disponível em: https://doi.org/10.1007/978-3-030-32161-1_2. Acesso em: 28 mar. 2024.

BARBERO, Alvaro; GROSSE-WENTRUP, Moritz. **Biased feedback in brain-computer interfaces.** *Journal of neuroengineering and rehabilitation*, 2010. Disponível em: <http://dx.doi.org/10.1186/1743-0003-7-34>. Acesso em: 10 mai. 2024.

BARKER, Anthony T. *et al.* **Non-invasive magnetic stimulation of human motor cortex.** *The Lancet*, Volume 325, Issue 8437, 1106 - 1107, 1985. Disponível em: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(85\)92413-4/](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(85)92413-4/). Acesso em: 2 set. 2024.

BARNOVA, Katerina *et al.* **Implementation of artificial intelligence and machine learning-based methods in brain–computer interaction.** *Computers in Biology and Medicine*, Volume 163, 2023. Disponível em: <https://doi.org/10.1016/j.compbiomed.2023.107135>. Acesso em 4 set. 2024.

BAROCAS, Solon; SELBST, Andrew D. **Big Data's Disparate Impact**. *California Law Review*, 2016. Disponível em: <https://doi.org/10.2139/ssrn.2477899>. Acesso em: 9 set. 2024.

BAYSARI, Melissa Therese; HILMER, Sarah Nicole *et al.* **Effectiveness of computerised alerts to reduce drug–drug interactions (DDIs) and DDI-related harm in hospitalised patients: a quasi-experimental controlled pre–post study**. *BMJ Qual Saf*, 2025. Disponível em: <https://qualitysafety.bmj.com/content/qhc/early/2025/06/02/bmjqs-2024-018243.full.pdf>. Acesso em: 22 abr. 2025.

BEAUCHAMP, Tom L.; CHILDRESS, James F. **Principles of Biomedical Ethics**. 7. ed. New York: Oxford University Press, 2013.

BENABID, Alim-Louis *et al.* **Combined (Thalamotomy and Stimulation) Stereotactic Surgery of the VIM Thalamic Nucleus for Bilateral Parkinson Disease**. *Applied Neurophysiology*, v. 50, n. 1-6., 1987. Disponível em: <https://doi.org/10.1159/000100803>. Acesso em: 2 set. 2024.

BENABID, Alim-Louis. **Treatment for Parkinson's disease**. European Inventor Award, European Patent Office, 2016. Disponível em <https://www.epo.org/en/news-events/european-inventor-award/meet-the-finalists/alim-louis-benabid>. Acesso em: 2 out. 2024.

BENDER, Emily M. *et al.* **On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?** *In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery, New York, NY, USA, 610–623, 2021. Disponível em: <https://doi.org/10.1145/3442188.3445922>. Acesso em: 27 jun. 2024.

BERTONI, Eduardo; IENCA, Marcello. **The privacy and data protection implication of the use of neurotechnology and neural data from the perspective of Convention 108**. Paper prepared for presentation to the 46th Plenary meeting, 2024. Disponível em: <https://rm.coe.int/expert-report-neuroscience/1680b12eaa>. Acesso em: 20 nov. 2024.

BEZERRA, Pedro Miranda Vieira *et al.* [organizadores]. **Medicina personalizada [livro eletrônico]: genética, ambiente, estilo de vida e suas influências nas novas abordagens**. 1.ed. Campo Grande, MS: Editora Inovar, 2024. PDF. Disponível em: <https://doi.org/10.36926/editorainovar-978-65-5388-279-9>. Acesso em: 10 abr. 2025.

BHARGAVA, Akhil; LOPEZ-ESPINA, Carlos; SCHMALZ, Lee *et al.* **FDA-Authorized AI/ML Tool for Sepsis Prediction: Development and Validation**. *NEJM AI*, 2024. Disponível em: <https://ai.nejm.org/doi/full/10.1056/AIoa2400867>. Acesso em: 20 nov. 2024.

BHASKARANAND, Malavika, RAMACHANDRA, Chaithanya *et al.* **The Value of Automated Diabetic Retinopathy Screening with the EyeArt System: A Study of More Than 100,000 Consecutive Encounters from People with Diabetes**. *Diabetes technology & therapeutics*, 2019. Disponível em: <https://doi.org/10.1089/dia.2019.0164>. Acesso em: 19 fev. 2024.

BIGNAMI, Elena Giovanna, BERDINI, Michele *et al.* **Artificial Intelligence in Sepsis Management: An Overview for Clinicians**. *Journal of clinical medicine*, 14(1), 286, 2025. Disponível em: <https://doi.org/10.3390/jcm14010286>. Acesso em: 14 ago. 2025.

BLACKMAN, Justin; VEERAPEN, Richard. **On the practical, ethical, and legal necessity of clinical Artificial Intelligence explainability: an examination of key arguments.** BMC Medical Informatics and Decision Making, 2025. Disponível em: <https://doi.org/10.1186/s12911-025-02891-2>. Acesso em: 30 jul. 2025.

BLAJ, Andrei. **Explainable AI for Personalized Treatment: A New Era of Trustworthy Healthcare.** Medicaï, 2025. Disponível em: <https://blog.medicaï.io/en/explainable-ai-for-personalized-treatment/>. Acesso em: 25 jul. 2025.

BLAY, Jean-Yves *et al.* **Artificial Intelligence Applied to Oncology.** *In Healthcare and Artificial Intelligence.* Cham: Springer, 2020. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-32161-1>. Acesso em: 15 jun. 2024.

BOMMASANI, Rishi; HUDSON, Drew; ADELI, Ehsan *et al.* **On the Opportunities and Risks of Foundation Models.** Stanford: Stanford Institute for Human-Centered Artificial Intelligence (HAI), 2022. Disponível em: <https://doi.org/10.48550/arXiv.2108.07258>. Acesso em: 10 jul. 2024.

BOSELLI, Gisele Machado Figueiredo. **A Explicabilidade da IA para o Consentimento do Paciente.** Medicina S/A, 2024. Disponível em: <https://medicinasa.com.br/ia-aspectos-legais/>. Acesso em: 10 jun. 2024.

BOTTINI, Massimo; RYU, Seung-Jun; TERANDER, Adrian Elmi *et al.* **The Ever-Evolving Regulatory Landscape Concerning Development and Clinical Application of Machine Intelligence: Practical Consequences for Spine Artificial Intelligence Research.** Neurospine, 2025. Disponível em: <https://doi.org/10.14245/ns.2449186.593>. Acesso em: 15 ago. 2025.

BRANCACCIO, Gabriella *et al.* **Artificial Intelligence in Skin Cancer Diagnosis: A Reality Check.** Journal of Investigative Dermatology, Volume 144, Issue 3, 492 – 499, 2024. Disponível em: <https://doi.org/10.1016/j.jid.2023.10.004>. Acesso em: 10 ago. 2025.

BRANDÃO, Rodrigo. **Inteligência Artificial na saúde: uma visão da literatura e diretrizes para o Brasil.** *In: COMITÊ GESTOR DA INTERNET NO BRASIL. Inteligência Artificial na saúde: potencialidades, riscos e perspectivas para o Brasil.* São Paulo: Comitê Gestor da Internet no Brasil, 2024. Disponível em: https://cetic.br/media/docs/publicacoes/7/20240903150639/estudos_setoriais-ia-na-saude.pdf. Acesso em: 18 jun. 2025.

BRASIL. **Constituição da República Federativa do Brasil de 1988.** Brasília, DF: Presidência da República, 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 29 jul. 2024.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990.** Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF, 1990. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 29 jul. 2024.

BRASIL. **Lei nº 13.146, de 6 de julho de 2015. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência).** *Diário Oficial da União:* seção 1, Brasília, DF, 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em 7 mar. 2024.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD).** Brasília, DF: Presidência da República, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 29 jul. 2024.

BRASIL. **Lei nº 14.874, de 28 de maio de 2024. Dispõe sobre a pesquisa com seres humanos e institui o Sistema Nacional de Ética em Pesquisa com Seres Humanos.** Brasília, DF: Presidência da República, 2024. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/lei/114874.htm. Acesso em: 29 jul. 2025.

BRASIL. Câmara dos Deputados. **Projeto de Lei nº 1.229, de 2021. Modifica a Lei nº 13.709, de 14 de agosto de 2018 (Lei Geral de Proteção de Dados Pessoais), a fim de conceituar dado neural e regulamentar a sua proteção.** Brasília, DF: Câmara dos Deputados, 2021. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2276604>. Acesso em: 12 ago. 2025.

BRASIL. Câmara dos Deputados. **Projeto de Lei n.º 522/2022. Modifica a Lei n.º 13.709/2018 (LGPD) para conceituar dado neural e regulamentar sua proteção.** Brasília, 9 mar. 2022. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2317524>. Acesso em: 12 ago. 2025.

BRASIL. Senado Federal. **Projeto de Lei nº 2.338, de 2023.** Dispõe sobre o uso da Inteligência Artificial. Brasília, DF: Senado Federal, 2023a. Disponível em: <https://www.congressonacional.leg.br/materias/materias-bicamerais/-/ver/pl-2338-2023>. Acesso em: 12 ago. 2025.

BRASIL. Câmara dos Deputados. **Projeto de Lei nº 2.174, de 2023.** Estabelece as normas e princípios para proteção dos direitos fundamentais relacionados ao cérebro e ao sistema nervoso humano, objetivando garantir a proteção e promoção dos neurodireitos dos indivíduos. Brasília, DF: Câmara dos Deputados, 2023b. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2358605>. Acesso em: 12 ago. 2025.

BRASIL. Senado Federal. **Proposta de Emenda à Constituição nº 29, de 2023.** Altera a Constituição Federal para incluir, entre os direitos e garantias fundamentais, a proteção à integridade mental e à transparência algorítmica. Brasília, DF: Senado Federal, 2023c. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/158095>. Acesso em: 12 ago. 2025.

BRASIL. Senado Federal. **Senado aprova regulamentação da inteligência artificial; texto vai à Câmara.** Agência Senado, Brasília, DF, 10 dez. 2024. Disponível em: <https://www12.senado.leg.br/noticias/materias/2024/12/10/senado-aprova-regulamentacao-da-inteligencia-artificial-texto-vai-a-camara>. Acesso em: 12 ago. 2025.

BRASIL. **Superior Tribunal de Justiça, Recurso Especial nº 1.540.580/DF.** Rel. Min. Lázaro Guimarães (Convocado TRF-5ª Região), Rel. p/ acórdão Min. Luis Felipe Salomão. 4ª Turma. Julgado em 02/08/2018, DJe 04/09/2018. Disponível em:

<https://www.stj.jus.br/websecstj/cgi/revista/REJ.cgi/ATC?seq=87116219&tipo=0&nreg=&SeqCgrmaSessao=&CodOrgaoJgdr=&dt=&formato=PDF&salvar=false>. Acesso em: 29 jul. 2024.

BRASIL. **Superior Tribunal de Justiça, Recurso Especial nº 1.662.338/SP**. Relatora: Ministra Nancy Andrigh, DJe: 12 de dezembro de 2017. Disponível em: https://processo.stj.jus.br/processo/revista/documento/mediado/?componente=ITA&sequencia=1654017&num_registro=201503075580&data=20180202&formato=PDF. Acesso em: 14 mar. 2024.

BRASIL. **Supremo Tribunal Federal, Recurso Extraordinário nº 1.212.272. Tema 1069, Repercussão Geral**. Rel. Min. Gilmar Mendes, DJe: 25 de agosto de 2025. Disponível em: <https://portal.stf.jus.br/processos/detalhe.asp?incidente=5703626>. Acesso em: 29 ago. 2025.

BRIGNULL, Harry *et al.* **Deceptive patterns – user interfaces designed to trick you**. Deceptive design. Retrieved April 25, 2023. Disponível em: <https://www.deceptive.design/>. Acesso em: 7 mar. 2024.

BRITO, Marcos Toscano Siebra. **Governança antecipatória da inteligência artificial no setor de saúde: o que dizem os policy papers internacionais**. Nota Técnica IPEA, n. 135, maio 2024. p. 7. Disponível em: <https://repositorio.ipea.gov.br/entities/publication/a4b39a45-066f-4592-b28b-8fdaf98c1617>. Acesso em: 10 mar. 2025.

BROWN, Cohen Marcus Lionel. **Neurorights, Mental Privacy, and Mind Reading**. Neuroethics, 2024. Disponível em: <https://doi.org/10.1007/s12152-024-09568-z>. Acesso em: 17 mar. 2025.

BROWN, Tom B.; MANN, Benjamin; RYDER, Nick *et al.* **Language Models are Few-Shot Learners**. *Advances in Neural Information Processing Systems* (NeurIPS 2020), v. 33, p. 1877-1901, 2020. Disponível em: <https://arxiv.org/abs/2005.14165>. Acesso em: 21 ago. 2024.

BROWNING, John G.. **Are Technology and the Law on the Same “Wavelength”?: Examining the New Frontier of Brainwaves and Data Privacy**. *North Carolina Journal of Law & Technology*, 2025. Disponível em: <https://scholarship.law.unc.edu/ncjolt/vol26/iss3/3>. Acesso em: 10 jun. 2025.

BUOLAMWINI, Joy; GEBRU, Timnit. **Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification**. *Proceedings of Machine Learning Research*, 2018. Disponível em: <https://proceedings.mlr.press/v81/buolamwini18a.html>. Acesso em: 11 ago. 2024.

BURWELL, Sasha; SAMPLE, Matthew *et al.* **Ethical aspects of brain computer interfaces: a scoping review**. *BMC Med Ethics*, 2017. Disponível em: <https://doi.org/10.1186/s12910-017-0220-y>. Acesso em: 27 ago. 2024.

CABRAL, Anna Cecília Moreira. **Regulação da proteção de dados no âmbito da neurotecnologia e dos neurodireitos**. Repositório FGV: Rio de Janeiro, 2024. Disponível em: <https://repositorio.fgv.br/server/api/core/bitstreams/2cd5dceb-114a-4e8d-98d2-c1a32480b256/content>. Acesso em: 20 nov. 2024.

CALIFÓRNIA. **Senate Bill nº 1223. An act to amend Section 1798.140 of the Civil Code, relating to consumer privacy.** Sacramento, CA: California Legislature, 2024. Disponível em: <https://trackbill.com/bill/california-senate-bill-1223-consumer-privacy-sensitive-personal-information-neural-data/2519511/>. Acesso em: 13 ago. 2025.

CALIXTO, Marcelo Junqueira; SCHULZE, Sandro Coutinho. **Responsabilidade civil por ausência de informação.** *Civilistica.com*, Rio de Janeiro, v. 14, n. 1, p. 1–20, 2025. Disponível em: <https://civilistica.emnuvens.com.br/redc/article/view/1045>. Acesso em: 17 ago. 2025.

CAMÕES; Diana. **Os Direitos Fundamentais na Internet: Uma Nova Era de Constitucionalismo Digital.** *In: Revista Científica sobre Cyberlaw do Centro de Investigação Jurídica do Ciberespaço – CIJIC – Da Faculdade de Direito da Universidade de Lisboa.* CIJIC, Portugal. ed. XIII, junho de 2025. Disponível em: https://www.iuris.edu.pt/xms/files/Cyberlaw_by_CIJIC_ed_XIII_versao_final.pdf. Acesso em: 10 ago. 2025.

CAMPANELLA, Gabriele *et al.* **A clinical benchmark of public self-supervised pathology models.** *PubMed Central*, 2025. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12003829/>. Acesso em: 13 ago. 2025.

CAMPOS, Ricardo. **Novos direitos para uma nova sociedade: a introdução dos neurodireitos na reforma do Código Civil.** *In: A reforma do Código Civil: artigos sobre a atualização da Lei nº 10.406/2002 / org. Rodrigo Pacheco.* Brasília, DF: Senado Federal, Coordenação de Edições Técnicas, 2025. Disponível em: https://www2.senado.leg.br/bdsf/bitstream/handle/id/685736/Reforma_codigo_civil_1ed.pdf. Acesso em: 13 ago. 2025.

CANADÁ. **Pre-market guidance for machine learning-enabled medical devices.** Health Canada, 2025. Disponível em: <https://www.canada.ca/en/health-canada/services/drugs-health-products/medical-devices/application-information/guidance-documents/pre-market-guidance-machine-learning-enabled-medical-devices.html>. Acesso em: 28 jul. 2025.

CARD, Nicholas S.; WAIRAGKAR, Maitreyee *et al.* **An Accurate and Rapidly Calibrating Speech Neuroprosthesis.** *New England Journal of Medicine*, 2024. Disponível em: <https://www.nejm.org/doi/10.1056/NEJMoa2314132>. Acesso em: 20 nov. 2024.

CAZANI, Luana; ALPINO, Pedro; SANTIAGO, Chelsya *et al.* **Dermatologia Digital: Uso de IA no Diagnóstico de Doenças de Pele.** *Brazilian Journal of Implantology and Health Sciences*, 2025. Disponível em: <http://dx.doi.org/10.36557/2674-8169.2025v7n3p1312-1324>. Acesso em: 7 jul. 2025.

CENOLLI, Ilona. **Literature Review: Neurotechnology.** Harvard Medical School Center for Bioethics, 2024. Disponível em: <https://www.nuffieldbioethics.org/wp-content/uploads/Neurotechnology-Literature-Review-WEB-FINAL.pdf>. Acesso em: 10 ago. 2025.

CFM. **Resolução nº 1.986, de 22 de março de 2012.** Reconhecer a Estimulação Magnética Transcraniana (EMT) superficial como ato médico privativo e cientificamente válido para utilização na prática médica nacional, com indicação para depressões uni e bipolar, alucinações auditivas nas esquizofrenias e planejamento de neurocirurgia. *Diário Oficial da União*, Brasília, 2 mai. 2012. Disponível em:

<https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2012/1986>. Acesso em: 16 ago. 2024.

CFM. **Resolução nº 2.057, de 20 de setembro de 2013.** Consolida as diversas resoluções da área da Psiquiatria e reitera os princípios universais de proteção ao ser humano, à defesa do ato médico privativo de psiquiatras e dá outras providências. Diário Oficial da União, Brasília, 12 nov. 2013. Disponível em: https://sistemas.cfm.org.br/normas/arquivos/resolucoes/BR/2013/2057_2013.pdf. Acesso em: 16 ago. 2024.

CFM. **Recomendação nº 1, de 21 de janeiro de 2016.** Dispõe sobre o processo de obtenção de consentimento livre e esclarecido na assistência médica. Diário Oficial da União, Brasília, 1 jan. 2016. Disponível em: https://portal.cfm.org.br/images/Recomendacoes/1_2016.pdf. Acesso em: 16 ago. 2024.

CFM. **Resolução nº 2.217, de 27 de setembro de 2018.** Aprova o Código de Ética Médica. Diário Oficial da União, Brasília, 1 nov. 2018. Disponível em: <https://sistemas.cfm.org.br/normas/visualizar/resolucoes/BR/2018/2217>. Acesso em: 16 ago. 2024.

CFM. **Resolução n.º 2.314, de 10 de agosto de 2022.** Define e regulamenta a telemedicina, como forma de serviços médicos mediados por tecnologias de comunicação. Diário Oficial da União, Brasília, 5 mai. 2022. Disponível em: https://sistemas.cfm.org.br/normas/arquivos/resolucoes/BR/2022/2314_2022.pdf. Acesso em: 16 ago. 2024.

CFM. **CFM avança na elaboração de resolução sobre uso da Inteligência Artificial na Medicina.** Brasília, DF: CFM, 13 ago. 2025. Disponível em: <https://portal.cfm.org.br/noticias/cfm-avanca-na-elaboracao-de-resolucao-sobre-uso-da-inteligencia-artificial-na-medicina/>. Acesso em: 13 ago. 2025.

CFM. **Departamento de Inteligência Artificial do CFM intensifica trabalhos para aprimorar a prática médica no Brasil.** Disponível em: <https://portal.cfm.org.br/noticias/departamento-de-inteligencia-artificial-do-cfm-inten-sifica-trabalhos-para-aprimorar-a-pratica-medica-no-brasil>. Acesso em: 12 abr. 2025.

CHAN, Kenneth; WAHOME, Elizabeth *et al.* **Inflammatory risk and cardiovascular events in patients without obstructive coronary artery disease: the ORFAN multicentre, longitudinal cohort study.** The Lancet, 2024. Disponível em: [https://doi.org/10.1016/S0140-6736\(24\)00596-8](https://doi.org/10.1016/S0140-6736(24)00596-8). Acesso em: 10 jul. 2024.

CHAR, Danton S.; SHAH, Nigan H.; MAGNUS, David. **Implementing Machine Learning in Health Care - Addressing Ethical Challenges.** The New England journal of medicine, 2018. Disponível em: <https://doi.org/10.1056/NEJMp1714229>. Acesso em: 22 abr. 2024.

CHEN, Jeniffer. **Deep Brain Stimulation Helps Parkinson's Patients Regain Independence.** Yale Medicine News, 2019. Disponível em: <https://ym.care/m5z>. Acesso em: 27 jul. 2024.

CHEN, Shugeng; CHEN, Mingyi; WANG, Xu *et al.* **Brain-computer interfaces in 2023–2024.** *Brain-X*, 2025. Disponível em: <http://dx.doi.org/10.1002/brx2.70024>. Acesso em: 10 jul. 2025.

CHEN, Xi *et al.* **Enhancing diagnostic capability with multi-agents conversational large language models.** *npj Digital Medicine*, v. 8, n. 159, 2025. Disponível em: <https://doi.org/10.1038/s41746-025-01550-0>. Acesso em: 12 ago. 2025.

CHENG, Jun; PAN, Joshua *et al.* **Accurate proteome-wide missense variant effect prediction with AlphaMissense.** *Science*, 2023. Disponível em: <https://doi.org/10.1126/science.adg7492>. Acesso em: 15 mar. 2024.

CHILE. **Ley 21.383, de 14 de octubre de 2021.** Modifica la Constitución para proteger la integridad mental y los datos neuronales. *Diario Oficial de la República de Chile*, Santiago, 15 out. 2021. Disponível em: <https://www.bcn.cl/leychile/navegar?idNorma=1166983>. Acesso em: 8 abr. 2025.

CHOPLIN, R. H.; BOEHME, J. M. *et al.* **Picture archiving and communication systems: an overview.** *Radiographics: a review publication of the Radiological Society of North America, Inc*, 12(1), 127–129, 1992. Disponível em: <https://doi.org/10.1148/radiographics.12.1.1734458>. Acesso em: 15 jun. 2024.

CILAR BUDLER, Leona; STIGLIC, Gregor. **Ethical approval and informed consent in mental health research: a scoping review.** *AI & Soc*, 2025. Disponível em: <https://doi.org/10.1007/s00146-025-02364-0>. Acesso em: 10 jul. 2025.

CIRESAN, Dan; GIUSTI, Alessandro; GAMBARDELLA, Luca M.; SCHMIDHUBER, Jürgen. **Deep neural networks segment neuronal membranes in electron microscopy images.** *In: Advances in Neural Information Processing Systems (NIPS)* 25, 2012. p. 2843–2851. Disponível em: <https://dl.acm.org/doi/10.5555/2999325.2999452>. Acesso em: 11 mai. 2024.

CLINICALTRIALS.GOV. **PRIME Study (Precise Robotically Implanted Brain-Computer Interface).** 2024. Disponível em: <https://www.clinicaltrials.gov/study/NCT06429735>. Acesso em: 12 jul. 2025.

CLINICALTRIALS.GOV. **ADvance II Study: DBS-f in Patients With Mild Alzheimer's Disease.** 2024. Disponível em: <https://www.clinicaltrials.gov/study/NCT06429735>. Acesso em: 12 jul. 2025.

CLOTET, Joaquim. **O consentimento informado nos comitês de ética em pesquisa e na prática médica: conceituação, origens e atualidade.** *Bioética*, Brasília, 2009. Disponível em: https://revistabioetica.cfm.org.br/revista_bioetica/article/view/430. Acesso em: 10 mar. 2024.

CNJ - Conselho da Justiça Federal. **V Jornada de Direito Civil (Enunciado nº 403) – Liberdade de crença e recusa de tratamento.** *Jornadas de Direito Civil, CJF*, 2011. Disponível em: <https://www.cjf.jus.br/enunciados/enunciado/207>. Acesso em: 10 ago. 2024.

COALIZÃO DIREITOS NA REDE. **Carta de apoio à tramitação do PL 2338/2023 na Câmara dos Deputados.** São Paulo: Coalizão Direitos na Rede, 20 fev. 2025. Disponível em:

<https://direitosnarede.org.br/2025/02/20/carta-de-apoio-a-tramitacao-do-pl-2338-2023-na-camara-dos-deputados/>. Acesso em: 12 ago. 2025.

COHEN, I. Glenn. **Patients with Passports: Medical Tourism, Law, and Ethics**. New York: Oxford University Press, 2014. Disponível em: <https://doi.org/10.1093/acprof:oso/9780199975099.001.0001>. Acesso em: 10 mar. 2024.

COHEN, I. Glenn. **Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?** *Georgetown Law Journal*, 2019. Disponível em: <https://www.law.georgetown.edu/georgetown-law-journal/in-print/volume-108/volume-108-issue-6-june-2020/informed-consent-and-medical-artificial-intelligence-what-to-tell-the-patient/>. Acesso em: 20 out. 2024.

COHEN, I. Glenn et al. **Regulation of Health and Health Care Artificial Intelligence**. *Journal of the American Medical Association*, 2025. Disponível em: <https://jamanetwork.com/journals/jama/fullarticle/2831831>. Acesso em: 24 mar. 2025.

COHEN, I. Glenn; SLOTTJE, Andrew. **Artificial intelligence and the law of informed consent**. In: *Research Handbook on Health, AI and the Law*. EE Elgar Online, 2024. Disponível em: <https://doi.org/10.4337/9781802205657.ch10>. Acesso em: 22 abr. 2025.

COLE, Eleanor J.; PHILLIPS, Angela. L. *et al.* **Stanford Neuromodulation Therapy (SNT): A Double-Blind Randomized Controlled Trial**. *American Journal of Psychiatry*, 2022. Disponível em: <https://doi.org/10.1176/appi.ajp.2021.20101429>. Acesso em: 10 out. 2024.

CONSELHO NACIONAL DE SAÚDE. **Comissão Nacional de Ética em Pesquisa (CONEP)**. Brasília, DF: Ministério da Saúde, [s.d.]. Disponível em: <https://www.gov.br/conselho-nacional-de-saude/pt-br/camaras-tecnicas-e-comissoes/conep>. Acesso em: 12 ago. 2025.

CORNEJO, Yan. **Neurorights, Neurotechnologies and Personal Data: Review of the Challenges of Mental Autonomy**. *Journal of Digital Technologies and Law*, 2024. Disponível em: <http://dx.doi.org/10.21202/jdtl.2024.36>. Acesso em 20 jul. 2025.

CORNEJO-PLAZA, Isabel; CIPPITANI, Roberto. **Neurodata and its protection from Artificial Intelligence and Neurotechnologies**. *Cad. Ibero-amer. Dir. Sanit.*, Brasília, 2024. Disponível em: <https://doi.org/10.17566/ciads.v13i4.1289>. Acesso em: 10 jan. 2025.

COUTINHO, Patricia B.; JOHNSON, Kara A. *et al.* **Elevated Mood Induced by Subthalamic Nucleus Deep Brain Stimulation: A Video-Recorded Case Report. Tremor and other hyperkinetic movements**. *New York*, 2024. Disponível em: <https://doi.org/10.5334/tohm.900>. Acesso em: 20 nov. 2024.

CRAWFORD, Kate. **Atlas of AI: power, politics, and the planetary costs of artificial intelligence**. New Haven: Yale University Press, eBook Kindle, 2021.

CREMERS - Conselho Regional de Medicina do Estado do Rio Grande do Sul. **Resolução nº 6, de 12 de março de 2025**. Dispõe sobre o uso da Inteligência Artificial (IA) na prática médica. Porto Alegre: CREMERS, 2025. Disponível em: https://cremers.org.br/wp-content/uploads/2025/05/Resolucao_CREMERS_6.2025-Inteligencia-Artificial-.pdf. Acesso em: 12 ago. 2025.

CREMERS - Conselho Regional de Medicina do Estado do Rio Grande do Sul. **CREMERS publica resolução sobre uso da Inteligência Artificial na Medicina**. Porto Alegre: CREMERS, 28 maio 2025. Disponível em: <https://cremers.org.br/cremers-publica-resolucao-sobre-uso-da-inteligencia-artificial-na-medicina/>. Acesso em: 12 ago. 2025.

DADALTO, Luciana. **Testamento vital [recurso eletrônico]**. 7. ed. - Indaiatuba, SP: Editora Foco, 2025.

DADALTO, Luciana; BARUFFI, Priscila Demari; ARANTES, Alexandra Mendes Barreto. **Diretivas antecipadas de vontade em pacientes com doença de Alzheimer**. Revista Bioética, v. 29, n. 2, p. 291-301, 2021. Disponível em: https://revistabioetica.cfm.org.br/index.php/revista_bioetica/article/view/2791. Acesso em: 10 nov. 2024.

DANKWA-MULLAN, Irene. **Health Equity and Ethical Considerations in Using Artificial Intelligence in Public Health and Medicine**. Prev Chronic Dis, 2024. Disponível em: <http://dx.doi.org/10.5888/pcd21.240245>. Acesso em: 8 abr. 2025.

DANTAS, Carlos Henrique Félix; JEREISSATI, Régis Gurgel do Amaral. **Corpo, neurotecnologias e dados neurais: o consentimento como um parâmetro para preservar o direito à identidade**. In: DANTAS, Carlos Henrique Félix; JEREISSATI, Régis Gurgel do Amaral (org.). Desafios da Interface Neurodireito e Inteligência Artificial. Porto Alegre: Livraria do Advogado, 2024. p. 99-114. Disponível em: <https://neurorights.com.br/wp-content/uploads/2024/02/Desafios-da-Interface-Neurodireito-e-Inteligencia-Artificial.pdf>. Acesso em: 12 abr. 2025.

DANTAS, Eduardo; NOGAROLI, Rafaella. **Consentimento informado do paciente frente às novas tecnologias da saúde (telemedicina, cirurgia robótica e inteligência artificial)**. Lex Medicinæ: Revista Portuguesa de Direito da Saúde, v. 33, n. 1, p. 25-63, 2020. Disponível em: https://www.researchgate.net/publication/356172107_Consentimento_informado_do_paciente_frente_as_novas_tecnologias_da_saude_telemedicina_cirurgia_robotica_e_inteligencia_artif. Acesso em: 10 mar. 2024.

DAVIES, Ben. **The right not to know and the obligation to know**. Journal of medical ethics, 2020. Disponível em: <https://doi.org/10.1136/medethics-2019-106009>. Acesso em: 19 jul. 2024.

DAVTYAN, Susanna; MESROPYAN, Milena. **'The Ethical Issues of Neurotechnologies'**. Global Bioethics - Current Challenges, New Developments, and Future Directions. IntechOpen, 2025. Disponível em: <https://doi.org/10.5772/intechopen.1010599>. Acesso em: 11 ago. 2025.

DI SALVO, Michele. **The protection of neural rights in the age of neurotechnologies and AI. the ethical challenge for law and neuroscience**. Russian Journal of Economics and Law, 2025. Disponível em: <http://dx.doi.org/10.21202/2782-2923.2025.1.202-233>. Acesso em: 10 jul. 2025.

DIAS, Fernanda Moura Vargas; ANHOQUE, Carolina Fiorin *et al.* **Neuromodulação não-invasiva: uma revisão narrativa sobre aplicabilidade clínica, acesso ao tratamento e direções futuras**. Revista Brasileira de Pesquisa em Saúde, 2024. Disponível em: <https://doi.org/10.47456/rbps.v26i1.46477>. Acesso em: 27 jul. 2025.

DIRECTOR, Samuel. **“Does Black Box AI In Medicine Compromise Informed Consent?”**. *Philosophy & Technology*. Springer Nature, 2025. Disponível em: <https://doi.org/10.1007/s13347-025-00860-1>. Acesso em: 30 jul. 2025.

DONEDA, Danilo Cesar Maganhoto; SOUZA, Carlos Affonso Pereira de *et al.* **Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal**. *Pensar*, v. 23, n. 4, pp. 1-17, out.-dez./2018. Disponível em: <https://doi.org/10.5020/2317-2150.2018.8257>. Acesso em: 12 dez. 2023.

DONEDA, Danilo. **Da privacidade à proteção de dados pessoais**. 3. ed. São Paulo: Thomson Reuters Brasil, eBook Kindle, 2021.

DONEDA, Danilo; MENDES, Laura Schertel (coord.). **Tratado de proteção de dados pessoais**. 2 ed. Rio de Janeiro: Forense, eBook Kindle, 2023.

DOSOVITSKIY, Alexey *et al.* **An image is worth 16x16 words: Transformers for image recognition at scale**. *In: International Conference on Learning Representations (ICLR)*, 2021. Disponível em: <https://doi.org/10.48550/arXiv.2010.11929>. Acesso em: 15 out. 2024.

DREW, Lian. **Neuralink brain chip: advance sparks safety and secrecy concerns**. *Nature*, 2024. Disponível em: <https://doi.org/10.1038/d41586-024-00550-6>. Acesso em: 15 out. 2024.

DURAIVEL, Suseendrakumar *et al.*, **High-resolution neural recordings improve the accuracy of speech decoding**. *Nat Commun* 14, 6938, 2023. Disponível em: <https://doi.org/10.1038/s41467-023-42555-1>. Acesso em: 5 set. 2024.

EBSWORTH, Jonathan *et al.* **‘Atlas of AI’ by Kate Crawford: a review**. *Evangelical Focus – TechHuman*, 2023. Disponível em: <https://evangelicalfocus.com/techhuman/23173/atlas-of-ai-by-kate-crawford-a-review>. Acesso em: 29 jun. 2024.

EISEMANN, Nora; BUNK, Stefan *et al.* **Nationwide real-world implementation of AI for cancer detection in population-based mammography screening**. *Nature Medicine*, 2025. Disponível em: <https://doi.org/10.1038/s41591-024-03408-6>. Acesso em: 27 jul. 2025.

ENGELMANN, Alana. **A Transparência Algorítmica como Direito Fundamental e a Necessidade de Explicabilidade do Processo de Tomada de Decisões no Estado Democrático de Direito**. *Juridicamente*, 2023. Disponível em: <https://juridicamente.info/a-transparencia-algoritmica-como-direito-fundamental-e-a-necessidade-de-explicabilidade-do-processo-de-tomada-de-decisoes-no-estado-democratico-de-direito>. Acesso em: 20 abr. 2025.

ESHKAGHI, Adrien A. *et al.* **The cochlear implant: historical aspects and future prospects**. *Anat Rec (Hoboken)*, 295(11):1967-80, 2012. Disponível em: <https://doi.org/10.1002/ar.22580>. Acesso em: 3 set. 2024.

ESPAÑA. **Declaración de León sobre la Neurotecnología Europea: un enfoque centrado en los derechos humanos**. León, 24 oct. 2023. Disponível em: <https://spanish-presidency.consilium.europa.eu/media/o4rh53jr/le%C3%B3n-declaration.pdf>. Acesso em: 3 set. 2024.

ESPAÑA. Secretaría de Estado de Digitalización e Inteligencia Artificial. **Carta de Derechos Digitales**. Madrid: Gobierno de España, 2021. Disponível em: https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/participacion_publica/audienca/ficheros/SEDIACartaDerechosDigitales.pdf. Acesso em: 12 ago. 2025.

ESPAÑA. Presidencia del Gobierno. **Sánchez presenta la Carta de Derechos Digitales, con la que "España se sitúa a la vanguardia internacional en la protección de derechos de la ciudadanía"**. Madrid: La Moncloa, 14 jul. 2021. Disponível em: <https://www.lamoncloa.gob.es/presidente/actividades/paginas/2021/140721-derechos-digitales.aspx>. Acesso em: 12 ago. 2025.

EUA. **Brain Research through Advancing Innovative Neurotechnologies (BRAIN 2025)**. Working Group Report to the Advisory Committee to the Director NIH, 2014. Disponível em: https://braininitiative.nih.gov/sites/default/files/documents/brain2025_508c_2.pdf. Acesso em: 10 jun. 2025.

EUA. **S.2892 - Algorithmic Accountability Act of 2023**. Washington, DC: U.S. Congress, 2023a. Disponível em: <https://www.congress.gov/bill/118th-congress/senate-bill/2892>. Acesso em: 13 ago. 2025.

EUA. **H.R.5628 - Algorithmic Accountability Act of 2023**. Washington, DC: U.S. Congress, 2023b. Disponível em: <https://www.congress.gov/bill/118th-congress/house-bill/5628>. Acesso em: 13 ago. 2025.

FARAHANY, Nita. **The Battle for Your Brain: Defending the Right to Think Freely in the Age of Neurotechnology**. New York: St. Martin's Press, eBook Kindle, 2023.

FAZANO, Humberto; MARTINS, Amanda Smith. **Atualização do Código Civil como oportunidade para codificação dos neurodireitos**. Legal Grounds Institute, 2024. Disponível em: <https://legalgroundsinstitute.com/blog/atualizacao-do-codigo-civil-como-oportunidade-para-codificacao-dos-neurodireitos>. Acesso em: 4 jul. 2025.

FDA. **Implanted Brain-Computer Interface Devices for Patients with Paralysis or Amputation — Non-Clinical Testing and Clinical Considerations: Guidance for Industry and FDA Staff**, 2021. Disponível em: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/implanted-brain-computer-interface-bci-devices-patients-paralysis-or-amputation-non-clinical-testing>. Acesso em: 13 mar. 2025.

FDA. **Predetermined Change Control Plans for Medical Devices - Draft Guidance for Industry and FDA Staff**, 2024. Disponível em: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/predetermined-change-control-plans-medical-devices>. Acesso em: 13 mar. 2025.

FDA. **Marketing Submission Recommendations for a Predetermined Change Control Plan for Artificial Intelligence-Enabled Device Software Functions**, 2025a. Disponível em: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligence>. Acesso em: 13 mar. 2025.

FDA. **Artificial Intelligence/Machine Learning-Enabled – Software as a Medical Device (SaMD)**. Silver Spring, 25 mar. 2025b (atualizada). Disponível

em: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device>. Acesso em: 25 ago. 2025.

FDA. Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry and Other Interested Parties. January, 2025c. Disponível em: <https://www.fda.gov/media/184830/download>. Acesso em: 25 ago. 2025.

FDA. Artificial Intelligence-Enabled Device Software Functions: Lifecycle 2 Management and Marketing 3 Submission Recommendations. January, 2025d. Disponível em: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/artificial-intelligence-enabled-device-software-functions-lifecycle-management-and-marketing>. Acesso em: 25 ago. 2025.

FEHR, Jana; CITRO, Brian *et al.* **A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare.** *Frontiers in Digital Health*, v. 6, 1267290, 2024. Disponível em: <https://doi.org/10.3389/fdgth.2024.1267290>. Acesso em: 9 fev. 2025.

FERNANDEZ, Eduardo; ROBLES, Jose Antonio. **Advances and challenges in the development of visual prostheses.** *PLOS Biology*, v. 22, n. 10, e3002896, 2024. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002896>. Acesso em: 22 fev. 2025.

FERREIRA, Dominic; SEGRAVE, Rebecca A. *et al.* **"He's Back so I'm Not Alone": The Impact of Deep Brain Stimulation on Personality, Self, and Relationships in Parkinson's Disease.** *Qualitative health research*, 2020. Disponível em: <https://doi.org/10.1177/1049732320951144>. Acesso em: 10 fev. 2024.

FERREIRA, Versalhes Enos Nunes; FERREIRA, Vanessa Rocha; LEAL, Pastora do Socorro Teixeira. **NEUROTECNOLOGIAS E NEURODIREITOS: a tutela jurídica da mente humana.** *Revista de Direitos Humanos e Efetividade*, 2025. Disponível em: <http://dx.doi.org/10.26668/IndexLawJournals/2526-0022/2024.v10i2.10836>. Acesso em 20 jul. 2025.

FILIPPIS, Rocco de; FOYSAL, Abdullah Al *et al.* **The Risk Perspective of AI in Healthcare: GDPR and GELSI Framework (Governance, Ethical, Legal and Social Implications) and the New European AI Act.** *Italian Journal of Psychiatry*, 10, 12-16, 2024. Disponível em: <https://doi.org/10.36180/2421-4469-2024-4>. Acesso em: 10 mai. 2025.

FINE, Ione; BOYNTON, Geoffrey M. **A virtual patient simulation modeling the neural and perceptual effects of human visual cortical stimulation, from pulse trains to percepts.** *Scientific Reports*, v. 14, 17400, 2024. Disponível em: <https://www.nature.com/articles/s41598-024-65337-1>. Acesso em: 22 fev. 2025

FLEUREN, Lucas M.; KLAUSCH, Thomas L. T.; ZWAGER, Charlotte L. *et al.* **Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy.** *Intensive care medicine*, 2020. Disponível em: <https://doi.org/10.1007/s00134-019-05872-y>. Acesso em: 20 ago. 2024.

FLORIDI, Luciano; COWLS, Josh; BELTRAMETTI, Monica *et al.* **AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and**

Recommendations. Minds & Machines, 2018. Disponível em: <https://doi.org/10.1007/s11023-018-9482-5>. Acesso em: 19 jul. 2024.

FORD, Roger; PRICE II, W. Nicholson. **Privacy and Accountability in Black-Box Medicine.** 23 Mich. Telecomm. & Tech. L. Rev. 1, 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2758121. Acesso em: 24 abr. 2025.

FUTURO DA SAÚDE. **CFM vai criar resolução de Inteligência Artificial na medicina.** São Paulo: Futuro da Saúde, 23 abr. 2025. Disponível em: <https://futurodasaude.com.br/cfm-resolucao-inteligencia-artificial/>. Acesso em: 13 ago. 2025.

GALA, Dhir; BEHL, Haditya; SHAH, Mili; MAKARYUS, Amgad N. **The Role of Artificial Intelligence in Improving Patient Outcomes and Future of Healthcare Delivery in Cardiology: A Narrative Review of the Literature.** Healthcare, Basel, v. 12, n. 4, p. 481, fev. 2024. Disponível em: <https://doi.org/10.3390/healthcare12040481>. Acesso em: 16 ago. 2024.

GAO, Weyhang; YAN, Zineng; ZHOU, Hong *et al.* **Revolutionizing brain–computer interfaces: overcoming biocompatibility challenges in implantable neural interfaces.** Journal of nanobiotechnology, 2025. Disponível em: <https://doi.org/10.1186/s12951-025-03573-x>. Acesso em: 24 jul. 2025.

GARCIA, Christine; UZBELGER, Georges. **Artificial Intelligence to Help the Practitioner Choose the Right Treatment: Watson for Oncology.** In: *Healthcare and Artificial Intelligence*, Springer Cham, 2020. Disponível em: https://doi.org/10.1007/978-3-030-32161-1_11. Acesso em: 12 fev. 2024.

GENOVESE, Ariana *et al.* **From Promise to Practice: Harnessing AI's Power to Transform Medicine.** *Journal of Clinical Medicine*, v. 14, n. 1225, fev. 2025. p. 2. Disponível em: <https://doi.org/10.3390/jcm14041225>. Acesso em: 10 jun. 2025.

GERARDI, Chiara; XINARIS, Christodoulos. **Beyond human limits: the ethical, social, and regulatory implications of human enhancement.** *Frontiers in medicine*, 2025. Disponível em: <https://doi.org/10.3389/fmed.2025.1595213>. Acesso em: 5 ago. 2025.

GIANFRANCESCO, Milena A.; TAMANG, Suzanne *et al.* **Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data.** *JAMA Intern Med.* 2018. Disponível em: <https://doi.org/10.1001/jamainternmed.2018.3763>. Acesso em: 15 ago. 2024.

GIL, Antônio Carlos, 1946 - **Como elaborar projetos de pesquisa.** 7. ed. – São Paulo: Atlas, eBook Kindle, 2022.

GLANNON, Walter. **Ethical issues with brain–computer interfaces.** *Frontiers in Systems Neuroscience*, v. 8, art. 36, 2014. Disponível em: <https://www.frontiersin.org/articles/10.3389/fnsys.2014.00136/full>. Acesso em: 17 ago. 2025.

GLEASON, Alec *et al.* **Detection of neurologic changes in critically ill infants using deep learning on video data: a retrospective single center cohort study.** *eClinicalMedicine*, Volume 78, 102919, 2024. Disponível em: [https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370\(24\)00498-X](https://www.thelancet.com/journals/eclinm/article/PIIS2589-5370(24)00498-X). Acesso em: 7 jul. 2024.

GOODFELLOW, Ian et al. **Generative Adversarial Nets**. In: Advances in Neural Information Processing Systems (NIPS) 27, 2014. Disponível em: <https://arxiv.org/pdf/1406.2661>. Acesso em: 7 jul. 2024.

GOOGLE. **Google's Impact on Health – Helping everyone, everywhere live a longer, healthier life**. Relatório corporativo, fevereiro 2025. Disponível em: https://services.google.com/fh/files/misc/google_health_impact.pdf. Acesso em: 20 abr. 2025.

GORDON, Emma C.; SETH, Anil K. **Ethical considerations for the use of brain–computer interfaces for cognitive enhancement**. PLoS Biol, 2024. Disponível em: <https://doi.org/10.1371/journal.pbio.3002899>. Acesso em: 14 fev. 2025.

GRAAFSMA, Jetske; MURPHY, Rachel M; GARDE, Ewoudt M W van de. *et al.* **The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: a scoping review**. *Journal of the American Medical Informatics Association*, 2024. Disponível em: <https://doi.org/10.1093/jamia/ocae076>. Acesso em: 20 ago. 2024.

GREENBAUM, Dove; CABRERA, Laura Y. **Editorial: ELSI in Human Enhancement: What Distinguishes It From Therapy?**. *Frontiers in genetics*, 2020. Disponível em: <https://doi.org/10.3389/fgene.2020.00618>. Acesso em: 15 ago. 2024.

GUANDALINI, Camila Carolina Valero et al. **Inteligência Artificial na Detecção de Câncer de Pele: Benefícios e Desafios para a Prática Dermatológica**. *Revista Ibero-Americana de Humanidades, Ciências e Educação*, [S. l.], v. 10, n. 12, p. 14–25, 2024. Disponível em: <https://periodicorease.pro.br/rease/article/view/17209>. Acesso em: 26 jul. 2025.

GUERRA, Carla; SCHAEFER, Fernanda. **Consentimento informado e recusa terapêutica: limites da autonomia privada na esfera da saúde**. *Revista Bioética*, impressa, v. 17, n. 1, p. 43-56, 2009.

GUERRA, Rodrigo Barbosa; SATHLER, Maria Vitória Moreita *et al.* **Estimulação Cerebral Profunda no Tratamento da Doença de Alzheimer, Parkinson e Outras Demências**. *Brazilian Journal of Implantology and Health Sciences*, 2025. Disponível em: <https://doi.org/10.36557/2674-8169.2025v7n2p2106-2119>. Acesso em: 20 jul. 2025.

GULSHAN, Varun *et al.* **Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs**. *JAMA*, v. 316, n. 22, p. 2402–2410, 2016. Disponível em: <https://jamanetwork.com/journals/jama/fullarticle/2588763>. Acesso em: 7 jul. 2024.

GUNDLACK, Jana; THIEL, Carolin; NEGASH, Sara *et al.* **Patients' Perceptions of Artificial Intelligence Acceptance, Challenges, and Use in Medical Care: Qualitative Study**. *J Med Internet Res*, 2025. Disponível em: <https://doi.org/10.2196/70487>. Acesso em: 25 jul. 2025.

GUNNING, David et al. **XAI – Explainable artificial intelligence**. *Science Robotics*, v. 4, n. 37, 2019. Disponível em: <https://doi.org/10.1126/scirobotics.aay7120>. Acesso em: 11 nov. 2024.

GYMREK, Melissa *et al.* **Identifying Personal Genomes by Surname Inference.** *Science*, v. 339, n. 6117, p. 321–324, 2013. Disponível em: <https://www.science.org/doi/10.1126/science.1229566>. Acesso em: 15 ago. 2025.

HAAG, Lea; IENCA, Marcelo *et al.* **Ethical gaps in closed-loop neurotechnology: a scoping review.** *NPJ Digital Medicine*, 2025. Disponível em: <https://doi.org/10.1038/s41746-025-01908-4>. Acesso em: 10 ago. 2025.

HAENLEIN, Michael; KAPLAN, Andreas. **A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence.** *California Management Review*, v.61, n.4, 2019. Disponível em <https://doi.org/10.1177/0008125619864925>. Acesso em: 26 nov. 2024.

HAENSSLE, Holger A. *et al.* **Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists.** *Annals of Oncology*, v. 29, n. 8, p. 1836-1842, ago. 2018. Disponível em: <https://doi.org/10.1093/annonc/mdy166>. Acesso em: 24 mai. 2024.

HALE, Conor. **FDA clears Precision Neuroscience's minimally invasive brain-computer interface implant.** *Fierce Biotech*, 17 abr. 2025. Disponível em: <https://www.fiercebiotech.com/medtech/fda-clears-precision-neurosciences-minimally-invasive-brain-computer-interface-implant>. Acesso em: 13 ago. 2025.

HANNUN, Awni Y. *et al.* **Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network.** *Nature Medicine*, v. 25, p. 65–69, 2019. Disponível em: <https://doi.org/10.1038/s41591-018-0268-3>. Acesso em: 13 ago. 2024.

HASANZADEH, Farzad. **Bias recognition and mitigation strategies in healthcare AI: reviewing origins and stakeholder responsibilities.** *NPJ Digital Medicine*, v. 8, art. 503, 2025. Disponível em: <https://www.nature.com/articles/s41746-025-01503-7>. Acesso em: 12 ago. 2025.

HENRIQUES; Cristina Reis. **Responsabilidade Civil na Inteligência Artificial - Uma Análise da Legislação Existente.** *In: Revista Científica sobre Cyberlaw do Centro de Investigação Jurídica do Ciberespaço – CIJIC – Da Faculdade de Direito da Universidade de Lisboa.* CIJIC, Portugal. ed. XIII, junho de 2025. Disponível em: https://www.iuris.edu.pt/xms/files/Cyberlaw_by_CIJIC_ed_XIII_versao_final.pdf. Acesso em: 10 ago. 2025.

HERNANDEZ-BOUSSARD, Tina *et al.* **Promoting transparency in AI for biomedical and behavioral research.** *Journal of Biomedical Informatics*, v. 142, 2025. Disponível em: <https://doi.org/10.1038/s41591-025-03680-0>. Acesso em: 12 ago. 2025.

HERTZ, Nora. **Neurorights – Do we Need New Human Rights? A Reconsideration of the Right to Freedom of Thought.** *Neuroethics*, 2023. <https://doi.org/10.1007/s12152-022-09511-0>. Acesso em: 11 ago. 2025.

HERZOG, Christian. **On the risk of confusing interpretability with explicability.** *AI Ethics* 2, 219–225, 2022. Disponível em <https://doi.org/10.1007/s43681-021-00121-9>. Acesso em: 26 nov. 2024.

HL7 INTERNATIONAL. **FHIR – Overview**. HL7 International, 2023. Disponível em: <https://hl7.org/fhir/overview.html>. Acesso em: 10 ago. 2024.

HOCHBERG, Leigh R. *et al.* **Neuronal ensemble control of prosthetic devices by a human with tetraplegia**. *Nature* 442, 164–171, 2006. Disponível em: <https://doi.org/10.1038/nature04970>. Acesso em: 27 jul. 2024.

HOCHBERG, Leigh R. *et al.* **Reach and grasp by people with tetraplegia using a neurally controlled robotic arm**. *Nature* 485, 372–375 (2012). Disponível em: <https://doi.org/10.1038/nature11076>. Acesso em: 27 jul. 2024.

HOLDSWORTH, James. **What is AI bias?** IBM, 2023. Disponível em: <https://www.ibm.com/think/topics/ai-bias>. Acesso em: 10 ago. 2024.

HONFUNG POON, Hoifung; USUYAMA, Naoto. **GigaPath: Whole-Slide Foundation Model for Digital Pathology**. Microsoft Research Blog, 22 maio 2024. Disponível em: <https://www.microsoft.com/en-us/research/blog/gigapath-whole-slide-foundation-model-for-digital-pathology/>. Acesso em: 15 ago. 2024.

HOOPER, Vincent James. **The Mind in the Machine: Legal Implications of Brain-Computer Interfaces**. *The Times of Israel*, 2025. Disponível em: <https://blogs.timesofisrael.com/the-mind-in-the-machine-legal-implications-of-brain-computer-interfaces/>. Acesso em: 5 ago. 2025.

HURLEY, Meghan E.; LANG, Benjamin H.; KOSTICK-QUENET, Kristin M. *et al.* **Patient consent and the right to notice and explanation of AI systems used in health care**. *The American Journal of Bioethics*, v. 25, n. 3, p. 102-114, 2025. Disponível em: <https://doi.org/10.1080/15265161.2024.2399828>. Acesso em: 4 jun. 2025.

IENCA, Marcello. **On Neurorights**. *Front. Hum. Neurosci.*, 2021. Disponível em: <https://doi.org/10.3389/fnhum.2021.701258>. Acesso em: 20 jul. 2024.

IENCA, Marcello; ADORNO, Roberto. **Towards new human rights in the age of neuroscience and neurotechnology**. *Life Sciences, Society and Policy*, v. 13, n. 5, p. 1-27, 2017. <https://www.doi.org/10.1186/s40504-017-0050-1>. Acesso em: 13 ago. 2024.

IENCA, Marcello; IGNATIADIS, Karolina. **Artificial Intelligence in Clinical Neuroscience: Methodological and Ethical Challenges**. *AJOB neuroscience*, 2020. Disponível em: <https://doi.org/10.1080/21507740.2020.1740352>. Acesso em: 15 jul. 2024.

INFECTION CONTROL TODAY. **Balancing Regulation and Risk of AI and Machine Learning Software in Medical Devices**. Cranbury, NJ: MJH Life Sciences, 13 ago. 2025. Disponível em: <https://www.infectioncontrolday.com/view/balancing-regulation-risk-ai-machine-learning-software-medical-devices>. Acesso em: 13 ago. 2025.

INSTITUTO PENSI. **Avanço tecnológico da Inteligência Artificial x Comprometimento Ético na Saúde**. São Paulo: Instituto Pensi, 23 jun. 2023. Disponível em: <https://institutopensi.org.br/avanco-tecnologico-da-inteligencia-artificial-x-comprometimento-etico-na-saude/>. Acesso em: 12 ago. 2025.

INTERNATIONAL MEDICAL DEVICE REGULATORS FORUM (IMDRF). **Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations.** IMDRF Software as a Medical Device (SaMD) Working Group, 18 set. 2014. Disponível em: <https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf>. Acesso em: 15 ago. 2024.

ISTACE, Timo; TRASCASAS, Milena Costas. **Innovation and Ethics in Neurotechnology.** The Geneva Academy of International Humanitarian Law and Human Rights, 2024. Disponível em: <https://www.geneva-academy.ch/joomlatools-files/docman-files/Between%20Science-Fact%20and%20Science-Fiction%20Innovation%20and%20Ethics%20in%20Neurotechnology.pdf>. Acesso em: 10 ago. 2025.

JAMES, Ted A. **Confronting the Mirror: Reflecting on Our Biases Through AI in Health Care.** Harvard Medical Scholl, 2024. Disponível em: <https://learn.hms.harvard.edu/insights/all-insights/confronting-mirror-reflecting-our-biases-through-ai-health-care>. Acesso em: 10 mar. 2025.

JELEMBI, Armindo Gideão K. **Proteção de Dados Pessoais em Saúde.** Centro de Direito Biomédico | Faculdade de Direito da Universidade de Coimbra. Portugal, 2025. Disponível em: <https://www.centrodedireitobiomedico.org/publicacao/proteccao-de-dados-pessoais-em-saude/106/>. Acesso em: 4 ago. 2025.

JOSHI, Geeta.; JAIN, Aditi *et al.* **FDA - Approved Artificial Intelligence and Machine Learning (AI/ML) - Enabled Medical Devices: An Updated Landscape.** *Electronics*. 13, 498, 2024. Disponível em: <https://doi.org/10.3390/electronics13030498>. Acesso em: 4 jul. 2025.

KAHNEMAN, Daniel. **Rápido e devagar: duas formas de pensar.** Tradução Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012.

KASZNAR LEONARDOS. **ANVISA aprova nova RDC relacionada aos dispositivos médicos.** Rio de Janeiro: Kasznar Leonardos, 2022. Disponível em: <https://www.kasznarleonardos.com/anvisa-aprova-nova-rdc-relacionada-aos-dispositivos-medicos/>. Acesso em: 12 ago. 2025.

KAYE, Jane; Whitley; Edgar A. *et al.* **Dynamic consent: a patient interface for twenty-first century research networks.** *European Journal of Human Genetics*, 2015. Disponível em: <https://doi.org/10.1038/ejhg.2014.71>. Acesso em: 10 set. 2024.

KETRYX. **A Complete Guide to the FDA's AI/ML Guidance for Medical Devices.** Boston, MA: Ketryx, 2025. Disponível em: <https://www.ketryx.com/blog/a-complete-guide-to-the-fdas-ai-ml-guidance-for-medical-devices>. Acesso em: 13 ago. 2025.

KFOURI NETO, Miguel. **Responsabilidade civil do médico.** 5. ed., São Paulo: Thomson Reuters Brasil, eBook Kindle, 2024.

KFOURI NETO, Miguel; SILVA, Rodrigo da Guia; NOGAROLI, Rafaella. **Inteligência artificial e big data no diagnóstico e tratamento da COVID-19 na América Latina: novos desafios à proteção de dados pessoais.** Direitos Fundamentais & Justiça, Belo Horizonte, ano

14, Número especial, p. 149-178, nov. 2020. Disponível em: <https://doi.org/10.30899/dfj.v0i0.974>. Acesso em: 13 abr. 2024.

KILIAN, Robert; JÄCK, Linda; EBEL, Dominik. *European AI Standards - Technical Standardization and Implementation Challenges under the EU AI Act*. March, 2025. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5155591. Acesso em: 15 jun. 2025.

KIM, Yubin *et al.*, 2024. **MDAgents: an adaptive collaboration of LLMs for medical decision-making**. In: *Proceedings of the 38th International Conference on Neural Information Processing Systems (NIPS '24)*, Vol. 37. Curran Associates Inc., Red Hook, NY, USA, Article 2522, 79410–79452. Disponível em: <https://doi.org/10.48550/arXiv.2404.15155>. Acesso em: 10 mar. 2025.

KLEIN, Eran. **Informed Consent in Implantable BCI Research: Identifying Risks and Exploring Meaning**. *Sci Eng Ethics*, 2016. Disponível em: <https://doi.org/10.1007/s11948-015-9712-7>. Acesso em: 11 jan. 2025.

KOLFSCHOOTEN, Hannah van. **EU regulation of artificial intelligence: Challenges for patients' rights, Common Market Law Review, 2022**. Amsterdam Law School Legal Studies Research Paper N° 2022-60, Law Centre for Health & Life Research Paper N° 2022-01. Disponível em: <http://dx.doi.org/10.2139/ssrn.3997366>. Acesso em: 14 nov. 2024.

KOLFSCHOOTEN, Hannah van; OIRSCHOT, Janneke van. **The EU Artificial Intelligence Act (2024): Implications for healthcare**. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0168851024001623>. Acesso em: 12 abr. 2025.

KULIKOWSKI, Casimir A. **Beginnings of Artificial Intelligence in Medicine (AIM): Computational Artifice Assisting Scientific Inquiry and Clinical Art – with Reflections on Present AIM Challenges**. *Yearbook of Medical Informatics*, 2019. Disponível em: <https://doi.org/10.1055/s-0039-1677895>. Acesso em: 14 nov. 2024.

KULIKOWSKI, Casimir A.; WEISS, Sholom M. **Representation of an expert knowledge for consultation: The CASNET and EXPERT projects**. In: *SZOLOVITS, Peter (Ed.). Artificial intelligence in medicine*. Boulder, CO: Westview Press, 1982. p. 21-55. Disponível em: <https://doi.org/10.4324/9780429052071>. Acesso em: 14 nov. 2024.

LA FISCA, Luca; VANDENBULCKE, Virginie *et al.* **Biases in BCI experiments: Do we really need to balance stimulus properties across categories?** *Front. Comput. Neurosci.*, 2022. Disponível em: <https://doi.org/10.3389/fncom.2022.900571>. Acesso em: 10 mai. 2024.

LA SCALA, Shayna; MULLINS, Jordan L. *et al.* **Equity, diversity, and inclusion in developmental neuroscience: Practical lessons from community-based participatory research**. *Frontiers in integrative neuroscience*, 2023. Disponível em: <https://doi.org/10.3389/fnint.2022.1007249>. Acesso em: 10 mai. 2024.

LAUTERBUR, Paul. **Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance**. *Nature* 242, 190–191 (1973). Disponível em: <https://doi.org/10.1038/242190a0>. Acesso em 10 ago. 2024.

LAVAZZA, Andrea; BALCONI, Michela; IENCA, Marcello *et al.* **Neuralink's brain-computer interfaces: medical innovations and ethical challenges.** *Front. Hum. Dyn.*, 23 March 2025. Disponível em: <https://doi.org/10.3389/fhumd.2025.1553905>. Acesso em: 11 jul. 2025.

LEDERBERG, Joshua. **How DENDRAL was conceived and born.** *In: ACM Symposium on the History of Medical Informatics*, 5 Nov 1987, Rockefeller University, New York. *National Library of Medicine*. Disponível em: <https://dl.acm.org/doi/abs/10.1145/41526.41528>. Acesso em: 15 mai. 2025

LEITE, Cicília Raquel Maia; REIS, Célia Aparecida dos *et al.* (Organizadores). **Novas tecnologias aplicadas à saúde [recurso eletrônico]: inovação, internet das coisas, horizontes e desafios.** Mossoró, RN: EDUERN, 2021. Disponível em: <https://portal.uern.br/wp-content/uploads/sites/14/2024/09/E-book-Novas-Tecnologias-Aplicadas-a-Saude.pdf>. Acesso em: 11 out. 2024.

LEONEL, Jhadson Silva; LEONEL, Camila Ferreira Silva *et al.* **Inteligência artificial: desafios éticos e futuros.** *Revista Bioética*, 2025. Disponível em: <https://doi.org/10.1590/1983-803420243739PT>. Aceso em: 7 jun. 2025.

LESSIG, Lawrence. **Code: version 2.0 (English Edition).** Ed. 2.0.1, eBook Kindle, 2011.

LEUTHARDT, Eric C. *et al.* **Defining surgical terminology and risk for brain-computer interfaces.** *Journal of Neural Engineering*, 2021. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8044752>. Acesso em: 17 ago. 2025.

LEVINE, Joseph. **Materialism and Qualia: The Explanatory Gap.** *Pacific Philosophical Quarterly*, v. 64, n. 4, p. 354-361, 1983. Disponível em: <https://www.newdualism.org/papers/J.Levine/Levine-PPQ1983.pdf>. Acesso em: 10 mar. 2024.

LI, Shizhe; GAO, Shutao *et al.* **Brain-Computer Interfaces in Spinal Cord Injury: A Promising Therapeutic Strategy.** *The European journal of neuroscience*, 2025. Disponível em: <https://doi.org/10.1111/ejn.70183>. Acesso em: 20 jul. 2025.

LI, Yuru; CHANG; Xiaowei; Wu, Jianlin *et al.* **Machine learning in early diagnosis of neurological diseases: Advancing accuracy and overcoming challenges.** *Brain Network Disorders*, 2025. Disponível em: <https://doi.org/10.1016/j.bnd.2025.04.001>. Acesso em: 17 ago. 2025.

LIGTHART, Nicholas; BUBLITZ, Christoph; ALEGRE, Miguel. **Neurotechnology: we need new laws, not new rights.** *Nature*, 2023. Disponível em: <http://dx.doi.org/10.1038/d41586-023-02698-z>. Acesso em: 13 ago. 2024.

LITJENS, Geert *et al.* **A survey on deep learning in medical image analysis.** *Medical Image Analysis*, v. 42, p. 60-88, 2017. Disponível em: <https://doi.org/10.1016/j.media.2017.07.005>. Acesso em: 15 fev. 2024.

LIU, Xiu-Yun, WANG, Wen-Long, LIU, Miao *et al.* **Recent applications of EEG-based brain-computer-interface in the medical field.** *Military Med Res* 12, 14, 2025. Disponível em: <https://doi.org/10.1186/s40779-025-00598-z>. Acesso em: 7 jul. 2025.

LIVANIS, Efstratios; VOULTSOS, Polychronis; VADIKOLIAS, Konstantinos *et al.* **Understanding the Ethical Issues of Brain-Computer Interfaces (BCIs): A Blessing or the Beginning of a Dystopian Future?.** *Cureus*, 2024. Disponível em: <https://doi.org/10.7759/cureus.58243>. Acesso em: 7 jul. 2025.

LOPES, Ana Maria D'Ávila *et al.* (organizadores); Aline Passos Maia *et al.* **Neurodireito, neurotecnologia e direitos humanos.** Porto Alegre: Livraria do Advogado, 2022. Disponível em: <https://neurorights.com.br/wp-content/uploads/2023/09/Neurodireito-Neurotecnologia-e-Direitos-Humanos.pdf>. Acesso em: 20 out. 2024.

LOPES, Ana Maria D'Ávila. **Iniciativas internacionais de proteção aos neurodireitos humanos: transhumanismo ou bioconservadorismo?.** *REVISTA QUAESTIO IURIS*, 2024. Disponível em: <http://dx.doi.org/10.12957/rqi.2024.75061>. Acesso em: 20 mai. 2025.

LORACH, Henri; GALVEZ, Andrea; SPAGNOLO, Valeria *et al.* **Walking naturally after spinal cord injury using a brain–spine interface.** *Nature* 618, 126–133, 2023. Disponível em: <https://doi.org/10.1038/s41586-023-06094-5>. Acesso 4 set. 2024.

LOZANO, Andres M., LIPSMAN, Nir *et al.* **Deep brain stimulation: current challenges and future directions.** *Nat Rev Neurol*, 2019. Disponível em: <https://doi.org/10.1038/s41582-018-0128-2>. Acesso em: 20 mai. 2024.

LOZANO, Andres; FOSDICK, Lisa *et al.* **A Phase II Study of Fornix Deep Brain Stimulation in Mild Alzheimer's Disease.** *Journal of Alzheimer's Disease*, 2024. Disponível em: <https://www.doi.org/10.3233/JAD-160017>. Acesso em: 10 nov. 2024.

LU, Guanghan; LIU, Jingxuan; WEI, Xiang *et al.* **The Conservative Treatment of a Rare Postoperative Complication of DBS-Brain Abscess: Case Series.** *Infection and drug resistance*, 2025. Disponível em: <https://doi.org/10.2147/IDR.S497124>. Acesso em: 10 jun. 2025.

MACHADO, André Mendonça. **O impacto de vieses cognitivos sobre a imparcialidade do conteúdo de inteligência.** *Revista Brasileira de Inteligência*, Brasília, Brasil, n. 13, p. 1–16, 2018. DOI: 10.58960/rbi.2018.13.147. Disponível em: <https://rbi.abin.gov.br/RBI/article/view/147>. Acesso em: 26 ago. 2025.

MALERBI, Fernando; ANDRADE, Rafael *et al.* **Manejo da retinopatia diabética.** *Diretriz Oficial da Sociedade Brasileira de Diabetes*, 2021. Disponível em: <https://doi.org/10.29327/557753.2022-17>. Acesso em: 4 jul. 2025.

MALGIERI, Gianclaudio; PASQUALE, Frank. **Licensing high-risk artificial intelligence: Toward ex ante justification for a disruptive technology.** *Computer Law & Security Review*, v. 49, 105899, 2024. Disponível em: <https://doi.org/10.1016/j.clsr.2023.105899>. Acesso em: 8 jul. 2025.

MANRAI, Arjun K. *et al.* **Genetic misdiagnoses and the potential for health disparities.** *New England Journal of Medicine*, 2016. Disponível em: <https://www.nejm.org/doi/full/10.1056/NEJMsa1507092>. Acesso em: 27 ago. 2024.

MANSFIELD, Peter; MAUDSLEY, AA. **Medical imaging by NMR**. Br J Radiol. 1977 Mar; 50(591):188-94. Disponível em: <https://doi.org/10.1259/0007-1285-50-591-188>. Acesso em: 10 ago. 2024.

MARIN, Heimar de Fátima. Prólogo: Inteligência Artificial e saúde. In: COMITÊ GESTOR DA INTERNET NO BRASIL. **Inteligência Artificial na saúde: potencialidades, riscos e perspectivas para o Brasil**. São Paulo: CGI.br, 2024. Disponível em: <https://cetic.br/pt/publicacao/inteligencia-artificial-na-saude-potencialidades-riscos-e-perspectivas-para-o-brasil/>. Acesso em: 10 abr. 2025.

MARKS, Robin; KURTZMAN, Laura. **How Artificial Intelligence Gave a Paralyzed Woman Her Voice Back**. UCSF News Center, 2023. Disponível em: <https://www.ucsf.edu/news/2023/08/425986/how-artificial-intelligence-gave-paralyzed-woman-her-voice-back>. Acesso em: 28 jul. 2024.

MARTINI, Vitor Galioti. **Análise de Equidade em Algoritmos de IA na área da Saúde: Um Estudo Sobre Viés de Dados, Medidas de Pós-Processamento e Correlações de Atributos**. Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) – Universidade Federal de São Paulo, São José dos Campos, 2023. Disponível em: <https://repositorio.unifesp.br/items/a34aa04b-11ee-43b2-8a82-583d42e5ff1a>. Acesso em: 10 ago. 2024.

MATHUR, Arunesh; GUNES, Acar *et al.* **Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites**. *Proceedings of the ACM on Human-Computer Interaction*, 2019. Disponível em: <https://doi.org/10.1145/3359183>. Acesso em: 7 abr. 2024.

MCGRATH, Amanda; JONKER, Alexandra. **What is AI interpretability?** IBM, 2024. Disponível em: <https://www.ibm.com/think/topics/interpretability#:~:text=Intrinsic%20vs.%20post>. Acesso em: 12 ago. 2025.

MEHRABI, Ninareh. *et al.* **A Survey on Bias and Fairness in Machine Learning**. *ACM Computing Surveys*, 2022. Disponível em: <https://doi.org/10.48550/arXiv.1908.09635>. Acesso em: 15 ago. 2025.

MELLO, Michelle M.; CHAR, Danton; XU, Sonnet H. **Ethical obligations to inform patients about use of AI tools**. *JAMA*, 2025. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/40690211/>. Acesso em: 7 ago. 2025.

MESKE, Christian; BRENNE, Justin *et al.* **From Explainable to Explanatory Artificial Intelligence: Toward a New Paradigm for Human-Centered Explanations through Generative AI**. ArXiv, 2025. Disponível em: www.doi.org/10.48550/arXiv.2508.06352. Acesso em: 25 ago. 2025.

METZGER, Sean L. *et al.* **Generalizable spelling using a speech neuroprosthesis in an individual with severe limb and vocal paralysis**. *Nature Communications*, Londres, v. 13, 2022. Disponível em: <https://www.nature.com/articles/s41467-022-33611-3>. Acesso em: 24 jun. 2025.

METZGER, Sean L. *et al.* Author Correction: **A high-performance neuroprosthesis for speech decoding and avatar control.** *Nature* 631, E13, 2024. Disponível em: <https://doi.org/10.1038/s41586-024-07735-z>. Acesso em: 5 set. 2024.

MILEKOVIC, Tomislav; MORAUD, Eduardo Martin *et al.* **A spinal cord neuroprosthesis for locomotor deficits due to Parkinson's disease.** *Nature Medicine*, 2023. Disponível em: <https://doi.org/10.1038/s41591-023-02584-1>. Acesso em: 20 nov. 2024.

MILLER, R. A. **Medical diagnostic decision support systems--past, present, and future: a threaded bibliography and brief commentary.** *Journal of the American Medical Informatics Association*: *JAMIA*, 1(1), 8–27, 1994. Disponível em: <https://doi.org/10.1136/jamia.1994.9523614>. Acesso em: 12 out. 2024.

MINSKY, Marvin; PAPERT, Seymour. **Perceptrons: An Introduction to Computational Geometry.** Cambridge, MA: MIT Press, 1969. Disponível em: <https://doi.org/10.7551/mitpress/11301.001.0001>. Acesso em: 20 nov. 2024.

MITCHELL, Melante. **Artificial Intelligence: a guide for thinking humans.** New York: Farrar, Straus and Giroux, eBook Kindle, 2019.

MITTERMAIER, Mirja; RAZA, Mariam M.; KVEDAR, Joseph C. **Bias in AI-based models for medical applications: challenges and mitigation strategies.** *npj Digit. Med.*, 2023. Disponível em: <https://doi.org/10.1038/s41746-023-00858-z>. Acesso em: 14 mar. 2024.

MITTELSTADT, Brent. **The impact of artificial intelligence on the doctor-patient relationship.** Council of Europe, 2021. Disponível em: <https://www.coe.int/en/web/human-rights-and-biomedicine/report-impact-of-ai-on-the-doctor-patient-relationship>. Acesso em: 14 nov. 2024.

MOLNÁR-GÁBOR, Fruzsina; IENCA, Marcello; FINS, Joseph *et al.* **Towards a Governance Framework for Brain Data.** arXivLabs, 2021. Disponível em: <https://doi.org/10.48550/arXiv.2109.11960>. Acesso em: 27 ago. 2024.

MOLNÁR-GÁBOR, Fruzsina. **Artificial Intelligence in Healthcare: Doctors, Patients and Liabilities.** In: Wischmeyer, T., Rademacher, T. (eds) *Regulating Artificial Intelligence.* Springer Nature, 2020. Disponível em: https://doi.org/10.1007/978-3-030-32361-5_15. Acesso em: 10 fev. 2024.

MOLNÁR-GÁBOR, Fruzsina; LÓPEZ-SILVA, Pablo *et al.* **Neurotechnological Applications and the Protection of Mental Privacy: An Assessment of Risks.** *Neuroethics*, 2024. Disponível em: <http://dx.doi.org/10.1007/s12152-024-09565-2>. Acesso em: 10 fev. 2025

MOLNÁR-GÁBOR, Fruzsina. **Implementing the human right to science in the context of health-related data processing.** *Journal of Law and the Biosciences*, 2024. Disponível em: <http://dx.doi.org/10.1093/jlb/lxae004>. Acesso em: 10 fev. 2025

MOSES, David A. *et al.* **Neuroprosthesis for deciphering spoken words in a paralyzed person with anarthria.** *New England Journal of Medicine*, Boston, v. 385, n. 3, p. 217-227, 2021. Disponível em: <https://www.nejm.org/doi/full/10.1056/NEJMoa2027540>. Acesso em: 24 jun. 2025.

MUIRHEAD, William R.; HORSFALL, Hugo Layard *et al.* **Implanted cortical neuroprosthetics for speech and movement restoration.** *Journal of Neurology*, v. 271, n. 11, p. 7156-7168, 2024. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/39446156/>. Acesso em: 22 jul. 2024.

MUÑOZ, José M.; MARINARO, José Ángel. **Neurorights as reconceptualized human rights.** *Frontiers in Political Science*, 2023. Disponível em: <https://doi.org/10.3389/fpos.2023.1322922>. Acesso em: 20 mai. 2025.

MUSOLE, Edward. **The Role of Neurotechnology in Shaping Business Decisions: A Comprehensive Literature Review.** *Open Journal of Business and Management*, 2025. Disponível em: <https://doi.org/10.4236/ojbm.2025.132053>. Acesso em: 7 jul. 2025.

NALIN, Paulo; NOGAROLI, Rafaella. **Perspectivas sobre ética e responsabilidade civil no contexto dos robôs inteligentes de assistência à saúde.** In: *CAMPOS, Aline França; BERLINI Luciana Fernanda (coord.). Temas contemporâneos de responsabilidade civil: teoria e prática.* Belo Horizonte: D'Plácido; 2020, p. 61-94.

NAN, Ju Gang *et al.* **AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population: A Randomized Controlled Trial.** *Radiological Society of North America, Radiology*, 2023. Disponível em: <https://doi.org/10.1148/radiol.221894>. Acesso em: 10 mai. 2025.

NASCIMENTO, Sabrina Maciel *et al.* **Inteligência artificial e suas implicações éticas e legais: revisão integrativa.** *Revista Bioética*, v. 32, 2024. Disponível em: https://revistabioetica.cfm.org.br/revista_bioetica/article/view/3729. Acesso em: 15 abr. 2025.

NATIONAL ELECTRICAL MANUFACTURERS ASSOCIATION (NEMA). **Digital Imaging and Communications in Medicine (DICOM®) Standard.** Rosslyn, VA: NEMA, 2024. Disponível em: <https://www.dicomstandard.org/>. Acesso em: 10 ago. 2024.

NAZER, Lama H.; ZATARAH, Razan *et al.* **Bias in artificial intelligence algorithms and recommendations for mitigation.** *PLOS Digit Health*, 2023. Disponível em: <https://doi.org/10.1371/journal.pdig.0000278>. Acesso em: 10 ago. 2024.

NEGRI, Sergio M. C. A. **Personalidade, responsabilidade e classificação dos riscos na Inteligência Artificial e na robótica.** Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/347862/personalidade-responsabilidade-e-classificacao-dos-riscos-na-ia>. Acesso em: 10 nov. 2024.

NEGRI, Sergio M. C. Ávila; FELIPE, Bruno Farage da Costa; GIOVANINI, Carolina Fiorini Ramos. **Neurodireitos e a crise da autodeterminação informativa.** In: *O futuro dos neurodireitos. Coord. Amanda Cunha e Mello Smith Martins, Bruno Farage da Costa Felipe, Mariana Ruzzi; coord. institucional: Francisco Cavalcante de Sousa.* 1. ed. São Paulo: Legal Grounds Institute, eBook Kindle, 2025.

NEURORIGHTS FOUNDATION. **Our Approach: The Five Neurorights.** Disponível em: <https://www.neurorightsfoundation.org/mission/impact>. Acesso em: 10 nov. 2024.

FACCHINI NETO, Eugênio; BARBOSA, Rodrigo Mambrini Sandoval. **Viés da automação e responsabilidade civil médica por erro de diagnóstico realizado com auxílio da**

inteligência artificial. Civilistica.com, Rio de Janeiro, v. 12, n. 3, p. 1–22, 2023. Disponível em: <https://civilistica.emnuvens.com.br/redc/article/view/922>. Acesso em: 2 jun. 2025.

NEVES, Juliana Guimarães de Castro; BARROS, Maria Carolina Mendonça de. **Neurodireitos - A tutela da nossa mente.** Migalhas, 2024. Disponível em: <https://www.migalhas.com.br/depeso/405714/neurodireitos--a-tutela-do-nossa-mente>. Acesso em: 2 mai. 2025.

NEVES, Maria do Céu Patrão [coordenadora]. **Relatório sobre o Estado da Aplicação das Novas Tecnologias à Vida Humana - NEUROTECNOLOGIAS.** Conselho Nacional de Ética para as Ciências da Vida, 2024. Disponível em: https://www.cneqv.pt/pt/deliberacoes/relatorios-de-novas-tecnologias/relatorio-sobre-o-estado-da-aplicacao-das-novas-tecnologias-a-vi?download_document=12913&token=2f113df23d90f3f400e9198c3c155d71. Acesso em: 4 jul. 2025.

NGUYEN, Tram N.; LEE, Tyrone *et al.* **AlphaMissenseR: an integrated framework for investigating missense mutations in human protein-coding genes.** Bioinformatics Advances, Volume 5, Issue 1, 2025. Disponível em: <https://doi.org/10.1093/bioadv/vbaf093>. Acesso em: 4 jul. 2025.

NICOLÁS-ALONSO, Luis F.; GÓMEZ-GIL, Jaime. **Brain–computer interfaces: a review.** Sensors, Basel, v. 12, n. 2, p. 1211-1279, 2012. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC3304110>. Acesso em 24 jun. 2025.

NIST. **Artificial Intelligence Risk Management Framework (AI RMF 1.0).** Gaithersburg, MD: National Institute of Standards and Technology, 2023. Disponível em: <https://www.nist.gov>. Acesso em: 13 ago. 2025.

NIXON LAW GROUP. **FDA's New Playbook for AI in Medical Devices: What Developers Need to Know.** Fairfax, VA: Nixon Law Group, 25 jun. 2025. Disponível em: <https://nixonlawgroup.com/nlg-blog/fdas-new-playbook-for-ai-in-medical-devices-what-developers-need-to-know>. Acesso em: 13 ago. 2025.

NOBEL ASSEMBLY AT KAROLINSKA INSTITUTET. **Press Release: The 1979 Nobel Prize in Physiology or Medicine (Cormack & Hounsfield – development of computed tomography).** Estocolmo, 11 out. 1979, p.130-138. Disponível em: <https://www.nobelprize.org/prizes/medicine/1979/press-release>. Acesso em: 10 ago. 2024.

NOBEL ASSEMBLY AT KAROLINSKA INSTITUTET. **Press Release: The 2003 Nobel Prize in Physiology or Medicine (Lauterbur & Mansfield – discoveries concerning MRI).** Estocolmo, 6 out. 2003, p.119-124. Disponível em: <https://www.nobelprize.org/prizes/medicine/2003/press-release>. Acesso em: 10 ago. 2024.

NOBLE, Safiya Umoja. **Algorithms of Oppression: How Search Engines Reinforce Racism.** NYU Press, 2018. Disponível em: <https://doi.org/10.2307/j.ctt1pwt9w5>. Acesso em: 9 de set. 2024.

NOGAROLI, Rafaella. **Responsabilidade civil médica na inteligência artificial: culpa médica e deveres de conduta no século XXI.** São Paulo, Thomson Reuters Brasil, eBook Kindle, 2023.

NOGAROLI, Rafaela; DANTAS, Eduardo. **Consentimento informado do paciente frente às novas tecnologias da saúde (telemedicina, cirurgia robótica e inteligência artificial)**. *Lex Medicinae*, Ano 17, n.º 33, 2020. Disponível em: https://www.researchgate.net/publication/356172107_Consentimento_informado_do_paciente_frente_as_novas_tecnologias_da_saude_telemedicina_cirurgia_robotica_e_inteligencia_artificial. Acesso em: 9 de set. 2024.

NOGAROLI, Rafaela; FALEIROS JÚNIOR, Luiz de Moura Faleiros. **Do consentimento informado ao processo de escolha esclarecida: uma resenha à obra 'Consentimento do paciente no direito médico', de Flaviana Rampazzo Soares (Indaiatuba: Foco, 2021)**. *Revista IBERC*, Belo Horizonte, v. 4, n. 2, p. 179–185, 2021. DOI: 10.37963/iberc.v4i2.177. Disponível em: <https://revistaiberc.responsabilidadecivil.org/iberc/article/view/177>. Acesso em: 18 ago. 2024.

NORDLINGER, Bernard; VILLANI, Cédric; RUS, Daniela (Ed.). **Healthcare and Artificial Intelligence**. Cham: Springer, 2020. p. vii - x. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-32161-1>. Acesso em: 15 jun. 2024.

NORORI, Natalia; HU, Qiyang *et al.* **Addressing bias in big data and AI for health care: A call for open science**. *Patterns*, New York, 2021. Disponível em: <https://doi.org/10.1016/j.patter.2021.100347>. Acesso em: 10 mar. 2024.

OBERMEYER, Ziad; POWERS, Brian; VOGELI, Christine; MULLAINATHAN, Sendhil. **Dissecting racial bias in an algorithm used to manage the health of populations**. *Science*, 2019. Disponível em: <https://doi.org/10.1126/science.aax2342>. Acesso em: 2 set. 2024.

OCDE - Organização para a Cooperação e Desenvolvimento Econômico. **Recommendation of the Council on Responsible Innovation in Neurotechnology**. OECD/LEGAL/0457, 2019. Disponível em: <https://legalinstruments.oecd.org/fr/instruments/OECD-LEGAL-0457>. Acesso em: 24 jun. 2025.

OCDE - Organização para a Cooperação e Desenvolvimento Econômico. **Neurotechnology Toolkit. To support policymakers in implementing the OECD Recommendation on Responsible Innovation in Neurotechnology**. OECD, 2025. Disponível em: <https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/emerging-technologies/neurotech-toolkit.pdf>. Acesso em: 24 ago. 2025.

OEA. **CJI/RES. 281 (CII-O/23) - Declaración de Principios Interamericanos en Materia de Neurociencias, Neurotecnologías y Derechos Humanos**. OAS.ORG, Rio de Janeiro, Brasil, 2023. Disponível em: https://www.oas.org/es/sla/cji/docs/CJI-RES_281_CII-O-23_corr1_ESP.pdf. Acesso em: 2 set. 2024.

O'NEIL, Cathy. **Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democracia**. Trad. Rafael Abraham. São Paulo: Rua do Sabão, eBook Kindle, 2020.

O'NEILL, Onora. **A Question of Trust**. Cambridge: Cambridge University Press, 2002. Disponível em: <https://assets.cambridge.org/97805215/29969/sample/9780521529969ws.pdf>. Acesso em: 11 abr. 2025.

OLIVEIRA JR., Paulo Cosmo de. **Discriminação algorítmica e a violação dos direitos fundamentais.** Migalhas, 2025. Disponível em: <https://www.migalhas.com.br/depeso/431950/discriminacao-algoritmica-e-a-violacao-dos-direitos-fundamentais>. Acesso em: 5 jul. 2025.

OLIVEIRA, Luiz Eduardo Silva de; SILVA, Marcelo Colares da *et al.* **Diagnosis of diabetic retinopathy by artificial intelligence using smartphone.** *Rev. bras. oftalmol.* [online]. 2024, vol. 83, Disponível em: <https://www.rbojournal.org/en/article/diagnosis-of-diabetic-retinopathy-by-artificial-intelligence-using-smartphone/>. Acesso em: 4 jul. 2025.

OLIVEIRA, Marcela. C. A. de; LIMA, Gabriel C. P. B. *et al.* **Inteligência Artificial na cardiologia: uma revisão das aplicações no diagnóstico e prevenção de doenças cardiovasculares.** *CONTRIBUCIONES A LAS CIENCIAS SOCIALES*, 18(3), e16021, 2025. Disponível em: <https://doi.org/10.55905/revconv.18n.3-032>. Acesso em: 5 jul. 2025.

OLVERA-BARRIOS, Abraham; HEEREN, Tjebo FC *et al.* **Diagnostic accuracy of diabetic retinopathy grading by an artificial intelligence-enabled algorithm compared with a human standard for wide-field true-colour confocal scanning and standard digital retinal images.** *British Journal of Ophthalmology*, 2021. Disponível em: <https://doi.org/10.1136/bjophthalmol-2019-315394>. Acesso em: 4 jul. 2025.

OMS - ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Ética e governança da inteligência artificial para a saúde.** Genebra: OMS, 2021. Disponível em: <https://www.who.int/publications/i/item/9789240029200>. Acesso em: 6 abr. 2025.

OMS - ORGANIZAÇÃO MUNDIAL DA SAÚDE; UNIÃO INTERNACIONAL DE TELECOMUNICAÇÕES. **Regulatory considerations on Artificial Intelligence for health.** Geneva: WHO, 2023. Disponível em: <https://www.who.int/publications/i/item/9789240078871>. Acesso em: 10 mai. 2024

OMS - ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Ethics and governance of artificial intelligence for health. Guidance on large multi-modal models.** Geneva: World Health Organization, 2024. Licence: CC BY-NC-SA 3.0 IGO. Disponível em: <https://www.who.int/publications/i/item/9789240084759>. Acesso em: 13 ago. 2025.

OPEN FUTURE. **AI Act and Open Source.** Varsóvia: Open Future Foundation, 12 jul. 2024. Disponível em: <https://openfuture.eu/observatory/aia-open-source/>. Acesso em: 13 ago. 2025.

PAGE, Katie. **The four principles: Can they be measured and do they predict ethical decision making?.** *BMC Med Ethics*, 2012. Disponível em: <https://doi.org/10.1186/1472-6939-13-10>. Acesso em: Acesso em: 16 ago. 2024.

PACHECO, Rodrigo (org.). **A reforma do Código Civil: artigos sobre a atualização da Lei nº 10.406/2002.** Brasília, DF: Senado Federal, Coordenação de Edições Técnicas, 2025. Disponível em: <https://livraria.senado.leg.br/a-reforma-do-codigo-civil>. Acesso em: 14 jul. 2025.

PAIXÃO, Gabriela Miana de Mattos; SANTOS, Bruno Campos *et al.* **Machine Learning in Medicine: Review and Applicability.** *Arq. Bras. Cardiol.*, v. 118, n. 1, p. 95-102, jan. 2022. Disponível em: www.doi.org/10.36660/abc.20200596. Acesso em: 16 fev. 2024.

PARIKH, Purvish Mahendra; VENNIYOOR, Ajit. **Neuralink and Brain–Computer Interface—Exciting Times for Artificial Intelligence.** *South Asian Journal of Cancer*, v. 13, n. 1, p. 63-65, 2024. Disponível em: <https://www.thieme-connect.de/products/ejournals/abstract/10.1055/s-0043-1774729>. Acesso em: 22 set. 2024.

PASSARINI. **ANVISA RDC nº 657: Regularização de software como dispositivo médico.** São Paulo: Passarini Regulatory Affairs, 2022. Disponível em: <https://passarini.com.br/anvisa-rdc-n-657-regularizacao-de-software-como-dispositivo-medico/>. Acesso em: 12 ago. 2025.

PASQUALE, Frank. **The black box society: the secret algorithms that control money and information.** Cambridge: Harvard University Press, eBook Kindle, 2015.

PASQUALE, Frank. **New laws of robotics: defending human expertise in the age of AI.** Belknap Press, eBook Kindle, 2020.

PASTERNAK, Alex. **How philosopher Shannon Vallor delivered the year’s best critique of AI. The author of ‘The AI Mirror’ worries about what we’re doing to our moral muscles as we grow more reliant on machines.** Fast Company, 2024. Disponível em: <https://www.fastcompany.com/91240425/how-philosopher-shannon-vallor-delivered-the-years-best-critique-of-ai>. Acesso em: 27 jun. 2024.

PATIL, Ganesh, PANPATIL, Chetan *et al.* **A Review on Neuralink – Achieve AI Symbiosis with Medical Field.**, *Int. J. of Pharm. Sci.*, 2024, Vol 2. Disponível em: <https://doi.org/10.5281/zenodo.14531117>. Acesso em: 10 set. 2025.

PAUKER, Stephen G.; GORRY, G. Anthony *et al.* **Towards the simulation of clinical cognition: Taking a present illness by computer.** *The American Journal of Medicine*, v. 60, n. 7, p. 981-996, 1976. Disponível em: [https://doi.org/10.1016/0002-9343\(76\)90570-2](https://doi.org/10.1016/0002-9343(76)90570-2). Acesso em: 27 jun. 2024.

PENFIELD, Wilder; RASMUSSEN, Theodore. **The Cerebral Cortex of Man, a Clinical Study of Localization of Function.** JAMA, 1950. Disponível em: <https://jamanetwork.com/journals/jama/article-abstract/307907>. Acesso em: 2 set. 2024.

PHAM, Tuan. **Ethical and legal considerations in healthcare AI: innovation and policy for safe and fair use.** Royal Society Open Science, v. 12, 241873, 2025. Disponível em: <https://doi.org/10.1098/rsos.241873>. Acesso em: 8 jul. 2025.

PHELPS, Michael E. *et al.* **Application of annihilation coincidence detection to transaxial reconstruction tomography.** *J Nucl Med.* 1975 Mar;16(3):210-24. PMID: 1113170. Disponível em: <https://jnm.snmjournals.org/content/16/3/210.long>. Acesso em: 10 ago. 2024.

PICCOLINO, Marco. **Luigi Galvani's path to animal electricity.** *Comptes Rendus Biologies*, Volume 329, Issues 5–6, 2006, Pages 303-318, ISSN 1631-0691. Disponível em: <https://doi.org/10.1016/j.crv.2006.03.002>. Acesso em: 27 jul. 2024.

PIETRZYKOWSKI, Tomas; SMILOWSKA, Katarzyna. **The reality of informed consent: empirical studies on patient comprehension - systematic review.** *Trials*, 2021. Disponível em: <https://doi.org/10.1186/s13063-020-04969-w>. Acesso em: 11 ago. 2024.

PINTO; Adriano Roberto. **Software Livre e Direitos Fundamentais**. In: *Revista Científica sobre Cyberlaw do Centro de Investigação Jurídica do Ciberespaço – CIJIC – Da Faculdade de Direito da Universidade de Lisboa*. CIJIC, Portugal. ed. XIII, junho de 2025. Disponível em: https://www.iuris.edu.pt/xms/files/Cyberlaw_by_CIJIC_ed_XIII_versao_final.pdf. Acesso em: 10 ago. 2025.

PLOEM, Martine C.; GIESBERTZ, Noor A. A. *et al.* **Duty to recontact in genomic cancer care: A tool helping to assess the professional's responsibility**. *European journal of câncer*. Oxford, England: 2023. Disponível em: <https://doi.org/10.1016/j.ejca.2023.03.004>. Acesso em: 4 jul. 2024.

PLOUG, Thonas; HOLM, Soren. **The right to a second opinion on Artificial Intelligence diagnosis – Remediating the inadequacy of a risk-based regulation**. *Bioethics*, 2023. Disponível em: <https://doi.org/10.1111/bioe.13124>. Acesso em: 13 dez. 2024.

PLOUG, Thomas; HOLM, Soren. **The right to refuse diagnostics and treatment planning by artificial intelligence**. *Medicine, health care, and philosophy*, 2020. Disponível em: <https://doi.org/10.1007/s11019-019-09912-8>. Acesso em: 19 jul. 2024.

POMES, Mattia Vittorio; D'URSO, Giordano; BOVE, Ilaria *et al.* **Psychiatric Outcomes of Subthalamic Nucleus Deep Brain Stimulation: A Systematic Review of Short - and Long - Term Effects**. *Brain Sciences*, 2025. Disponível em: <https://doi.org/10.3390/brainsci15060566>. Acesso em: 20 jul. 2025.

POPEJOY, Alice B.; FULLERTON, Stephanie M. **Genomics is failing on diversity**. *Nature*, 2016. Disponível em: <https://www.nature.com/articles/538161a>. Acesso em: 27 ago. 2024.

PORTO, Camille; WERNECK, Alexandre. **O valor de uma existência: uma análise pragmática de valorizações da vida humana em situações envolvendo dinheiro. Dossiê: Sociologia pragmática e pragmatismo**. Scielo Brasil, 2021. Disponível em: <https://doi.org/10.1590/s0102-6992-202136020009>. Acesso em: 20 jul. 2024.

PORTNOW, Lea H *et al.* **The history of cerebral PET scanning: from physiology to cutting-edge technology**. *Neurology*. 2013. Disponível em: <https://doi.org/10.1212/WNL.0b013e318285c135>. Acesso em: 10 ago. 2024.

PORTUGAL. Ministério da Saúde. **Serviços Partilhados do Ministério da Saúde (SPMS). White Paper: Inteligência Artificial na Saúde em Portugal: Regulamentação, Impactos e Perspetivas de Futuro**. Lisboa: SPMS, fev. 2025, p. 15. Disponível em: https://www.spms.min-saude.pt/wp-content/uploads/2025/03/White-Paper_Inteligencia-Artificial-na-Saude-em-Portugal_Final2-1.pdf. Acesso em: 10 mai. 2025.

PRACTICAL NEUROLOGY. **FDA Clears Brain-Computer Interface Device for the Measurement and Stimulation of Cortical Brain Activity**. Wayne, PA: Practical Neurology, 21 abr. 2025. Disponível em: <https://practicalneurology.com/news/fda-clears-braincomputer-interface-device-for-the-measurement-and-stimulation-of-cortical-brain-activity/2474229/>. Acesso em: 13 ago. 2025.

PRICE II, W. Nicholson. **Regulating Black-Box Medicine**. *Michigan Law Review*, 2017; 116(3):421-74. PMID: 29240330. Disponível em:

<https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1683&context=mlr>. Acesso em: 13 ago. 2024.

PRICE II, W. Nicholson; COHEN, I. Glenn. **Privacy in the Age of Medical Big Data**. *Nature Medicine*, 2019. Disponível em: <https://repository.law.umich.edu/articles/2764/>. Acesso em: 10 set. 2024.

PRICE II, W. Nicholson; GERKE, Sara; COHEN, I. Glenn. **Potential Liability for Physicians Using Artificial Intelligence**. *JAMA*, 2019. Disponível em: <https://www.doi.org/10.1001/jama.2019.15064>. Acesso em: 10 set. 2024.

PRICE II, W. Nicholson. **Health Care AI: Law, Regulation, and Policy**. Douglas Mcnair, co-author. In: *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*, edited by Michael Matheny et al., 181-213. Washington, D.C.: The National Academy of Medicine, 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK605945/>. Acesso em: 17 abr. 2025.

PRICE II, W. Nicholson; COHEN, I. Glenn *et al.* **Liability for use of artificial intelligence in medicine**. In *B. Solaiman (Eds.) et al., Research Handbook on Health, AI and the Law*. (pp. 150–166). Edward Elgar Publishing Ltd., 2024. Disponível em: <https://doi.org/10.4337/9781802205657.ch09>. Acesso em: 10 fev. 2025.

PRICE II, W. Nicholson. **Medical AI and Contextual Bias**. *Harvard Journal of Law & Technology*, v. 33, n. 1, p. 65-116, Fall 2019. Disponível em: <https://repository.law.umich.edu/articles/2239/>. Acesso em: 11 mar. 2024.

PRIEWASSER, Tina. **Regulatory requirements for medical devices with machine learning**. *Johner Institute*, 14 abr. 2025. Disponível em: <https://blog.johner-institute.com/regulatory-affairs/regulatory-requirements-for-medical-devices-with-machine-learning/>. Acesso em: 15 ago. 2025.

QIANG, Yan-Rui; ZHANG, Shao-Wu *et al.* **Diagnosis of Alzheimer's disease by joining dual attention CNN and MLP based on structural MRIs, clinical and genetic data**. *Artificial intelligence in medicine*, 2023. Disponível em: <https://doi.org/10.1016/j.artmed.2023.102678>. Acesso em: 10 ago. 2024.

RAINEY, Stephen *et al.* **Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?** *Journal of Law and the Biosciences*, 2020. Disponível em: <https://doi.org/10.1093/jlb/ljaa051>. Acesso em: 10 ago. 2024.

RAJPURA, Param; CECOTTI, Hubert; MEENA, Yogesh Kumar. **Explainable artificial intelligence approaches for brain-computer interfaces: a review and design space**. *Arxiv*, 2023. Disponível em: <https://doi.org/10.48550/arXiv.2312.13033>. Acesso em: 20 ago. 2024.

RAMACHANDRAN, Anand. **Revolutionizing Brain-Computer Interfaces The Transformative Impact of Advanced AI Across Functional Areas**. *IBM Research*, 2025. Disponível em: https://www.researchgate.net/publication/387766083_Revolutionizing_Brain-Computer_Interfaces_The_Transformative_Impact_of_Advanced_AI_Across_Functional_Areas. Acesso em: 24 jul. 2025.

RAMOS, Matheus Rassi F.; GOERIGK, Stephan *et al.* **Accelerated Theta-Burst Stimulation for Treatment-Resistant Depression: A Randomized Clinical Trial.** *JAMA Psychiatry*, 2025 Disponível em: <https://doi.org/10.1001/jamapsychiatry.2025.0013>. Acesso em 20 jul. 2025.

RATTI, Emanuele; MORRISON, Michael *et al.* **Ethical and social considerations of applying artificial intelligence in health care: a two-strand scoping review.** *BMC Medical Ethics*, v. 26, 2025. Disponível em: <https://doi.org/10.1186/s12910-025-01198-1>. Acesso em: 14 jun. 2025.

REDDY, N. Sudhir; KHANAA, V. **Diagnosing and categorizing of pulmonary diseases using Deep learning conventional Neural network.** *International Journal of Experimental Research and Review, [S. l.]*, v. 31, n. Spl Volume, p. 12–22, 2023. Disponível em: <https://qtanalytics.in/journals/index.php/IJERR/article/view/2328>. Acesso em: 20 ago. 2024.

REIS, Gabriella; REIS, Zilma Silveira Nogueira. **Proteção de dados em pesquisa na saúde.** Centro de Inovação em Inteligência Artificial para a Saúde da UFMG, 1 ed – Belo Horizonte: 2024. Disponível em: <https://doi.org/10.5281/zenodo.14677910>. Acesso em: 14 mar. 2025.

REUTERS. **Elon Musk's Neuralink implants brain chip in first human.** 30 jan. 2024. Disponível em: <https://www.reuters.com/technology/neuralink-implants-brain-chip-first-human-musk-says-2024-01-29>. Acesso em: 21 jun. 2024.

REUTERS. **Musk's Neuralink says second trial implant went well, no thread retraction issue.** 21 ago. 2024. Disponível em: <https://www.reuters.com/business/healthcare-pharmaceuticals/musks-neuralink-says-second-trial-implant-went-well-no-thread-retraction-issue-2024-08-21/>. Acesso em: 22 set. 2024.

RODOTÀ, Stefano. **A vida na sociedade da vigilância: a privacidade hoje.** Org. Maria Celina Bodin de Moraes; Danilo Doneda. Trad. Danilo Doneda; Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

RODOTÀ, Stefano. **Il diritto di avere diritti.** 4 ed. Roma-Bari: Laterza, 2015.

RODRIGUES, Pedro Henrique de Oliveira; PASSOS, Ana Paula Vieira *et al.* **O Papel da Inteligência Artificial como Ferramenta Diagnóstica na Medicina: Uma Revisão Sistemática.** *Brazilian Journal of Implantology and Health Sciences*, 6(9), 3308–3319. 2024. Disponível em: <https://doi.org/10.36557/2674-8169.2024v6n9p3308-3319>. Acesso em 4 jul. 2025.

ROSBACH, Emely; GANZ, Jonathan *et al.* **Automation Bias in AI-Assisted Medical Decision-Making under Time Pressure in Computational Pathology.** arXiv, 2024. Disponível em: <https://arxiv.org/pdf/2411.00998>. Acesso em: 10 mai. 2025.

ROSENVALD, Nelson; MARTINS, Guilherme Magalhães (coord.). **Responsabilidade Civil e Novas Tecnologias.** 2. ed. - Indaiatuba, SP: Editora Foco, eBook Kindle, 2024.

ROY, Yannick; BANVILLE; Hubert *et al.* **Deep learning-based electroencephalography analysis: a systematic review.** *Journal of Neural Engineering*, v. 16, n. 5, 051001, 2019. Disponível em: <https://doi.org/10.1088/1741-2552/ab260c>. Acesso em: 15 jun. 2024.

ROY, Eleanor; MALAFA, Sara; ADWER, Lina M. *et al.* **Artificial Intelligence in Healthcare: Medical Students' Perspectives on Balancing Innovation, Ethics, and Patient-Centered Care.** *International Journal of Medical Students*, 2025. Disponível em: <https://doi.org/10.5195/ijms.2025.3344>. Acesso em: 20 mai. 2025.

RÜGER, André. **Conflitos familiares em genética humana: o profissional da saúde diante do direito de saber e do direito de não saber.** Belo Horizonte, 2007. Disponível em: http://www.biblioteca.pucminas.br/teses/Direito_RugerA_1.pdf. Acesso em: 13 ago. 2024.

RUIZ, Sergio; VALERA, Luca *et al.* **Neurorights in the Constitution: from neurotechnology to ethics and politics.** Royal Society, 2024. Disponível em: <https://doi.org/10.1098/rstb.2023.0098>. Acesso em: 10 abr. 2025.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: A Modern Approach.** Global Edition. 4 ed., Hoboken: Pearson Education, eBook Kindle, 2021.

SADEGHI, Zahra; ALIZADEHSANI, Roohallah *et al.* **A review of Explainable Artificial Intelligence in healthcare.** *Computers and Electrical Engineering*, 2024. Disponível em: <https://doi.org/10.1016/j.compeleceng.2024.109370>. Acesso em: 27 jul. 2025.

SAENZ, Agustina D.; MASS GENERAL BRIGHAM AI GOVERNANCE COMMITTEE *et al.* **Establishing responsible use of AI guidelines: a comprehensive case study for healthcare institutions.** *npj Digital Medicine*, v. 7, art. 348, 2024. Disponível em: <https://www.nature.com/articles/s41746-024-01300-8>. Acesso em: 7 ago. 2025.

SALLAM, Malik. **ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns.** *Healthcare*, 2023. Disponível em: <https://doi.org/10.3390/healthcare11060887>. Acesso em: 10 mar. 2025.
SALLES, Arleen; ROMMELFANGER, Karen *et al.* **Towards Inclusive EU Governance of Neurotechnologies.** ICFG, 2024. Disponível em: <file:///Users/christianoqueiroz/Downloads/Towards-Inclusive-EU-Governance-of-Neurotechnologies-Full-Report.pdf>. Acesso em: 14 jan. 2025.

SANTOS, Bruno Cavalcante Leitão. **Integração de neurotecnologias e inteligência artificial: implicações para a proteção dos neurodireitos como direitos fundamentais.** *Revista Jurídica da Presidência*, v. 27 n. 141 Jan – Abr 2025. Disponível em: <https://doi.org/10.20499/2236-3645.RJP2025v27e141-3227>. Acesso em 20 jun. 2025.

SARMENTO, Daniel. **Os princípios constitucionais da liberdade e da autonomia privada.** *Boletim Científico ESMPU*, Brasília, n. 14, p. 15-40, jan./mar. 2005. Disponível em: <https://escola.mpu.mp.br/publicacoescientificas/index.php/boletim/article/view/179>. Acesso em: 10 mar. 2024.

SCHIFF, Daniel; BORENSTEIN, Jason. **How Should Clinicians Communicate With Patients About the Roles of Artificially Intelligent Team Members?.** *AMA journal of ethics*, 2019. Disponível em: <https://doi.org/10.1001/amajethics.2019.138>. Acesso em: 10 mar. 2024.

SCHRODER, Tyler; SIRBU, Renée; PARK, Sohee *et al.* **Cyber Risks to Next-Gen Brain-Computer Interfaces: Analysis and Recommendations.** *Neuroethics*, 2025. Disponível em: <http://dx.doi.org/10.1007/s12152-025-09607-3>. Acesso em: 20 jul. 2025.

SCHULZ, Raymond A. *et al.* **How CT happened: the early development of medical computed tomography.** *Journal of Medical Imaging*, Vol. 8, Issue 5, 2021. Disponível em: <https://doi.org/10.1117/1.JMI.8.5.052110>. Acesso em: 10 ago. 2024.

SCHULZE-BONHAGE, Andreas; HIRSCH, Martin; KNAKE, Susanne *et al.* **Focal Cortex Stimulation With a Novel Implantable Device and Antiseizure Outcomes in 2 Prospective Multicenter Single-Arm Trials.** *JAMA Neurology*, 2023. Disponível em: <https://doi.org/10.1001/jamaneurol.2023.0066>. Acesso em: 4 jul. 2025.

SEGAL, Yoram; HOD, Adi. **The Integration of AI Technologies in Modern Healthcare: A Paradigm Shift in Data Security and Patient Care. Artificial Intelligence in Medicine and Surgery - An Exploration of Current Trends, Potential Opportunities, and Evolving Threats.** IntechOpen, 2025. Disponível em: <https://www.doi.org/10.5772/intechopen.1009317>. Acesso em: 10 ago. 2025.

SCHÖNWEITZ, Franziska Britta *et al.* **Where is the exit? The ethical importance of exit plans in clinical trials with neural implants.** *Brain Stimulation*, v. 17, n. 5, p. 1145–1154, 2024. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/39321914/>. Acesso em: 13 abr. 2025

SHORTLIFFE, Edward H.; SEPÚLVEDA, Martin J. **Clinical decision support in the era of artificial intelligence.** *New England Journal of Medicine*, v. 379, n. 13, p. 1206-1209, 2018. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/30398550>. Acesso em: 17 ago. 2025.

SI, Jia-Yue; LIN, Zi-Yan *et al.* **Informed consent competency assessment for brain-computer interface clinical research and application in psychiatric disorders: a systematic review.** *World Journal of Psychiatry*, v. 15, n. 8, 107593, 2025. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC12362645/>. Acesso em: 20 ago. 2025.

SILVA, Fabiane Rodrigues da; MAC-KAY, Ana Paula Machado Goyano *et al.* **Estimulação transcraniana por corrente contínua: estudo sobre respostas em tarefas de nomeação em afásicos.** *CoDAS* 30 (5), 2018. Disponível em: <https://doi.org/10.1590/2317-1782/20182017242>. Acesso em: 17 ago. 2024.

SILVA, Caroline Oliveira da; CRIPPA, Anelise; BONHEMBERGER, Marcelo. **Diretivas antecipadas de vontade: busca pela autonomia do paciente.** *Rev. Bioética*, 2021. Disponível em: https://revistabioetica.cfm.org.br/revista_bioetica/article/view/2572. Acesso em: 16 ago. 2024.

SILVA, Fayllen Lemos Da. **A qualificação do dado neural como dado pessoal sensível.** *LUME UFRGS*, 2024. Disponível em: <http://hdl.handle.net/10183/279008>. Acesso em: 20 jul. 2025.

SILVA, Gabriela Gomes da; SILVA, Hugo Paixão da; RODRIGUES, Mariana Luiza de Acioly. **Desafios do uso da inteligência artificial nos diagnósticos de saúde: uma revisão integrativa.** *Cadernos Ibero-Americanos de Direito Sanitário*, Brasília, v. 13, n. 2, p. 1-17, 2024. Disponível em: <https://www.cadernos.prodisa.fiocruz.br/index.php/cadernos/article/view/1241>. Acesso em: 10 mar. 2025.

SILVA, Jeniffer Gomes da. **Discriminação algorítmica no direito brasileiro: violação do direito fundamental de proteção ao consumidor pela negação de acesso ao crédito.** *Revista Internacional da Academia Paulista de Direito*, 2022. Disponível em: <https://apd.org.br/wp-content/uploads/2023/05/TEXT0-7.pdf>. Acesso em: 10 mar. 2024.

SILVA, Maria Cecília Carneiro da; RIGUETTI, Sara Lima *et al.* **Inteligência artificial no combate à sepse: apoiando o diagnóstico e tratamento clínico.** *Revista Eletrônica Acervo Saúde*, 2025. Disponível em: <https://doi.org/10.25248/reas.e19761.2025>. Acesso em 20 jul. 2025.

SIMONASSI, Guilherme Samad; BARROS, Luca Parreira de *et al.* **O impacto da inteligência artificial no diagnóstico médico: avanços, desafios e oportunidades.** *Revista Ibero-Americana De Humanidades, Ciências E Educação*, 10(10), 2233–2242, 2024. Disponível em: <https://doi.org/10.51891/rease.v10i10.16047>. Acesso em: 4 jul. 2025.

SINYAVSKAYA, Oxana, BIRYUKOVA, Svetlana; SELEZNEVA, Elena. **Human Enhancement Technologies as a New Factor of Social Inequality in Modern Society.** *Journal of Economic Sociology*, 2025. Disponível em: <https://demreview.hse.ru/index.php/ecsoc/article/view/27244>. Acesso em: 7 jul. 2025.

SIRBU, Renee; MORLEY, Jessica; SCHRODER, Tyler *et al.* **Regulating Next-Generation Implantable Brain-Computer Interfaces: Recommendations for Ethical Development and Implementation.** *arXiv*, 2025. Disponível em: <https://doi.org/10.48550/arXiv.2506.12540>. Acesso em: 11 ago. 2025.

SMITH, Jeffrey; HOLDER, Andre *et al.* **Detecting algorithmic bias in medical-AI models.** *arXiv*, 2024. Disponível em: <https://arxiv.org/pdf/2312.02959v3>. Acesso em: 10 ago. 2024.

SMOLENAARS, Claudine Costa; GONÇALVES, Thaís Leal Vieira. **A tardia responsabilidade civil na discriminação algorítmica e a regulação estatal para sistemas de governança e compliance.** UFC, 2021. Disponível em: https://repositorio.ufc.br/bitstream/riufc/74286/1/2021_art_ccsmolenaars.pdf. Acesso em: 10 mai. 2024.

SNOMED INTERNATIONAL. **SNOMED CT Starter Guide.** London: SNOMED International, 2023. Disponível em: <https://www.snomed.org/snomed-ct/five-step-briefing>. Acesso em: 10 ago. 2024.

SOARES, Flaviana Rampazzo. **Inteligência artificial e consentimento do paciente: Desafios e limites.** Migalhas, 2025. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/436506/ia-e-consentimento-do-paciente-desafios-e-limites>. Acesso em: 12 ago. 2025.

SOLAIMAN, Barry; MEKKI, Yosra Magdi; AL-ANSARI, Abdulla. **A “True Lifecycle Approach” towards governing healthcare AI with the GCC as a global governance model.** *npj Digital Medicine*, 2025. Disponível em: <https://www.nature.com/articles/s41746-025-01614-1>. Acesso em: 7 ago. 2025.

SOLOVE, Daniel J. **Understanding Privacy.** Cambridge: Harvard University Press, 2008. *Last revised: 6 Mar 2013.* Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1127888#. Acesso em: 19 set. 2024.

SOLOVE, Daniel J. **Artificial Intelligence and Privacy**. Florida Law Review, Vol. 77, Issue 1, 2025. Disponível em: <https://www.floralawreview.com/article/129976-artificial-intelligence-and-privacy>. Acesso em: 10 ago. 2025.

STAHL, Bernd Carsten; AUCOUTURIER, Etienne; LEKSTUTIENE, Jurate *et. al.* **Knowledge needs of research ethics committees for the integration of ethics in research and technology development**. Journal of Responsible Innovation, 2025. Disponível em: <https://doi.org/10.1080/23299460.2025.2466904>. Acesso em: 20 jul. 2025.

SURESH, Harini; GUTTAG, John V. **A framework for understanding sources of harm throughout the machine learning life cycle**. In: *EAAMO '21: PROCEEDINGS OF THE 2021 ACM CONFERENCE ON EQUITY AND ACCESS IN ALGORITHMS, MECHANISMS, AND OPTIMIZATION, 2021, [S.l.]. Anais [...]*. New York: ACM, 2021. Disponível em: <https://www.doi.org/10.1145/3465416.3483305>. Acesso em: 14 mar. 2024.

TABASSI, Elham. **Artificial Intelligence Risk Management Framework (AI RMF 1.0)**. Gaithersburg, MD: National Institute of Standards and Technology, 2023. (NIST AI 100-1). DOI: 10.6028/NIST.AI.100-1. Disponível em: <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10>. Acesso em: 13 ago. 2025.

TANG, Jarry *et al.* **Semantic reconstruction of continuous language from non-invasive brain recordings**. Nat Neurosci 26, 858–866, 2023. Disponível em: <https://doi.org/10.1038/s41593-023-01304-9>. Acesso em: 7 jul. 2024.

TEFFÉ, Chiara Spadaccini de; CEIA, Eleonora Mesquita. **Os neurodireitos e a atualização do Código Civil Brasileiro**. Série Brasil em Foco, nº 04, 2024. Disponível em: <https://www.kas.de/pt/web/brasilen/brasil-em-foco/detail/-/content/serie-brasil-em-foco-04-2024>. Acesso em: 7 jul. 2025.

TELLES, Vanessa de J.; ALCÂNTARA, Matheus. S. **Impacts and challenges of Artificial Intelligence in Medicine: An updated literature review**. Research, Society and Development, [S. l.], v. 13, n. 12, p. e215131247962, 2024. Disponível em: <https://rsdjournal.org/index.php/rsd/article/view/47962>. Acesso em: 5 jul. 2025.

TEPEDINO, Gustavo. **Fundamentos do Direito Civil - Vol. 4 - Responsabilidade Civil**. 5ª Ed. Rio de Janeiro: Forense, eBook Kindle, 2024.

THALER, Richard H.; SUNSTEIN, Cass R. **Nudge. O Empurrão Para A Escolha Certa**. Rio de Janeiro: Elsevier, 2008.

THE SOCIABLE. **ONU finaliza projeto de ética em neurotecnologia, a ser adotado na Conferência Geral**. The Sociable, 2025. Disponível em: <https://sociable.co/government-and-policy/un-neurotech-ethics/>. Acesso em 12 ago. 2025.

THE WHITE HOUSE. **Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People**. Washington, DC: White House Office of Science and Technology Policy, out. 2022. Disponível em: <https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/>. Acesso em: 13 ago. 2025.

THIRUNAVUKARASU, Arun James; TING, Darren Shu Jeng, *et al.* **Large language models in medicine.** *Nat Med* 29, 1930–1940, 2023. Disponível em: <https://doi.org/10.1038/s41591-023-02448-8>. Acesso em: 10 mar. 2025.

TOPOL, Eric. **Deep medicine: how artificial intelligence can make healthcare human again.** New York: Basic Books, eBook Kindle, 2019.

TORGERSEN, Leanne N. S.; SCHULZ, Stefan M. *et al.* **Patient informed consent, ethical and legal considerations in the context of digital vulnerability with smart, cardiac implantable electronic devices.** *PLOS Digital Health*, 2024. Disponível em: <https://doi.org/10.1371/journal.pdig.0000507>. Acesso em: 13 mar. 2025

TOOSI, Amirhosein. **A brief history of AI: how to prevent another winter (a critical review).** *PET clinics*, 2021. Disponível em: <https://doi.org/10.1016/j.cpet.2021.07.001>. Acesso em: 15 jun. 2024.

TREMBLAY-MCGAW, Alexandra G. *et al.* **Best practices for clinical trials of deep brain stimulation for neuropsychiatric indications.** *Frontiers in Human Neuroscience*, v. 19, 2025. Disponível em: <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2025.1572972/ful>. Acesso em: 11 jul. 2025.

TUDOR, Mario; TUDOR, Lorainne; TUDOR, Katarina Ivana. **Hans Berger (1873-1941) – the history of electroencephalography.** *Acta medica Croatica: casopis Hrvatske akademije medicinskih znanosti*, 59(4), 307–313, 2005. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/16334737/>. Acesso em: 20 mar. 2024.

TURING, Alan M. **Computing machinery and intelligence.** *Mind*, 49: 433-460, 1950. Disponível em: <https://courses.cs.umbc.edu/471/papers/turing.pdf>. Acesso em: 12 abr. 2024.

UGUR, Muhammed *et al.* **Towards Forever Access for Implanted Brain-Computer Interfaces.** *arXiv*, 2024. Disponível em: <https://arxiv.org/abs/2409.17496>. Acesso em: 14 ago. 2025.

UNESCO - ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA. **Declaração Universal sobre Bioética e Direitos Humanos.** Paris: UNESCO, 2005. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000146180>. Acesso em: 11 ago. 2024.

UNESCO. **Draft text of the Recommendation on the Ethics of Neurotechnology.** Conference: Intergovernmental Meeting of Experts (Category II) Related to a Draft UNESCO Recommendation on the Ethics of Neurotechnology, Paris, 2025. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000393395>. Acesso em: 27 jul. 2025.

UNIÃO EUROPEIA. **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data.** Official Journal of the European Union, L 119, 4 maio 2016. Disponível em: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Acesso em: 14 jan. 2025.

UNIÃO EUROPEIA. **Regulamento (UE) 2017/745 do Parlamento Europeu e do Conselho, de 5 de abril de 2017, relativo aos dispositivos médicos.** Jornal Oficial da União Europeia, L

117, p. 1-175, 5 maio 2017a. Disponível em: <https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng>. Acesso em: 12 ago. 2025.

UNIÃO EUROPEIA. **Regulamento (UE) 2017/746 do Parlamento Europeu e do Conselho, de 5 de abril de 2017, relativo aos dispositivos médicos de diagnóstico *in vitro* (IVDR)**. Jornal Oficial da União Europeia, 2017b. Disponível em: <https://eur-lex.europa.eu>. Acesso em: 13 ago. 2025.

UNIÃO EUROPEIA. **Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL))**. Jornal Oficial da União Europeia, 2018a. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52017IP0051&from=SV>. Acesso em: 16 fev. 2024.

UNIÃO EUROPEIA. **Regulamento nº 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024**. Cria regras harmonizadas em matéria de inteligência artificial – “AI Act”). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689>. Acesso em: 12 de abril de 2025.

UNIÃO EUROPEIA. **Regulation (EU) 2025/327 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU)**, 2025. Disponível em: https://health.ec.europa.eu/publications/regulation-eu-2025327-european-health-data-space-and-amending-directive-201124eu-and-regulation-eu_en. Acesso em: 10 mai. 2025.

UNIÃO EUROPEIA. **Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, European Commission, 2018b**. Disponível em: https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/automated-decision-making-and-profiling_en. Acesso em: 29 jul. 2025.

UNIÃO EUROPEIA. **Charter of Fundamental Rights of the European Union**. Official Journal of the European Union, 2012/C 326/02, 26 out. 2012. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:12012P/TXT>. Acesso em: 14 jan. 2025.

UNIVERSITY OF OXFORD. **AI analysis of routine heart scans can predict risk of a developing heart problems ten years in advance, new research finds**. News, 2024. Disponível em: <https://www.rdm.ox.ac.uk/news/ai-analysis-of-routine-heart-scans-can-predict-risk-of-a-developing-heart-problems-ten-years-in-advance-new-research-finds#:~:text=A%20research%20team%20led%20by,to%20ten%20years%20in%20advance>. Acesso em: 15 ago. 2024.

VASERINO, Débora Karina Gonçalves. **A tutela jurídica nas ações de responsabilidade civil por wrongful conception, wrongful birth e wrongful life: estudo de casos estrangeiros e análise da aplicabilidade no Brasil**. Unesp, 2020. Disponível em: <http://hdl.handle.net/11449/202397>. Acesso em: 19 ago. 2024.

VIDAL, Jacques J. **Toward Direct Brain-Computer Communication**. Annual review of biophysics and bioengineering, 2, 157–180, 1973. Disponível em: <https://doi.org/10.1146/annurev.bb.02.060173.001105>. Acesso em: 15 jan. 2024.

VILAÇA, Murilo Mariano; MARQUES, Maria Clara Dias. **Tratar, sim; melhorar, não? Análise crítica da fronteira terapia/melhoramento.** *Rev. Bioética*, 2015. Disponível em: <https://doi.org/10.1590/1983-80422015232065>. Acesso em: 15 ago. 2024.

WACHTER, Sandra; MITTELSTADT, Brent; RUSSELL, Chris. **Bias Preservation in Machine Learning: The Legality of Fairness Metrics Under EU Non-Discrimination Law.** *West Virginia Law Review*, 2021. Disponível em: <http://dx.doi.org/10.2139/ssrn.3792772>. Acesso em: 15 abr. 2024.

WAGNER-ALTENDORF, Tobias A. **Progress in understanding consciousness? Easy and hard problems, and philosophical and empirical perspectives.** *Acta Analytica*, 2024. Disponível em: <https://link.springer.com/article/10.1007/s12136-024-00584-5>. Acesso em: 23 mar. 2025.

WEISS, Sholom M.; KULIKOWSKI, Casimir A.; SAFIR, Aran. **Glaucoma consultation by computer.** *Computers in Biology and Medicine*, v. 8, n. 1, p. 25-40, 1978. Disponível em: [https://doi.org/10.1016/0010-4825\(78\)90011-2](https://doi.org/10.1016/0010-4825(78)90011-2). Acesso em: 16 abr. 2024.

WILLETT, Francis R. *et al.* **High-performance brain-to-text communication via handwriting.** *Nature* 593, 249–254, 2021. Disponível em: <https://doi.org/10.1038/s41586-021-03506-2>. Acesso em: 7 jul. 2024.

WILLETT, Francis R.; KUNZ, Erin M. *et al.* **A high-performance speech neuroprosthesis.** *Nature* 620, 2023. Disponível em: <https://doi.org/10.1038/s41586-023-06377-x>. Acesso em: 11 ago. 2024.

WMA – World Medical Association. **Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants.** WMA, 2025. Disponível em: <https://www.wma.net/policies-post/wma-declaration-of-helsinki/>. Acesso em: 15 ago. 2025.

WÓJCIK, Malwina Anna. **Algorithmic Discrimination in Health Care: An EU Law Perspective.** *Health and human rights*, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/35747275/>. Acesso em: 10 mai. 2024.

WORLD ECONOMIC FORUM. **Intelligent Clinical Trials: Using Generative AI to Fast-Track Therapeutic Innovations.** White Paper, Dec. 2024. Disponível em: https://reports.weforum.org/docs/WEF_Intelligent_Clinical_Trials_2024.pdf. Acesso em: 10 mai. 2025.

XIE, Qianqian *et al.* **Medical foundation large language models for comprehensive text analysis and beyond.** *npj Digit. Med.* 8, 141, 2025. Disponível em: <https://doi.org/10.1038/s41746-025-01533-1>. Acesso em: 18 jun. 2025.

XU, Hanwen *et al.* **In new Nature paper, Allen School researchers ‘slide’ into the future of cancer diagnosis with groundbreaking AI model for digital pathology.** *Washington CS News*, 29 maio 2024. Disponível em: <https://news.cs.washington.edu/2024/05/29/in-new-nature-paper-allen-school-researchers-slide-into-the-future-of-cancer-diagnosis-with-groundbreaking-ai-model-for-digital-pathology/>. Acesso em: 15 ago. 2024.

XU, Hanwen; USUYAMA, Naoto *et al.* **A whole-slide foundation model for digital pathology from real-world data.** *Nature* 630, 181–188, 2024. <https://doi.org/10.1038/s41586-024-07441-w>. Acesso em: 4 jul. de 2025.

XU, Junpeng; LIU, Bin; SHANG, Guosong *et al.* **Deep brain stimulation of the nucleus basalis of Meynert in severe Alzheimer's disease.** *Journal of Alzheimer's disease reports*, 2024. Disponível em: <https://doi.org/10.1177/25424823241296780>. Acesso em: 10 nov. 2024.

XU, Junpeng; LIU, Bin; SHANG, Guosong *et al.* **Efficacy and Safety of Bilateral Deep Brain Stimulation (DBS) for Severe Alzheimer's Disease: A Comparative Analysis of Fornix Versus Basal Ganglia of Meynert.** *CNS neuroscience & therapeutics*, 2025. Disponível em: <https://doi.org/10.1111/cns.70285>. Acesso em: 20 jul. 2025.

YAZAKI, Shu; PEI, Xin *et al.* **AlphaMissense for Identifying Pathogenic Missense Mutations in DNA Damage Repair Genes in Cancer.** *JCO precision oncology*, 2025. Disponível em: <https://doi.org/10.1200/PO-24-00908>. Acesso em: 4 jul. 2025.

YE, Jiarui. TANG, Hao. **Multimodal Large Language Models for Medicine: A Comprehensive Survey.** *Arxiv*, 2025. Disponível em: <https://arxiv.org/html/2504.21051v1>. Acesso em: 27 jul. 2025.

YI, Xin; WALIA, Ekta; BABYN, Paul. **Generative adversarial network in medical imaging: a review.** *Medical Image Analysis*, v. 58, p. 101552, 2019. Disponível em: <https://www.doi.org/10.1016/j.media.2019.101552>. Acesso em: 7 abr. 2024.

YEUNG, Karen. **'Hypernudge': Big Data as a mode of regulation by design.** *Information, Communication & Society*, 2017. Disponível em: https://www.researchgate.net/publication/303479231_'Hypernudge'_Big_Data_as_a_mode_of_regulation_by_design. Acesso em: 7 abr. 2024.

YUSTE, Rafael; GOERING, Sara *et al.* **Four ethical priorities for neurotechnologies and AI.** *Nature*, Londres, 2017. Disponível em: <https://doi.org/10.1038/551159a>. Acesso em: 23 nov. 2024.

YUSTE, Rafael, *et al.* **"It's Time for Neuro-Rights: New Human Rights for the Age of Neurotechnology."** *Horizons: Journal of International Relations and Sustainable Development*, no. 18, 2021, pp. 154–65. *JSTOR*, <https://www.jstor.org/stable/48614119>. Acesso em: 31 jul. 2025.

YUSTE, Rafael, *et al.* **International Human Rights Protection Gaps in the Age of Neurotechnology.** *Neurorights Foundation*, 2022. Disponível em: <https://ntc.columbia.edu/wp-content/uploads/2022/05/NeurorightsFoundationPUBLICAnalysis5.6.22.pdf>. Acesso em: 31 jul. 2025.

YUSTE, Rafael. **Advocating for neurodata privacy and neurotechnology regulation.** *Nature Protocols*, 2023. Disponível em: <https://doi.org/10.1038/s41596-023-00873-0>. Acesso em: 5 jun. 2024.

VILLANI, Cédric; RONDEPIERRE, Bertrand. **Artificial Intelligence and Tomorrow's Health.** In *Healthcare and Artificial Intelligence*. Cham: Springer, 2020. Disponível em: <https://link.springer.com/book/10.1007/978-3-030-32161-1>. Acesso em: 15 jun. 2024.

WAISBERG, Ethan; ONG, Joshua; LEE, Andrew G. **The potential to restore vision with Neuralink's "Blindsight" neural interface technology.** *Pan-American Journal of Ophthalmology*, v. 6, n. 3, p. 84, 2024. Disponível em: https://journals.lww.com/pajo/fulltext/2024/05310/the_potential_to_restore_vision_with_neuralink_s.5.aspx. Acesso em: 22 fev. 2025.

WINKLER, Julia K.; BLUM, Andreas *et al.* **Assessment of Diagnostic Performance of Dermatologists Cooperating With a Convolutional Neural Network in a Prospective Clinical Study: Human With Machine.** *JAMA Dermatology*, 2023. Disponível em: <https://jamanetwork.com/journals/jamadermatology/fullarticle/2804568>. Acesso em: 15 mar. 2024.

WINSTON, Patrick H. **Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.** In: BUCHANAN, B. G.; SHORTLIFFE, E. H. (Ed.). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, MA: Addison-Wesley, 1984. (Capítulo 36, p. 671–697). Disponível em: <https://people.dbmi.columbia.edu/~ehs7001/Buchanan-Shortliffe-1984/Chapter-36.pdf>. Acesso em: 15 jun. 2024.

ZHAO, Wayne Xim *et al.* **A Survey of Large Language Models**, 2023. Disponível em: <https://arxiv.org/pdf/2303.18223>. Acesso em: 10 mai. 2024.

ZHANG, Hongyu *et al.* **Brain-computer interfaces: the innovative key to unlocking neurological conditions.** *International Journal of Surgery*, Londres, v. 110, p. 5745-5762, 2024. Disponível em: <https://doi.org/10.1097/JS9.0000000000002022>. Acesso em: 17 fev. 2025.

ZHOU, S. Kevin; Greenspan, Hayit *et al.* **A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises.** *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, 109(5), 820–838, 2021. Disponível em: <https://doi.org/10.1109/JPROC.2021.3054390>. Acesso em: 16 fev. 2024.

ZOU, Ke; ZHIHAO, Chenb *et al.* **A review of uncertainty estimation and its application in medical imaging.** *Meta-Radiology*, 2023. Disponível em: <https://doi.org/10.1016/j.metrad.2023.100003>. Acesso em: 14 set. 2024.

ZUBOFF, Shoshana. **A Era do Capitalismo de Vigilância: A Luta por um Futuro Humano na Nova Fronteira de Poder.** Rio de Janeiro: Editora Intrínseca, eBook Kindle, 2021.

ZUK, Peter; SANCHEZ, Clarissa E.; KOSTICK-QUENET, Kristin *et al.* **Researcher Views on Changes in Personality, Mood, and Behavior in Next-Generation Deep Brain Stimulation.** *AJOB neuroscience*, 2023. Disponível em: <https://doi.org/10.1080/21507740.2022.2048724>. Acesso em: 20 mai. 2025.