**UNIVERSIDADE SANTA CECÍLIA**
**PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO DA SAÚDE**

**KLAUSS CARVALHO DE MALTA**

# A SUPERESTRUTURA PRINCIPIOLÓGICA DA IA: FUNDAMENTOS ÉTICO-JURÍDICOS APLICADOS À SAÚDE

**SANTOS/SP**
**2025**

**KLAUSS CARVALHO DE MALTA**

# A SUPERESTRUTURA PRINCIPIOLÓGICA DA IA: FUNDAMENTOS ÉTICO-JURÍDICOS APLICADOS À SAÚDE

**DISSERTAÇÃO APRESENTADA A UNIVERSIDADE SANTA CECÍLIA COMO PARTE DOS REQUISITOS PARA OBTENÇÃO DE TÍTULO DE MESTRE NO PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO DA SAÚDE**, sob a orientação do Prof. Dr. Marcelo Lamy.

**Santos/SP**

**2025**

Autorizo a reprodução parcial ou total deste trabalho, por qualquer que seja o processo, exclusivamente para fins acadêmicos e científicos.

344.041 Malta, Klauss Carvalho De.
M225s A Superestrutura Principiológica da IA: Fundamentos Ético-Jurídicos aplicados à Saúde / Klauss Carvalho De Malta.
2025.
195 f.

Orientador: Dr. Marcelo Lamy.

Dissertação (Mestrado) - Universidade Santa Cecília, Programa de pós-graduação em PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO DA SAÚDE, Santos, SP, 2025.

1. Inteligência Artificial na Saúde. 2. Regulamentação da IA. 3. Ética Aplicada. 4. Governança Algorítmica. 5. Responsabilidade Civil da IA. I. Lamy, Marcelo . II. A Superestrutura Principiológica da IA: Fundamentos Ético-Jurídicos aplicados à Saúde.

Elaborada via formulário eletrônico, desenvolvido pelo SIBi Unisanta, com dados informados pelo autor(a). Bibliotecário responsável: Ana Maria Racioppi Silveira - CRB/8 - 1890

**KLAUSS CARVALHO DE MALTA**

# A SUPERESTRUTURA PRINCIPIOLÓGICA DA IA: FUNDAMENTOS ÉTICO-JURÍDICOS APLICADOS À SAÚDE

Dissertação apresentada à Universidade Santa Cecília como parte dos requisitos para obtenção do título de Mestre em Direito da Saúde, no Programa de Pós-Graduação em Direito da Saúde, sob orientação do Prof. Dr. Marcelo Lamy.

Aprovada em: _____ de _____ de _____

BANCA EXAMINADORA:

Prof. Dr. Marcelo Lamy
Universidade Santa Cecília – UNISANTA
(Orientador)

Prof.(a) Dr.(a) _____
Instituição: _____

Prof.(a) Dr.(a) _____
Instituição: _____

Prof.(a) Dr.(a) _____
Instituição: _____

**SANTOS/SP**
**2025**

# RESUMO

**Objetivo**: Esta dissertação tem como objetivo analisar os fundamentos éticos e jurídicos que devem orientar a regulação da inteligência artificial na saúde, com base nas diretrizes da Organização Mundial da Saúde. Busca-se identificar princípios que assegurem a autonomia do paciente, a transparência dos sistemas e a responsabilização por danos, propondo parâmetros normativos que conciliem inovação tecnológica com a proteção dos direitos fundamentais no contexto clínico-assistencial. **Metodologia**: A pesquisa adota abordagem qualitativa, com caráter exploratório e analítico, fundamentada em revisão bibliográfica crítica e não sistemática. Foram analisados documentos internacionais, especialmente as diretrizes da OMS, além dos marcos regulatórios desenvolvidos (União Europeia), e em desenvolvimento: Estados Unidos e Brasil. O estudo articula referenciais éticos, jurídicos e técnicos, com foco na construção de um modelo normativo aplicável à inteligência artificial em saúde. **Resultados**: indicam a inexistência de um marco regulatório consolidado que contemple de forma adequada os desafios éticos e jurídicos da inteligência artificial na saúde. Evidenciaram-se lacunas normativas relacionadas à transparência, à responsabilidade civil e à proteção da autonomia do paciente. A análise comparada permitiu identificar diretrizes internacionais emergentes e apontar parâmetros estruturantes para a formulação de um modelo regulatório nacional coerente com os direitos fundamentais. **Discussão**: evidencia-se que a incorporação da inteligência artificial na saúde demanda um arranjo regulatório específico, capaz de equilibrar inovação tecnológica com garantias fundamentais. A dissertação demonstra que princípios como transparência, explicabilidade, equidade e responsabilização devem integrar a estrutura normativa, orientando a atuação dos sistemas algorítmicos em conformidade com a dignidade da pessoa humana e a ética assistencial. **Conclusão**: Conclui-se que a construção de um marco regulatório da inteligência artificial na saúde deve partir de fundamentos ético-jurídicos sólidos, com foco na proteção da autonomia do paciente e na responsabilização adequada dos agentes envolvidos. A dissertação propõe parâmetros normativos que visam assegurar um uso seguro, transparente e justo da IA, contribuindo para o avanço do debate acadêmico e institucional sobre governança algorítmica em contextos assistenciais.



**Palavras Chave:** Inteligência Artificial na Saúde. Regulamentação da IA. Ética Aplicada. Governança Algorítmica. Responsabilidade Civil da IA.

## ABSTRATCT

**Objective**: This dissertation aims to analyze the ethical and legal foundations that should guide the regulation of artificial intelligence in health, based on the guidelines of the World Health Organization. It seeks to identify principles that ensure patient autonomy, transparency of systems and accountability for damages, proposing normative parameters that reconcile technological innovation with the protection of fundamental rights in the clinical and care context. **Methodology**: The research adopts a qualitative, exploratory and analytical approach, based on a critical and non-systematic literature review. International documents were analyzed, especially WHO guidelines, as well as developed (European Union) and developing (United States and Brazil) regulatory frameworks. The study articulates ethical, legal and technical references, with a focus on building a normative model applicable to artificial intelligence in health. **Results**: indicate the lack of a consolidated regulatory framework that adequately addresses the ethical and legal challenges of artificial intelligence in health. Regulatory gaps were found in relation to transparency, civil liability and the protection of patient autonomy. The comparative analysis made it possible to identify emerging international guidelines and point to structuring parameters for the formulation of a national regulatory model that is consistent with fundamental rights. **Discussion**: It is clear that the incorporation of artificial intelligence into healthcare requires a specific regulatory arrangement, capable of balancing technological innovation with fundamental guarantees. The dissertation shows that principles such as transparency, explainability, fairness and accountability should be part of the regulatory framework, guiding the actions of algorithmic systems in accordance with the dignity of the human person and the ethics of care. **Conclusion**: The conclusion is that the construction of a regulatory framework for artificial intelligence in healthcare must be based on solid ethical and legal foundations, with a focus on protecting patient autonomy and ensuring that the agents involved are held properly accountable. The dissertation proposes normative parameters that aim to ensure the safe, transparent and fair use of AI, contributing to the advancement of the academic and institutional debate on algorithmic governance in healthcare contexts.



**Key words:** Artificial Intelligence in Health. AI regulation. Applied Ethics. Algorithmic Governance. AI Liability.

## LISTA DE SIGLAS

AILD - *AI Liability Directive* (Diretiva de Responsabilidade da Inteligência Artificial)
ANPD - Autoridade Nacional de Proteção de Dados
AUROC - *Area Under the Receiver Operating Characteristic Curve* (Área sob a Curva ROC)
CDSS - *Clinical Decision Support System* (Sistema de Suporte à Decisão Clínica)
CE - *Conformité Européenne* (Conformidade Europeia)
CNN - *Convolutional Neural Network* (Rede Neural Convolucional)
Curva ROC - Receiver Operating Characteristic (Característica de operação do receptor)
Curva PR - *Precision-Recall* (Curva de Precisão e Revocação ou Sensibilidade)
DL - *Deep Learning* (Aprendizado Profundo)
EHDS - *European Health Data Space* (Espaço Europeu de Dados de Saúde)
EHR - Electronic Health Record (Registro Eletrônico de Saúde)
EMA - Entropia Modulada da Acurácia
EMA - European Medicines Agency (Agência Europeia de Medicamentos)
FDA - Food and Drug Administration (Administração de Alimentos e Medicamentos dos Estados Unidos)
GDPR - *General Data Protection Regulation* (Regulamento Geral de Proteção de Dados)
GHTF - *Global Harmonization Task Force* (Força-Tarefa de Harmonização Globa)
GMLP - *Good Machine Learning Practices* (Boas Práticas de Aprendizado de Máquina)
GOFAI - *Good Old-Fashioned Artificial Intelligence* (A boa e velha inteligência artificial)
HHS - *Department of Health and Human Services* (Departamento de Saúde e Serviços Humanos dos Estados Unidos)
HIPAA - *Health Insurance Portability and Accountability Act* (Lei de Portabilidade e Responsabilidade do Seguro de Saúde)
HITECH - *Health Information Technology for Economic and Clinical Health Act* (Lei de Tecnologia da Informação em Saúde para a Eficiência Econômica e Clínica)
PHI - *Protected Health Information* (Informação de Saúde Protegida)
IA - Inteligência Artificial
IBC - International Bioethics Committee (Comitê Internacional de Bioética)
IMDRF - *International Medical Device Regulators Forum* (Fórum Internacional de Reguladores de Dispositivos Médicos)
ISO - *International Organization for Standardization* (Organização Internacional de Padronização)

LIME - *Local Interpretable Model-agnostic Explanations* (Explicações Localmente Interpretáveis Independentes do Modelo)
LLM - *Large Language Model* (Modelo de Linguagem de Grande Escala)
LGPD - Lei Geral de Proteção de Dados
MDCG - *Medical Device Coordination Group* (Grupo de Coordenação de Dispositivos Médicos)
MDD - *Medical Devices Directive* (Diretiva de Dispositivos Médicos)
MDR - *Medical Device Regulation* (Regulamento de Dispositivos Médicos)
MINIMAR - *MINimum Information for Medical AI Reporting* (Informação Mínima para Relato de IA Médica)
ML - *Machine Learning* (Aprendizado de Máquina)
MLMD - *Machine Learning-enabled Medical Devices* (Dispositivos Médicos com Capacitação por Aprendizado de Máquina)
MLP - *Multilayer Perceptron* (Perceptron Multicamadas)
NIT - Normalized Information Transfer (Fator Normalizado de Transferência de Informação)
NNH - *Number Needed to Harm* (Número Necessário para Causar Dano)
NNT - *Number Needed to Treat* (Número Necessário para Tratar)
NLP - *Natural Language Processing* (Processamento de Linguagem Natural)
OCDE - Organização para a Cooperação e Desenvolvimento Econômico
OMS - Organização Mundial da Saúde
PDP - *Parallel Distributed Processing* (Processamento Paralelo Distribuído)
PDP Law - *Personal Data Protection Law* (Lei de Proteção de Dados Pessoais da Indonésia)
PIPL - *Personal Information Protection Law* (Lei de Proteção de Informações Pessoais da China)
PL - Projeto de Lei
PMA - *Premarket Approval* (Aprovação Pré-Comercialização)
PROBAST-AI - *Prediction model Risk Of Bias Assessment Tool – Artificial Intelligence extension* (Ferramenta de Avaliação de Risco de Viés em Modelos de Predição – Extensão para Inteligência Artificial)
QMS - *Quality Management System* (Sistema de Gestão da Qualidade)
RNN - *Recurrent Neural Network* (Rede Neural Recorrente)
SHAP - *SHapley Additive exPlanations* (Explicações Aditivas de Shapley)
SaMD - *Software as a Medical Device* (Software como Dispositivo Médico)
SVM - *Support Vector Machines* (Máquinas de vetores de suporte)

TRIPOD-AI - *Transparent Reporting of a multivariable prediction model for Individual* Prognosis Or Diagnosis (Relato Transparente de Modelos de Predição Multivariada para Diagnóstico ou Prognóstico – Extensão para Inteligência Artificial)
UDI - *Unique Device Identification* (Identificação Única de Dispositivo)
UE - União Europeia
XAI - Explainable Artificial Intelligence (Inteligência Artificial Explicável)

## LISTA DE ABREVIATURAS

art. - artigo

et al. - et alii (e outros)

n.º - número

p. - página

v. - volume

## SUMÁRIO


**INTRODUÇÃO ..... 14**

**1 FUNDAMENTOS E CLASSIFICAÇÕES DA IA ..... 20**

1.1 CONCEITUAÇÃO E ABORDAGENS DA INTELIGÊNCIA ARTIFICIAL ..... 20

*1.1.1 Delimitação conceitual da inteligência artificial ..... 20*

*1.1.2 Estruturas e mecanismos de inteligência artificial ..... 21*

*1.1.3 Evolução histórica e complexidades na definição da inteligência artificial ..... 23*

1.2 CATEGORIAS E CLASSIFICAÇÕES DA INTELIGÊNCIA ARTIFICIAL ..... 25

*1.2.1 Classificações da IA: especialização e generalização cognitiva ..... 25*

*1.2.2 Paradigmas estruturais da IA: abordagens simbólicas e conexionistas ..... 26*

1.3 FUNDAMENTOS E APLICAÇÕES DO APRENDIZADO DE MÁQUINA ..... 29

*1.3.1 Conceito e caracterização do aprendizado de máquina ..... 29*

*1.3.2 Mecanismos operacionais do Aprendizado de Máquina ..... 30*

*1.3.3 Estruturas conceituais do Aprendizado de Máquina ..... 31*

*1.3.4 Desafios na Aprendizagem Contínua em Sistemas de Inteligência Artificial . 34*

*1.3.5 Avanços e limitações das redes neurais profundas ..... 35*

1.4 SÍNTESE REFLEXIVA ..... 38

**2 IMPLICAÇÕES DA IA NA SAÚDE ..... 40**

2.1 TRAJETÓRIA HISTÓRICA E CONTEMPORÂNEA ..... 41

*2.1.1 Apoio à tomada de decisão ..... 42*

*2.1.2 Aprendizado de Máquina na Saúde ..... 43*

*2.1.3 Análise automatizada de imagens médicas ..... 46*

*2.1.4 Abordagens individualizadas no cuidado à Saúde com apoio da IA ..... 48*

*2.1.5 Modelagem antecipatória e inferência de tendências em Saúde ..... 49*

*2.1.6 Criação de fármacos com apoio da inteligência artificial ..... 51*

*2.1.7 Gestão de serviços de Saúde e cuidados médicos ..... 53*

*2.1.8 Novos usos da inteligência artificial ..... 54*

2.2 ESTRUTURAÇÃO DE SOLUÇÕES INTELIGENTES PARA A SAÚDE ..... 55

*2.2.1 Delimitação do desafio e especificação das funções da IA ..... 55*

*2.2.2 Princípios do treinamento supervisionado em inteligência artificial ..... 56*

*2.2.3 Etapas de coleta, curadoria e tratamento inicial dos dados ..... 57*

*2.2.4 Estratégias e processos para o desenvolvimento de modelos de IA ..... 58*

*2.2.5 Critérios e métodos para análise de desempenho de modelos ..... 60*

2.3 OBSTÁCULOS DA IA NA SAÚDE ..... 64

*2.3.1 Critérios na seleção de dados ..... 65*

*2.3.2 Aplicabilidade prática ..... 67*

*2.3.3 A resiliência da IA ... 69*
*2.3.4 Atenção aos cuidados com a IA ... 71*
*2.3.5 A importância da transparência e compreensão dos processos ... 72*
*2.3.6 Dinâmica entre profissionais de saúde e sistemas inteligentes ... 73*
*2.3.7 Supervisão permanente e atualização em tempo real ... 76*
2.4 SÍNTESE REFLEXIVA ... 77

**3 BASES NORMATIVAS PARA O USO DA IA ... 80**

3.1 PRINCÍPIOS MORAIS E REGULATÓRIOS ... 80
*3.1.1 Concretização de um direito humano essencial ... 81*
*3.1.2 Mecanismos normativos na promoção do direito à saúde ... 82*
*3.1.3 A importância da regulação da inteligência artificial ... 84*
*3.1.4 Proteção da dignidade humana nas aplicações de IA ... 86*
*3.1.5 Diretrizes para o uso responsável da inteligência artificial ... 87*
3.2 DIRETRIZES ÉTICAS APLICADAS À INTELIGÊNCIA ARTIFICIAL NA SAÚDE ... 89
*3.2.1 Preservação da autonomia do paciente ... 91*
*3.2.2 Entre a proteção do indivíduo e o interesse público ... 93*
*3.2.3 Clareza algorítmica e compreensão acessível no uso da IA ... 95*
*3.2.4 Mecanismos de accountability na aplicação da IA ... 97*
*3.2.5 Justiça distributiva e acesso equitativo às inovações ... 98*
*3.2.6 Ética na adaptabilidade tecnológica e a sustentabilidade sistêmica ... 100*
3.3 GESTÃO E RESPONSABILIDADE SOBRE DADOS NA SAÚDE ... 101
*3.3.1 Noções fundamentais ... 102*
*3.3.2 A proteção de dados pessoais como pilar regulatório ... 103*
*3.3.3 Proteção da intimidade e dos dados pessoais sensíveis ... 104*
*3.3.4 Modelo regulatório dos Estados Unidos para a inteligência artificial ... 107*
*3.3.5 Modelo regulatório adotado pela União Europeia ... 110*
*3.3.6 Direito à privacidade e proteção de dados ... 113*
*3.3.7 Instrumentos voltados à salvaguarda das informações em saúde ... 117*
3.4 SÍNTESE REFLEXIVA ... 121

**4 REFERENCIAIS PARA O MARCO REGULATÓRIO DA IA ... 123**

4.1 GARANTIA DE CONFIABILIDADE NAS APLICAÇÕES COM IA ... 123
*4.1.1 Soluções digitais na Saúde com função de dispositivo médico ... 124*
*4.1.2 Tecnologias médicas com Aprendizado de Máquina integrado ... 127*
*4.1.3 Modelo regulatório dos Estados Unidos ... 129*
*4.1.4 Estrutura normativa europeia para a IA ... 131*
4.2 CLAREZA NOS PROCESSOS E DECISÕES ALGORÍTMICAS ... 134
*4.2.1 Entre a transparência operacional e a clareza de decisão ... 135*
*4.2.2 Explicabilidade como direito fundamental na era da IA ... 136*
*4.2.3 Instrumentos de explicabilidade aplicados à IA na Saúde ... 138*

*4.2.4 A Complexidade da explicação algorítmica na prática clínica ..... 140*
*4.2.5 Registros técnicos como instrumentos de prestação de contas ..... 142*
4.3 *ACCOUNTABILITY* EM SISTEMAS DE IA ..... 143
*4.3.1 Fundamentos jurídicos da responsabilidade civil aplicada à IA ..... 143*
*4.3.2 Definição de responsáveis pelo uso da IA na área da Saúde ..... 146*
*4.3.3 Instrumentos viáveis de imputação de responsabilidade ..... 148*
*4.3.4 Novas abordagens para a responsabilidade civil diante da IA ..... 150*
4.4 SÍNTESE REFLEXIVA ..... 152
**5 RUMOS DA INTELIGÊNCIA ARTIFICIAL ..... 154**
5.1 ENTRE DIRETRIZES GERAIS E NORMAS ESPECÍFICAS ..... 155
5.2 ESTRUTURAS DE TESTAGEM JURÍDICA PARA TECNOLOGIAS EMERGENTES ..... 157
5.3 RUMOS DA IA NA REALIDADE BRASILEIRA ..... 159
*5.3.1 Inteligência Artificial na Proposta de Reforma do Código Civil. ..... 162*
*5.3.2 Plano Brasileiro de Inteligência Artificial (PBIA) ..... 162*
*5.3.3 Observatório Brasileiro de Inteligência Artificial (OBIA) ..... 163*
5.4 SÍNTESE REFLEXIVA. ..... 164
**CONSIDERAÇÕES FINAIS ..... 167**
**REFERÊNCIAS ..... 172**
**APÊNDICE ..... 199**

## Introdução

Nas últimas décadas, a inteligência artificial (IA) deixou de ser uma expectativa futurista para se tornar uma presença concreta e influente nas diversas esferas da vida humana, incluindo áreas tão sensíveis quanto o cuidado na saúde. A ascensão de modelos algorítmicos capazes de processar grandes volumes de dados, reconhecer padrões complexos e tomar decisões com graus variados de autonomia representa uma das transformações mais profundas já observadas no campo tecnológico. No entanto, essa revolução, marcada por inovações técnicas de elevada complexidade, levanta uma série de desafios de ordem ética, jurídica e institucional, sobretudo quando aplicada a contextos que envolvem direitos fundamentais, como o direito à saúde, à privacidade e à dignidade da pessoa humana.

A aplicação da inteligência artificial nos sistemas de saúde tem promovido avanços significativos em áreas como diagnóstico por imagem, modelagem preditiva, desenvolvimento de fármacos e gestão hospitalar. Contudo, tais benefícios não são isentos de riscos. A opacidade dos processos decisórios algorítmicos, a possibilidade de discriminação por vieses de treinamento, a indefinição de critérios de responsabilidade civil e a violação da autonomia do paciente são apenas alguns dos pontos críticos que demandam reflexão rigorosa e regulação adequada. Em resposta a esse conteúdo, organismos internacionais, como a Organização Mundial da Saúde (OMS), vêm desempenhando papel central na formulação de diretrizes ético-normativas voltadas à orientação do uso responsável e humanizado dessas tecnologias.

Diante desse panorama, a presente dissertação tem como objetivo analisar a superestrutura principiológica que deve orientar a utilização da inteligência artificial no campo da saúde, com base nas diretrizes formuladas pela OMS. Trata-se de uma investigação jurídica, de natureza interdisciplinar, que parte da premissa de que o desenvolvimento tecnológico, por mais promissor que seja, não pode se descolar dos fundamentos axiológicos que estruturam o Estado Democrático de Direito. Assim, busca-se compreender de que forma princípios como equidade, autonomia, explicabilidade, justiça distributiva e responsabilidade podem e devem ser integrados aos modelos

normativos que regulam a atuação da inteligência artificial em ambientes clínico-assistenciais.

Para isso, o trabalho está estruturado em cinco capítulos, além desta introdução e das considerações finais. O Capítulo 1 apresenta os fundamentos conceituais e as classificações da inteligência artificial, situando as abordagens técnicas que compõem essa estrutura e delimitando os principais paradigmas operacionais em jogo. O Capítulo 2 analisa as implicações práticas da IA na saúde, destacando suas aplicações clínicas, administrativas e tecnocientíficas. O Capítulo 3 propõe um exame detalhado das diretrizes normativas e éticas da OMS, enquanto o Capítulo 4 aprofunda os referenciais jurídicos e os mecanismos regulatórios aplicáveis à governança da IA na saúde. Por fim, o Capítulo 5 apresenta as novas tendências regulatórias a título global, discutindo a transição entre modelos principiológicos e normas jurídicas vinculantes, com ênfase em mecanismos de testagem, como as *sandboxes* regulatórias, e marcos legais emergentes.

A relevância do tema decorre não apenas da novidade tecnológica ou do interesse acadêmico, mas, sobretudo, da urgência em construir uma inteligência artificial que seja compatível com os valores fundamentais da vida humana. A saúde, como direito social constitucionalmente garantido, não pode ser colocada a reboque de soluções automatizadas desprovidas de controle, legitimidade e orientação ética. Neste contexto, investigar a estrutura principiológica que deve guiar a IA em saúde significa contribuir para a consolidação de um modelo de governança tecnológica que priorize o bem comum, a justiça social e a proteção de direitos.

A hipótese central que guia esta pesquisa é a de que os princípios éticos e normativos estabelecidos por organismos internacionais, como a OMS, constituem alicerces fundamentais para a construção de uma regulação eficiente, transparente e humanizada da inteligência artificial em saúde. A metodologia adotada é qualitativa, de base exploratória e analítica, com ênfase na revisão bibliográfica de fontes especializadas, documentos internacionais e referenciais jurídicos comparados.

O desenvolvimento da presente dissertação fundamenta-se em um marco teórico de natureza interdisciplinar, articulando conceitos e categorias oriundos da ética, do direito e da governança tecnológica. Tal abordagem é

essencial diante da complexidade envolvida na regulação da inteligência artificial aplicada à saúde, campo em que os avanços tecnológicos desafiam diretamente os limites tradicionais das garantias jurídicas e éticas.

O primeiro eixo teórico está centrado na ética da inteligência artificial, com destaque para os princípios estabelecidos por organismos internacionais, em especial pela Organização Mundial da Saúde (OMS), que propõe um conjunto de diretrizes voltadas à proteção da autonomia, à promoção da equidade, à transparência algorítmica e à responsabilização. Esses princípios fornecem uma base normativa fundamental para orientar o uso ético da IA em ambientes clínicos e assistenciais.

O segundo eixo envolve os fundamentos jurídico-constitucionais, com ênfase no direito à saúde, na dignidade da pessoa humana, na autodeterminação informacional e na responsabilidade civil. A análise parte do reconhecimento de que a regulação da IA deve estar ancorada em garantias fundamentais, exigindo que a inovação tecnológica se submeta a balizas normativas compatíveis com a proteção dos direitos individuais e coletivos no âmbito da saúde.

Por fim, o terceiro eixo do marco teórico refere-se à governança algorítmica e à regulação da inovação tecnológica, com foco nos modelos regulatórios emergentes e nos instrumentos normativos como o AI Act da União Europeia, os projetos de lei brasileiros e os mecanismos de regulação experimental, como as *sandboxes* regulatórias. Nesse contexto, são exploradas abordagens baseadas em risco, diferenciações entre sistemas de IA explicáveis e interpretáveis, e estratégias regulatórias orientadas por princípios de proporcionalidade, justiça algorítmica e segurança jurídica.

Assim, o marco teórico adotado nesta dissertação fornece a sustentação conceitual necessária para uma análise crítica e propositiva da inteligência artificial na saúde, contribuindo para a formulação de um paradigma normativo que concilie o avanço tecnológico com a proteção dos direitos fundamentais.

Espera-se, com este estudo, oferecer uma contribuição crítica à reflexão sobre os rumos da inteligência artificial na saúde, enfatizando a necessidade de se instituírem marcos normativos que articulem inovação tecnológica, segurança institucional e respeito irrestrito à dignidade humana.

## Metodologia

Partindo de uma perspectiva interdisciplinar e jurídico-analítica, este trabalho buscou aprofundar a compreensão dos fundamentos éticos e normativos que devem orientar o uso da inteligência artificial na área da saúde, com especial atenção às diretrizes emitidas pela Organização Mundial da Saúde (OMS).

A pesquisa adotou uma abordagem qualitativa, de natureza exploratória e analítica, fundamentada em revisão bibliográfica crítica. Foram utilizados como principais referenciais documentos internacionais, em especial os relatórios e orientações técnicas da OMS, além de artigos científicos especializados, normas jurídicas nacionais e estrangeiras, bem como marcos regulatórios emergentes no campo da inteligência artificial aplicada à saúde.

O levantamento teórico foi realizado com base em publicações indexadas em bases de dados acadêmicas de relevância como Scielo, mais de 100 artigos (106 documentos científicos) do PBi (Portal de Busca integrada) da Universidade de São Paulo (USP), além de documentos encontrados nos portais JSTOR, Pubmed, Springer Nature Link, SSRN, Nature, Nature Madicine, NIH, Science Direct, Revista de Direito Sanitário, The New England Journal of Medicine, HHS Public Access, frontiers, Journal of Medical Internet Research, The Journal of Law, Medicine & Ethics e repositórios institucionais de organismos internacionais, priorizando fontes publicadas nos últimos cinco anos.

A análise normativa comparada abrangeu os modelos regulatórios dos Estados Unidos, da União Europeia e do Brasil, buscando identificar convergências e tensões entre os princípios universais de direitos humanos e as inovações regulatórias específicas para sistemas algorítmicos.

Foram utilizados como descritores temáticos os seguintes termos: ética da inteligência artificial, governança algorítmica, autonomia do paciente, explicabilidade, responsabilidade civil, privacidade e proteção de dados em saúde, todos conectados à centralidade da dignidade da pessoa humana nos contextos clínico-assistenciais.

A dissertação foi estruturada para permitir uma análise sistemática e crítica das implicações da IA em saúde, com ênfase na integração entre princípios axiológicos, normatividade internacional e exigências regulatórias contemporâneas. O método adotado possibilita não apenas a identificação dos desafios ético-jurídicos atuais, mas também a formulação de premissas para a construção de marcos legais capazes de garantir inovação com justiça e segurança institucional.

**Objetivos da pesquisa**

O presente trabalho teve como objetivo identificar e explicar os fundamentos ético-jurídicos que devem orientar a formulação de um paradigma regulatório aplicável ao uso da inteligência artificial na saúde, em conformidade com os princípios da Organização Mundial da Saúde (OMS) e com foco na preservação da autonomia do paciente.

Para alcançar esse propósito, definiram-se os seguintes objetivos específicos:

a) Identificar os princípios éticos e jurídicos fundamentais para a regulação da inteligência artificial em ambientes clínico-assistenciais;
b) Analisar diretrizes internacionais, recomendações técnicas e documentos normativos que tratam da governança algorítmica na saúde;
c) Avaliar as implicações da utilização de sistemas de inteligência artificial sobre a autonomia, a privacidade e a responsabilização civil no âmbito da prática médica;
d) Propor parâmetros normativos que sirvam de base para a construção de um marco regulatório nacional alinhado à inovação tecnológica, à proteção dos direitos do paciente e à segurança jurídica nas relações em saúde.

E assim, ofereceu-se uma contribuição teórica ao debate contemporâneo sobre a regulação da inteligência artificial na saúde, articulando os referenciais normativos internacionais com os desafios concretos

enfrentados nos contextos clínicos e assistenciais. Pretendeu-se, assim, fundamentar a necessidade de um marco regulatório específico que não apenas promova a inovação tecnológica, mas que também assegure a centralidade da dignidade humana e a efetivação dos direitos fundamentais no uso de sistemas algorítmicos em saúde.

## Aderência à linha e ao projeto de pesquisa do programa

A presente dissertação está inserida na linha de pesquisa **Estado e Planejamento Social da Saúde**, à medida que analisou a necessidade de regulação ética e jurídica da inteligência artificial no campo da saúde, com foco na proteção dos direitos fundamentais, especialmente o direito à saúde em sua dimensão coletiva. A proposta normativa aqui desenvolvida considerou os deveres estatais no enfrentamento dos riscos tecnológicos, com ênfase na responsabilidade civil, na transparência dos processos algorítmicos e na preservação da autonomia dos pacientes. Nesse sentido, o trabalho dialogou com as desigualdades estruturais e com os desafios de acesso equitativo aos serviços de saúde mediados por sistemas de IA, evidenciando o risco de novas formas de exclusão.

Além disso, a dissertação aderiu ao projeto de pesquisa **Direitos Humanos, Desenvolvimento Sustentável e a Tutela Jurídica da Saúde**, ao investigar criticamente os fundamentos e as lacunas da proteção jurídica à saúde diante da incorporação de tecnologias emergentes. O estudo adotou uma abordagem interdisciplinar que articula os princípios dos direitos humanos à análise normativa da efetividade dos instrumentos de tutela da saúde — tanto no plano das políticas públicas quanto no âmbito regulatório e judicial. A proposta buscou não apenas compreender a evolução do direito à saúde diante das transformações tecnológicas, mas também propôs mecanismos inovadores e alinhados aos compromissos internacionais de desenvolvimento sustentável e de governança ética, contribuindo para a progressiva realização do direito à saúde no contexto digital.

# 1 Fundamentos e classificações da IA

## 1.1 Conceituação e abordagens da inteligência artificial

### 1.1.1 Delimitação conceitual da inteligência artificial

A expressão "inteligência artificial" surgiu formalmente em uma proposta elaborada em 1955 por um grupo de pesquisadores liderado por John McCarthy para um projeto de pesquisa realizado no verão seguinte no Dartmouth College. Na proposta original, o grupo sugeria que aspectos diversos da inteligência poderiam ser descritos tão precisamente a ponto de permitir que máquinas fossem programadas para simular tais características. Dessa forma, o objetivo inicial da inteligência artificial estava diretamente associado à tentativa de compreender e replicar artificialmente processos considerados típicos da inteligência humana, tais como aprender, raciocinar, abstrair e solucionar problemas antes restritos ao domínio humano (McCarthy et al., 1955, p.2).

Embora o conceito formal de inteligência artificial tenha se consolidado posteriormente, o campo já apresentava questionamentos fundamentais antes desse período. Em 1950, Alan Turing publicou um artigo influente onde sugeriu substituir a tradicional questão "As máquinas podem pensar?" por uma abordagem prática e testável: a questão deveria ser se "uma máquina pode ser indistinguível linguisticamente de um humano". Essa proposta deu origem ao famoso "Jogo da Imitação", conhecido hoje como "Teste de Turing", utilizado para avaliar se uma máquina consegue exibir comportamentos inteligentes indistinguíveis daqueles manifestados por seres humanos (Turing, 1950, p. 433).

Mais de cinquenta anos após participar da criação do campo conhecido como inteligência artificial (IA), John McCarthy ainda definia essa área como o desenvolvimento científico e tecnológico de máquinas e programas dotados de inteligência. Apesar disso, ele reconhecia que essa definição não era consensual, sobretudo pela ausência de uma caracterização universalmente aceita do que seja inteligência, conceito crucial nessa discussão. Segundo McCarthy, inteligência não é uma capacidade homogênea, mas se manifesta em diferentes formas e

intensidades em humanos, animais e máquinas. Sob essa perspectiva, a inteligência na IA é entendida como a dimensão computacional que permite às máquinas alcançar objetivos específicos no mundo (McCarthy, 2007, p. 2).

A inteligência artificial pode ser compreendida como a disciplina dedicada ao desenvolvimento de sistemas artificiais capazes de desempenhar ações que, em circunstâncias apropriadas, se assemelham às realizadas por seres humanos ou animais. Mais precisamente, trata-se de um campo interdisciplinar cujo propósito central é criar agentes ou máquinas que sejam capazes de simular aspectos específicos da inteligência humana, tais como a capacidade de aprender com experiências, adaptar-se a novas circunstâncias e desempenhar atividades que normalmente exigem habilidades cognitivas humanas (Russell et al., 2010, p. 34).

### 1.1.2 Estruturas e mecanismos de inteligência artificial

Para Russell et al. (2010, p. 2) a inteligência artificial pode ser analisada sob quatro abordagens fundamentais, que a descrevem como: 1) sistemas que reproduzem o pensamento humano; 2) sistemas que realizam raciocínio lógico-racional; 3) sistemas que agem à semelhança dos humanos; e 4) sistemas que atuam de forma racional, buscando resultados ótimos dentro de contextos específicos.

Uma das formas pelas quais a inteligência artificial pode ser abordada é a construção de sistemas que procuram simular o pensamento humano. Essa perspectiva implica desenvolver programas e máquinas capazes de reproduzir processos cognitivos específicos que ocorrem em seres humanos ao resolver problemas e tomar decisões, especialmente através de manipulação simbólica e análise heurística. Nessa abordagem, o foco é colocado não apenas no resultado das ações executadas pelos sistemas, mas especialmente nos métodos internos utilizados para alcançar tais resultados, refletindo padrões semelhantes ao pensamento humano (Newell et al., 1976, p. 116).

A modelagem cognitiva constitui uma abordagem fundamental na área da inteligência artificial, que busca reproduzir em sistemas artificiais os

processos mentais complexos característicos da cognição humana. Essa perspectiva parte do pressuposto de que compreender como o cérebro opera pode oferecer caminhos promissores para o desenvolvimento de sistemas inteligentes mais robustos. Nesse contexto, o estudo das estruturas e mecanismos neurais é considerado essencial, pois fornece tanto inspiração para o desenvolvimento de novos algoritmos quanto validação para técnicas já existentes, aproximando o funcionamento artificial do biológico (Hassabis et al., 2017, p. 245).

A perspectiva conhecida como “IA lógica” está centrada no desenvolvimento de sistemas computacionais capazes de raciocinar com base em estruturas formais e regras de inferência. Nessa abordagem, a máquina é programada para empregar lógica dedutiva com o intuito de tomar decisões e resolver problemas, utilizando conjuntos de premissas previamente estabelecidas. A lógica formal serve como o alicerce para representar o conhecimento e derivar conclusões válidas, permitindo que esses sistemas simulem a capacidade humana de inferência racional em contextos específicos (Genesereth, 2012, p. 9).

Dentro da abordagem lógica da inteligência artificial, destaca-se o uso de representações simbólicas estruturadas em linguagens formais, como o cálculo de predicados, com o objetivo de organizar o conhecimento. Essa representação permite que sistemas computacionais manipulem tais símbolos por meio de algoritmos, inferindo conclusões a partir de um conjunto de premissas. A proposta de McCarthy ao descrever o programa “advice taker” ilustra como essas operações lógicas poderiam ser utilizadas para gerar conhecimento adicional, tomando ações com base em deduções feitas a partir de informações previamente armazenadas (McCarthy, 1959, p. 1-2).

A terceira abordagem no estudo da inteligência artificial propõe a construção de sistemas que se comportam de maneira semelhante aos seres humanos. Essa linha de pesquisa ganhou notoriedade com a proposta de Alan Turing, que substituiu a questão "máquinas podem pensar?" por um experimento prático conhecido como o "jogo da imitação". Nesse teste, a máquina busca interagir linguisticamente de forma tão convincente que se torne indistinguível de um ser humano, servindo como critério empírico para

avaliar comportamentos inteligentes em contextos específicos (Turing, 1950, p. 434).

As técnicas utilizadas para aproximar o comportamento das máquinas ao humano incluem áreas como o processamento de linguagem natural, a visão computacional e a robótica. Essas ferramentas permitem que os sistemas artificiais interpretem dados sensoriais, compreendam e gerem linguagem, além de interagirem com o ambiente físico de maneira coordenada. Na geração automatizada de discursos, por exemplo, essas tecnologias viabilizam formas complexas de comunicação e organização textual, aproximando as interações computacionais das humanas (Hovy, 1993, p. 342).

A quarta vertente da inteligência artificial está centrada na construção de sistemas que atuam de maneira racional, frequentemente denominada como abordagem do “agente racional”. Nessa perspectiva, a máquina deve ser capaz de tomar decisões que maximizem os resultados esperados com base na sua percepção do ambiente e na definição de metas. Essa concepção não se limita apenas à inferência lógica, mas inclui qualquer mecanismo que leve a uma ação apropriada frente às circunstâncias e incertezas do ambiente em que está inserida (Russell et al., 2010, p. 5).

Tal concepção compreende vertentes como a elaboração estratégica automatizada, os mecanismos de aprendizado por reforço e as dinâmicas decisórias em cenários marcados por incertezas.

### 1.1.3 Evolução histórica e complexidades na definição da inteligência artificial

A evolução da inteligência artificial tem sido caracterizada por ciclos alternados de otimismo e frustração. Nos chamados "verões da IA", o entusiasmo com descobertas tecnológicas impulsiona investimentos e expectativas. Em contrapartida, os "invernos da IA" representam períodos de desilusão, em que promessas não cumpridas levam à retração de recursos e apoio institucional. Apesar dessas oscilações, a área voltou a ganhar grande destaque na década de 2010, impulsionada pelos avanços em aprendizado profundo, um subtipo do aprendizado de máquina baseado em redes neurais,

que possibilitou conquistas relevantes, como o desempenho de sistemas em aplicações robustas da visão computacional e o reconhecimento de padrões (Haenlein et al., 2019, p. 4).

O crescimento acelerado da inteligência artificial nas últimas décadas está diretamente vinculado à ampliação da capacidade tecnológica de armazenar, transmitir e processar informações em larga escala. Esses avanços possibilitaram o surgimento do que se denomina *Big Data*, um conjunto de práticas e tecnologias que lidam com volumes massivos de dados, fundamentais para alimentar algoritmos de aprendizado de máquina e demais técnicas de IA. Estima-se que, entre 1986 e 2007, a capacidade mundial de computação em dispositivos de uso geral aumentou em média 58% ao ano, enquanto a de telecomunicação e armazenamento evoluiu a taxas anuais de 28% e 23%, respectivamente, consolidando uma infraestrutura essencial para o desenvolvimento contemporâneo da IA (Hilbert et al., 2011, p. 60-61).

Atualmente, o conceito de inteligência artificial é utilizado para designar um domínio vasto e multidisciplinar que reúne conhecimentos provenientes de diferentes áreas do saber, como ciência da computação, filosofia, psicologia, linguística, neurociência e matemática. Essa diversidade reflete a complexidade do fenômeno que se busca compreender e reproduzir por meio de sistemas computacionais. Ainda assim, a definição precisa da IA continua sendo objeto de divergência entre estudiosos e organizações, dado que o termo "inteligência", por si só, já carrega múltiplas interpretações. Em razão disso, instituições como a Comissão Europeia têm buscado elaborar definições operacionais que auxiliem na padronização conceitual e no desenvolvimento de políticas públicas sobre o tema (UE, 2019, p. 1-2).

A definição de inteligência artificial continua sendo um dos maiores desafios enfrentados por autoridades regulatórias em todo o mundo. Embora haja consenso sobre a necessidade de estabelecer parâmetros conceituais para regular esse campo, ainda não se chegou a uma formulação universalmente aceita. Organismos internacionais como a OCDE e o Conselho da Europa adotaram definições amplas, enquanto outras instituições, como a UNESCO, optaram por não se comprometer com uma definição precisa. Essa diversidade reflete a complexidade do tema e a velocidade com que as tecnologias evoluem. No âmbito europeu, o debate em torno da definição de IA

foi um tema central para a formulação do AI Act, já que uma abordagem muito restrita poderia ser rapidamente superada, enquanto uma definição excessivamente ampla poderia gerar insegurança jurídica. (Renda et al., 2023, p. 1-2).

## 1.2 Categorias e classificações da inteligência artificial

A inteligência artificial costuma ser categorizada, de modo predominante, com base em dois critérios: o nível de semelhança com as capacidades cognitivas humanas e as funções que é capaz de desempenhar.

### 1.2.1 Classificações da IA: especialização e generalização cognitiva

Considerando o nível de aproximação com as capacidades intelectivas humanas, a inteligência artificial costuma ser segmentada em duas categorias centrais: a inteligência artificial Estreita, também denominada Restrita, e a Inteligência Artificial Geral.

A chamada Inteligência Artificial Estreita, também conhecida como IA Fraca, refere-se a sistemas desenvolvidos para executar funções específicas com elevado grau de eficiência, como reconhecimento de voz, análise de imagens ou algoritmos de recomendação. Apesar de superarem os seres humanos em determinadas tarefas, esses sistemas não possuem consciência ou capacidade de aprendizado fora de seus domínios delimitados, operando apenas conforme os parâmetros e objetivos estabelecidos para a função para a qual foram projetados (Kurzweil, 2005, p. 260).

Até o momento, todas as modalidades de inteligência artificial concebidas inserem-se nessa tipologia.

A Inteligência Artificial Geral, também chamada de IA Forte, diz respeito à concepção de sistemas artificiais com capacidades cognitivas amplas,

comparáveis ou até superiores às humanas. Diferentemente da IA Estreita, que atua de forma limitada e especializada, a IA Geral busca desenvolver agentes capazes de aprender, raciocinar, planejar e adaptar-se a diferentes contextos, mesmo aqueles não previstos em seu treinamento inicial. Esses sistemas devem ser capazes de generalizar conhecimentos adquiridos em um domínio para aplicá-los em outros, demonstrando flexibilidade semelhante à inteligência humana, incluindo elementos como consciência, autocontrole, raciocínio abstrato e tomada de decisão em ambientes complexos (Goertzel, 2014, p. 1-2).

A concepção de uma inteligência artificial geral que seja capaz de replicar, com autonomia, todas as capacidades intelectuais humanas, como a compreensão, o raciocínio, o aprendizado e a tomada de decisões conscientes, ainda permanece no campo da hipótese. Embora essa ideia alimente parte da literatura futurista e esteja presente em agendas de pesquisa ambiciosas, não há, até o momento, qualquer sistema que alcance esse nível de complexidade cognitiva. Autores como Fjelland (2020) argumentam, inclusive, que esse objetivo pode ser inalcançável em termos fundamentais, dada a impossibilidade de uma máquina, desprovida de corpo, cultura e vivência no mundo, desenvolver uma cognição verdadeiramente humana (Fjelland, 2020, p. 1-2).

Apesar dos avanços notáveis na inteligência artificial estreita, ainda estamos distantes da concretização da IA geral, cuja complexidade cognitiva permanece como um ideal teórico.

### 1.2.2 Paradigmas estruturais da IA: abordagens simbólicas e conexionistas

Outra distinção relevante no campo da inteligência artificial decorre de sua funcionalidade, a partir da qual se delineiam duas vertentes principais: a Inteligência Artificial Simbólica (Symbolic AI) e a Inteligência Artificial Conexionista (Connectionist AI).

A Inteligência Artificial Simbólica, também chamada de IA clássica ou baseada em regras, fundamenta-se no uso de estruturas simbólicas e na manipulação lógica dessas representações por meio de algoritmos e

heurísticas. Essa abordagem, conhecida também pela sigla GOFAI (*Good Old-Fashioned Artificial Intelligence*), utiliza sistemas formais para modelar o raciocínio, estruturando o conhecimento em símbolos que podem ser processados e combinados para simular comportamentos inteligentes. A hipótese do sistema físico de símbolos, formulada por Newell e Simon (1976), sustenta que toda ação inteligente pode ser realizada por um sistema computacional que armazena e manipula símbolos segundo regras precisas (Newell et al., 1976, p. 116-117).

A IA simbólica parte da ideia de que símbolos podem ser usados como unidades fundamentais do raciocínio, assim como os humanos recorrem a representações abstratas para entender e interagir com o mundo. Essa abordagem funciona muito bem em cenários bem definidos, com regras claras e estáveis, como em jogos de tabuleiro, por exemplo. No entanto, quando aplicada a contextos mais imprevisíveis ou abertos, ela se mostra limitada: o mundo real é demasiadamente variável para que todas as situações possam ser previstas e codificadas em regras explícitas (Santoro et al., 2021, p. 1-3).

A chamada Inteligência Artificial Conexionista, também conhecida por redes neurais artificiais ou processamento paralelo distribuído (PDP), busca inspiração direta no funcionamento do cérebro humano. Diferente da IA simbólica, que opera com regras explícitas e processamento sequencial, essa abordagem estrutura-se em redes compostas por unidades interconectadas, análogas a neurônios, que processam informações de forma simultânea e distribuída. O processamento ocorre através de ajustes graduais nos pesos das conexões, com base em algoritmos de aprendizado, o que confere às redes uma grande capacidade de adaptação e generalização (Rumelhart et al., 1986, p. 77-109).

Diferentemente da abordagem simbólica, que depende fortemente de regras explícitas programadas por humanos, os modelos conexionistas ganham sofisticação à medida que são expostos a grandes volumes de dados, aprendendo por meio da identificação de padrões e correlações. A inteligência, nessa perspectiva, é concebida como um produto emergente da interação entre numerosas unidades de processamento simples. Em vez de armazenar informações em locais fixos, esses sistemas distribuem o conhecimento através da rede, codificando-o na força das conexões entre os nós. Essa

dinâmica distribuída permite uma representação mais flexível e adaptativa, especialmente útil em tarefas complexas e pouco estruturadas (Smolensky, 1988, p. 1-2).

Tais sistemas operam com base em vastas redes compostas por inúmeros elementos de processamento, ou neurônios artificiais, cada qual estruturado por unidades ponderadas, uma função de ativação e um canal de saída. As unidades ponderadas são responsáveis por atribuir níveis de importância e direção às informações recebidas, enquanto a função de ativação (ou de transferência) tem a finalidade de sintetizar múltiplas entradas em um único valor de saída, traduzindo o comportamento coletivo da rede em respostas computacionais.

O debate sobre qual abordagem de inteligência artificial seria mais adequada, a simbólica ou a conexionista, acompanha a história do campo há décadas. Embora ambas tenham alcançado avanços significativos em suas respectivas áreas, continuam apresentando limitações consideráveis quando aplicadas isoladamente. A IA simbólica se destaca em contextos estruturados, mas mostra rigidez diante de situações não previstas. Já os sistemas conexionistas apresentam maior flexibilidade e adaptabilidade, porém carecem de mecanismos eficazes de raciocínio abstrato e representação de conhecimento. Minsky (1991) argumenta que essa dicotomia é, em si, um equívoco: em vez de optar por um único caminho, o futuro da IA está na integração dessas abordagens, combinando suas virtudes e superando suas deficiências (Minsky, 1991, p. 35-36).

A distinção entre as abordagens simbólica e conexionista revela que nenhuma delas, isoladamente, é capaz de abarcar toda a complexidade dos desafios enfrentados pela inteligência artificial. Enquanto a primeira se destaca em contextos estruturados, a segunda oferece maior adaptabilidade em ambientes incertos. Nesse cenário, a tendência atual aponta para modelos híbridos que integrem os pontos fortes de ambas, visando ampliar a eficácia e a aplicabilidade de ambos.

## 1.3 Fundamentos e aplicações do aprendizado de máquina

O principal vetor do avanço contemporâneo da inteligência artificial é representado pelo domínio do aprendizado de máquina (*Machine Learning*). Embora os termos 'inteligência artificial' (IA) e '*Machine Learning*' (ML) sejam frequentemente utilizados como sinônimos no discurso corrente, é importante destacar que ML constitui, na realidade, uma subcategoria da IA, notadamente a de maior aplicação prática na atualidade e com projeções mais promissoras para o futuro. Diferentemente da IA, cuja definição se ancora em um conceito amplo e impreciso de 'inteligência', o aprendizado de máquina possui contornos conceituais mais delimitados e tecnicamente rigorosos.

### 1.3.1 Conceito e caracterização do aprendizado de máquina

O aprendizado de máquina pode ser entendido como uma área que combina métodos estatísticos e computacionais para que sistemas sejam capazes de melhorar seu desempenho com base em experiências anteriores, sem necessidade de reprogramação explícita. A partir da análise de dados, essas técnicas extraem padrões, constroem modelos e fazem previsões quando confrontadas com novas informações, processo conhecido como generalização. Em essência, algoritmos de aprendizado de máquina transformam conjuntos de dados e exemplos em conhecimento operacional, frequentemente na forma de modelos matemáticos que são utilizados para realizar tarefas que, de outro modo, seriam complexas demais para serem programadas manualmente (Mitchell, 1997, p. 2-3).

Sob a ótica da ciência da computação, o aprendizado de máquina distingue-se da programação convencional, usualmente denominada “sistema baseado em regras” (rule-based system), na medida em que esta última se fundamenta na codificação manual de funções específicas que correlacionam entradas predefinidas a saídas esperadas. De modo simplificado, toda

interação computacional pode ser modelada como uma relação funcional expressa pela equação y = f(x), em que x representa a entrada, f a função de processamento, e y o resultado obtido. Nesse paradigma tradicional, o processamento consiste na aplicação de regras explicitamente formuladas com o objetivo de transformar insumos em respostas específicas, conforme determinado pelo programador.

### 1.3.2 Mecanismos operacionais do Aprendizado de Máquina

De forma geral, o aprendizado de máquina (*Machine Learning*) pode ser entendido como o desenvolvimento de métodos que capacitam sistemas computacionais a extrair padrões de dados e a realizar inferências ou previsões sobre novos dados, sem que tenham sido explicitamente programados para tal. O processo envolve a construção de funções que mapeiam entradas para saídas, com o intuito de generalizar esse mapeamento para exemplos ainda não vistos. Os modelos resultantes são, geralmente, expressos de forma matemática e ajustados por meio de técnicas estatísticas, permitindo que o sistema se torne progressivamente mais preciso com base na experiência adquirida a partir dos dados observados (Hastie et al., 2009, p. 30-31).

Em outras palavras, embora ainda seja necessário desenvolver um programa, a finalidade no contexto do aprendizado de máquina difere substancialmente da programação tradicional: em vez de definir previamente uma função de mapeamento, busca-se permitir que o sistema a descubra com precisão a partir dos dados. Assim, apesar de frequentemente ser atribuído um caráter antropomórfico ao termo “aprendizado”, trata-se, na verdade, de um processo essencialmente matemático, fundamentado em algoritmos que inferem parâmetros a partir de conjuntos de dados, orientando a construção e a posterior calibragem do modelo computacional.

Dentro do universo da inteligência artificial, o aprendizado de máquina (*Machine Learning*) representa um campo específico voltado à criação de modelos e algoritmos que permitem que sistemas computacionais evoluam com base na experiência adquirida a partir de dados. Essa abordagem, que

começou a ganhar forma ainda na década de 1970, teve seu desenvolvimento impulsionado nas décadas seguintes, especialmente com o avanço de métodos estatísticos e probabilísticos aplicados à computação. Ao longo desse percurso, buscou-se superar as deficiências dos primeiros sistemas de IA, muitas vezes limitados e inflexíveis, direcionando esforços para soluções mais especializadas e adaptáveis ao tratamento de dados complexos (Bishop, 2006, p. 147).

O avanço significativo da área de aprendizado de máquina (*Machine Learning*) nas últimas décadas pode ser atribuído, em grande parte, à combinação entre a crescente capacidade computacional e a vasta disponibilidade de dados. A partir dos anos 2000, esse cenário impulsionou uma expansão sem precedentes no desenvolvimento e aplicação de algoritmos de ML, tornando-os ferramentas centrais para decisões baseadas em evidências nos mais diversos setores, como saúde, indústria, comércio e ciências sociais (Jordan et al., 2015, p. 255), desta forma, a operação do aprendizado de máquina baseia-se em princípios matemáticos e estatísticos que permitem a construção de modelos preditivos capazes de aprender com os dados. Esse paradigma representa um afastamento da lógica determinística tradicional, viabilizando sistemas adaptativos e autônomos.

### 1.3.3 Estruturas conceituais do Aprendizado de Máquina

No estágio atual de desenvolvimento, o aprendizado de máquina configura-se como um subcampo da inteligência artificial que se vale, primordialmente, de métodos estatísticos para construir modelos preditivos baseados na identificação de padrões complexos nos dados. Essa abordagem não depende de instruções explícitas para operar, mas sim da capacidade dos algoritmos de extrair relações subjacentes entre variáveis, a partir de grandes volumes de informações.

Dentro desse espectro, destacam-se quatro orientações teórico-práticas distintas, que se diferenciam pela forma como os dados são utilizados no processo de treinamento. O aprendizado supervisionado, por exemplo, é

caracterizado pelo uso de conjuntos de dados previamente rotulados, nos quais as respostas corretas já são conhecidas. Nessa configuração, o sistema é conduzido a reconhecer e internalizar a correlação entre variáveis de entrada e suas respectivas saídas, com o intuito de generalizar esse conhecimento para dados ainda não analisados. Durante o treinamento, o modelo ajusta seus parâmetros com base em exemplos representativos, aprendendo a inferir, de maneira autônoma, as saídas mais prováveis para novas informações que venham a ser processadas no futuro.

Ainda que frequentemente personificado pelo uso do termo “aprendizado”, esse processo não envolve compreensão no sentido humano, mas sim cálculos matemáticos (a matemática das redes neurais[1]) complexos que resultam em uma capacidade preditiva refinada, orientada estatisticamente.

No contexto do aprendizado supervisionado, as tarefas são tradicionalmente classificadas em dois tipos principais: regressão e classificação. A regressão está voltada para a predição de variáveis com valores contínuos, como por exemplo a temperatura ou o preço de um bem, enquanto a classificação lida com a previsão de categorias distintas, como rótulos de diagnóstico médico ou tipos de objetos em imagens. Essa distinção fundamental orienta o uso de diferentes modelos e técnicas conforme a natureza da variável alvo (Murphy, 2012, p. 2).

Diferentemente do aprendizado supervisionado, o aprendizado não supervisionado (unsupervised learning) não conta com rótulos ou saídas previamente conhecidas durante a fase de treinamento. Nesse contexto, os algoritmos trabalham apenas com os dados de entrada, buscando identificar padrões ocultos ou estruturas subjacentes sem uma orientação explícita. Tais técnicas são amplamente utilizadas para extrair informações latentes dos dados, sendo aplicadas em tarefas como agrupamento de instâncias semelhantes (*clustering*)[2] e redução de dimensionalidade, cuja finalidade é

[1] A matemática das redes neurais. Disponível em: https://www.youtube.com/watch?v=qZ9xuPcoWSA. Acesso em: 05 de abril de 2025.

[2] Agrupamento no aprendizado de máquina: 5 Algoritmos de agrupamento essenciais. Disponível em: https://www.datacamp.com/pt/blog/clustering-in-machine-learning-5-essential-clustering-algorithms#:~:text=Clustering%20%C3%A9%20o%20processo%20de,objetos%20de%20qualquer%20outro%20grupo. Acesso em: 05 de abril de 2025.

simplificar a representação dos dados preservando ao máximo suas características essenciais (Hastie et al., 2009, p. 486).

O aprendizado semi-supervisionado[3] (semi-supervised learning) representa uma abordagem intermediária entre os modelos supervisionado e não supervisionado, sendo frequentemente classificado como uma forma de "supervisão fraca". Essa técnica se mostra especialmente útil em cenários onde há uma grande disponibilidade de dados não rotulados e apenas uma pequena amostra de dados rotulados. Nessa configuração, o algoritmo recebe informações parciais de supervisão e busca, a partir delas, estender a classificação aos dados não rotulados ou agrupá-los com base em padrões compartilhados. O crescimento exponencial de dados disponíveis, associado aos altos custos de rotulagem manual, tem impulsionado o uso de estratégias semi-supervisionadas, sobretudo no treinamento de modelos de linguagem (Chapelle et al., 2006, p. 2).

O aprendizado por reforço[4] (*reinforcement learning*) configura-se como uma abordagem de *Machine Learning* em que um agente é capaz de aprender comportamentos por meio da interação contínua com o ambiente em que está inserido. Nessa dinâmica, o agente executa ações e observa as consequências, as quais resultam em recompensas que podem ser positivas ou negativas, a depender dos resultados alcançados. Trata-se de um processo iterativo baseado em tentativa e erro, cujo objetivo é a construção de uma política ideal, isto é, uma estratégia que permita ao agente selecionar, diante de diferentes situações, as ações mais adequadas para maximizar os ganhos cumulativos ao longo do tempo. Essa modalidade é especialmente relevante em sistemas de apoio à decisão baseados em inteligência artificial, nos quais não há necessariamente uma base de conhecimento pré-programada, mas sim uma adaptação contínua fundamentada em padrões estatísticos e aprendizado autônomo (Sutton et al., 2020, p. 2).

As metodologias comumente classificadas como aprendizado de máquina tradicional correspondem, em sua maioria, a algoritmos empregados em cenários de aprendizado supervisionado. Esse conjunto abrange uma

[3] Desvendando o aprendizado semi-supervisionado. Disponível em: https://www.youtube.com/watch?v=3rU1H1q0vKI. Acesso em: 05 de abril de 2025.

[4] Aprendizado por Reforço: conceitos, aplicações e desafios. Disponível em: https://www.youtube.com/watch?v=HZ2CoPuMUB0. Acesso em: 05 de abril de 2024.

variedade de técnicas consolidadas, entre as quais se destacam a regressão linear, a regressão logística[5], as máquinas de vetores de suporte (SVM - *support vector machines*)[6], as árvores de decisão e as florestas aleatórias (random forests), dentre outras. Tais métodos recebem a designação de "tradicionais" justamente para diferenciá-los das abordagens mais recentes, representadas pelas técnicas de aprendizado profundo (*Deep Learning*), uma vertente emergente e cada vez mais proeminente no campo do aprendizado de máquina.

### 1.3.4 Desafios na Aprendizagem Contínua em Sistemas de Inteligência Artificial

No âmbito do aprendizado de máquina, sobressai-se a abordagem denominada “aprendizado contínuo” (continual learning), também referida como “aprendizado incremental” (incremental learning), que vem ganhando relevância crescente nas pesquisas atuais.

O **aprendizado contínuo** (ou contínuo incremental) baseia-se na capacidade de um modelo adaptar-se progressivamente a partir de fluxos de dados não estacionários, ou seja, cujas distribuições mudam ao longo do tempo. Essa abordagem busca mitigar o chamado “esquecimento catastrófico”, fenômeno comum em redes neurais profundas, no qual o modelo tende a esquecer conhecimentos adquiridos anteriormente quando exposto a novos dados ou tarefas. O aprendizado contínuo ocorre de maneira incremental, atualizando o modelo conforme novas informações se tornam disponíveis. Três formas distintas de aprendizagem incremental são identificadas: aprendizado incremental por tarefa (task-incremental), em que diferentes tarefas são aprendidas sequencialmente; aprendizado incremental por domínio (domain-incremental), em que o modelo aprende a resolver a mesma tarefa em

[5] A Regressão Logística - Algoritmos de Aprendizado de Máquinas. Disponível em: https://www.youtube.com/watch?v=WKvrpEa-Hl0. Acesso em: 05 de abril de 2025.

[6] SVM - Support Vector Machines: Fundamentos e prática. Disponíem em: https://www.youtube.com/watch?v=b8nP9g0p8X4. Acesso em: 05 de abril de 2025.

contextos variados; e aprendizado incremental por classe (class-incremental), em que o sistema aprende progressivamente a discriminar entre um número crescente de classes (Van de ven et al., 2022, p. 1185).

A área de aprendizado contínuo busca aproximar as capacidades da inteligência artificial das habilidades naturais de aquisição de conhecimento ao longo da vida, como observado em humanos e animais. Inspirado por mecanismos neurocognitivos, esse campo investiga como modelos computacionais podem aprender progressivamente, refinando e transferindo conhecimento sem perder informações anteriormente adquiridas. Um dos principais desafios é manter o equilíbrio entre a plasticidade, a capacidade de adaptação a novos dados, e a estabilidade, a preservação do conhecimento consolidado. Esse dilema, conhecido como stability-plasticity, exige que os modelos aprendam de forma eficiente sem incorrer em esquecimentos catastróficos, assegurando simultaneamente a retenção e a incorporação contínua de novos saberes (Parisi et al., 2019, p. 55).

Uma modalidade especialmente relevante dentro do aprendizado contínuo é o chamado “aprendizado on-line” (online learning), concebido para contextos em que os dados são gerados ao longo do tempo. Nessa abordagem, os dados chegam sequencialmente e são imediatamente utilizados para atualizar o modelo, permitindo ajustes contínuos e em tempo real. Isso contrasta com o modelo tradicional de “aprendizado em lote” (batch learning), no qual todo o conjunto de dados de treinamento é disponibilizado antecipadamente, exigindo reprocessamentos periódicos. Ao possibilitar atualizações instantâneas com base em novas instâncias, o aprendizado on-line se mostra uma alternativa eficiente e escalável para aplicações em larga escala e de alta velocidade, especialmente em cenários com fluxos contínuos de dados (Hoi et al., 2021, p. 2).

### 1.3.5 Avanços e limitações das redes neurais profundas

A expressão "aprendizado profundo" (*Deep Learning*) designa uma categoria específica dentro do universo do aprendizado de máquina, notável

por sua arquitetura diferenciada e pela aptidão em captar e representar funções altamente complexas a partir dos dados.

Em contraste com os métodos tradicionais de *Machine Learning*, nos quais o número de parâmetros costuma estar restrito à quantidade de variáveis de entrada, os modelos de *Deep Learning* (DL) superam essa limitação ao organizar os parâmetros em múltiplas camadas hierárquicas. Essa arquitetura permite que as representações dos dados sejam aprendidas de forma progressiva e abstrata, com cada camada processando os dados recebidos da anterior e fornecendo entradas mais complexas para a próxima etapa do modelo. Essa estrutura em profundidade potencializa interações mais sofisticadas entre os parâmetros e as características dos dados, promovendo a aprendizagem de funções altamente não lineares e complexas, o que amplia substancialmente a capacidade de representação dos sistemas computacionais baseados em redes neurais profundas (Goodfellow et al., 2016, p. 6-7).

A concepção do *Deep Learning* tem origem no paradigma conexionista da inteligência artificial, inspirado pela estrutura e pelo funcionamento do cérebro humano. Os modelos de *Deep Learning* utilizam redes neurais artificiais compostas por múltiplas camadas ocultas, em que cada camada atua como uma unidade de processamento responsável por transformar, de forma não linear, os dados recebidos da camada anterior. O número de camadas ocultas pode variar conforme a arquitetura e a complexidade do problema a ser solucionado, redes com mais de duas camadas são consideradas profundas. A composição sucessiva dessas transformações possibilita à rede neural aprender representações de dados altamente abstratas e complexas, ampliando sua capacidade de generalização e desempenho em tarefas como reconhecimento de imagem, processamento de linguagem natural e predição de padrões em grandes volumes de dados[7] (Lecun et al., 2015, p. 436).

As redes neurais artificiais aprendem por meio de erros, semelhantemente a outros sistemas baseados em aprendizado supervisionado, ajustando iterativamente os pesos das conexões entre os neurônios a fim de melhorar seu desempenho ao longo do tempo. Esse processo de ajuste é

[7] Uma Visão sobre a Mineração de Grandes Volumes de Dados - Prof. Dr. Murilo Coelho Naldi. Disponível em: https://www.youtube.com/watch?v=TIm_kECd_po. Acesso em: 05 de abril de 2025.

conduzido por um algoritmo conhecido como “*backpropagation*”, que consiste na propagação dos erros da saída de volta para as camadas anteriores da rede. Por meio desse mecanismo, cada conexão é ajustada proporcionalmente à sua contribuição para o erro em relação ao valor esperado, com o objetivo de minimizar a função de custo. O aprendizado ocorre de forma contínua, com a rede sendo refinada a cada ciclo de treinamento com base nos erros cometidos anteriormente, o que a torna progressivamente mais eficiente em suas previsões (Rumelhart et al., 1986, p. 533).

As redes neurais recorrentes (RNNs), por sua natureza projetada para lidar com sequências temporais, são amplamente utilizadas em tarefas como tradução automática e reconhecimento de fala. Já as redes convolucionais (CNNs), com suas camadas organizadas para identificar padrões espaciais, são especialmente eficazes em aplicações de reconhecimento de imagem e vídeo. As perceptrons multicamadas (MLPs) representam uma evolução das redes neurais sem realimentação (feedforward), agregando profundidade e capacidade de aprender representações mais complexas. Esses modelos, integrados à abordagem de *Deep Learning*, têm sido centrais para avanços notáveis no processamento de linguagem natural, visão computacional e outras áreas que dependem da análise de grandes volumes de dados (LeCun et al., 2015, p. 436).

Recentemente, a arquitetura *Transformer* tem se destacado de maneira significativa, sobretudo nas aplicações voltadas ao processamento de linguagem natural. Diferentemente das redes recorrentes ou *convolucionais* tradicionais, o Transformer propõe uma estrutura inteiramente baseada em mecanismos de atenção, eliminando a necessidade de recorrência ou *convoluções*. Essa abordagem permite a modelagem de dependências entre elementos da sequência sem considerar sua distância, o que, além de aumentar a eficiência paralela, melhora consideravelmente o desempenho em tarefas como tradução automática e modelagem de linguagem. O *Transformer* não só supera os modelos anteriores em termos de qualidade, como também reduz substancialmente o tempo de treinamento (Vaswani et al., 2017, p. 1).

Apesar de ainda se encontrarem em contínuo processo de aprimoramento, as redes neurais artificiais já demonstraram elevado potencial em uma diversidade de aplicações práticas. À medida que sua evolução

tecnológica avança, é plausível supor que se tornem progressivamente mais eficientes e ampliem significativamente sua adoção em distintos contextos.

## 1.4 Síntese reflexiva

O panorama apresentado nesse capítulo evidencia a complexidade estrutural e conceitual que caracteriza a inteligência artificial. A pluralidade de abordagens, simbólica, conexionista, supervisionada, não supervisionada e contínua, revela um ecossistema dinâmico, em constante transformação, cuja evolução é impulsionada tanto por inovações tecnológicas quanto por desafios epistemológicos.

As distinções entre modelos clássicos e contemporâneos, bem como entre arquiteturas de processamento baseadas em regras e redes neurais profundas, indicam que a inteligência artificial transcende uma concepção única e linear, configurando-se como um domínio multifacetado. Essa multiplicidade se reflete na variedade de classificações funcionais e estruturais que buscam ordenar um campo em permanente expansão e redefinição.

Compreender tais fundamentos é essencial não apenas para o delineamento técnico das aplicações da IA, mas também para a formulação de modelos regulatórios e princípios éticos que serão discutidos a seguir.

Ao reconhecer os limites e potencialidades dos diferentes paradigmas, torna-se possível contextualizar o papel da inteligência artificial em setores sensíveis, como a saúde, onde suas implicações exigem atenção especial às dimensões humanas, jurídicas e sociais.

Para sistematizar os principais conceitos discutidos até aqui, apresenta-se a seguir a:

**Quadro 1 – Síntese dos fundamentos e classificações da inteligência artificial**

| | |
|---|---|
| **Inteligência Artificial** | Campo interdisciplinar voltado à simulação computacional de capacidades cognitivas humanas. |
| **IA Estreita** | Sistemas especializados em tarefas específicas, sem capacidade de generalização. |
| **IA Geral** | Modelo teórico de IA com raciocínio e adaptação |

| | |
|---|---|
| | comparáveis à inteligência humana. |
| **IA Simbólica** | Abordagem baseada em regras explícitas e manipulação de símbolos para tomada de decisão. |
| **IA Conexionista** | Modelo inspirado no cérebro humano, baseado em redes neurais e aprendizado por padrões. |
| **Aprendizado de Máquina** | Técnica da IA que permite aos sistemas aprenderem com dados e aprimorarem seu desempenho. |
| **Aprendizado Supervisionado** | Modelo de aprendizado que utiliza dados rotulados para ensinar relações entre entradas e saídas. |
| **Aprendizado Não supervisionado** | Técnica que identifica padrões ocultos em dados não rotulados. |
| **Aprendizado Semi-supervisionado** | Combina dados rotulados e não rotulados para otimizar o aprendizado com poucos exemplos. |
| **Aprendizado por reforço** | Método baseado em recompensas e penalidades para orientar decisões em ambientes dinâmicos. |
| **Racionalidade** | Princípio que orienta sistemas a maximizarem resultados com base em metas e contexto. |
| **Modelagem cognitiva** | Simulação de processos mentais humanos em sistemas computacionais. |
| ***Big Data*** | Conjuntos massivos de dados que viabilizam o funcionamento e evolução dos sistemas de IA. |
| **Paradigmas históricos** | Alternância entre períodos de otimismo (verões) e estagnação (invernos) na evolução da IA. |
| **Generalização** | Capacidade do modelo de aplicar conhecimentos aprendidos a novos dados ou contextos. |

**Fonte:** Elaborado pelo autor

Conforme sintetizado no Quadro 1, os principais conceitos abordados no capítulo sobre fundamentos e classificações da inteligência artificial envolvem diferentes abordagens técnicas, categorias tipológicas e mecanismos de aprendizado.

# 2 IMPLICAÇÕES DA IA NA SAÚDE

A evolução das tecnologias de inteligência artificial (IA) na saúde também se estende a áreas inovadoras e com grande potencial de impacto. No campo da cirurgia robótica, sistemas dotados de aprendizado de máquina vêm sendo treinados para realizar procedimentos com elevada precisão, utilizando abordagens minimamente invasivas e aperfeiçoando continuamente seu desempenho com base em dados de operações anteriores. Simultaneamente, dispositivos vestíveis (wearables), integrados a algoritmos de reconhecimento de fala e análise preditiva, têm sido aplicados no contexto da telemedicina para monitoramento remoto de sinais vitais, apoio a pacientes em reabilitação e intervenções personalizadas em tempo real. Essas tecnologias, ainda em expansão, apontam para uma ampliação do papel da IA na promoção de cuidados mais precisos, acessíveis e centrados no paciente.

A teleconsulta, modalidade de atendimento remoto facilitada por sistemas de IA e demais tecnologias de informação, tem se consolidado como ferramenta essencial para ampliar o acesso ao cuidado, especialmente em contextos de pandemia e isolamento social. No entanto, sua regulamentação segue fragmentada, com normas emergenciais que não oferecem diretrizes claras sobre padrões de comunicação, consentimento e segurança dos dados, o que expõe pacientes e médicos ao risco de falhas informacionais e à possibilidade de negligência médica (Calado; Lamy, 2020, p. 115).

Calado e Lamy destacam que, embora a Lei nº 13.989/2020 tenha permitido a teleconsulta em caráter excepcional durante a crise de COVID-19, sua aplicação “diferida - e nunca dispensada - em situações emergenciais” reforça a urgência de uma regulação definitiva, que contemplaria a formalização eletrônica do consentimento informado, a garantia de integridade e sigilo das informações e a definição de protocolos homogêneos de atuação para todo o país (Calado; Lamy, 2020, p. 116).

Essas inovações, ainda em franca expansão, que vão desde cirurgia robótica e wearables para monitoramento remoto até sistemas de prescrição assistida por IA, ampliam o potencial da inteligência artificial para oferecer cuidados mais precisos, acessíveis e centrados no paciente (Roski et al., 2019,

p. 7-9).

## 2.1 Trajetória histórica e contemporânea

A saúde figura entre os primeiros campos em que a inteligência artificial encontrou aplicação prática, sobretudo por seu potencial de apoiar decisões clínicas complexas. Ainda na década de 1950, pesquisadores começaram a conceber os primeiros sistemas voltados à assistência diagnóstica, lançando as bases para os modernos sistemas de suporte à decisão clínica. Tais iniciativas, surgidas nos primórdios da IA, refletiam uma tentativa pioneira de formalizar e automatizar o raciocínio médico, sendo que desde 1959 já se discutiam princípios lógicos e probabilísticos aplicáveis ao diagnóstico médico computacional (Miller, 1994, p. 8).

Desde a década de 1970, começaram a surgir diversos sistemas especializados que utilizam a lógica computacional para apoiar o raciocínio médico em tarefas específicas. Esses sistemas foram concebidos para oferecer suporte a decisões clínicas, como interpretação de exames, recomendação de tratamentos e formulação de hipóteses diagnósticas em contextos clínicos complexos. Tais programas simulavam, em alguma medida, o raciocínio de especialistas humanos, contribuindo em áreas nas quais a tomada de decisão médica exigia um alto nível de conhecimento técnico e análise de múltiplos dados clínicos. A consolidação desses sistemas marcou uma nova fase na informatização da saúde, direcionada à complementação da expertise médica com o uso da inteligência artificial (Szolovits et al., 2018, p. 2).

Tais programas representaram aplicações expressivas da, então dominante, abordagem simbólica da inteligência artificial, ganhando ampla difusão ao longo das décadas de 1980 e 1990, período marcado pela ascensão dos sistemas de apoio à decisão clínica (Clinical Decision Support Systems - CDSS).

### 2.1.1 Apoio à tomada de decisão

Ao invés de substituírem o julgamento clínico, os sistemas de suporte à decisão clínica (CDSS) atuam como ferramentas de auxílio direto ao profissional da saúde. O funcionamento típico desses sistemas baseia-se na comparação entre dados específicos do paciente, como sinais, sintomas e histórico clínico, com um banco de conhecimento médico estruturado. A partir dessa análise, o CDSS oferece recomendações personalizadas, como possíveis diagnósticos ou propostas terapêuticas, a fim de informar e aprimorar a tomada de decisão do profissional de saúde. Em essência, trata-se de um software cuja principal finalidade é potencializar a qualidade e a segurança da atenção clínica, incorporando informações de saúde, diretrizes médicas e características individuais do paciente (Sutton et al., 2020, p. 1).

A inserção dos sistemas de apoio à decisão clínica de primeira geração, concebidos nas décadas de 1980 e 1990, revelou-se bastante limitada no ambiente assistencial. Estruturados com base em modelos determinísticos e sistemas de regras fixas, esses instrumentos evidenciaram diversas fragilidades, entre as quais se destacam os elevados custos operacionais e a constante necessidade de atualizações manuais, que exigiam revisão minuciosa por especialistas humanos. A complexidade inerente à integração de múltiplos saberes médicos, oriundos de diferentes áreas e profissionais, tornava a codificação de interações clínicas sofisticadas uma tarefa árdua e, por vezes, inviável. Ademais, a eficácia desses sistemas permanecia diretamente vinculada à precisão e à completude do conhecimento médico previamente incorporado, o que restringia seu desempenho em situações clínicas mais dinâmicas ou pouco estruturadas (Yu et al., 2018, p. 720).

### 2.1.2 Aprendizado de Máquina na Saúde

A medicina clínica tem historicamente lidado com uma complexidade crescente de informações, desde sinais fisiológicos até dados genéticos e comportamentais. Nesse cenário, a capacidade de transformar grandes volumes de dados em conhecimento prático tornou o *Machine Learning* (ML) uma ferramenta indispensável na prática médica contemporânea. Diferente dos sistemas especialistas que operam com regras pré-definidas, o ML aprende diretamente com os dados, identificando padrões complexos e não lineares que ajudam a prever desfechos clínicos, aperfeiçoar diagnósticos e personalizar tratamentos. Sua recepção entusiástica no campo da saúde decorre de sua habilidade de processar e interpretar volumes massivos de informações que seriam inviáveis à análise humana tradicional (Obermeyer et al., 2016, p. 1216).

Com a rápida expansão do poder computacional e a disponibilidade massiva de dados clínicos digitais a partir dos anos 2000, criou-se um terreno fértil para a ascensão da inteligência artificial na medicina. As técnicas de *Machine Learning* (ML), ao invés de exigirem a codificação explícita de regras para cada tarefa, funcionam a partir do aprendizado direto com exemplos, associando entradas (como sinais clínicos, imagens ou textos) a saídas esperadas (diagnósticos, prognósticos ou decisões terapêuticas). Esse processo permite aos modelos reconhecerem padrões complexos e, muitas vezes, até então desconhecidos, sem intervenção humana para cada decisão, superando limitações de abordagens estatísticas tradicionais (Rajkomar et al., 2019, p. 1348).

As recentes conquistas da inteligência artificial na medicina têm sido impulsionadas, sobretudo, pelos notáveis resultados obtidos com as técnicas de *Deep Learning* (DL) em campos como interpretação de imagens médicas, reconhecimento de padrões e tradução de linguagem. Inicialmente, os esforços em IA voltavam-se à construção de sistemas baseados em lógica simbólica e regras fixas; contudo, foi a abordagem baseada em redes neurais profundas que trouxe avanços decisivos. Esses sistemas aprendem representações complexas a partir de dados brutos, superando muitas vezes especialistas

humanos em tarefas específicas, como na classificação de lesões dermatológicas. O desempenho superior e a versatilidade do DL provocaram uma nova onda de entusiasmo quanto à aplicação da IA na saúde (Hinton, 2018, p. 1101).

O avanço das redes neurais profundas na classificação de imagens se consolidou com um ponto de inflexão em 2012, ano em que o desempenho obtido por uma rede convolucional profunda superou significativamente os métodos anteriores no competitivo desafio ImageNet[8]. O sucesso dessa abordagem não apenas demonstrou a superioridade técnica do *Deep Learning* (DL) para lidar com grandes volumes de dados visuais, mas também impulsionou sua adoção em diversos setores, como medicina, indústria automotiva e segurança. O modelo AlexNet[9], treinado sobre milhões de imagens utilizando GPUs e técnicas como "*dropout*" para evitar *overfitting*[10], marcou o início de uma nova era na inteligência artificial aplicada à visão computacional (Krizhevsky et al., 2012, p. 1).

As especialidades médicas que tradicionalmente dependem de imagens para diagnóstico foram as primeiras a explorar, com sucesso, o potencial do *Deep Learning* (DL) na prática clínica. Dermatologia e radiologia têm se destacado pela integração precoce de algoritmos de redes neurais profundas, que demonstraram desempenho comparável, e por vezes superior, ao de especialistas humanos em tarefas como a detecção de câncer de pele, classificação de lâminas patológicas e interpretação de radiografias. Essa adoção precoce decorre da natureza altamente visual dessas áreas, nas quais a identificação precisa de padrões em imagens é essencial para a decisão médica (Topol, 2019, p. 45).

A rápida expansão da inteligência artificial na medicina, após o sucesso do *Deep Learning* em visão computacional, abriu caminho para outras aplicações igualmente promissoras. Técnicas como o processamento de linguagem natural (NLP) vêm sendo aplicadas à análise de prontuários eletrônicos de saúde, permitindo extrair significados relevantes de registros

[8] https://www.image-net.org/

[9] CNN e Alexnet - apresentando o projeto. Disponível em: https://www.youtube.com/watch?v=_cb0Kd4anTA. Acesso em: 05 de abril de 2025.

[10] How does dropout help to avoid overfitting in neural networks?. Disponível em: https://medium.com/@upendravijay2/how-does-dropout-help-to-avoid-overfitting-in-neural-networks-91b90fd86b20. Acesso em: 05 de abril de 2025.

médicos em texto livre. Simultaneamente, métodos de aprendizado por reforço (reinforcement learning) têm sido explorados para tarefas complexas e dinâmicas, como a cirurgia robótica assistida, na qual agentes aprendem a partir da demonstração de especialistas. Essas abordagens revelam como a IA tem superado os limites iniciais e se espalhado para múltiplas frentes do cuidado em saúde (Esteva et al., 2019, p. 25-26).

Sistemas baseados em *Deep Learning* vêm superando modelos tradicionais ao lidar com a complexidade e o volume dos dados clínicos, especialmente na predição de desfechos relevantes em diversos contextos hospitalares. A aplicação dessas técnicas permitiu, por exemplo, prever com alta acurácia eventos como mortalidade hospitalar, tempo prolongado de internação e reinternações não planejadas, considerando milhões de pontos de dados oriundos de registros eletrônicos de saúde (EHR)[11]. Esse desempenho promissor reforça o potencial da IA em transformar a medicina preditiva, destacando-se sua capacidade de gerar insights personalizados sem a necessidade de seleção manual de variáveis clínicas (Rajkomar et al., 2018, p. 1).

Com a chegada dos anos 2020, o cenário da inteligência artificial em saúde passou a ser cada vez mais influenciado pelo surgimento dos modelos de fundação, especialmente os grandes modelos de linguagem (LLMs)[12]. Esses modelos, inicialmente desenvolvidos para tarefas de linguagem natural, têm sido progressivamente adaptados para aplicações biomédicas, dada sua capacidade de aprender representações complexas a partir de grandes volumes de dados textuais e multimodais. A evolução tecnológica tem sido tão veloz que os limites entre modalidades, como texto, imagem e sinais fisiológicos, vêm sendo progressivamente dissolvidos por arquiteturas como *transformers*[13], capazes de operar sobre diferentes tipos de entrada. Com isso,

[11] Automação de registros eletrônicos de saúde (EHR). Disponível em: https://www.blueprism.com/pt/resources/blog/electronic-health-record-ehr-automation/#:~:text=de%20assist%C3%AAncia%20m%C3%A9dica.-,O%20que%20%C3%A9%20o%20processamento%20de%20EHRs%20(registros%20eletr%C3%B4nicos%20de,m%C3%A9dicos%2C%20hospitais%20e%20assist%C3%AAncia%20m%C3%A9dica. Acesso em: 05 de abril de 2025.

[12] O que é LLM (grandes modelos de linguagem)?. Disponível em: https://www.ibm.com/br-pt/think/topics/large-language-models. Acesso em: 05 de abril de 2025.

[13] O que é um Modelo Transformer?. Disponível em: https://blog.nvidia.com.br/blog/o-que-e-um-modelo-transformer/. Acesso em: 05 de abril de 2025.

o campo da IA em saúde vive uma fase de transformação contínua e acelerada, abrindo caminho para aplicações como gêmeos digitais, assistentes médicos virtuais e ensaios clínicos digitais (Acosta et al., 2022, p. 1773-1774).

A inteligência artificial tem revelado um potencial crescente nos mais variados domínios da saúde e da prática médica. Entre suas aplicações mais relevantes, destacam-se os avanços em diagnósticos por imagem, estratégias de medicina personalizada, análises preditivas de condições clínicas, desenvolvimento de fármacos e otimização de processos administrativos no setor.

### 2.1.3 Análise automatizada de imagens médicas

A aplicação de redes neurais profundas na medicina tem alcançado um nível de sofisticação considerável, particularmente no campo da análise de imagens médicas. Sistemas de *Deep Learning*, sobretudo as redes neurais convolucionais (CNNs)[14], vêm sendo amplamente empregados na interpretação automatizada de exames como tomografias, ressonâncias magnéticas, mamografias e radiografias. Esses modelos, treinados com grandes volumes de imagens rotuladas, são capazes de detectar anomalias com precisão muitas vezes comparável ou superior à de especialistas humanos. O desenvolvimento de arquiteturas mais profundas e a capacidade de aprendizado de características diretamente a partir dos dados consolidaram o DL como uma metodologia de escolha para essas tarefas (Litjens et al., 2017, p. 1).

Apesar do entusiasmo crescente e do elevado número de soluções comerciais disponíveis, os impactos mais profundos da inteligência artificial (IA) na radiologia ainda estão por se concretizar de forma ampla. A especialidade, no entanto, é considerada pioneira na incorporação dessas tecnologias, uma vez que seu fluxo de trabalho digital padronizado facilita a integração de algoritmos. Além disso, radiologistas têm experimentado benefícios em tarefas

[14] O que são redes neurais convolucionais?. Disponível em: https://www.ibm.com/br-pt/think/topics/convolutional-neural-networks. Acesso em: 05 de abril de 2025.

como triagem, detecção e quantificação de imagens. Ainda assim, desafios como a generalização dos modelos, a validação contínua e a colaboração homem-máquina permanecem como barreiras para uma adoção clínica realmente ampla e segura (Rajpurkar et al., 2023, p. 1982-1983).

Entre as especialidades médicas que têm se beneficiado do avanço das técnicas de inteligência artificial, a patologia destaca-se pela aplicação promissora de sistemas de *Deep Learning* no reconhecimento automatizado de padrões histológicos. Utilizando imagens de anatomia patológica digitalizadas, esses sistemas têm sido empregados para classificar tipos de câncer, prever prognósticos e identificar estruturas celulares com precisão significativa. A literatura recente demonstra que, em determinadas tarefas, a combinação entre patologistas e ferramentas baseadas em IA apresenta desempenho superior ao de cada um isoladamente. Tais evidências sugerem que a patologia está entre os campos mais dinâmicos na integração dessas tecnologias, ainda que muitos dos modelos estejam em fase experimental (Harrison et al., 2021, p. 1230-1231).

O uso de inteligência artificial na análise de imagens médicas tem gerado impactos notáveis em especialidades clínicas como oftalmologia e dermatologia. Em particular, sistemas baseados em *Deep Learning* já demonstraram capacidade de realizar triagem e avaliação automatizada de fundus ocular para detecção de retinopatias diabéticas, com sensibilidade e especificidade comparáveis às de médicos oftalmologistas. Ao mesmo tempo, modelos treinados para reconhecimento de lesões cutâneas suspeitas vêm se consolidando como ferramentas auxiliares no rastreamento de câncer de pele. Essas tecnologias, ao lidarem com grandes volumes de imagens, permitem a ampliação do acesso ao diagnóstico em regiões com escassez de especialistas e têm sido progressivamente integradas à prática clínica em centros de saúde ao redor do mundo (Yu et al., 2018, p. 720).

### 2.1.4 Abordagens individualizadas no cuidado à Saúde com apoio da IA

A personalização do cuidado médico tem se beneficiado substancialmente do uso de algoritmos de *Machine Learning* (ML), especialmente na análise integrada de dados complexos como perfil genético, histórico clínico, exames laboratoriais, dados comportamentais e estilo de vida. Tais técnicas têm possibilitado a identificação de padrões sutis e a associação entre características específicas dos pacientes e desfechos clínicos, permitindo a recomendação de estratégias terapêuticas adaptadas a cada indivíduo. Essa abordagem orientada por dados tem ampliado a precisão diagnóstica e a efetividade do tratamento em diversas especialidades médicas (Jiang et al., 2017, p. 231-232).

A medicina de precisão representa um novo paradigma assistencial, voltado à individualização das intervenções com base nas características biológicas, clínicas e contextuais de cada paciente. Nesse contexto, modelos de *Machine Learning* são aplicados à análise de uma ampla gama de dados, desde variações genômicas até fatores ambientais, como microbioma, metabolismo, idade, histórico familiar, sexo, perfil imunológico e vulnerabilidades sociais. Em vez de se basear em generalizações populacionais, essa abordagem busca adequar diagnósticos, prognósticos e tratamentos às particularidades de cada indivíduo, aumentando a efetividade terapêutica e minimizando efeitos adversos. Trata-se de uma aplicação direta da IA à prática clínica personalizada, já em expansão em centros médicos e programas de saúde de ponta (Bohr et al., 2020, p. 28).

A medicina molecular tem vivenciado uma transformação impulsionada pela convergência entre tecnologias ômicas e algoritmos de inteligência artificial, especialmente aqueles baseados em aprendizado de máquina. As aplicações chamadas de “multiômicas” passaram a integrar dados genômicos, transcriptômicos, proteômicos e metabolômicos, permitindo uma compreensão mais abrangente do estado de saúde e doença em nível individual. Técnicas como sequenciamento de DNA e RNA, bem como espectrometria de massas de alta resolução para proteínas e metabólitos, geram volumes massivos de dados cuja interpretação exige ferramentas computacionais avançadas. Nessa

fronteira, a IA tem demonstrado impacto direto na priorização de variantes genéticas, na predição de estruturas proteicas e na descoberta de biomarcadores clínicos, ampliando significativamente o potencial da medicina de precisão (Gomes et al., 2023, p. 2456-2464).

### 2.1.5 Modelagem antecipatória e inferência de tendências em Saúde

Algoritmos oriundos do aprendizado de máquina tradicional, e mais recentemente das técnicas de aprendizado profundo, vêm sendo empregados na previsão de tendências e desfechos clínicos em múltiplos contextos assistenciais. Tais previsões baseiam-se na análise de grandes volumes de dados em saúde, provenientes tanto de registros históricos quanto de informações coletadas em tempo real.

A possibilidade de antecipar desfechos clínicos relevantes por meio de análises preditivas tem transformado a forma como gestores de saúde planejam intervenções e estratégias terapêuticas. Técnicas avançadas de *Big Data Analytics*[15], que integram algoritmos de *Machine Learning*, têm sido utilizadas para prever a progressão de doenças, estimar o risco de readmissões hospitalares e identificar condições de risco iminente. Com isso, torna-se possível implementar ações preventivas, ajustadas ao perfil clínico de cada paciente e aos padrões extraídos de grandes volumes de dados. Tais abordagens ajudam a melhorar a qualidade assistencial e a eficiência operacional das instituições de saúde (Wang et al., 2018, p. 8-9).

A evolução dos sistemas de apoio à decisão clínica (CDSS)[16] tem incluído, de forma crescente, modelos não baseados em conhecimento, os quais utilizam técnicas de *Machine Learning* (ML) para identificar padrões complexos nos dados clínicos. Diferentemente dos sistemas tradicionais, que operam com regras programadas a partir do conhecimento médico especializado, essas novas abordagens são projetadas para fazer inferências

[15] O que é *Big Data Analytics*?. Disponível em: https://www.ibm.com/br-pt/topics/big-data-analytics. Acesso em: 05 de abril de 2025.

[16] Sistemas de Apoio à Decisão Clínica. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK543516/. Acesso em: 05 de abril de 2025.

com base em dados históricos, utilizando reconhecimento estatístico e algoritmos de inteligência artificial. Embora essa transição amplie o escopo e a flexibilidade dos CDSS, ela também impõe desafios quanto à transparência das decisões e à validação contínua dos modelos aplicados (Sutton et al., 2020, p. 2).

A aplicação de ferramentas analíticas em larga escala tem se mostrado promissora não apenas no cuidado individual, mas também na gestão coletiva de sistemas de saúde. O uso de técnicas de *Machine Learning* em saúde pública permite prever desfechos populacionais, identificar pacientes de alto custo, antecipar readmissões hospitalares, além de otimizar a distribuição de recursos humanos e financeiros. Tais modelos, ao estratificarem riscos e detectarem padrões emergentes em grandes volumes de dados, oferecem suporte à formulação de políticas mais eficazes e à prática de intervenções direcionadas a grupos vulnerabilizados, promovendo uma gestão mais racional e eficiente dos sistemas de saúde (Bates et al., 2014, p. 1124).

A vigilância epidemiológica tem sido significativamente aprimorada com o uso de ferramentas baseadas em inteligência artificial, que permitem identificar precocemente surtos infecciosos, rastrear sua disseminação e avaliar a eficácia das estratégias de contenção. Sistemas como o HealthMap[17] exemplificam essa aplicação, ao empregar algoritmos de processamento de linguagem natural e aprendizado de máquina para minerar dados de fontes diversas, como notícias, redes sociais e registros de saúde, em tempo real, antecipando sinais de eventos epidêmicos com agilidade superior aos métodos tradicionais. Essas tecnologias também têm sido aplicadas no mapeamento de hotspots[18], na diferenciação de patógenos e no apoio à alocação de recursos de saúde pública, consolidando a IA como ferramenta estratégica para

[17] Software de gestão de saúde. Disponível em: https://healthmap.com.br/. Acesso em 05 de abril de 2025.

[18] “Hotspots são locais que concentram alta biodiversidade, associada a uma grande ocorrência de endemismos (espécies que só ocorrem numa certa região) e sujeitas a grande pressão antrópica (ações do homem: atividades humanas como o desmatamento, as queimadas, a poluição, a caça e a pesca ilegais, o tráfico de animais, a destruição de habitats, a introdução de espécies exóticas, dentre outras.), à extinção de espécies animais e aos impactos severos das mudanças climáticas globais. (Estado de São Paulo, 2024)”, disponível em: https://semil.sp.gov.br/educacaoambiental/prateleira-ambiental/hotspots-de-biodiversidade/#:~:text=Hotspots%20s%C3%A3o%20locais%20que%20concentram,o%20tr%C3%A1fico%20de%20animais%2C%20a. Acesso em: 05 de abril de 2025.

respostas rápidas e eficazes a ameaças sanitárias (Brownstein et al., 2023, p. 1598-1599).

### 2.1.6 Criação de fármacos com apoio da inteligência artificial

A descoberta e o desenvolvimento de fármacos têm se beneficiado da aplicação de modelos de aprendizado de máquina (ML), especialmente aqueles baseados em redes neurais profundas e aprendizado por reforço. Essas técnicas têm se mostrado particularmente eficazes na predição de propriedades bioquímicas de moléculas, otimização de rotas de síntese e geração de compostos com características desejáveis, como solubilidade, seletividade e baixa toxicidade. Os modelos são treinados com grandes bases de dados químicos e estruturais, e têm possibilitado avanços significativos em etapas críticas da pesquisa farmacêutica, reduzindo custos, aumentando a precisão preditiva e acelerando o ciclo de desenvolvimento de novos medicamentos (Vamathevan et al., 2019, p. 469-470).

A previsão de propriedades farmacológicas e toxicológicas de compostos candidatos tem sido aprimorada com o uso de algoritmos de *Deep Learning* aplicados diretamente sobre representações estruturais moleculares. Esses modelos são treinados com grandes bancos de dados de compostos químicos e seus alvos biológicos, aprendendo a associar características estruturais com atributos como eficácia, seletividade, solubilidade, metabolismo e perfil toxicológico. Ao dispensar descritores pré-definidos e aprender diretamente das estruturas moleculares, as redes neurais profundas têm superado métodos tradicionais em tarefas como predição de bioatividade e desenho de novas moléculas com propriedades otimizadas para o desenvolvimento de fármacos (Chen et al., 2018, p. 1244).

Embora o processo de descoberta de novos medicamentos continue sendo longo, caro e incerto, ferramentas baseadas em inteligência artificial, especialmente modelos generativos e preditivos, vêm sendo apontadas como promissoras para reduzir os custos e acelerar etapas críticas do desenvolvimento farmacêutico. Relatórios recentes sugerem que a IA pode

proporcionar economias de tempo e recursos da ordem de 25% a 50% nas fases iniciais, que incluem a identificação de alvos moleculares, triagem de compostos e testes pré-clínicos. Apesar da necessidade de validação independente e rigor científico contínuo, exemplos práticos, como o da empresa Insilico Medicine[19], que alcançou a fase clínica em 30 meses com um fármaco desenvolvido por IA, ilustram o potencial disruptivo dessas tecnologias (Nature, 2023, p. 217).

A precisão na previsão da estrutura tridimensional de proteínas[20] teve um avanço significativo a partir de 2021, com a criação de modelos baseados em inteligência artificial que superaram limitações históricas do problema de dobramento proteico. Entre essas soluções, destaca-se o sistema AlphaFold[21], que demonstrou capacidade de prever estruturas com precisão atômica mesmo na ausência de estruturas homólogas previamente conhecidas. Essa abordagem inovadora permitiu prever, com elevado grau de acurácia, milhões de estruturas proteicas de organismos diversos, oferecendo enorme potencial para aplicações em biologia estrutural, farmacologia e biotecnologia. O AlphaFold combina redes neurais profundas com princípios físico-químicos e dados evolutivos, abrindo uma nova era para a bioinformática estrutural (Jumper et al., 2021, p. 583-584).

A abrangência do modelo AlphaFold, desenvolvido com técnicas de *Deep Learning*, expandiu-se com tamanha rapidez que, em apenas um ano após sua primeira publicação, ele foi capaz de prever a estrutura tridimensional de mais de 200 milhões de proteínas, praticamente todas conhecidas pela ciência até hoje, oriundas de cerca de um milhão de espécies. Essa façanha marca um ponto de inflexão na biologia estrutural, pois oferece um repositório

[19] “Somos uma empresa de biotecnologia orientada por IA generativa de ponta a ponta. (...) com a missão de acelerar a descoberta e o desenvolvimento de medicamentos, aproveitando nossa plataforma de Pharma.AI. Em rápida evolução em biologia, química e desenvolvimento clínico. (Insilico, 2025)”. Disponível em: https://insilico.com/. Acesso em: 05 de abril de 2025.

[20] “A estrutura tridimensional de uma proteína é a sua forma, determinada pela sequência de aminoácidos que a compõem. Essa estrutura é hierárquica, com quatro níveis de organização (2025)”. Disponível em: https://cesad.ufs.br/ORBI/public/uploadCatalago/11281716022012Bioquimica_aula_5.pdf. Acesso em: 05 de abril de 2025.

[21] AlphaFold é um sistema de IA desenvolvido pelo Google DeepMind que prevê a estrutura 3D de uma proteína a partir de sua sequência de aminoácidos. Ele regularmente atinge precisão competitiva com experimentos. Disponível em: https://alphafold.ebi.ac.uk/. Acesso em: 05 de abril de 2025.

acessível e gratuito de dados preditivos que anteriormente demandariam décadas de trabalho laboratorial. A utilidade prática vai além da simples previsão: pesquisadores já utilizam esse banco para descobrir novas famílias de proteínas e inferir propriedades funcionais, inclusive para aplicações ainda emergentes, como o design de proteínas artificiais (Callaway, 2022, p. 15).

As ciências naturais adentram um novo paradigma, impulsionado pela decodificação das complexas estruturas moleculares que regem os mecanismos biológicos. A configuração espacial das proteínas, elemento central na determinação de suas funções celulares, assume papel estratégico na engenharia de novos fármacos, servindo como base racional para o desenho de terapias mais precisas e eficazes.

### 2.1.7 Gestão de serviços de Saúde e cuidados médicos

Além das aplicações clínicas, técnicas de aprendizado de máquina (ML) têm desempenhado um papel cada vez mais relevante na melhoria da eficiência administrativa nos sistemas de saúde. A automação de tarefas rotineiras e operacionais, como o agendamento de consultas, a gestão de registros eletrônicos de saúde (EHRs), o faturamento, o processamento de autorizações de seguro e a alocação de recursos, tem sido facilitada por ferramentas analíticas avançadas, que analisam grandes volumes de dados em tempo real. Ao incorporar algoritmos preditivos nos fluxos de trabalho, é possível aumentar a produtividade, reduzir desperdícios e aprimorar a tomada de decisões organizacionais, o que representa uma importante contribuição da IA para a sustentabilidade dos sistemas de saúde (Bates et al., 2014, p. 1127).

A eficiência dos sistemas de saúde tem sido positivamente impactada pela incorporação de ferramentas baseadas em inteligência artificial, especialmente aquelas voltadas à automação de tarefas operacionais e análise de grandes volumes de dados. Com isso, reduz-se o tempo dedicado a processos repetitivos e administrativos, o que permite aos profissionais de saúde redirecionarem sua atenção para o cuidado direto ao paciente. A aplicação da IA nesse contexto tem o potencial de diminuir erros, otimizar

fluxos de trabalho e melhorar a produtividade das equipes clínicas, sem comprometer a qualidade da assistência (Topol, 2019, p. 44-45).

Dentre os diversos domínios da atenção à saúde, a incorporação da inteligência artificial (IA) nas funções administrativas e operacionais vem sendo apontada como a frente mais promissora em termos de impacto econômico. A aplicação de IA em processos como faturamento, codificação, autorização prévia e gerenciamento do ciclo de receita não apenas automatiza tarefas repetitivas, mas também gera economias significativas, melhora a experiência do paciente e libera recursos humanos para funções assistenciais. Em particular, algoritmos de aprendizado de máquina e processamento de linguagem natural vêm permitindo a antecipação de recusas de pagamento, a padronização de protocolos e a eliminação de falhas operacionais, o que pode gerar uma economia substancial para os sistemas de saúde (Sahni et al., 2023, p. 349-350).

### 2.1.8 Novos usos da inteligência artificial

A evolução das tecnologias de inteligência artificial (IA) na saúde também se estende a áreas inovadoras e com grande potencial de impacto. No campo da cirurgia robótica, sistemas dotados de aprendizado de máquina vêm sendo treinados para realizar procedimentos com elevada precisão, utilizando abordagens minimamente invasivas e aperfeiçoando continuamente seu desempenho com base em dados de operações anteriores. Simultaneamente, dispositivos vestíveis (*wearables*), integrados a algoritmos de reconhecimento de fala e análise preditiva, têm sido aplicados no contexto da telemedicina para monitoramento remoto de sinais vitais, apoio a pacientes em reabilitação e intervenções personalizadas em tempo real. Essas tecnologias, ainda em expansão, apontam para uma ampliação do papel da IA na promoção de cuidados mais precisos, acessíveis e centrados no paciente (Roski et al., 2019, p. 7-9).

## 2.2 Estruturação de soluções inteligentes para a Saúde

Embora cada aplicação de inteligência artificial na área da saúde apresente singularidades conforme sua finalidade específica e o contexto em que é inserida, é possível identificar etapas estruturais comuns aos distintos processos de desenvolvimento desses sistemas. Notadamente, a elaboração e a validação de modelos baseados em aprendizado de máquina, voltados ao diagnóstico clínico, à formulação de prognósticos ou à recomendação de intervenções preventivas e assistenciais, exigem a observância de fases metodológicas rigorosas que assegurem a eficácia, a segurança e a aplicabilidade dessas tecnologias no ambiente de cuidado.

### 2.2.1 Delimitação do desafio e especificação das funções da IA

Antes de se iniciar a construção de um modelo de *Machine Learning* (ML) para a área da saúde, é necessário compreender claramente quem será o usuário-alvo e qual impacto a intervenção baseada no modelo terá sobre o processo clínico. Deve-se considerar, por exemplo, quais ações serão desencadeadas pelas previsões do modelo, quais riscos estão envolvidos, como será operacionalizada a intervenção e qual mudança se espera alcançar. Essas reflexões preliminares são indispensáveis para avaliar se a aplicação de IA de fato agrega valor ao problema clínico proposto e se há recursos e viabilidade prática para esse desenvolvimeto. Somente após essa análise, define-se o problema a ser resolvido e verifica-se se ele é adequado para ser tratado com técnicas de IA (Liu et al., 2019, p. 15-16).

A definição da tarefa preditiva e a avaliação contextual da aplicação clínica são aspectos centrais para o desenvolvimento eficaz de modelos de inteligência artificial na saúde. A construção de uma solução baseada em IA requer, inicialmente, a identificação de um problema clínico que seja não apenas relevante, mas também viável de ser resolvido por meio de técnicas

computacionais. Essa viabilidade deve considerar elementos como a utilidade clínica, o custo de implantação, a aceitação por parte dos profissionais de saúde, o impacto na prática médica e a sustentabilidade da solução no longo prazo. Uma vez delimitado o problema, é essencial caracterizar o tipo de aprendizado esperado: se o modelo buscará replicar decisões humanas quantificáveis ou se será orientado à extração de padrões inéditos nos dados clínicos. Ambas as abordagens exigem cuidados metodológicos distintos, especialmente quanto à qualidade dos dados e à interpretação dos resultados (Chen et al., 2019, p. 411-412).

Trata-se da etapa em que se define a estratégia de aprendizado mais adequada ao problema em questão, seja ela supervisionada, não supervisionada ou baseada em reforço.

### 2.2.2 Princípios do treinamento supervisionado em inteligência artificial

No cenário atual da saúde digital, a maior parte das aplicações práticas de inteligência artificial (IA) concentra-se no paradigma do aprendizado supervisionado, que consiste na criação de modelos preditivos treinados com base em dados históricos rotulados. Esse tipo de abordagem permite que algoritmos identifiquem padrões em grandes conjuntos de dados clínicos, como registros eletrônicos, exames laboratoriais e imagens médicas, e os utilizem para antecipar desfechos clínicos, como risco de internação, complicações ou resposta a tratamentos. Por ser direcionado à previsão de eventos futuros e baseado em exemplos passados, o aprendizado supervisionado tem se consolidado como a principal forma de aplicação da IA na prática assistencial atual (Beam et al., 2018, p. 1).

No desenvolvimento de modelos de aprendizado de máquina (*Machine Learning* - ML) para a área da saúde, é essencial dominar alguns termos e conceitos técnicos fundamentais. A **acurácia** mede a proporção total de acertos entre todas as previsões realizadas. Já a **precisão** indica, entre todas as previsões positivas feitas, quantas são verdadeiramente positivas, enquanto a **sensibilidade** (revocação) representa a capacidade do modelo de identificar

corretamente os casos positivos. Complementarmente, a **especificidade** mede quantos negativos reais são corretamente classificados, capacidade de acertar quem está saudável, oposto a sensibilidade.

Outras ferramentas gráficas relevantes são a **curva ROC**, que relaciona sensibilidade e a taxa de falsos positivos (especificidade), e a **curva PR**, que associa precisão e sensibilidade (revogação).

Do ponto de vista da construção do modelo, os dados se dividem em três conjuntos: o **conjunto de treinamento** *(training set)*, que ajusta os parâmetros internos do modelo; o **conjunto de validação** *(validation set)*, usado para definir os hiperparâmetros (ajustes externos que influenciam a aprendizagem); e o **conjunto de teste** (test set), destinado a avaliar a performance final do modelo, devendo ser independente dos dados de treino, ainda que provenha da mesma distribuição estatística (Liu et al., 2019, p. 128).

### 2.2.3 Etapas de coleta, curadoria e tratamento inicial dos dados

Após a definição clara da tarefa de *Machine Learning* na área da saúde, inicia-se a preparação dos dados com a coleta de exemplos históricos nos quais aquela tarefa já foi executada. Esses dados brutos, muitas vezes heterogêneos e não estruturados, precisam passar por um processo rigoroso de pré-processamento, que consiste na limpeza, transformação e padronização das informações para torná-las utilizáveis pelos algoritmos. É esse procedimento que permite a extração de características (*features*) e a definição dos rótulos (*labels*) para o aprendizado supervisionado. A qualidade desse conjunto processado é fundamental, pois influencia diretamente a capacidade do modelo de aprender padrões significativos e generalizáveis a novos casos clínicos (Rajkomar et al., 2019, p. 1349).

O uso de dados heterogêneos é uma característica marcante das aplicações de inteligência artificial na saúde, e os modelos de *Deep Learning* têm se destacado por sua capacidade de processar diferentes formatos de entrada. Dados estruturados, como códigos de diagnósticos, resultados laboratoriais e informações demográficas, são organizados em tabelas com

colunas e linhas e geralmente extraídos de prontuários eletrônicos. Já os dados não estruturados, como imagens médicas ou relatórios clínicos redigidos em linguagem natural, demandam técnicas mais avançadas, como redes neurais convolucionais e processamento de linguagem natural, para serem interpretados e utilizados de forma eficaz pelos modelos. Essa flexibilidade torna os sistemas baseados em IA particularmente adequados para integrar múltiplos tipos de dados clínicos no suporte à decisão médica (Esteva et al., 2019, p. 25-26).

### 2.2.4 Estratégias e processos para o desenvolvimento de modelos de IA

O aprendizado supervisionado, uma das abordagens mais consolidadas em inteligência artificial aplicada à saúde, se baseia na associação entre entradas rotuladas (dados de entrada e seus respectivos desfechos) para treinar modelos preditivos. Cada exemplo, também chamado de instância, é composto por um conjunto de características (*features*) e um rótulo (*label*), que representa o resultado esperado. Esses exemplos são divididos em três subconjuntos:

(1) conjunto de treinamento, que serve para ajustar os parâmetros internos do modelo;

(2) conjunto de validação, que auxilia no ajuste de hiperparâmetros e na escolha da melhor configuração do modelo sem introduzir viés;

(3) conjunto de teste, que é mantido completamente separado e reservado para avaliação final do desempenho do modelo em dados inéditos.

Esse processo é fundamental para garantir que o modelo seja capaz de generalizar para novos cenários, e não apenas memorizar os dados com os quais foi treinado (Chen et al., 2019, p. 411-412).

No treinamento de modelos de aprendizado supervisionado, a meta principal é encontrar uma função que produza previsões tão próximas quanto possível dos rótulos verdadeiros associados aos dados de entrada. Essa discrepância entre a predição do modelo e o valor real é mensurada por uma função de perda, cujo valor o algoritmo tenta minimizar ajustando iterativamente

seus parâmetros (ou pesos). Isso é geralmente realizado por meio de um algoritmo de otimização, como o ***gradiente descendente***, que ajusta os pesos na direção em que a perda diminui mais rapidamente. Quando a saída prevista é um número contínuo, o problema é classificado como **regressão**. Já quando se trata de prever categorias, o modelo lida com uma tarefa de classificação, atribuindo probabilidades a cada classe possível com base nos recursos fornecidos (Murphy, 2012, p. 243-244).

Durante o desenvolvimento de modelos de aprendizado de máquina, sobretudo em contextos supervisionados, um dos maiores desafios está em alcançar o ponto de equilíbrio ideal entre subajuste (*underfitting*), quando o modelo é incapaz de capturar a complexidade dos dados, e sobreajuste (*overfitting*), quando o modelo se ajusta tão bem aos dados de treinamento que perde capacidade de generalização. Esse equilíbrio é formalizado como a compensação entre viés e variância (*bias-variance trade-off*): o viés corresponde ao erro introduzido por suposições simplificadas demais feitas pelo modelo, enquanto a variância representa o quanto o modelo é sensível a pequenas variações nos dados de entrada. Modelos com alto viés tendem a cometer erros sistemáticos, e os de alta variância apresentam desempenho instável. Escolher a arquitetura e o grau de complexidade de um modelo significa, portanto, minimizar tanto o viés quanto a variância de forma conjunta (Bishop, 2006, p. 147-148).

Durante o processo de desenvolvimento de modelos de aprendizado de máquina na área da saúde, um dos principais desafios está em evitar os dois extremos do desempenho inadequado: o underfitting, que se refere à incapacidade do modelo de capturar padrões nos dados devido à sua baixa complexidade, e o overfitting, que ocorre quando o modelo aprende de forma excessiva os detalhes e ruídos do conjunto de treinamento, comprometendo sua capacidade de generalização.

Esse equilíbrio é formalizado pela chamada compensação entre viés e variância (*bias-variance trade-off*), em que o viés representa os erros sistemáticos decorrentes de suposições simplificadas do modelo e a variância indica o quanto as previsões oscilam em resposta a pequenas alterações nos dados de entrada. Modelos excessivamente simples tendem ao underfitting, enquanto os modelos altamente complexos, com muitos parâmetros, são mais propensos ao overfitting, especialmente em cenários com dados escassos (Chen et al., 2019, p. 413).

O treinamento de um modelo supervisionado em aprendizado profundo segue uma dinâmica iterativa cuidadosamente estruturada, que visa encontrar a melhor configuração para resolver a tarefa proposta. O ciclo começa com a definição da arquitetura da rede, a configuração do ambiente computacional e a escolha inicial dos hiperparâmetros, como taxa de aprendizado, número de neurônios por camada, tipo de otimizador, entre outros.

A partir daí, o modelo é inicializado com parâmetros aleatórios e passa a aprender com os dados disponíveis no conjunto de treinamento. A cada nova iteração, ou época de treinamento, o desempenho da rede é medido em um conjunto de validação separado, usado exclusivamente para monitorar a capacidade de generalização do modelo durante o processo.

Caso essa nova rodada apresente resultados superiores às anteriores, a versão atualizada da rede é armazenada. Esse ciclo se repete enquanto houver ganhos sucessivos no desempenho. Quando esses ganhos cessam, isto é, quando o modelo entra em um platô, entende-se que a rede atingiu seu melhor desempenho dado aquele conjunto de hiperparâmetros.

Como o desempenho do modelo pode variar de forma significativa de acordo com os hiperparâmetros escolhidos, diversas configurações precisam ser testadas e comparadas. Somente após identificar a melhor configuração é que o modelo é finalmente submetido ao conjunto de teste, que permanece intocado durante todo o processo de treinamento e serve como critério final para avaliar sua performance em dados realmente inéditos (Goodfellow et al., 2016, p. 113-115).

### 2.2.5 Critérios e métodos para análise de desempenho de modelos

A etapa de avaliação do desempenho de um sistema baseado em aprendizado de máquina constitui um momento crítico dentro do ciclo de desenvolvimento, exigindo rigor metodológico e alinhamento com o objetivo prático do modelo. A seleção das métricas avaliativas deve refletir fielmente o propósito da aplicação, especialmente em contextos clínicos. Métricas como sensibilidade, especificidade e precisão dependem da definição prévia de um limiar de discriminação entre casos positivos e negativos, o que lhes confere

particular relevância na área da saúde, dada a natureza binária e decisiva das intervenções clínicas. Nessas situações, torna-se indispensável considerar o equilíbrio ou *trade-off* entre essas medidas, uma vez que otimizar uma delas pode implicar a redução da outra. A escolha entre priorizar a sensibilidade ou a precisão, por exemplo, representa uma decisão estratégica que deve ser orientada pelas implicações clínicas específicas do caso em análise.

A seleção da métrica de avaliação mais adequada para um modelo preditivo depende diretamente do seu propósito clínico. Em contextos de triagem, como em emergências ou rastreamentos populacionais, o maior risco está na perda de casos verdadeiros, ou seja, nos falsos negativos. Nesses cenários, é desejável priorizar a sensibilidade do modelo, métrica que expressa a capacidade de detectar corretamente os casos positivos, garantindo que pacientes com a condição não sejam erroneamente descartados.

Por outro lado, em situações diagnósticas, como decisões terapêuticas mais invasivas, o principal cuidado é evitar falsos positivo**s**, que podem induzir a tratamentos desnecessários e potencialmente danosos. Assim, nesses casos, a precisão (ou valor preditivo positivo) torna-se mais relevante, pois indica a proporção de resultados positivos que, de fato, correspondem a casos verdadeiros. O ideal, portanto, é que a escolha da métrica reflita as consequências clínicas de decisões incorretas, e não apenas critérios estatísticos abstratos (Wynants et al., 2019, p. 2-4).

Embora amplamente adotada por sua simplicidade, **a acurácia** nem sempre é uma medida confiável para avaliar modelos de aprendizado de máquina na saúde, especialmente em cenários com desequilíbrio de classes. Isso ocorre quando os dados disponíveis apresentam distribuição assimétrica entre os rótulos, por exemplo, com número muito maior de negativos do que positivos, o que é comum em tarefas clínicas, onde uma condição de interesse (como uma doença) tende a ocorrer com menor frequência na população geral.

Nesse contexto, um modelo que classifique todos os pacientes como “saudáveis” pode apresentar alta acurácia, mesmo sendo completamente inútil para detectar o desfecho clínico de interesse. Esse fenômeno é conhecido como *paradoxo da acurácia*, no qual modelos com elevado percentual de acertos globais podem falhar em capturar corretamente o padrão informacional relevante para a tarefa de classificação.

Com base em fundamentos da teoria da informação, autores propõem o uso de medidas alternativas, como a *Entropia Modulada da Acurácia (EMA)* e o *Fator Normalizado de Transferência de Informação (NIT)*, que avaliam a real capacidade do modelo de transmitir informação entre as variáveis de entrada e de saída, oferecendo critérios mais robustos e interpretáveis de desempenho (Valverde-Albacete et al., 2014, p. 1-4).

No âmbito da saúde, a utilização da acurácia como métrica primária para avaliar modelos de aprendizado de máquina deve ser tratada com cautela, principalmente em tarefas de classificação que envolvem desequilíbrio significativo entre as classes. Um modelo que acerte majoritariamente os casos negativos, mas falhe na identificação de casos positivos relevantes, pode exibir uma acurácia elevada e, ainda assim, ser clinicamente inútil.

Diante disso, ferramentas como a curva ROC (*Receiver Operating Characteristic*) se tornam fundamentais. Essa curva relaciona, ao longo de diferentes limiares, a taxa de verdadeiros positivos (sensibilidade) no eixo y com a taxa de falsos positivos (1 - especificidade) no eixo x.

Ela permite visualizar o desempenho global do classificador, destacando o compromisso inevitável entre aumentar a sensibilidade e sacrificar a especificidade, ou vice-versa. A partir dela, é possível selecionar o limiar ótimo de operação do modelo para uma determinada aplicação.

Mais do que isso, a área sob a curva ROC (AUROC) resume em um único número o desempenho discriminativo do modelo em todos os limiares possíveis. Um AUROC de 0,5 equivale a um modelo aleatório; valores próximos de 1,0 indicam excelente capacidade de distinguir corretamente entre as classes. Por ser independente do limiar de decisão, essa métrica é amplamente utilizada para comparar a eficácia de diferentes modelos, especialmente em contextos clínicos sensíveis (Liu et al., 2019, p. 136-138).

Em conjuntos de dados clínicos com forte desequilíbrio entre as classes, como frequentemente ocorre na área da saúde, em que eventos positivos (como o diagnóstico de uma doença) são proporcionalmente mais raros, o uso da curva PR (Precision-Recall) tem se mostrado uma estratégia mais eficaz que a tradicional curva ROC para avaliação do desempenho de classificadores binários.

A curva PR mapeia a precisão (*precision*) em função da revocação (*recall* ou sensibilidade) ao longo de diferentes limiares de decisão. Essa abordagem permite captar com maior clareza a capacidade de um modelo prever corretamente os casos positivos entre todas as predições positivas, o que é essencial em cenários em que o número de negativos é muito superior.

Ao contrário da curva ROC, cujo *baseline* é fixo, a curva PR possui um *baseline* dinâmico, que varia conforme a razão entre instâncias positivas e negativas, refletindo com mais realismo os desafios de classificação sob dados desbalanceados.

Assim, quando se busca avaliar o desempenho prático de modelos em contextos clínicos reais, o gráfico de precisão e *recall* (curva PR) é geralmente mais informativo, oferecendo uma visualização mais fiel da eficácia do modelo, sobretudo em faixas iniciais de recuperação de dados (Saito et al., 2015, p. 1-6).

A seleção do modelo mais apropriado para uma tarefa específica de aprendizado de máquina não depende apenas da métrica de desempenho isolada, mas sim de uma análise empírica e contextualizada de como diferentes algoritmos se comportam diante dos dados e objetivos clínicos em questão. Na prática, é comum testar múltiplas combinações de algoritmos e arquiteturas, já que cada técnica traz consigo conjuntos específicos de pressupostos, e um modelo que performa bem em um domínio pode ter resultados insatisfatórios em outro.

Essa constatação está ancorada no chamado teorema do “não há almoço grátis” (*no free lunch theorem*), que afirma que não existe um algoritmo universalmente superior para todos os problemas. A eficácia de um modelo depende da compatibilidade entre sua estrutura interna e as propriedades do conjunto de dados ao qual ele é aplicado.

Portanto, no desenvolvimento de soluções com *Machine Learning* aplicadas à saúde, é fundamental considerar esse aspecto e avaliar diversas abordagens de forma comparativa, sob o risco de comprometer a generalização e utilidade prática do modelo (Murphy, 2012, p. 38-39).

## 2.3 Obstáculos da IA na Saúde

Apesar do progresso técnico observado nas últimas décadas, o uso de inteligência artificial (IA) e aprendizado de máquina (ML) em ambientes clínicos ainda se encontra em estágio incipiente. Diversas aplicações, como leitura automatizada de eletrocardiogramas, reconhecimento de lesões cutâneas e análise de imagens retinais, já foram integradas à prática médica, mas permanecem limitadas a tarefas específicas e sob supervisão humana constante.

No entanto, especialistas destacam que a adoção generalizada de sistemas baseados em IA na medicina depende da superação de obstáculos técnicos, regulatórios e éticos. Os sistemas precisam demonstrar consistência, escalabilidade e utilidade clínica real em cenários diversos e com dados heterogêneos, além de respeitar princípios como equidade, explicabilidade e segurança.

Por essa razão, os modelos atualmente disponíveis são tratados como promessas com grande potencial, cuja consolidação ainda requer um ciclo contínuo de validação e adaptação ao ambiente clínico real (Haug et al., 2023, p. 1204-1206)

A introdução bem-sucedida de soluções de inteligência artificial (IA) na área da saúde depende de um processo estruturado e iterativo, que compreende quatro etapas essenciais. A primeira delas é o design e desenvolvimento, no qual se identifica o problema clínico relevante a ser resolvido, levando em conta a ação clínica que pode ser tomada a partir da previsão do modelo. Em seguida, realiza-se a fase de avaliação e validação, momento em que se testam o desempenho estatístico, a utilidade clínica e o impacto econômico do modelo, primeiramente com dados retrospectivos e, posteriormente, em ambientes reais de produção clínica.

A terceira etapa é a difusão e dimensionamento, que envolve expandir a aplicação do modelo para além do ambiente original de desenvolvimento, superando desafios como a interoperabilidade entre sistemas de prontuário eletrônico e o financiamento para implantação em larga escala. Por fim, entra-se na fase de monitoramento e manutenção contínuos, essencial para garantir

a robustez do modelo ao longo do tempo, considerando mudanças nos dados, nos fluxos de trabalho e nas definições clínicas utilizadas (Sendak et al., 2020, p. 1-6).

Observa-se que a efetiva incorporação dessa tecnologia na prática assistencial em saúde ainda requer a superação de obstáculos significativos, envolvendo desde limitações técnicas e operacionais até questões regulatórias, éticas e organizacionais que condicionam sua viabilidade e aceitação no cotidiano clínico.

### 2.3.1 Critérios na seleção de dados

Nos projetos de inteligência artificial aplicados à saúde, um dos aspectos mais críticos para o sucesso dos modelos de aprendizado de máquina reside na qualidade e na preparação dos dados. Isso se torna ainda mais relevante quando se considera a diversidade de fontes típicas desse setor: registros de prontuário eletrônico, exames laboratoriais, imagens médicas, dados genômicos e anotações clínicas em linguagem natural. Esses elementos compõem um cenário caracterizado por alta heterogeneidade e mistura entre dados estruturados e não estruturados.

A máxima “garbage in, garbage out” ou “entra lixo, sai lixo” é especialmente verdadeira nesse contexto. Dados mal coletados, inconsistentes ou incompletos comprometem diretamente a capacidade do modelo de extrair padrões úteis e de oferecer predições confiáveis. Como destacam especialistas, o pré-processamento e a curadoria de dados são, muitas vezes, mais determinantes para o desempenho do modelo do que o próprio algoritmo utilizado, sendo uma etapa essencial do pipeline de desenvolvimento (Kilkenny et al., 2018, p. 103-104).

Entre os diversos obstáculos enfrentados no desenvolvimento de aplicações de inteligência artificial voltadas à saúde, destaca-se de forma recorrente a dificuldade de acesso a grandes volumes de dados clínicos rotulados com qualidade suficiente para o treinamento de modelos robustos, especialmente aqueles baseados em arquiteturas de *Deep Learning* (DL).

Tais modelos requerem um número significativo de exemplos anotados com precisão para extrair padrões confiáveis, mas esse tipo de dado, na área médica, é escasso, sensível, disperso e frequentemente protegido por restrições éticas e regulatórias. Além disso, muitos conjuntos de dados disponíveis são pequenos, incompletos ou mal padronizados, o que limita a capacidade dos algoritmos de generalizar adequadamente em ambientes clínicos complexos.

Esse gargalo de dados tem sido apontado por especialistas como um dos principais entraves à consolidação da IA na prática clínica, ao lado de desafios computacionais, regulatórios e de interoperabilidade entre sistemas (Shah et al., 2019, p. 2).

Um dos grandes desafios na construção de modelos de aprendizado de máquina para aplicações clínicas é a preservação da validade e da representatividade dos dados utilizados para o treinamento. Em especial, os dados extraídos de prontuários eletrônicos de saúde (EHRs), embora amplamente acessíveis, são suscetíveis a diversas formas de viés sistemático.

Entre os principais riscos, destacam-se os erros de classificação de condições clínicas, derivados de registros imprecisos, preenchimento incompleto ou até mesmo interpretações divergentes entre profissionais. Soma-se a isso a seletividade das amostras, uma vez que populações vulnerabilizadas, como pacientes com baixa renda, baixa escolaridade, ou usuários de múltiplos sistemas de saúde, muitas vezes estão sub-representadas ou têm dados fragmentados e incompletos.

Essas falhas podem comprometer gravemente a capacidade do modelo de gerar predições válidas para todo o espectro da população-alvo, podendo inclusive reproduzir ou acentuar desigualdades já existentes nos cuidados em saúde. Por essa razão, é imprescindível que os desenvolvedores de algoritmos avaliem criticamente a qualidade e a origem dos dados, considerando tanto os aspectos técnicos quanto os determinantes sociais que influenciam a coleta e a documentação clínica (Gianfrancesco et al., 2018, p. 1545-1547).

Além de aspectos técnicos e éticos, é fundamental compreender que os dados clínicos, por sua própria natureza, são gerados em um ambiente altamente dinâmico, no qual as práticas médicas, os protocolos institucionais e os perfis populacionais sofrem alterações constantes ao longo do tempo. Essa

volatilidade implica que modelos preditivos baseados em dados de saúde precisam ser monitorados continuamente quanto à sua acurácia e validade.

A relevância de dados clínicos históricos para a tomada de decisão diminui com o passar do tempo, ou seja, esses dados “envelhecem” rapidamente, de modo que informações coletadas meses antes podem já não refletir as práticas clínicas vigentes. Estudos indicam que, em muitos cenários hospitalares, a "vida útil" dos dados clínicos pode ser estimada em cerca de quatro meses, sendo mais eficaz treinar modelos com quantidades menores de dados recentes do que com grandes volumes de dados mais antigos.

Portanto, a manutenção da performance preditiva exige o uso sistemático de dados atualizados e a adoção de estratégias automatizadas de re-treinamento e validação contínua dos modelos (Chen et al., 2017, p. 71-79).

### 2.3.2 Aplicabilidade prática

Outro aspecto crítico na avaliação de modelos de aprendizado de máquina (ML) aplicados à saúde diz respeito às métricas utilizadas na validação desses sistemas, que nem sempre refletem, de fato, a utilidade clínica das soluções propostas. É comum que estudos de IA na área médica concentrem-se em métricas tradicionais, como acurácia, sensibilidade, especificidade e área sob a curva ROC (Receiver Operating Characteristic). No entanto, essas medidas, embora estatisticamente válidas, nem sempre são intuitivas ou clinicamente informativas para os profissionais da saúde e não necessariamente traduzem o impacto real na tomada de decisão ou nos desfechos do paciente.

Como apresentado por especialistas, há um verdadeiro "abismo" entre desempenho técnico e relevância clínica, que precisa ser superado com o uso de métricas mais representativas, como valores preditivos positivos e negativos, análise de decisão clínica (*decision curve analysis*) e indicadores de melhoria na qualidade do cuidado. O sucesso de um modelo, portanto, não deve ser medido apenas por sua capacidade discriminativa, mas sim por sua

efetiva contribuição para melhores decisões clínicas e resultados em saúde (Kelly et al., 2019, p. 3).

Além das métricas estatísticas tradicionais usadas na validação de modelos preditivos, há indicadores mais diretamente ligados à utilidade clínica, cuja interpretação favorece a tomada de decisão na prática médica. Duas dessas medidas fundamentais são o número necessário para tratar (NNT) e o número necessário para causar danos (NNH).

O NNT (number needed to treat) expressa o número de pacientes que precisam receber uma intervenção para que um único paciente se beneficie dela. Em outras palavras, representa quantas pessoas precisam ser tratadas para evitar que uma delas sofra o desfecho adverso que se deseja prevenir. Já o NNH (number needed to harm) refere-se ao número de pacientes que devem ser submetidos a uma intervenção para que um deles sofra um efeito colateral ou evento adverso adicional.

Diferentemente das métricas baseadas exclusivamente em acertos preditivos, o NNT e o NNH incorporam uma dimensão pragmática e ética, pois possibilitam o balanceamento entre benefício clínico e risco associado à intervenção. Isso permite que médicos e gestores considerem, de forma mais clara, a relação entre efetividade e segurança no uso de intervenções recomendadas por modelos de IA ou por ensaios clínicos tradicionais (Laupacis et al., 1988, p. 1728-1733).

Mesmo que um modelo de inteligência artificial (IA) apresente métricas estatísticas promissoras, essas métricas isoladamente não garantem sua utilidade clínica prática. A questão central reside em compreender se, e de que forma, o uso do modelo pode efetivamente promover mudanças benéficas no cuidado ao paciente.

Nesse contexto, uma técnica proposta com esse objetivo é a análise da curva de decisão (*decision curve analysis*). Diferente de métodos baseados apenas em acurácia, essa abordagem considera os impactos clínicos reais das decisões informadas pelo modelo, levando em conta os benefícios líquidos para diferentes limiares de probabilidade. Em essência, a curva de decisão permite quantificar o valor agregado de um modelo preditivo em relação a estratégias clínicas padrão, como tratar todos os pacientes ou não tratar nenhum, ponderando os efeitos adversos e positivos de falsos positivos e falsos negativos.

Além de ser aplicável a modelos com saídas contínuas ou binárias, essa metodologia pode ser executada diretamente nos conjuntos de validação, sem a necessidade de coleta de dados adicionais, tornando-se uma ferramenta prática e eficiente para avaliar a real aplicabilidade clínica de modelos de predição médica (Vickers et al., 2006, p. 565-574).

A análise da curva de decisão considera o limiar de probabilidade em que se justifica agir clinicamente, levando em conta os custos relativos de falsos positivos e falsos negativos para calcular o benefício líquido do modelo.

No contexto da saúde, essa abordagem permite incorporar restrições reais do ambiente clínico, como a capacidade da equipe de atendimento, o custo e a eficácia das intervenções, e a probabilidade de adesão do paciente. Esses fatores devem ser considerados durante o treinamento do modelo, e não apenas após a seleção baseada em métricas estatísticas. Assim, garante-se que o modelo escolhido seja realmente útil e viável na prática clínica (Shah et al., 2019, p. 4).

Em síntese, uma das principais complexidades associadas à adoção de soluções baseadas em inteligência artificial na área da saúde reside na necessidade de se estabelecerem mecanismos eficazes para aferir a utilidade e a viabilidade dos sistemas de aprendizado de máquina.

### 2.3.3 A resiliência da IA

O bom desempenho de um modelo de aprendizado de máquina em conjuntos de validação restritos não garante que ele funcionará bem em populações distintas ou contextos clínicos diversos. Essa limitação reflete o que tem sido chamado de “abismo da IA”, ou seja, a distância entre um modelo desenvolvido de forma tecnicamente robusta e sua aplicação efetiva no mundo real.

A generalização limitada e os desafios na tradução de protótipos para soluções clinicamente úteis demonstram que o sucesso técnico não é suficiente: é necessário comprovar a eficácia do sistema em ambientes clínicos complexos e variados (Keane et al., 2018, p. 2).

Uma das razões mais recorrentes para a queda de desempenho de modelos de inteligência artificial na prática clínica é a ocorrência do chamado "*dataset shift"*, ou mudança de distribuição dos dados. Isso acontece quando os dados utilizados no treinamento do modelo não refletem mais a realidade dos novos dados aos quais ele é exposto durante sua aplicação.

Essa divergência pode surgir por múltiplos fatores, como alterações no perfil demográfico da população atendida, mudanças nas práticas clínicas, revisões de protocolos médicos ou transformações provocadas pela própria introdução do modelo no sistema de saúde.

Diante disso, torna-se essencial a monitorização contínua do desempenho do modelo, bem como sua recalibração periódica, para garantir que ele continue sendo útil e seguro nas novas condições em que está inserido (Finlayson et al., 2021, p. 283).

Em muitos casos, modelos de aprendizado de máquina (ML) obtêm bom desempenho em ambientes controlados, mas falham ao serem aplicados em contextos reais. Uma explicação importante é que esses sistemas podem acidentalmente se ajustar a ruídos ou confundidores ocultos nos dados, fatores que não têm relação causal com o desfecho, mas que acabam sendo usados como atalhos pela IA.

Além disso, a maioria dos sistemas atuais não atinge uma generalização confiável, o que compromete seu uso clínico. Por isso, a avaliação do desempenho de modelos deve incluir validação externa com conjuntos de dados independentes, coletados de instituições diferentes daquelas que forneceram os dados de treinamento, garantindo maior robustez e representatividade populacional (Kelly et al., 2019, p. 3-4).

Diante das limitações recorrentes na generalização de modelos de aprendizado de máquina entre diferentes contextos clínicos, muitos especialistas propõem que a prática de sistemas de IA na saúde seja direcionada a cenários específicos, onde sua utilidade clínica já foi comprovada empiricamente.

Segundo essa visão, modelos não precisam, necessariamente, funcionar bem em todos os ambientes para serem valiosos. Pelo contrário, podem ser altamente eficazes quando aplicados localmente, respeitando as peculiaridades institucionais, práticas assistenciais e perfis populacionais daquele centro específico.

Assim, a busca pela generalização universal pode ser contraproducente, devendo ceder espaço à construção de soluções contextuais, baseadas em dados locais e adaptadas à realidade dos usuários finais (Futoma et al., 2020, p. 491-492).

### 2.3.4 Atenção aos cuidados com a IA

Modelos de aprendizado de máquina (ML) podem gerar efeitos adversos ao longo de todo seu ciclo de vida, desde a coleta de dados até sua inauguração. A presença de vieses estruturais e contextuais levanta preocupações importantes sobre injustiças e danos potenciais no uso da inteligência artificial em saúde. Para entender e mitigar esses riscos, é útil considerar diferentes categorias de viés.

Entre eles, destaca-se o viés histórico, que decorre da reprodução de desigualdades já presentes nos dados, como o predomínio de informações provenientes de grupos historicamente privilegiados. Já o viés de representação ocorre quando certos segmentos da população não estão adequadamente representados no conjunto de dados de treinamento.

O viés de medição envolve o uso de proxies ruidosos ou inconsistentes para variáveis clínicas, enquanto o viés de agregação emerge ao tentar aplicar um único modelo a grupos populacionais com perfis distintos.

Durante a validação, pode ocorrer o viés de avaliação, quando o conjunto de teste não representa a população-alvo real ou quando métricas inadequadas são utilizadas. Por fim, o viés da implantação decorre do descompasso entre o uso pretendido e o uso real do modelo, muitas vezes agravado por fatores humanos ou institucionais. Reconhecer esses vieses é essencial para promover o uso responsável e equitativo da IA em qualquer área (Suresh et al., 2021, p. 2-6).

### 2.3.5 A importância da transparência e compreensão dos processos

Um dos grandes entraves à consolidação da inteligência artificial (IA) na saúde reside na persistente opacidade algorítmica, que compromete tanto a transparência quanto a auditabilidade dos sistemas utilizados. Essa opacidade manifesta-se em três níveis distintos, mas inter-relacionados. O primeiro é de natureza técnica, decorrente da complexidade estrutural de modelos como os de *deep learning*, que operam com múltiplas camadas matemáticas e tornam praticamente inviável a identificação de quais variáveis influenciaram determinada decisão clínica. Esses modelos funcionam, na prática, como "caixas-pretas", já que o raciocínio interno permanece indecifrável até mesmo para seus desenvolvedores.

O segundo nível de opacidade é informacional, derivado da ausência de letramento computacional por parte dos usuários finais. Médicos, pacientes e gestores, muitas vezes, não dominam os fundamentos estatísticos e operacionais dos algoritmos, o que os impede de compreender, avaliar criticamente ou questionar os resultados gerados por tais sistemas.

O terceiro nível diz respeito à opacidade intencional, motivada por interesses comerciais ou regulatórios. Aqui, o sigilo sobre código-fonte, parâmetros de modelagem e bases de dados é justificado sob o argumento de proteção à propriedade intelectual ou segurança institucional. No entanto, esse bloqueio ao acesso compromete diretamente a supervisão ética e jurídica dos algoritmos em contextos clínicos, impactando a confiança e a responsabilização (BURRELL, 2016, p. 2–4).

Em aplicações de IA generativa (IAG) voltadas à saúde, como a elaboração automatizada de relatórios clínicos, sumários de atendimento ou orientações terapêuticas personalizadas, o problema da opacidade assume contornos ainda mais preocupantes. Isso porque, como observa Cebrian et al. (2023), embora esses sistemas demonstrem capacidade de construir inferências consistentes, carecem, em geral, de registros sistemáticos sobre o processo de geração da informação, o que compromete a possibilidade de verificar a coerência interna, apurar falhas e atribuir responsabilidade diante de eventuais danos (Cebrian et al., 2023, p. 36).

Como resposta a essa lacuna crítica, Cebrian et al. (2023) defende que tais sistemas passem a incorporar mecanismos que permitam algum grau de rastreabilidade, viabilizando a recuperação dos dados de entrada, da versão do modelo utilizado e das estruturas lógicas envolvidas no raciocínio algorítmico. Tais instrumentos possibilitariam não apenas a reconstituição técnica de cada inferência, mas também o exercício do controle ético, clínico e jurídico sobre a prática assistida por IA, assegurando maior responsabilização (Cebrian et al., 2023, p. 38).

Dessa forma, a adoção dessas práticas de rastreabilidade e de registro sistemático se apresenta não apenas como boa-prática técnica, mas como requisito jurídico-ético imprescindível para assegurar a confiança e a responsabilização em sistemas de IA na saúde.

### 2.3.6 Dinâmica entre profissionais de saúde e sistemas inteligentes

A aplicação de sistemas de suporte à decisão clínica (CDSS) baseados em IA na saúde é fortemente influenciada por fatores organizacionais, como cultura, fluxo de trabalho, capacitação da equipe e integração com sistemas existentes.

Esses fatores são determinantes para o sucesso ou fracasso da adoção de soluções de IA na assistência médica. Segundo Sutton et al. (2020, p. 2), a introdução de CDSS pode melhorar a segurança do paciente, apoiar o gerenciamento clínico, otimizar custos e agilizar fluxos de trabalho, mas também pode gerar desafios como sobrecarga de alertas, interrupção de rotinas clínicas e desconfiança dos usuários.

A aceitação e eficácia de sistemas de suporte à decisão clínica (CDSS) dependem fortemente de sua compatibilidade com o ambiente institucional, incluindo usabilidade, integração com prontuários eletrônicos (EHRs) e treinamento adequado dos profissionais de saúde. Além disso, o desempenho dos modelos de aprendizado de máquina (ML) nesse contexto é influenciado por fatores mais complexos, como mudanças comportamentais provocadas

pela introdução dos sistemas, a capacidade dos usuários de interpretar as recomendações geradas e o nível de confiança na tecnologia.

Como aplicado por Emanuel e Wachter (2019), o maior desafio não está nas previsões geradas pelos algoritmos, mas na transformação dos hábitos e rotinas de trabalho de médicos, enfermeiros e pacientes, o que exige mudanças estruturais e culturais profundas dentro das instituições (Emanuel et al., 2019, p. 2282).

A resistência de profissionais de saúde à adoção de soluções baseadas em inteligência artificial (IA) pode surgir por múltiplos fatores. Entre eles, estão o ceticismo em relação ao desempenho dos sistemas, a falta de treinamento adequado para sua utilização eficiente, e o temor de perda de autonomia clínica. Para que a realização da IA na prática assistencial seja viável e confiável, é fundamental investir na educação de profissionais e pacientes, bem como promover o desenvolvimento de modelos mais interpretáveis e revisar os processos de trabalho vigentes.

Como alertam Yu e Kohane (2019), a confiança do público e da comunidade médica depende diretamente do grau de transparência e compreensibilidade dos sistemas, além da atenção às dinâmicas institucionais e culturais envolvidas em sua adoção (Yu et al., 2019, p. 240).

Por outro lado, os profissionais de saúde também podem ser influenciados pelo chamado viés da automação (*automation bias*), que consiste na tendência de confiar excessivamente em sistemas automatizados, mesmo quando suas recomendações estão equivocadas. Esse viés pode levar tanto a erros de comissão (seguir conselhos incorretos) quanto a erros de omissão (deixar de agir por ausência de recomendação). A confiança desproporcional na tecnologia pode comprometer decisões previamente corretas feitas pelos clínicos, e sua ocorrência é particularmente crítica no contexto da saúde, onde erros de decisão têm alto impacto. Fatores como nível de confiança no sistema, sobrecarga cognitiva, experiência limitada com a tarefa ou com o próprio sistema, e até o posicionamento da recomendação na tela, podem aumentar a suscetibilidade ao viés (Goddard et al., 2012, p. 122-124).

Esse fenômeno pode se tornar particularmente grave à medida que os sistemas de inteligência artificial (IA) apresentem desempenho cada vez melhor, o que pode levar profissionais da saúde a um excesso de confiança e, eventualmente, à negligência no julgamento clínico humano. Mesmo quando o

sistema de IA é projetado apenas para fornecer recomendações, e não para realizar diagnósticos ou indicar tratamentos diretamente, ainda assim podem surgir consequências indesejadas na interação com os usuários, como o viés de confirmação e a fadiga de alerta, esta última se refere à desatenção ou ao descarte sistemático de avisos clínicos devido à frequência excessiva e à baixa relevância percebida.

Esse tipo de sobrecarga cognitiva reduz a efetividade dos sistemas de apoio à decisão clínica, comprometendo sua finalidade original de aumentar a segurança do paciente (Phansalkar et al., 2013, p. 489-490).

Uma alternativa prudente para reduzir os riscos da adoção de sistemas de inteligência artificial (IA) na saúde é o uso do chamado “modo silencioso”. Nessa fase, o sistema é operado em tempo real no ambiente clínico, produzindo previsões conforme faria em uso normal, mas sem que essas previsões sejam utilizadas na tomada de decisões médicas.

Esse período de observação tem múltiplas finalidades: permite a validação temporal do modelo, a adaptação do fluxo de trabalho, e a identificação de possíveis falhas antes que qualquer impacto clínico ocorra. Além disso, contribui para aumentar o entendimento da equipe sobre o funcionamento da ferramenta e fortalece a confiança institucional na solução antes de sua ativação oficial (Sendak et al., 2020, p. 13).

Essa modalidade de validação em modo silencioso é estratégica para avaliar não apenas o desempenho técnico do modelo, mas também a interação entre humanos e sistemas de IA. Ela permite verificar se os profissionais compreendem corretamente as previsões geradas, se as aplicam de maneira adequada e se estão direcionando o uso para a população certa.

Além disso, esse tipo de trabalho controlado é útil para testar a usabilidade da interface, organizar fluxos de trabalho e adaptar as configurações do produto às rotinas clínicas. Trata-se de uma abordagem essencial para garantir que a integração do sistema no ambiente real de cuidado ocorra com segurança (Wiens et al., 2019, p. 1339).

### 2.3.7 Supervisão permanente e atualização em tempo real

Garantir a segurança dos sistemas de inteligência artificial (IA) na saúde exige que seu desempenho seja monitorado continuamente ao longo do tempo. Mesmo após o treinamento adequado à rotina clínica, modelos de *Machine Learning* estão sujeitos a mudanças em sua performance, seja por flutuações naturais nos dados de entrada, seja por alterações inesperadas nas características da população atendida ou nos processos assistenciais.

Essas variações, que podem ser sutis ou abruptas, demandam a prática de mecanismos de controle estatístico rigorosos, voltados à detecção precoce de desvios e à realização de ajustes quando necessário. O monitoramento, portanto, não é uma etapa pontual, mas um processo contínuo e crítico para a confiabilidade clínica das soluções baseadas em IA (Feng et al., 2022, p. 1-2).

Esse desafio torna-se ainda mais complexo quando se trata de sistemas de inteligência artificial com capacidade de aprendizado contínuo. Embora esses modelos sejam projetados para evoluir com novos dados, eles introduzem dificuldades adicionais de monitoramento e validação, já que seu comportamento pode mudar dinamicamente ao longo do tempo.

A ausência de diretrizes consolidadas para validar e ajustar continuamente esses sistemas em ambientes clínicos reais compromete sua segurança e eficácia. Assim, há uma necessidade urgente de desenvolvimento de boas práticas específicas para o acompanhamento e a atualização desses modelos em uso (Walsh et al., 2022, p. 2).

Apesar do grande potencial das técnicas de aprendizado incremental ou aprendizado contínuo para otimizar decisões clínicas em tempo real, seu uso na prática médica ainda exige cautela. Um dos principais obstáculos está na complexidade envolvida no monitoramento desses modelos, que evoluem dinamicamente com base em novos dados, dificultando a avaliação contínua de desempenho e a prevenção de erros.

Essa dificuldade se agrava em contextos de alta sensibilidade clínica, nos quais resultados equivocados podem ter impacto direto na segurança do paciente. Assim, embora promissora, a aplicação de modelos de aprendizado

contínuo na saúde demanda estratégias rigorosas de controle, validação e acompanhamento (Lee et al., 2020, p. e279).

A manutenção do desempenho de sistemas de IA na saúde não termina com a fase de implantação. Modelos que não se atualizam automaticamente após o treinamento, os chamados modelos estáticos, precisam passar por atualizações regulares para evitar degradações causadas por mudanças nos dados clínicos ou na prática assistencial.

Essa tarefa depende da existência de dados clínicos atualizados, confiáveis e bem estruturados, além de processos consistentes de coleta e preparação dos dados. Na prática, é indispensável que as instituições de saúde definam estruturas claras de responsabilidade para acompanhar e revisar esses sistemas.

Embora desenvolvedores de software continuem exercendo papel importante nesse processo, pode ser necessário criar equipes especializadas dentro das organizações de saúde para conduzir avaliações periódicas. Órgãos reguladores também devem atuar ativamente, promovendo a supervisão contínua desses modelos e garantindo sua adaptação segura ao longo do tempo (Lee et al., 2020, p. e280-e281).

## 2.4 Síntese reflexiva

A análise das implicações da inteligência artificial na saúde demonstra que o desenvolvimento tecnológico, quando orientado por objetivos clínicos, pode representar um vetor de aprimoramento substancial na qualidade do cuidado prestado. A IA tem se consolidado como instrumento de suporte ao diagnóstico, à terapêutica e à gestão de sistemas de saúde, oferecendo respostas mais céleres e precisas diante de um contexto marcado pela crescente complexidade dos desafios epidemiológicos e estruturais.

A diversidade de aplicações, que vai desde a análise automatizada de imagens médicas até a personalização de tratamentos e a antecipação de desfechos clínicos, revela um campo em expansão contínua, cuja influência já ultrapassa o estágio experimental. Ao mesmo tempo, a integração dessas

tecnologias à prática médica impõe uma série de desafios éticos, operacionais e regulatórios, exigindo que a adoção de sistemas inteligentes se dê de forma crítica, responsável e centrada no ser humano.

A incorporação progressiva de algoritmos de aprendizado de máquina e redes neurais profundas, bem como a ascensão dos modelos fundacionais na área da saúde, reforçam a necessidade de repensar os limites da atuação profissional, os critérios de responsabilidade e os referenciais normativos aplicáveis. Nesse sentido, os próximos capítulos buscarão examinar as bases éticas e jurídicas que devem orientar o uso seguro, equitativo e legítimo da inteligência artificial em contextos assistenciais, com especial atenção às diretrizes propostas por organismos internacionais como a Organização Mundial da Saúde.

Para facilitar a visualização dos principais impactos discutidos neste capítulo, o Quadro 2 a seguir sintetiza as aplicações, benefícios e desafios associados à presença crescente da inteligência artificial no campo da saúde, destacando os pontos centrais analisados sob uma perspectiva técnico-assistencial e ética.

**Quadro 2 – Síntese das implicações da inteligência artificial na saúde.**

| | |
|---|---|
| **Suporte à decisão clínica** | Sistemas que auxiliam médicos com recomendações baseadas em dados e evidências.. |
| **Diagnóstico por imagem** | IA aplicada à análise automatizada e precisa de exames de imagem médica. |
| **Medicina personalizada** | Uso de IA para adaptar tratamentos às características genéticas e clínicas de cada paciente. |
| **Modelagem preditiva** | Previsão de riscos e evolução de doenças com base em grandes volumes de dados. |
| **Criação de fármacos** | Emprego de algoritmos para acelerar a descoberta e desenvolvimento de medicamentos.. |
| **Gestão hospitalar** | Otimização de fluxos, leitos e recursos por meio da análise de dados em tempo real. |
| **Triagem automatizada** | Classificação inteligente de pacientes com base em sintomas e históricos clínicos. |
| **Transparência algorítmica** | Necessidade de compreensão e auditabilidade das decisões geradas por IA.. |
| **Supervisão contínua** | Requisito de monitoramento e atualização constante dos sistemas inteligentes. |
| **Resiliência da IA** | Capacidade dos sistemas de manter desempenho mesmo em contextos incertos ou adversos.. |
| **Ética assistencial** | Preocupações com equidade, justiça e respeito à autonomia no uso de tecnologias em saúde. |

**Fonte**: Elaborado pelo autor.

A sistematização apresentada no Quadro 2 reforça a amplitude e a complexidade das transformações provocadas pela inteligência artificial na saúde, evidenciando tanto seu potencial inovador quanto os desafios que emergem de sua aplicação. Ao integrar tecnologia aos processos clínicos e administrativos, a IA demanda uma atuação cuidadosa, que concilie eficiência com responsabilidade ética e jurídica, antecipando-se a riscos e promovendo uma assistência centrada na dignidade humana.

# 3 BASES NORMATIVAS PARA O USO DA IA

## 3.1 Princípios morais e regulatórios

Diante da crescente inserção da inteligência artificial na saúde, torna-se imprescindível analisar os fundamentos normativos que orientam seu desenvolvimento e sua aplicação sob a perspectiva dos direitos humanos e da proteção à saúde. A Organização Mundial da Saúde, atenta às transformações tecnológicas e aos impactos decorrentes da digitalização das práticas assistenciais, propôs diretrizes orientadas por princípios éticos e jurídicos que visam garantir o uso responsável, seguro e equitativo das soluções baseadas em IA.

Esse esforço normativo busca equilibrar os benefícios proporcionados pelas inovações tecnológicas com a necessidade de salvaguardar a dignidade da pessoa humana, a autonomia dos pacientes e a justiça social no acesso aos cuidados. Nesse contexto, as diretrizes da OMS se fundamentam em uma matriz principiológica que valoriza a transparência algorítmica, a proteção de dados pessoais, a responsabilidade institucional e a inclusão social na implantação dos sistemas inteligentes.

Este capítulo tem por objetivo examinar os pilares morais e regulatórios delineados pela OMS, com ênfase na articulação entre os direitos fundamentais à saúde e à privacidade, os mecanismos de regulação do uso da IA em saúde, bem como os princípios éticos que devem nortear a relação entre inovação tecnológica e integridade assistencial. A análise foi estruturada a partir da leitura crítica dos documentos internacionais e dos desafios impostos à regulação contemporânea frente às tecnologias emergentes.

### 3.1.1 Concretização de um direito humano essencial

O direito à saúde é reconhecido no direito internacional como um direito humano fundamental. Esse reconhecimento remonta à adoção da Constituição da Organização Mundial da Saúde (OMS), em 1946, onde se afirmou que o gozo do mais alto nível possível de saúde é um direito fundamental de todos os seres humanos. Desde então, esse princípio foi incorporado em diversos tratados e declarações internacionais, consolidando-se como uma norma central nos sistemas globais de proteção aos direitos humanos (Leary, 1994, p. 25).

O reconhecimento do direito à saúde como um direito humano fundamental encontra respaldo direto no direito internacional, estando consagrado na Declaração Universal dos Direitos Humanos, no Artigo 25, a declaração estabelece que toda pessoa tem direito a um padrão de vida adequado à saúde e ao bem-estar, incluindo alimentação, moradia, cuidados médicos e serviços sociais essenciais (ONU, 1948, p. 7). Esse reconhecimento foi amplamente desenvolvido no Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais, adotado pela Assembleia Geral das Nações Unidas em 1966 em seu Artigo 12 detalhando uma série de obrigações dos Estados para sua efetivação, como a redução da mortalidade infantil, controle de doenças, melhoria das condições sanitárias e a garantia de acesso a serviços médicos em caso de enfermidade (ONU, 1966, p. 6).

O entendimento atual sobre o direito à saúde estabelece que ele não se refere a um ideal abstrato de “ser saudável”, mas sim ao direito de acesso ao mais elevado nível possível de saúde física e mental, considerando as condições e recursos disponíveis. Isso inclui não apenas liberdades individuais, como o controle sobre o próprio corpo e a recusa a tratamentos indesejados, mas também direitos concretos, como o acesso universal e igualitário a serviços, bens e condições essenciais à saúde, com base na equidade e na dignidade humana (CESCR, 2000, p. 3).

A formulação “a incorporação dos direitos econômicos e sociais feita pelos Estados contemporâneos para além dos direitos relacionados às liberdades individuais” está fundamentada na leitura de Norberto Bobbio sobre

a transição do Estado liberal para o Estado social. Esse fenômeno é associado ao processo de positivação e ampliação dos direitos fundamentais, marcado pela consolidação de uma nova geração de direitos que, ao lado dos direitos civis e políticos (liberdades clássicas), incorpora os direitos econômicos e sociais como expressão da justiça distributiva e da igualdade material (Bobbio, 2010, p. 407).

A incorporação dos direitos sociais ao catálogo dos direitos fundamentais promoveu uma transformação profunda na atuação do Estado, que passou de uma postura predominantemente passiva, voltada à não intervenção nas liberdades individuais, para uma função ativa, voltada à realização de prestações positivas.

Enquanto os direitos civis e políticos se caracterizam por sua aplicabilidade imediata, com o Estado obrigado a abster-se de interferir na esfera individual, os direitos sociais, como o direito à saúde, requerem atuação concreta do poder público, com a criação de políticas, serviços e garantias materiais que viabilizem o seu exercício. Essa mudança representa uma ampliação substancial do conteúdo dos direitos fundamentais e redefine a própria relação entre indivíduo e Estado no âmbito constitucional (Böckenförde, 2017, p. 266-289), pois é obrigação do Estado garantir os direitos sociais e o direito à saúde.

### 3.1.2 Mecanismos normativos na promoção do direito à saúde

A regulação em saúde constitui um conjunto estruturado de normas, procedimentos e instrumentos institucionais organizados por uma autoridade pública reconhecida, com o propósito de orientar, coordenar e controlar as atividades e decisões relacionadas à saúde coletiva. Essa intervenção estatal, que se manifesta tanto no setor público quanto no privado, se concretiza por meio de atos normativos, legislações, incentivos e sanções, buscando promover eficiência, equidade e proteção dos direitos fundamentais à saúde.

Nesse sentido, a atividade regulatória não deve ser vista como uma limitação arbitrária de liberdades, mas como uma forma legítima de afirmação

do direito social à saúde, compatível com os princípios internacionais de direitos humanos (Gostin, 2003, p. 545-549).

De forma mais ampla, a regulação econômica na área da saúde compreende um conjunto de políticas e intervenções estatais cujo objetivo é ajustar o comportamento dos diferentes agentes envolvidos, como prestadores, financiadores e usuários, a fim de lidar com falhas de mercado, promover eficiência e garantir o acesso equitativo aos serviços de saúde (Goodman et al., 2013, p. 407).

A regulação de produtos destinados à saúde, como dispositivos médicos, medicamentos e alimentos, constitui um aspecto central das funções dos sistemas de saúde, dado seu papel essencial na mitigação de riscos tanto à saúde coletiva quanto individual. Em paralelo, destaca-se a regulação da assistência à saúde, que inclui normas referentes à prestação dos serviços e à organização dos recursos disponíveis, como os profissionais da saúde, as instituições e os sistemas em que operam. Esses mecanismos regulatórios não apenas asseguram a qualidade e a segurança das intervenções em saúde, como também buscam garantir a eficiência e equidade no acesso aos serviços ofertados (Folland et al., 2013, p. 407).

A atuação regulatória no campo da saúde se expressa tanto por vias indiretas quanto diretas. A primeira ocorre, por exemplo, através do sistema de responsabilidade civil, o qual, ao lidar com casos de danos pessoais e lesões, influencia comportamentos de indivíduos e instituições e pode repercutir nos resultados em saúde ao promover maior cautela nas práticas adotadas. Já a regulação direta envolve o estabelecimento de critérios normativos para a certificação de clínicas, o licenciamento de profissionais e a autorização de comercialização de medicamentos e dispositivos médicos, assegurando padrões mínimos de eficácia, segurança e qualidade nos serviços e produtos disponibilizados à população. Ambas as formas visam, de maneira complementar, a promover o equilíbrio entre acesso, custo e qualidade na assistência à saúde (Folland et al., 2013, p. 410).

A regulação ocupa papel central na concretização do direito à saúde, uma vez que é por meio dela que os Estados assumem compromissos concretos em termos de proteção e promoção desse direito. Inserida no conjunto de obrigações estatais, tanto no dever de proteger quanto no de

cumprir, a atividade regulatória não apenas estabelece normas e diretrizes para o setor da saúde, mas também representa um mecanismo efetivo para operacionalizar esse direito fundamental em políticas públicas, legislação e práticas institucionais (OHCHR, 2008, p. 25).

O compromisso dos Estados com o direito ao mais elevado padrão possível de saúde requer, de forma inegociável, uma regulação eficaz dos sistemas de saúde, ancorada nos princípios fundamentais dos direitos humanos. Para que esse direito seja concretizado, é imprescindível que a ação estatal não se restrinja à prestação de serviços médicos, mas abarque também o fortalecimento de mecanismos regulatórios, fundados em valores como igualdade, participação, transparência e responsabilidade. A garantia de um sistema de saúde acessível, integrado e responsivo depende, assim, do compromisso com uma regulação que reflita obrigações jurídicas claras e universais (Backman et al., 2008, p. 82-83).

### 3.1.3 A importância da regulação da inteligência artificial

A crescente presença de sistemas automatizados e algoritmos na prestação dos serviços de saúde tem gerado novos desafios à estrutura regulatória vigente, antes orientada por um modelo tradicional centrado na ação humana direta. Diferentemente do passado, em que médicos licenciados e instituições acreditadas desempenhavam papel exclusivo no atendimento presencial, hoje assistimos à emergência de tecnologias que descentralizam e virtualizam esse processo. Não é mais incomum que informações médicas sejam produzidas, interpretadas e aplicadas sem a mediação de um profissional de saúde, por meio de softwares intuitivos que acessam vastas bases de dados. Essa realidade questiona os fundamentos da regulação atual, exigindo uma revisão que considere não só os riscos reais dessas inovações, mas também os benefícios e novos modelos de cuidado que elas propõem (Richman, 2018, p. 1694).

Apesar da crescente presença de tecnologias digitais na área da saúde, os atuais marcos regulatórios ainda demonstram significativa inércia diante da

velocidade vertiginosa com que essas ferramentas são desenvolvidas. Em vez de acompanhar a dinâmica fluida e multifacetada das inovações digitais, marcadas por constante mutabilidade, diversidade e aprendizado adaptativo, os sistemas de regulação permanecem ancorados em modelos tradicionais, pouco responsivos à variedade e à variabilidade técnica dessas soluções. Tal descompasso compromete a capacidade regulatória de garantir padrões mínimos de segurança, eficácia e confiabilidade aos usuários finais, expondo pacientes e profissionais a riscos potenciais decorrentes da ausência de um enquadramento normativo específico (IQBAL et al., 2022, p. 2).

Nesse cenário de transição tecnológica na área da saúde, os algoritmos baseados em aprendizado de máquina (*Machine Learning* - ML) impõem obstáculos particularmente complexos à regulação. Tais ferramentas, que integram bases de dados robustas e operam de modo autônomo com múltiplas variáveis, apresentam uma dinâmica de funcionamento que difere radicalmente dos modelos tradicionais de apoio à decisão clínica. Ao contrário de ferramentas mais simples, os sistemas de ML contemporâneos, dotados de alta capacidade preditiva em tempo real, são de difícil interpretação por parte dos profissionais e não se submetem facilmente aos mesmos critérios utilizados para validar intervenções clínicas ou terapias convencionais. Isso confere a esses sistemas uma posição de destaque entre os desafios mais significativos enfrentados pelos órgãos reguladores na atualidade (Parikh et al., 2019, p. 810).

Uma das razões que tornam os algoritmos de inteligência artificial na saúde tão desafiadores para a regulação está na sua capacidade de modificação contínua. Ao aprenderem com dados reais e se ajustarem progressivamente, esses sistemas não apenas refinam sua acurácia ao longo do tempo, mas também ultrapassam os limites dos modelos fixos, tradicionalmente utilizados em contextos clínicos. Esse dinamismo adaptativo, embora potencialize melhorias no desempenho, torna-se um entrave para modelos regulatórios que ainda operam sob lógicas estáticas ou cíclicas de validação, típicas de fármacos ou dispositivos convencionais (Bates, 2023, p. 26).

Regulamentar tecnologias que, por sua própria natureza, se transformam ao longo do tempo representa um dos maiores dilemas

contemporâneos no campo jurídico-sanitário. Diferentemente de medicamentos ou dispositivos médicos, cujas características permanecem pouco dinâmicas após sua aprovação, os sistemas de inteligência artificial, especialmente os baseados em aprendizado de máquina, continuam se moldando com base em dados e experiências do mundo real. Tentar normatizá-los equivale a legislar sobre algo em constante metamorfose. Por isso, a construção de uma governança normativa para a IA em saúde não pode ser adiada: trata-se de um passo imprescindível, tanto sob a ótica da inovação quanto sob a salvaguarda dos direitos fundamentais e dos valores éticos que sustentam a ordem democrática e os sistemas de proteção à vida e à dignidade humana (UNIÃO EUROPEIA, 2024, p. 3-4).

### 3.1.4 Proteção da dignidade humana nas aplicações de IA

Diante da rápida expansão da inteligência artificial em contextos sensíveis da vida social, surge um campo que busca compreender e orientar seu impacto com base em valores éticos: a ética da IA. Trata-se de um domínio que não opera isoladamente, mas em constante diálogo com áreas como o Direito, a filosofia, a ciência da computação e as ciências sociais. Em vez de focar apenas nos avanços técnicos, essa abordagem destaca a importância de submeter o desenvolvimento e uso da IA a critérios que respeitem direitos humanos, justiça social e responsabilidade institucional. Embora essas tecnologias possam servir a causas nobres como facilitar o acesso à justiça, identificar práticas discriminatórias ou monitorar violações de direitos, também podem reforçar desigualdades e ameaçar liberdades quando não reguladas adequadamente. A ética da IA, nesse contexto, é chamada a refletir criticamente sobre o poder dessas tecnologias, promovendo sua articulação com normas sociais, legais e morais de forma responsável e transparente (Dubber et al., 2020, p. 2-4).

O ponto nevrálgico do debate internacional acerca da inteligência artificial e dos direitos humanos está na compreensão de que princípios éticos e jurídicos universais, como a dignidade humana, a liberdade, a igualdade e a

não discriminação, devem servir como parâmetros orientadores e limitadores do uso dessas tecnologias emergentes. Como os direitos humanos reconhecidos em tratados internacionais ratificados pelos Estados possuem força obrigatória, o direito internacional já dispõe de uma base normativa sólida que impõe deveres concretos a governos e organizações. Embora essa base não dispense a formulação de marcos regulatórios específicos para o setor, ela delineia os contornos éticos fundamentais a partir dos quais qualquer regime jurídico sobre inteligência artificial deve ser construído.

### 3.1.5 Diretrizes para o uso responsável da inteligência artificial

Foi em 2019 que a comunidade internacional reconheceu oficialmente o primeiro conjunto de diretrizes éticas voltadas à inteligência artificial, com a Recomendação do Conselho sobre inteligência artificial da OCDE (OCDE, 2019, p. 7-8). Embora não vinculante juridicamente, essa Recomendação rapidamente se consolidou como um referencial político global, influenciando formulações legislativas e políticas nacionais. A adesão inicial pelos 36 países membros da OCDE, seguida pela aprovação dos princípios pelo G20, sinalizou uma convergência multilateral em torno da necessidade de promover uma IA confiável, centrada no ser humano e alinhada com os direitos fundamentais. Entre os princípios destacados estão: a promoção do bem-estar coletivo e da sustentabilidade; a garantia de que os sistemas de IA respeitem valores democráticos, como igualdade, liberdade e não discriminação; o incentivo à transparência e explicabilidade nos processos automatizados; a exigência de robustez e segurança ao longo de todo o ciclo de vida dos sistemas; e, por fim, a responsabilização de todos os atores envolvidos quanto ao funcionamento ético e seguro dessas tecnologias (OCDE, 2019, p. 7-8).

Também em 2019, a União Europeia deu um passo decisivo rumo à construção de uma estrutura normativa e ética sólida para o desenvolvimento da inteligência artificial, com a publicação das Diretrizes Éticas para uma IA de Confiança. Elaboradas por um grupo independente de especialistas de alto nível, com representantes da academia, da indústria e da sociedade civil, essas

orientações não apenas se baseiam em fundamentos éticos universais, mas os operacionalizam por meio de princípios claros e requisitos concretos. A proposta parte de quatro pilares éticos: o respeito à autonomia humana, a prevenção de danos, a equidade e a explicabilidade. A partir desses princípios, são delineados sete requisitos que os sistemas de IA devem cumprir para serem considerados confiáveis: (a) supervisão e controle humanos, garantindo a autonomia na tomada de decisões; (b) solidez técnica e segurança, com ênfase na resiliência e confiabilidade dos sistemas; (c) privacidade e governança de dados, com respeito à integridade, qualidade e legitimidade no uso dos dados; (d) transparência, com rastreabilidade e explicação dos resultados; (e) diversidade, não discriminação e justiça, assegurando acessibilidade e imparcialidade; (f) bem-estar social e ambiental, considerando impactos humanos e ecológicos; e (g) responsabilidade, exigindo auditoria, prevenção de danos e mecanismos de reparação (UE, 2019, p. 10-20).

Embora haja relativa uniformidade internacional quanto aos valores fundamentais que devem orientar o uso da inteligência artificial, como a justiça, a proteção dos direitos humanos e a promoção do bem-estar, ainda persistem diferenças expressivas nas formas de aplicar esses valores em práticas éticas concretas. Uma alternativa amplamente discutida no campo é a transposição dos princípios clássicos da bioética para o universo da IA: autonomia, não-maleficência, beneficência e justiça. A esses, propõe-se somar um quinto princípio essencial, adaptado às especificidades tecnológicas: a explicabilidade, entendida como a capacidade de compreender como os sistemas funcionam (inteligibilidade) e de atribuir responsabilidade por suas ações (*accountability)*. Este princípio atua como elemento articulador dos demais, sendo indispensável para tornar possível uma governança ética da IA. Ao integrar essas cinco diretrizes, constrói-se uma estrutura sólida capaz de orientar decisões sobre oportunidades e riscos, ancorando a IA em uma base ética compreensível e aplicável à realidade social (Floridi et al., 2018, p. 696-700)

Entre a segunda metade da década de 2010 e o início dos anos 2020, a governança da inteligência artificial começou a se estruturar, ainda que de forma preliminar e não vinculante. Nesse período, destacaram-se os códigos de conduta voluntários e diretrizes éticas, instrumentos típicos da chamada soft

law, como formas temporárias de orientação enquanto projetos de leis específicas para IA estavam em fase de debate e elaboração, sem ainda terem sido convertidos em normas jurídicas efetivas. Simultaneamente, no campo da proteção de dados, observou-se um avanço notável e mais sólido: as legislações de privacidade passaram a se tornar praticamente universais, com mais de 160 países adotando leis específicas sobre o tema até 2023. Esse movimento posicionou a regulação de dados como uma das mais amplamente difundidas no mundo, contrastando com a relativa lentidão da normatização específica para IA (Greenleaf, 2023, p. 1-3).

Apesar da enorme diversidade de instituições, desde órgãos governamentais e conselhos consultivos até empresas de tecnologia e centros de pesquisa, que têm elaborado documentos sobre ética na inteligência artificial, observa-se um padrão de convergência global emergente em torno de cinco princípios considerados fundamentais: transparência, justiça e equidade, não-maleficência, responsabilidade e privacidade. Esses princípios são amplamente mencionados em mais da metade dos documentos analisados, revelando um consenso inicial quanto aos valores centrais que devem nortear o desenvolvimento e uso responsável da IA. No entanto, essa aparente uniformidade esconde uma série de tensões: ainda há significativa disparidade quanto ao significado preciso desses princípios, aos motivos que justificam sua importância, aos atores e domínios a que se aplicam, bem como às estratégias práticas para sua efetiva implantação. Tal heterogeneidade demonstra que, embora o discurso ético esteja se consolidando, a tradução dos princípios em políticas concretas e práticas normativas ainda representa um desafio (Jobin et al., 2019, p. 6-7).

## 3.2 Diretrizes éticas aplicadas à inteligência artificial na saúde

Já na segunda metade da década de 2010, surgiram os primeiros debates estruturados sobre a necessidade de princípios éticos específicos para orientar o uso da inteligência artificial na área da saúde. Esses debates ganharam força à medida que o potencial revolucionário do *Machine Learning*

(ML) se tornava evidente nesse campo. Com a popularização das tecnologias de *Big Data* e o avanço acelerado dos algoritmos, cresceu a percepção de que os instrumentos tradicionais de proteção à dignidade humana e aos direitos fundamentais eram insuficientes diante da complexidade e da opacidade desses sistemas. A UNESCO, por meio do seu Comitê Internacional de Bioética (IBC), foi uma das primeiras entidades globais a alertar para a urgência de refletir criticamente sobre as implicações éticas, jurídicas e sociais do uso intensivo de dados na saúde, incluindo aspectos como autonomia, consentimento, governança e justiça no acesso e tratamento das informações (UNESCO, 2017, p. 1-3).

É importante destacar que em setembro de 2023, ela publicou o "Relatório de Avaliação de Preparação para Inteligência Artificial no Brasil", resultado da aplicação da metodologia RAM (Readiness Assessment Methodology), voltada para a implementação ética da IA. O documento apresenta 17 recomendações estruturais, que envolvem áreas como regulação, educação, pesquisa, infraestrutura, diversidade e proteção de direitos (UNESCO, 2023).

Em 2021, a Organização Mundial da Saúde apresentou sua primeira orientação global dedicada à ética e à governança da inteligência artificial na saúde, fruto de um extenso trabalho conduzido ao longo de dois anos por especialistas de diversas áreas. O documento consolidou os primeiros princípios amplamente aceitos para nortear o uso ético da IA no setor da saúde, com foco em garantir que essas tecnologias avancem de forma alinhada aos direitos humanos e aos compromissos sociais. Os seis princípios fundamentais propostos são: (a) proteção da autonomia humana; (b) promoção do bem-estar, da segurança e do interesse público; (c) garantia de transparência, explicabilidade e inteligibilidade; (d) incentivo à responsabilidade e prestação de contas; (e) promoção da inclusão e da equidade; e (f) desenvolvimento de uma IA responsiva e sustentável. Com base na bioética tradicional e nos valores da ética digital contemporânea, essas diretrizes foram elaboradas para guiar desenvolvedores, reguladores e usuários ao longo de todas as fases de concepção, aplicação e avaliação contínua de sistemas baseados em IA no campo da saúde (OMS, 2021).

### 3.2.1 Preservação da autonomia do paciente

O princípio da autonomia, considerado uma das pedras angulares da bioética moderna, está intrinsecamente ligado à valorização da liberdade individual e da capacidade de autodeterminação. Ele estabelece que cada pessoa tem o direito de tomar decisões fundamentadas sobre sua própria vida, corpo e cuidados de saúde, desde que esteja em condições de compreender adequadamente as informações e livre de coerções externas. Mais do que mera abstenção de interferência, respeitar a autonomia exige, em muitos casos, o dever ativo de criar condições para que o indivíduo possa exercer sua escolha com liberdade, compreensão e consentimento informado. Essa concepção ética surgiu como reação à tradição médica paternalista e foi consolidada nas últimas décadas como base normativa para a relação entre profissionais de saúde e pacientes, especialmente em contextos de tomada de decisão sensível, como recusa de tratamento, consentimento para pesquisas e definição de planos terapêuticos (Beauchamp, 2010, p. 36-37).

Na ética da inteligência artificial aplicada à saúde, respeitar a autonomia significa garantir que os seres humanos permaneçam no controle das decisões clínicas e dos sistemas de saúde. A tecnologia deve apoiar, e não substituir, a capacidade de escolha de pacientes e profissionais. Isso inclui assegurar que os usuários compreendam o papel da IA nos cuidados prestados e tenham acesso às informações necessárias para seu uso responsável (OMS, 2021, p. 12).

Os modelos de *Machine Learning* aplicados à saúde vêm demonstrando um imenso potencial para ampliar a precisão de diagnósticos, prever desfechos clínicos e sugerir condutas terapêuticas com base em padrões extraídos de grandes volumes de dados. No entanto, ao mesmo tempo que esses recursos tecnológicos podem otimizar a qualidade da assistência, também representam um desafio ético central: o risco de que suas recomendações sejam seguidas de forma automática, sem a devida consideração pelos valores, preferências e autonomia dos pacientes. Quando a lógica de funcionamento desses sistemas é opaca ou de difícil compreensão, há um enfraquecimento do processo de decisão compartilhada entre profissionais e usuários do sistema de saúde. Por

isso, garantir a **interpretabilidade** e a **transparência** das decisões geradas por IA na saúde é essencial para que a tecnologia atue como apoio e não substituição à deliberação humana (Mittelstadt, 2023, p. 2-4).

No caso dos sistemas de inteligência artificial voltados ao autocuidado, como os dispositivos vestíveis (*wearables*), o fornecimento direto de informações ao paciente pode comprometer sua capacidade de decisão autônoma. Embora essas tecnologias sejam frequentemente apresentadas como ferramentas de empoderamento individual, capazes de ampliar o controle sobre a saúde pessoal, estudos mostram que elas também podem impor normas, metas e padrões que pressionam os usuários a tomarem decisões baseadas em critérios externos, muitas vezes comerciais ou biomédicos, em vez de refletirem suas próprias preferências ou valores (Owens et al., 2019, p. 24-26).

A autonomia dos pacientes e profissionais está intimamente ligada à transparência dos sistemas de inteligência artificial aplicados à saúde. Para que a autonomia seja efetivamente respeitada, é necessário que os prestadores de cuidado tenham acesso às informações necessárias para compreender os sistemas de IA. Do mesmo modo, os próprios pacientes devem ser capazes de entender o papel da IA em seu processo de cuidado. A ausência de transparência impede o julgamento crítico e enfraquece a capacidade de decisão autônoma dos envolvidos (Char et al., 2020, p. 9).

A proteção da autonomia dos pacientes no uso da inteligência artificial requer mais do que consentimento formal: exige que as pessoas sejam devidamente informadas sobre o papel da IA nas decisões clínicas que as afetam, além de poderem compreender, questionar e, se necessário, contestar tais decisões. A opacidade característica de muitos sistemas baseados em *Machine Learning* pode comprometer esse processo. Como destacam Grote et al. (2020), a ausência de explicações claras e acessíveis por parte dos algoritmos pode transformar o consentimento informado em um mero procedimento formal, esvaziado de significado. Para que o direito à autodeterminação seja respeitado, é indispensável garantir a transparência das decisões automatizadas (Grote et al., 2020, p. 210).

A proteção da autonomia no uso da inteligência artificial na saúde está diretamente ligada ao respeito à privacidade, confidencialidade e o

consentimento informado. É imprescindível que a utilização de modelos de *Machine Learning*, sobretudo em funções como diagnóstico, prognóstico e formulação de planos terapêuticos, esteja claramente inserida no processo de consentimento válido. Isso significa ir além de acordos genéricos de usuário: os pacientes devem ser efetivamente informados sobre o uso de IA e sobre como seus dados serão processados. (Gerke et al., 2020, p. 301-303).

### 3.2.2 Entre a proteção do indivíduo e o interesse público

Na ética da inteligência artificial, especialmente quando aplicada à saúde, o compromisso com o bem-estar humano, a segurança e o interesse público se expressa na integração dos princípios tradicionais da bioética: não-maleficência e beneficência. O primeiro, centrado na exigência de que as tecnologias não causem danos. Já o segundo vai além da mera prevenção de riscos, exigindo que os sistemas de IA sejam desenvolvidos com a finalidade de gerar impactos positivos concretos, tanto para os indivíduos quanto para a coletividade e o meio ambiente. A ética aplicada à IA, portanto, deve operar simultaneamente na contenção dos potenciais danos e na ampliação de benefícios ao paciente (Floridi et al., 2018, p. 696-697).

A prática segura da inteligência artificial na saúde exige não apenas validações iniciais, mas também um arcabouço regulatório robusto e contínuo. Para tanto, é necessário estabelecer requisitos de segurança voltados a usos bem definidos, bem como instituir mecanismos de controle de qualidade para o monitoramento contínuo do desempenho das aplicações de IA. Reguladores como o FDA (nos EUA) e a EMA (na União Europeia) vêm discutindo adaptações nos modelos de avaliação para incorporar abordagens de ciclo de vida, com ênfase em vigilância pós-comercialização, rastreabilidade e auditoria técnica dos algoritmos. O objetivo é garantir que os sistemas de IA permaneçam confiáveis mesmo após sua introdução no mercado (Minssen et al., 2020, p. 5-6).

A aplicação ética de sistemas de inteligência artificial na saúde requer que esses sistemas operem dentro de parâmetros bem definidos de segurança,

precisão e eficácia clínica. Isso implica que, desde o desenvolvimento até a utilização, a IA deve seguir normas que evitem danos aos pacientes e garantam previsões corretas para a maioria dos contextos. Essa exigência se traduz na necessidade de validações contínuas, assegurando que os algoritmos sejam calibrados adequadamente para reduzir falsos positivos e negativos (Char et al., 2020, p. 8).

Dada a natureza evolutiva dos modelos de *Machine Learning* utilizados na saúde, cuja capacidade de adaptação se apoia em dados em constante atualização, a avaliação de segurança não pode ser pontual: ela precisa ser feita de forma contínua e, idealmente, em tempo real. Isso exige que os sistemas regulatórios adotem protocolos específicos de monitoramento pós-comercialização e mecanismos automatizados de rastreamento de mudanças nos algoritmos. Como demonstrado por Gilbert et al. (2021), é essencial que as atualizações dos modelos não comprometam a eficácia clínica nem a segurança do paciente, o que demanda estruturas regulatórias ágeis e mecanismos de vigilância contínua que estejam integrados ao próprio ciclo de vida da tecnologia (Gilbert et al., 2021, p. 1-3).

Em aplicações de inteligência artificial na saúde, é essencial adotar estratégias que previnam qualquer tipo de dano. Isso envolve não apenas o cumprimento de padrões técnicos de segurança, mas também o cuidado ético com os efeitos sociais dessas tecnologias. Os sistemas devem ser projetados com atenção especial aos contextos em que serão utilizados, prevendo mecanismos que evitem a exposição dos pacientes a situações de vulnerabilização, como o risco de discriminação e exclusão social decorrente de informações sensíveis. Ao equilibrar a utilidade das respostas automatizadas com salvaguardas robustas, cria-se um ambiente seguro para o uso da IA (OM S, 2021, p. 13).

Isso implica que os desenvolvedores devem garantir autorizações válidas para uso e reuso dos dados, respeitando os diferentes regimes jurídicos, tipos de dados e propósitos envolvidos. Incorporar essas obrigações no ciclo de vida dos algoritmos, desde a coleta até a aplicação clínica, não é apenas uma exigência legal, mas também uma medida fundamental para preservar a confiança e a legitimidade social do uso da IA em saúde (Vayena et al., 2018, p. 2).

A promoção do interesse público constitui um princípio ético fundamental para a regulação da inteligência artificial, especialmente quando aplicada em contextos sensíveis como a saúde. Isso implica uma obrigação compartilhada entre desenvolvedores, reguladores e usuários de assegurar que os benefícios decorrentes do uso da IA ultrapassem o nível individual e sejam distribuídos de forma justa em toda a sociedade. O objetivo não é apenas gerar avanços técnicos, mas sim orientar esses avanços de forma a contribuir para a coesão social, a equidade no acesso a benefícios tecnológicos e a proteção dos bens coletivos, pois, de acordo com Floridi et al. (2018), essa é uma forma de alinhar o desenvolvimento da IA com os valores da dignidade humana (Floridi et al., 2018, p. 693).

A equidade no acesso aos avanços tecnológicos é uma condição indispensável para que a inteligência artificial cumpra seu papel no fortalecimento dos sistemas de saúde em escala global (Schwalbe et al., 2020, p. 1580-1582).

É essencial que os benefícios proporcionados pela IA sejam amplamente compartilhados, considerando não apenas os indivíduos, mas as necessidades coletivas das populações, especialmente em países de baixa e média renda. Isso significa evitar a concentração dessas tecnologias em ambientes privilegiados, promovendo seu uso também em contextos onde há maior escassez de recursos.

### 3.2.3 Clareza algorítmica e compreensão acessível no uso da IA

Entre os diversos princípios éticos presentes nas diretrizes internacionais sobre inteligência artificial, destaca-se a transparência como o mais recorrente, frequentemente visto como condição indispensável para viabilizar outros valores éticos, como a responsabilidade, e a proteção à privacidade. Sua presença em praticamente todas as diretrizes analisadas sugere uma convergência global sobre a centralidade desse princípio, ainda que haja divergências quanto à forma de aplicá-la na prática (Jobin et al., 2019, p. 6-8).

Para que a inteligência artificial seja eticamente aplicada na área da saúde, é essencial que seu funcionamento não permaneça restrito a especialistas técnicos. O princípio da transparência exige que os sistemas sejam apresentados de forma que sejam compreendidos por todos os envolvidos, incluindo médicos, pacientes, reguladores e desenvolvedores. (OMS, 2021, p. 13).

Além disso, para que possam ser auditáveis e operem de forma segura, é fundamental que seus pressupostos, limitações e métodos de funcionamento estejam bem documentados e acessíveis. Isso inclui não apenas os protocolos operacionais, mas também os critérios utilizados na seleção, no processamento e na rotulagem dos dados, além das condições sob as quais os modelos foram desenvolvidos e validados (Liu et al., 2022, p. e386-e390).

A explicabilidade dos sistemas deve ser adaptada ao grau de entendimento dos diferentes usuários, profissionais de saúde, pacientes, reguladores e desenvolvedores, permitindo que todos compreendam, ainda que em níveis distintos, os fundamentos e condições das decisões automatizadas (Watson et al., 2019, p. 2-4). No atual cenário da inteligência artificial em saúde, tornar os modelos explicáveis representa um dos desafios mais relevantes e controversos da regulação.

Embora a demanda por explicabilidade tenha ganhado amplo apoio de pesquisadores, formuladores de políticas e sociedade civil, sua efetiva operacionalização ainda enfrenta limites conceituais e técnicos. A principal dificuldade reside no fato de que muitos sistemas utilizados são baseados em modelos de "caixa-preta", como redes neurais profundas, cujas decisões não são diretamente compreensíveis (Babic et al., 2021a, p. 1-3).

Por fim, explicabilidade pode ser entendido como a combinação entre inteligibilidade e responsabilização, é um princípio fundamental para garantir a ética no uso da IA em saúde. (Floridi et al., 2018, p. 700), assim, garantir que princípios como autonomia, justiça distributiva, responsabilização e controle institucional sejam efetivamente observados implica assegurar a inteligibilidade e a transparência dos mecanismos que regem o funcionamento da inteligência artificial, possibilitando que seus impactos sejam compreendidos.

### 3.2.4 Mecanismos de *accountability* na aplicação da IA

Desenvolvedores e prestadores de serviços devem assegurar que a IA opere dentro dos limites para os quais foi projetada, respeitando os objetivos clínicos e éticos definidos. Para isso, é recomendável que haja supervisão humana contínua, permitindo acompanhar o desempenho dos sistemas, prevenir usos indevidos e garantir que qualquer impacto negativo possa ser identificado. (OMS, 2021, p. 13).

Além da obrigação de garantir a segurança e a funcionalidade dos sistemas, deve-se assegurar meios eficazes de atribuição de responsabilidades e de reparação em casos de danos causados por decisões automatizadas. Isso é especialmente relevante em contextos nos quais os impactos das decisões de IA afetam direta ou indiretamente indivíduos e grupos sociais.

Ainda que os termos “responsabilidade” e “prestação de contas” (*accountability)* muitas vezes sejam usados como equivalentes, trata-se de conceitos distintos na ética da IA. A responsabilidade diz respeito ao vínculo entre os agentes humanos e suas decisões, sendo anterior à ação: é o dever de agir com consciência de suas implicações. Já a *accountability* se manifesta no momento posterior, exigindo que os responsáveis justifiquem suas escolhas. Na IA, isso implica mapear as decisões dos sistemas até os dados e as pessoas que os construíram, permitindo rastrear a cadeia de ações em processos complexos e distribuídos.

Essa distinção se torna ainda mais relevante à medida que aumenta o número de atores envolvidos no desenvolvimento, treinamento e aplicação de modelos algorítmicos, criando o risco de lacunas de responsabilidade (“*accountability gaps*”), especialmente quando não se consegue mais apontar com clareza quem responde por decisões automatizadas (Lechterman, 2022, p. 165-168).

Esse princípio envolve a expectativa contínua de conformidade com critérios estabelecidos, o que exige o acompanhamento do desempenho da IA, a produção de relatórios sobre suas decisões, a possibilidade de revisão por parte de órgãos competentes e, quando necessário, a aplicação de sanções. Na prática, *accountability* opera como uma estrutura que exige justificativas

claras e mecanismos formais para monitoramento e controle das ações realizadas por meio da IA (Novelli et al, 2024, p. 1872-1874).

Além disso, os responsáveis pelo desenvolvimento e aplicação desses sistemas devem ser capazes de justificar suas decisões de forma clara, tanto aos usuários quanto ao público em geral. Essa exigência se articula diretamente com a transparência e exige conformidade com valores éticos. Da mesma forma, o princípio ético da responsabilidade está intimamente ligado à noção de responsabilização civil, pois fundamenta a reparação de eventuais danos causados pelo uso inadequado de sistemas algorítmicos na assistência à saúde (Hatherley, 2023, p. 2-3).

### 3.2.5 Justiça distributiva e acesso equitativo às inovações

Na ética da saúde, a promoção de **equidade e inclusão** está diretamente vinculada ao princípio da justiça, entendido como o compromisso com a distribuição equilibrada de recursos. Esse princípio exige que os benefícios na saúde, inclusive os trazidos pela inteligência artificial, sejam oferecidos de forma proporcional às necessidades dos diferentes grupos sociais. Assim, torna-se fundamental garantir que nenhuma população seja negligenciada, e que as decisões que envolvem o uso de IA contribuam para a redução das desigualdades existentes (Beauchamp, 2010, p. 40-41).

Esse princípio repousa sobre os fundamentos éticos da justiça distributiva, conforme delineado por Rawls, e está ligado à ideia de que os benefícios e encargos sociais devem ser compartilhados de maneira justa entre todos os membros da sociedade. Em sua formulação, Rawls argumenta que uma sociedade justa é aquela em que a distribuição dos recursos não é determinada por circunstâncias arbitrárias, como classe social ou talentos naturais, mas por princípios que todos aceitariam em condições de equidade. Assim, os conceitos de justiça e equidade, quando aplicados à regulação da inteligência artificial em saúde, exigem que as vantagens trazidas por essas tecnologias estejam acessíveis a diferentes grupos sociais, e não apenas aos mais privilegiados (Rawls, 1999, p. 3-5).

É importante que as tecnologias sejam desenvolvidas de modo a garantir acesso justo, independentemente de quaisquer fatores ou qualquer outra característica pessoal. O objetivo é garantir que os benefícios da IA não se limitem a grupos privilegiados, mas sejam distribuídos de forma justa entre todos os segmentos da sociedade. (OMS, 2021, p. 13).

Além do compromisso com a equidade, o uso ético da inteligência artificial também pressupõe que essas tecnologias sejam acessíveis em escala global, especialmente nos países em desenvolvimento. Nesse cenário, os algoritmos precisam ser desenvolvidos com dados representativos e diversidade suficiente para evitar a reprodução de preconceitos já presentes nas sociedades. O objetivo deve ser não apenas impedir a exclusão de determinados grupos sociais, mas reduzir disparidades, promovendo resultados positivos de forma inclusiva e sustentável (Schwalbe et al., 2020, p. 1583-1585).

Para garantir que os sistemas de inteligência artificial não reforcem injustiças históricas, é indispensável estabelecer mecanismos permanentes de avaliação. Essa supervisão contínua é necessária para corrigir distorções que possam reproduzir preconceitos nos dados em que a IA é aplicada. Avaliar o impacto dos modelos é uma estratégia essencial para alinhar o desenvolvimento tecnológico com os princípios de equidade na saúde (Rajkomar et al., 2018, p. 867-869).

Quanto ao desenvolvimento na saúde, é fundamental que todos os envolvidos estejam ativamente comprometidos em reduzir possíveis vieses. Esse cuidado deve ser incorporado desde as etapas iniciais, a fim de evitar que decisões automatizadas causem danos ou desigualdades, especialmente em grupos vulnerabilizados (Ntoutsi et al., 2020, p. 2-4).

Além do mais, a transparência ao longo de todo o ciclo de vida dos sistemas de IA é essencial para permitir que todos os envolvidos, especialmente pacientes e profissionais da saúde, tenham acesso equitativo às explicações das decisões automatizadas. Essa exigência conecta-se diretamente ao conceito de "direito à justificação", que representa o reconhecimento de que cada pessoa tem o direito de questionar e receber respostas sobre decisões que a afetam (Giovanola et al., 2023, p. 557).

### 3.2.6 Ética na adaptabilidade tecnológica e a sustentabilidade sistêmica

Na ética da IA aplicada à saúde, ***responsividade*** e ***sustentabilidade*** são considerados aspectos complementares de um mesmo princípio. Isso significa que as tecnologias devem ser continuamente avaliadas quanto à sua capacidade de responder de forma adequada às necessidades dos usuários e aos objetivos declarados de sua aplicação. Ao mesmo tempo, devem ser projetadas com foco na sustentabilidade dos sistemas de saúde, respeito ao meio ambiente e na preparação dos trabalhadores para lidar com mudanças tecnológicas, **garantindo que os benefícios sejam duradouros** (OMS, 2021, p. 14).

Para que as tecnologias de IA cumpram seu papel social, devem ser concebidas com capacidade de resposta às necessidades das populações locais em termos de saúde. Ao mesmo tempo, esses sistemas precisam ser sustentáveis do ponto de vista econômico e ambiental, promovendo benefícios duradouros sem comprometer recursos futuros. (UNESCO, 2021, p. 21), assim a Recomendação sobre a Ética da Inteligência Artificial da UNESCO, aprovada por unanimidade em novembro de 2021 por seus 193 Estados-membros, representa um importante instrumento normativo global sobre o tema da ética na IA. Seu objetivo é assegurar que o desenvolvimento e a aplicação dessas tecnologias estejam alinhados com os direitos humanos, a dignidade da pessoa, a justiça social e o desenvolvimento sustentável.

A Recomendação é estruturada em torno de quatro valores fundamentais: 1) Respeito, proteção e promoção dos direitos humanos, das liberdades fundamentais e da dignidade humana; 2) Promoção da justiça e do bem-estar; 3) Inclusão e diversidade; 4) Ambientes sustentáveis e prósperos. Ela propõe que esses princípios sejam traduzidos em políticas públicas nacionais, legislações e estratégias institucionais. A Recomendação também apresenta ferramentas práticas como a Metodologia de Avaliação de Prontidão Ética (RAM), que já começou a ser aplicada em países como o Brasil, reforçando a aderência prática à recomendação (UNESCO, 2021).

No intuito de garantir a responsividade dos sistemas de IA, é fundamental que sua adequação às necessidades locais sejam avaliadas de

forma contínua. Ao mesmo tempo, o desenvolvimento e a implantação de soluções de IA devem contar com a participação ativa de setores, como governos, universidades, ONGs e centros de pesquisa, especialmente em países com menos recursos. Essa abordagem colaborativa é essencial para garantir que a tecnologia atenda de fato às demandas de forma justa e sustentável (Alami et al., 2020, p. 4), além do compromisso de governos e empresas na capacitação profissional e na gestão das transformações no mercado de trabalho, promovendo treinamento específico para os profissionais de saúde, a fim de prepará-los para o uso dessas tecnologias, e também adotar estratégias para minimizar impactos negativos da automação em sistemas de saúde com infraestrutura limitada (Hosny et al., 2019, p. 955-956).

## 3.3 Gestão e responsabilidade sobre dados na saúde

Nos últimos anos, tem-se repetido com frequência a analogia de que os dados representam o novo “petróleo” do século XXI, dado seu enorme valor estratégico, econômico e político. Seguindo essa metáfora, pode-se dizer que a inteligência artificial desempenha hoje um papel transformador análogo ao motor a combustão no século passado.

Assim como os motores redefiniram padrões industriais, urbanos e sociais, a IA tende a reconfigurar profundamente os sistemas de produção, os serviços públicos e, em especial, os cuidados em saúde, baseando-se no tratamento intensivo de dados para gerar eficiência, inovação e novas formas de organização social (Nascimento, 2020, p. 34).

Seja qual for a analogia empregada para ilustrar a ascensão da inteligência artificial, permanece evidente que sua consolidação enquanto motor da revolução digital está diretamente condicionada à disponibilidade de grandes volumes de dados confiáveis. Para que essa dinâmica se desenvolva de maneira legítima, é imprescindível o estabelecimento de arcabouços institucionais robustos voltados à governança responsável desses dados.

### 3.3.1 Noções fundamentais

Controlar quem decide sobre os dados, como essas decisões são tomadas e com quais critérios é fundamental para garantir o uso responsável e estratégico das informações. Essa estrutura organizacional de autoridade e gestão envolve definir responsabilidades claras, estabelecer regras para o acesso e garantir que as tecnologias associadas aos dados sejam utilizadas de forma ética, segura e eficaz ao longo do tempo (Plotkin, 2021, p. 1).

Definição clara de regras, responsabilidades e padrões para a gestão dos dados são imprescindíveis para o controle da IA. Organizações envolvidas devem aplicar e fiscalizar continuamente normas que garantam a responsabilidade e a prestação de contas ao longo de todo o ciclo de vida dos dados e dos algoritmos. Esse processo envolve a administração criteriosa da disponibilidade, usabilidade, qualidade, integridade, privacidade e segurança da informação. Por isso, a governança de dados é reconhecida como alicerce indispensável para o desenvolvimento de uma IA confiável (Janssen et al., 2020, p. 3-5).

Nas últimas duas décadas, a ideia do que constitui um “dado de saúde” deixou de se restringir aos registros clínicos tradicionais. Hoje, informações geradas em redes sociais, aplicativos, dispositivos vestíveis e mecanismos de busca passaram a ser integradas a esse campo, compondo um ecossistema complexo conhecido como megadados biomédicos. Esse conjunto reúne, além dos dados clínicos e administrativos, informações sociais, comportamentais, econômicas e ambientais que, embora inicialmente externas ao setor da saúde, tornam-se relevantes quando processadas por tecnologias de inteligência artificial. Diante dessa ampliação, ganha força o consenso sobre a necessidade de novas formas de governança, que garanta o acesso ético a esses dados (Vayena et al., 2017, p. 502-503).

### 3.3.2 A proteção de dados pessoais como pilar regulatório

A crescente utilização de dados para treinar sistemas de inteligência artificial aplicados à saúde traz à tona preocupações importantes. Questões relacionadas à qualidade dos dados, sobretudo aqueles provenientes de países pobres, podem comprometer a eficácia dos modelos. Além disso, a sub-representação de determinados grupos sociais pode perpetuar preconceitos nos resultados fornecidos pelos algoritmos. Em paralelo, a proteção da privacidade permanece como uma das principais inquietações, considerando os riscos de discriminação (OMS, 2023, p. 27-33).

O uso excessivo de dados de forma inadequada levantam sérios dilemas éticos, sobretudo no que diz respeito à proteção dos direitos humanos. Um dos riscos mais apontados nesse contexto é o da concentração de poder nas mãos de quem controla os dados, em detrimento daqueles que os fornecem. Esse cenário, por vezes chamado de “colonialismo de dados”, tende a acirrar desigualdades entre países, especialmente quando não há salvaguardas robustas para assegurar consentimento válido, privacidade e autonomia no tratamento dessas informações (OMS, 2021, p. 37).

A construção de políticas e estruturas normativas para governança de dados, em diferentes países, tem sido fortemente influenciada pelos marcos jurídicos de proteção de dados pessoais, com destaque para os paradigmas europeu e norte-americano. Nos últimos anos, observa-se predominância crescente do modelo europeu, especialmente com a difusão do Regulamento Geral sobre a Proteção de Dados (RGPD). A ampla disseminação de leis nacionais sobre privacidade consolidou uma espécie de consenso regulatório global. Segundo Greenleaf et al. (2023), mais de 160 países já possuem legislações específicas sobre o tema, o que demonstra a consolidação da privacidade de dados como princípio jurídico quase universal (Greenleaf et al., 2023, p. 1-3).

Essa consolidação se deu, em grande medida, a partir da convergência de princípios oriundos da proteção de dados pessoais e da privacidade, fortemente influenciada pelos modelos regulatórios europeus. Ainda que inexistam tratados internacionais vinculantes dedicados exclusivamente à

proteção de dados, a multiplicidade de legislações nacionais em vigor, somada à adoção de instrumentos normativos não obrigatórios por organizações como a OCDE e o Conselho da Europa, permitiu a formação de um sistema global normativo funcionalmente integrado, capaz de operar mesmo diante da diversidade normativa. Esse sistema sustenta-se, em boa medida, pela chamada interoperabilidade normativa, que permite o diálogo entre diferentes regimes jurídicos (Bernier et al., 2022, p. 2-4).

A proteção da privacidade e dos dados pessoais figura como uma das dimensões já consolidadas do arcabouço jurídico. Regras, como a HIPAA nos Estados Unidos e a RGPD na União Europeia, estabelecem obrigações específicas para garantir a confidencialidade das informações de saúde, exigindo que seu uso, mesmo para fins de desenvolvimento e validação de sistemas de IA, seja pautado por critérios éticos. Tais exigências se refletem, por exemplo, na necessidade de anonimização de dados em aplicações secundárias (Mcnair et al., 2019, p. 220-221).

A salvaguarda dos dados pessoais emerge como um dos alicerces normativos indispensáveis à regulação da inteligência artificial no domínio da saúde, articulando-se com princípios éticos. Desta forma, o direito à privacidade ocupa posição central na arquitetura da governança informacional, especialmente diante da natureza sensível dos dados clínicos e da crescente complexidade dos sistemas tecnológicos que os manipulam.

### 3.3.3 Proteção da intimidade e dos dados pessoais sensíveis

No campo da regulação da inteligência artificial, especialmente em sua aplicação à saúde, a proteção da privacidade desponta como um ponto de partida incontornável. Mais do que uma exigência normativa, ela constitui um valor estruturante que influencia todo o desenho do arcabouço regulatório, servindo como base para o estabelecimento de limites éticos e jurídicos ao uso de dados sensíveis.

A construção desse conceito não é recente. Já no final do século XIX, começou a ganhar força a noção de que as pessoas deveriam ter o direito de

controlar sua exposição ao mundo, protegendo aspectos da vida pessoal contra interferências indevidas. Foi nesse contexto que Warren e Brandeis (1890) formularam a célebre ideia de um “direito de estar só”, uma resposta ao crescimento da vigilância social e ao avanço das tecnologias de comunicação da época. Essa concepção inaugurou uma nova forma de compreender a liberdade individual, que hoje se projeta sobre os desafios da era digital e da inteligência artificial (Warren et al., 1890, p. 195). Essa perspectiva aproxima-se da concepção contemporânea de privacidade tal como delineada nos marcos atuais da governança de dados.

Com a transição para o Estado Social no pós-guerra, a privacidade foi reconhecida como um direito humano fundamental, sendo incorporada à Declaração Universal dos Direitos Humanos de 1948. O artigo 12 estabelece que ninguém deve sofrer interferências arbitrárias em sua vida privada, garantindo proteção legal contra abusos à intimidade, honra e reputação (ONU, 1948, p. 5).

Ao longo das últimas décadas, a privacidade foi gradualmente consolidada como um direito fundamental, impulsionada tanto por desenvolvimentos doutrinários quanto por decisões de cortes internacionais de direitos humanos. Esse direito é hoje compreendido como a garantia de que cada pessoa deve dispor de um espaço pessoal livre de interferências indevidas, sendo respeitada como sujeito autônomo em suas escolhas. No entanto, assim como ocorre com outros direitos, a privacidade não é absoluta: ela deve ser equilibrada com outros interesses da sociedade, como a segurança pública, a saúde coletiva ou os direitos de terceiros (Townend, 2022, p. 74-75).

Em 1988, o Comitê de Direitos Humanos da ONU reconheceu a urgência de proteger a privacidade diante dos avanços tecnológicos e postulou a necessidade de legislações específicas de proteção de dados pessoais para salvaguardar esse direito. A recomendação foi fundamentada na interpretação do Pacto Internacional sobre Direitos Civis e Políticos, destacando que os Estados devem adotar medidas legislativas eficazes para evitar interferências arbitrárias na vida privada. Essa orientação consolidou-se como referência normativa internacional e continua a embasar as iniciativas de regulação da inteligência artificial na atualidade (ONU, 1988, p. 1; ONU, 1966, p. 9).

Apesar do avanço significativo das legislações nacionais sobre proteção de dados ao longo das últimas duas décadas, a governança internacional de dados ainda carece de um tratado vinculante que estabeleça normas comuns em nível global. O cenário atual é marcado por uma multiplicidade de iniciativas regionais e diretrizes voluntárias, mas a ausência de um marco jurídico internacional uniforme deixa lacunas importantes na proteção transnacional de dados (Bernier et al., 2022, p. 3).

Garantir que as pessoas tenham domínio sobre suas informações pessoais é uma exigência central na abordagem dos direitos humanos à proteção de dados. Para isso, as legislações contemporâneas estabeleceram obrigações específicas para quem coleta e utiliza esses dados, além de preverem formas de responsabilização. Em linhas gerais, essas normas protegem direitos como o consentimento e a transparência, e buscam evitar abusos por parte de instituições públicas ou privadas (Privacy International, 2018, p. 11-13).

O respeito à privacidade está diretamente ligado à observância de normas que regulam o uso adequado de informações em contextos específicos. Quando essas normas são desconsideradas, seja pelo envolvimento de pessoas não autorizadas, pela adoção de procedimentos incorretos ou pelo uso com objetivos inadequados, ocorre a violação desse direito. A ética da proteção de dados parte do princípio de que tais transgressões não se justificam apenas pelos danos que causam, como exclusões sociais. Em muitos casos, a simples transgressão do limite previamente acordado já representa uma falha moral, mesmo na ausência de efeitos práticos (Price II et al, 2019, p. 38-39).

Atualmente, os dados de saúde são reconhecidos como uma categoria especialmente sensível de informações pessoais, dada a sua natureza confidencial e o potencial significativo de danos em caso de mau uso. Existem dois principais modelos regulatórios sobre privacidade e proteção de dados pessoais. O sistema norte-americano se baseia em legislações específicas por setor, como saúde e finanças. Já o europeu adota uma abordagem integrada, com normas gerais e abrangentes, sendo esse modelo cada vez mais adotado como referência por outras nações (Bernier et al., 2022, p. 7-8).

A construção de um modelo robusto de governança para os dados em saúde pressupõe a articulação equilibrada entre essas duas concepções, cuja convergência se mostra indispensável para garantir legitimidade, segurança e efetividade no tratamento das informações sensíveis.

### 3.3.4 Modelo regulatório dos Estados Unidos para a inteligência artificial

O primeiro marco relevante de governança de dados de saúde com impacto global foi estabelecido pelos Estados Unidos com a promulgação da *Health Insurance Portability and Accountability Act* (HIPAA), em 1996. Essa norma representou um avanço importante ao estabelecer padrões específicos para o tratamento de informações de saúde, sendo amplamente referenciada internacionalmente (Greenleaf, 2023, p. 4), modelo regulatório criado pela *Health Insurance Portability and Accountability Act* (HIPAA) nos Estados Unidos, foi desenvolvido com o objetivo de estabelecer padrões federais que assegurem a privacidade, a confidencialidade e a segurança das informações de saúde dos cidadãos.

Ao definir requisitos mínimos para o uso e o compartilhamento de dados médicos, a HIPAA se tornou um marco na proteção das informações sensíveis no setor da saúde, funcionando como referência para políticas institucionais e regulatórias (EUA, 2015, p. 5-6). A legislação norte-americana HIPAA estabelece diretrizes fundamentais para proteger as informações de saúde, assegurando que esses dados estejam disponíveis quando necessário, mas sem comprometer sua confidencialidade e integridade. Para tanto, permite o acesso controlado por prestadores de serviços, operadoras e outras entidades do sistema de saúde, desde que respeitadas normas rigorosas de uso e proteção de dados.

Entre os principais dispositivos regulatórios estão: a ***Privacy Rule***, que define os direitos dos pacientes e os limites para o compartilhamento de dados; a ***Security Rule***, que obriga a adoção de medidas técnicas e organizacionais para proteger as informações em formato digital; e a ***Enforcement Rule***, que

disciplina os mecanismos de fiscalização e sanção em caso de infrações (EUA, 2015, p. 7-9).

A **Regra de Privacidade** da HIPAA estabelece critérios rígidos sobre como informações de saúde identificáveis, denominadas informações de saúde protegidas (PHI), podem ser utilizadas e compartilhadas por organizações autorizadas, como hospitais, operadoras de planos de saúde, seguradoras e intermediárias de dados. De modo geral, qualquer uso ou divulgação dessas informações requer autorização expressa e por escrito da pessoa a quem os dados se referem, salvo nas situações excepcionais previstas na legislação. Tais exceções incluem, entre outras, o acesso do próprio paciente aos seus dados, o uso para finalidades assistenciais, administrativas ou financeiras relacionadas ao cuidado em saúde, e situações específicas de interesse público, como emergências sanitárias, investigações legais, ou pesquisas científicas autorizadas.

Além disso, a norma permite o compartilhamento de conjuntos de dados limitados, desde que atendidos critérios de anonimização. Quando as informações são devidamente desidentificadas, mediante a retirada de 18 categorias de identificadores pessoais, como nome, número de documentos e datas de atendimento, deixam de estar sujeitas às restrições da HIPAA, podendo ser utilizadas livremente inclusive para fins comerciais ou de pesquisa (EUA, 2015, p. 8-10).

Publicada em 2003, a **Regra de Segurança** da HIPAA foi criada para complementar a Regra de Privacidade, estabelecendo padrões específicos de proteção para as informações de saúde protegidas em formato eletrônico (e-PHI). A norma exige que prestadores de serviços, operadoras e demais entidades abrangidas adotem medidas administrativas, técnicas e físicas voltadas à proteção da confidencialidade, integridade e disponibilidade desses dados digitais. Essas exigências aplicam-se exclusivamente às informações eletrônicas, não se estendendo àquelas transmitidas por meio oral ou físico (EUA, 2015, p. 11-12).

A **Regra de Aplicação** (Enforcement Rule) da HIPAA trata dos mecanismos de fiscalização e responsabilização das entidades reguladas. Ela estabelece os procedimentos a serem adotados em casos de suspeita de descumprimento das normas administrativas da lei, incluindo a condução de

investigações, auditorias e a imposição de sanções civis, como multas financeiras. Essa regra confere autoridade ao Departamento de Saúde e Serviços Humanos dos Estados Unidos (HHS) para supervisionar a conformidade com a legislação e aplicar penalidades proporcionais às infrações cometidas (EUA, 2015, p. 13).

Esse modelo regulatório norte-americano para governança de dados em saúde tem como base a legislação HIPAA e suas atualizações promovidas pela Lei HITECH de 2009. No entanto, essa estrutura apresenta lacunas importantes. Um dos principais mecanismos adotados para equilibrar a proteção da privacidade com a viabilidade de uso dos dados é a anonimização das informações, por meio da retirada de 18 identificadores definidos pela HIPAA. Contudo, esse procedimento não é suficiente para impedir que os dados venham a ser reidentificados. À medida que aumenta a quantidade e a interconexão entre bases de dados, torna-se possível realizar triangulações e cruzamentos que permitem associar novamente os dados a indivíduos específicos, comprometendo assim a segurança informacional e a confiança no sistema (Price II et al., 2019, p. 4).

Apesar de a HIPAA ser considerada o principal marco legal norte-americano para proteção de dados de saúde, sua abrangência é limitada. A maioria dos dados gerados atualmente fora dos ambientes clínicos tradicionais como por aplicativos, redes sociais, mecanismos de busca ou comércio eletrônico, não está sujeita a essa legislação, pois não envolve entidades cobertas ou seus parceiros diretos. Essa limitação é especialmente crítica no contexto atual, em que dados relacionados à saúde podem ser produzidos em múltiplos ambientes digitais, inclusive pelos próprios usuários, muitas vezes sem proteção regulatória adequada (Cohen et al., 2018, p. 2).

As rápidas transformações digitais no setor de saúde evidenciaram importantes lacunas na legislação HIPAA. Dispositivos vestíveis, tecnologias de monitoramento remoto, aplicações baseadas na internet das coisas e o crescente uso de dados genômicos ainda não são devidamente contemplados pela estrutura normativa atual (Theodos et al., 2020, p. 5-6), sobretudo quando essas ferramentas são operadas por empresas que não se enquadram como entidades abrangidas pela lei. Isso se agravou com a intensificação do uso de ferramentas de atendimento remoto, revelando a urgência de modernizar o

arcabouço regulatório norte americano para refletir as práticas contemporâneas de coleta e uso de dados de saúde.

### 3.3.5 Modelo regulatório adotado pela União Europeia

No contexto europeu, a governança de dados pessoais, incluindo os dados de saúde, inicialmente foi regulada de forma centralizada e uniforme por meio do Regulamento Geral sobre a Proteção de Dados (GDPR - *General Data Protection Regulation*) (RGPD). Aprovado em 2016 pelo Parlamento Europeu e pelo Conselho da União Europeia, o regulamento entrou em vigor em maio de 2018 e passou a ser diretamente aplicável em todos os Estados-membros, sem necessidade de transposição nacional. O RGPD representa um marco na consolidação de uma abordagem holística e robusta para proteção de dados, estabelecendo regras claras sobre direitos dos titulares, deveres dos controladores e critérios para o tratamento lícito de dados pessoais, inclusive os sensíveis, como os relacionados à saúde (UE, 2016).

Marco regulatório robusto e abrangente: o Regulamento Geral sobre a Proteção de Dados (RGPD). Criado para substituir a antiga diretiva de 1995, o RGPD entrou em vigor em 2018 com o propósito de uniformizar as legislações dos países membros e modernizar a governança de dados em face dos desafios da era digital. Esse regulamento estabelece princípios fundamentais que orientam todo o tratamento de dados pessoais, os quais devem ser processados com base em legalidade, transparência e respeito aos limites das finalidades previamente estabelecidas.

Também impõe que se colete apenas o necessário (minimização), que as informações sejam exatas e atualizadas, que não sejam armazenadas por tempo superior ao necessário e que medidas técnicas e organizacionais garantam sua integridade e confidencialidade (UE, 2016, p. 35).

Pelo RGPD, considera-se dado pessoal qualquer informação que identifique ou possa identificar uma pessoa, de forma direta ou indireta, levando em conta os recursos tecnológicos disponíveis no momento do tratamento. Em contrapartida, dados verdadeiramente anônimos, aqueles que

não permitem a reidentificação do titular, não se submetem às exigências do regulamento, conforme prevê o “Considerando 26” (UE, 2016, p. 11-33).

O consentimento é amplamente reconhecido como um dos fundamentos essenciais do regime europeu de proteção de dados pessoais. Para que o tratamento de dados seja considerado lícito, o consentimento deve ser livre, específico, informado e inequívoco. Além disso, cabe ao controlador, ou seja, o responsável pelo tratamento, o ônus de demonstrar que o consentimento foi devidamente obtido. A qualquer momento, o titular dos dados pode revogar seu consentimento, e essa retirada não compromete a legalidade do uso anterior das informações. (Kuner et al., 2021, p. 45-46).

O Regulamento Geral sobre a Proteção de Dados (RGPD) classifica os dados de saúde entre as “categorias especiais de dados pessoais”, sujeitas a um regime jurídico mais rigoroso. Além de incluir informações sobre a saúde, também abrange dados genéticos e biométricos, bem como informações relacionadas a origem étnica, religião, filiação sindical, opiniões políticas e vida sexual. A coleta e o tratamento dessas informações são, como regra, proibidos, exceto quando preenchidos os requisitos legais estabelecidos, como consentimento explícito do titular, obrigações legais ou interesse público (Levitina, 2022, p. 231-232).

Um dos aspectos mais decisivos do modelo regulatório instituído pelo RGPD é o seu forte mecanismo de aplicação. O regulamento prevê sanções administrativas, com multas que podem alcançar até 4% do faturamento anual da organização infratora, a depender da natureza da violação. Além disso, as Autoridades de Proteção de Dados dos Estados-membros da União Europeia foram dotadas de extensas competências de supervisão, incluindo a possibilidade de investigar práticas e ordenar a suspensão de operações de tratamento de dados. Essa abordagem busca superar as limitações observadas na antiga Diretiva de 1995, cuja eficácia era comprometida pela fraca aplicação prática. Assim, o RGPD consolida o paradigma europeu de governança de dados, no qual os dados de saúde se inserem, reforçando o compromisso histórico da União Europeia com a proteção da privacidade (Hoofnagle et al., 2019, p. 66-68).

Dando continuidade ao modelo regulatório europeu, a União Europeia aprovou em março de 2024 o Regulamento da Inteligência Artificial (IA Act),

visando estabelecer um arcabouço jurídico uniforme para o desenvolvimento, comercialização e uso de sistemas de IA em todo o território da União. Esse regulamento complementa o RGPD ao abordar de forma específica os desafios éticos, sociais e técnicos associados à IA, com foco em garantir a proteção dos direitos fundamentais, a segurança, a transparência e a confiança pública. Ele segue uma abordagem baseada em risco, classificando os sistemas de IA conforme seu potencial de impacto sobre os direitos e a segurança das pessoas.

O IA Act estabelece quatro categorias principais de risco: risco inaceitável (proibido), risco elevado (regulado com rigor), risco limitado (com obrigações de transparência) e risco mínimo (livre). Os sistemas de IA de risco elevado, como os utilizados em áreas críticas como saúde, educação, segurança pública e administração da justiça, estão sujeitos a exigências rigorosas de conformidade, incluindo avaliação de impacto, supervisão humana, documentação técnica, qualidade dos dados e rastreabilidade. Sistemas que representem risco inaceitável, como os que envolvem manipulação subliminar ou categorização social discriminatória, são expressamente proibidos.

Entre as principais inovações do regulamento, destaca-se a proibição do uso de sistemas de identificação biométrica à distância em tempo real em espaços públicos para fins de manutenção da ordem pública, salvo em circunstâncias extremamente restritas e com autorização prévia. Também há exigências rigorosas quanto ao uso de sistemas de reconhecimento de emoções e categorização biométrica. O IA Act reforça ainda a necessidade de supervisão humana e responsabilidade nos processos automatizados, buscando evitar práticas abusivas e garantir que a tecnologia seja usada de maneira ética e segura.

Ele prevê mecanismos de governança e fiscalização robustos, incluindo a criação de autoridades competentes nos Estados-membros e a imposição de sanções administrativas em caso de descumprimento, com multas que podem ultrapassar 30 milhões de euros ou 6% do faturamento da empresa. Essa nova legislação representa um passo decisivo para a consolidação de um modelo europeu de governança tecnológica centrado na dignidade humana, na confiança e na inovação responsável (UNIÃO EUROPEIA, 2024).

A crescente complexidade das tecnologias de inteligência artificial de uso geral (general-purpose AI) exigiu uma mudança de paradigma regulatório, levando a União Europeia a adotar um modelo mais proativo com a promulgação do AI Act. Essa transição é percebida como uma tentativa de responder de maneira mais eficaz aos desafios trazidos por sistemas, como os modelos fundacionais e generativos.

Diferentemente do RGPD, cuja ênfase recai sobre os direitos individuais e proteção de dados, o AI Act incorpora uma abordagem híbrida que combina segurança de produto, padronização e, em menor escala, proteção de direitos fundamentais. Assim, enquanto o RGPD representava um modelo holístico de proteção à privacidade, o AI Act amplia o escopo para contemplar os riscos sistêmicos, a opacidade algorítmica e os efeitos sociais amplificados pela IA de propósito geral (Gstrein et al., 2024, p. 6-7).

O AI Act classifica os sistemas de IA em diferentes categorias de risco, inaceitável, alto, limitado e mínimo, e dedica um capítulo específico à regulação da IA de uso geral, cuja versatilidade dificulta seu enquadramento em categorias tradicionais.

Notadamente, os modelos com "risco sistêmico" passaram a ser objeto de exigências regulatórias mais rigorosas, incluindo transparência e avaliação de segurança. A legislação também impõe obrigações específicas para provedores, como o fornecimento de resumos de dados de treinamento e a garantia de conformidade com normas de direitos autorais (Gstrein et al., 2024, p. 10-12).

### 3.3.6 Direito à privacidade e proteção de dados

Até o início da década de 2020, diversas jurisdições ao redor do mundo, além dos Estados Unidos e da União Europeia, passaram a instituir legislações específicas sobre proteção e privacidade de dados, abrangendo também os dados de saúde. No caso brasileiro, o marco legal foi estabelecido com a promulgação da Lei nº 13.709/2018, conhecida como Lei Geral de Proteção de Dados Pessoais (LGPD). Inspirada em boa medida nos princípios da RGPD

europeia, a LGPD define as bases legais para o tratamento de dados pessoais, inclusive os sensíveis, como os relativos à saúde, e estabelece obrigações para agentes de tratamento e direitos para os titulares. A norma entrou em vigor em agosto de 2020, embora suas disposições sobre sanções administrativas só tenham se tornado efetivas em agosto de 2021 (BRASIL, 2018, p. 26).

Fortemente inspirada no modelo europeu consagrado pelo RGPD, a LGPD reproduzindo diversos institutos jurídicos centrais, como o conceito de “dado pessoal sensível”, que equivale à “categoria especial de dados” do regulamento europeu. De acordo com o artigo 5º, inciso II, da LGPD, são considerados sensíveis os dados relacionados à saúde, à origem racial, às convicções religiosas, à vida sexual, entre outros elementos que exigem tratamento diferenciado.

Pelo regime jurídico brasileiro, o tratamento desses dados só pode ocorrer nas hipóteses expressamente previstas no artigo 11, tais como: mediante consentimento específico do titular; para cumprimento de obrigações legais; em atividades de pesquisa; para exercício regular de direitos; em situações que envolvam proteção da vida; ou ainda, para tutela da saúde, desde que conduzido por profissionais ou instituições autorizadas.

A LGPD dispõe de regras específicas voltadas aos dados de saúde, estabelecidas nos parágrafos 4º e 5º do mesmo artigo. O §4º proíbe o compartilhamento de dados de saúde entre controladores com fins econômicos, salvo em casos de portabilidade a pedido do titular ou para a própria prestação de serviços de saúde. Já o §5º veda o uso dessas informações por operadoras de planos privados com o objetivo de seleção de risco, impedindo práticas discriminatórias em contratações e exclusões de beneficiários (BRASIL, 2018, p. 3-4). Além disso, impõe às entidades responsáveis pelo seu tratamento deveres de segurança e transparência, com sanções administrativas aplicáveis em caso de violação.

A legislação garante aos titulares direitos fundamentais, como o de revogar o consentimento, solicitar correções, eliminação e acesso aos próprios dados. É relevante notar que o uso desses dados exige, como regra, o consentimento livre, informado e inequívoco do titular, excetuando-se hipóteses legais específicas.

Ademais, institui-se a Autoridade Nacional de Proteção de Dados (ANPD), órgão encarregado de supervisionar a aplicação da norma e zelar pela observância de seus preceitos. Em linhas gerais, a Lei Geral de Proteção de Dados consolida um arcabouço normativo inicial para a governança das informações em saúde no contexto brasileiro. (Botelho et al., 2021, p. 20).

Tramitando no inicio de 2025 na Câmara dos Deputados Federais o Projeto de Lei nº 2.338/2023 representa um marco importante na tentativa de regulamentar o uso da inteligência artificial. O texto propõe uma abordagem centrada na proteção de direitos fundamentais, promovendo princípios como a transparência, a responsabilização, a segurança e a supervisão humana no uso de sistemas algorítmicos. A proposta busca estabelecer regras claras para aplicações de IA em setores sensíveis, como saúde, segurança e finanças, inspirando-se em experiências internacionais, como o AI Act europeu e sinalizando um esforço para equilibrar desenvolvimento tecnológico e preservação de direitos no contexto nacional.

Na China, o marco regulatório de governança de dados de saúde é definido principalmente pela Lei de Proteção de Informações Pessoais (mais conhecida pela sigla da tradução para o inglês PIPL - Personal Information Protection Law), que entrou em vigor em novembro de 2021.

A despeito de diferenças no plano político e institucional, o modelo regulatório da China referente ao uso de dados de saúde apresenta convergências significativas com os padrões internacionais, sobretudo o europeu. A legislação chinesa classifica os dados de saúde como altamente sensíveis e impõe medidas rigorosas para sua proteção, o que inclui o direito do titular à exclusão dos dados e à revogação do consentimento previamente fornecido. De maneira similar ao que se observa na União Europeia, o consentimento explícito do indivíduo é a base para o processamento dessas informações, excetuando-se apenas situações previstas em lei, como emergências sanitárias ou casos que envolvam a preservação da vida e da integridade de pessoas em risco. Esse alinhamento normativo evidencia que, mesmo sob um regime jurídico diverso, a China também incorpora diretrizes voltadas à proteção da privacidade e à segurança das informações em saúde, contribuindo para um movimento regulatório global mais uniforme nessa área (Wang et al., 2018, p. 6).

A consolidação do direito à privacidade como um direito fundamental na Índia ocorreu tardiamente e somente após intensos debates judiciais. Durante muitos anos, o arcabouço legal do país carecia de uma definição formal de dados pessoais e não reconhecia expressamente a privacidade como direito constitucional. A proteção de dados, especialmente na área digital, era regulada de forma dispersa, majoritariamente pela *Information Technology Act* de 2000, sendo a interpretação deixada a cargo dos tribunais caso a caso. Essa indefinição começou a ser superada em 2017, quando a Suprema Corte indiana, reunida em colegiado pleno, decidiu que o direito à privacidade está intrinsecamente ligado à dignidade e à liberdade pessoal, ambos protegidos pelo artigo 21 da Constituição. Essa decisão não apenas reforçou os fundamentos jurídicos para a proteção de dados na era digital, como também inseriu a Índia no rol dos países que reconhecem a privacidade como um pilar dos direitos fundamentais (Chatterjee, 2019, p. 182).

Essa legislação introduz princípios importantes para o tratamento de dados, entre eles a exigência de que qualquer operação envolvendo informações pessoais, definidas como dados que permitam identificar direta ou indiretamente uma pessoa, ocorra com base em consentimento livre e informado do titular. Embora garanta direitos como o acesso, correção e exclusão dos dados, a norma não incorporou o chamado “direito ao esquecimento”. O texto também institui o Conselho de Proteção de Dados da Índia, órgão responsável por fiscalizar a aplicação da lei e aplicar sanções em caso de descumprimento. Com isso, mesmo mantendo particularidades próprias, o país se alinha a uma tendência global de fortalecimento da governança digital e da proteção dos dados pessoais (ÍNDIA, 2023, p. 5-12).

Ainda que a nova legislação não contemple disposições específicas relativas aos dados de saúde, ela representa um avanço significativo ao estabelecer, pela primeira vez, um marco regulatório voltado à governança de dados pessoais no país mais populoso do planeta.

A Indonésia aprovou em outubro de 2022 a sua primeira lei nacional abrangente sobre proteção de dados pessoais, conhecida como Personal Data Protection Law (PDP Law). Inspirada no modelo europeu, a norma define dados pessoais como informações que identificam uma pessoa, direta ou indiretamente, e distingue entre dados comuns e sensíveis, como dados de

saúde, biométricos e financeiros. Reconhece direitos dos titulares, como acesso, correção, exclusão e revogação de consentimento, além de prever sanções para violações. Apesar dos avanços, a lei não cria uma autoridade independente de proteção de dados, o que pode comprometer sua efetividade (Firdaus, 2022).

Em 2023, a Nigéria deu um passo decisivo ao estabelecer um marco legal próprio para a proteção de dados com a entrada em vigor do Nigeria Data Protection Act. Essa iniciativa reflete uma tendência mundial de consolidação normativa sobre o tema (Greenleaf, 2023, p. 3-4).

### 3.3.7 Instrumentos voltados à salvaguarda das informações em saúde

Para que a inteligência artificial seja confiável na área da saúde, é essencial combinar medidas regulatórias, técnicas e de gestão que protejam os dados pessoais. O foco deve ser **garantir o controle dos indivíduos sobre suas informações e responsabilizar os agentes que as tratam**, sem dificultar pesquisas e ações de saúde pública. A experiência americana com o HIPAA mostra que, apesar de avanços, novas soluções são necessárias para lidar com os desafios do uso massivo de dados (Cohen et al., 2018, p. 2).

Nos modelos atuais de proteção de dados, a desidentificação é uma estratégia central. Quando uma informação não pode mais ser vinculada a uma pessoa, ela deixa de ser considerada dado pessoal e, portanto, sai do escopo das normas protetivas. O RGPD adota a pseudonimização como técnica relevante nesse processo, permitindo que os dados só possam ser reidentificados com informações adicionais mantidas separadamente, o que reduz riscos e reforça a segurança (UE, 2016, p. 87)

Embora inspirada no modelo europeu, a LGPD brasileira adota como conceito central a anonimização, definida como a aplicação de meios técnicos que impeçam a associação dos dados a um indivíduo (art. 5º, XI). Já a pseudonimização é mencionada apenas para fins específicos, como pesquisas em saúde pública (art. 13), o que evidencia diferenças conceituais relevantes entre os dois modelos (BRASIL, 2018).

Nos Estados Unidos, o termo “de-identification” é amplamente empregado para descrever qualquer técnica destinada a impedir a associação de dados a indivíduos identificáveis, com base nas diretrizes da HIPAA. Já na literatura biomédica, ainda prevalece confusão conceitual: os termos “anonymization” e “de-identification” são frequentemente usados de maneira intercambiável ou contraditória, revelando a ausência de consenso e a necessidade de maior padronização terminológica (Chevrier et al., 2019, p. 6-7).

A desidentificação é crucial em sistemas de IA aplicados à saúde, já que esses modelos usam grandes volumes de dados sensíveis. Essa técnica já vem sendo empregada em pesquisas multicêntricas para viabilizar o compartilhamento seguro de registros eletrônicos entre instituições (Kushida et al., 2012, p. 83-84).

Novas abordagens baseadas em redes neurais vêm sendo exploradas para desidentificar prontuários eletrônicos, especialmente aqueles em linguagem natural. Essas soluções usam mecanismos como atenção automática e redes recorrentes profundas, permitindo anonimização mais precisa e rápida, o que favorece seu reaproveitamento em aplicações de IA na saúde (Ahmed et al., 2020, p. 1-2).

Mesmo com esses métodos, dados de saúde podem ser reidentificados. Pesquisas mostram que poucas variáveis são suficientes para revelar a identidade de pessoas em bases desidentificadas, o que coloca em dúvida a segurança desses métodos (Rocher et al., 2019, p. 1-3).

Alguns dados, como o DNA, desafiam a desidentificação. Mesmo sem nome ou ID, informações genômicas podem revelar identidades por meio de parentes, tornando impossível garantir anonimato total nesses casos (May, 2018, p. 411-412).

Dados gerados por dispositivos vestíveis trazem novos desafios à desidentificação. Mesmo quando anonimizados, esses dados podem permitir a reidentificação de indivíduos, já que padrões biométricos únicos, como sinais cardíacos ou movimentos, persistem. Estudos mostram que poucos segundos de gravação já bastam para identificar alguém com alta precisão (Chikwetu et al., 2023, p. 240-244).

Em certos contextos, tornar os dados anônimos pode limitar benefícios para quem os forneceu, além de restringir a possibilidade de exercer controle sobre seu uso futuro (Vayena et al., 2017, p. 508).

Para proteger dados de saúde diante dos riscos trazidos pela IA, as normas devem exigir que os responsáveis usem métodos de desidentificação cada vez mais avançados. Isso é essencial para garantir a privacidade dos pacientes na evolução da tecnologia (Murdoch, 2021, p. 3).

Para garantir privacidade, a lei exige que o uso de dados tenha consentimento livre e informado. Isso pressupõe explicações claras sobre o uso, os fins e os riscos, algo crucial em aplicações com inteligência artificial (Gerke et al., 2020, p. 300-301).

Uma alternativa ao modelo tradicional de consentimento permite que as pessoas ajustem suas decisões ao longo do tempo, aceitando ou recusando novos usos de seus dados. Isso amplia o controle individual e torna o processo mais adaptável às mudanças (Vayena et al., 2017, p. 509).

Uma alternativa prática ao modelo tradicional tem sido o consentimento prévio para usos futuros, como adotado por instituições que armazenam dados em biobancos. Nesses casos, os participantes autorizam, de forma geral, a reutilização de seus dados e amostras em pesquisas futuras, incluindo possíveis colaborações com a indústria, sem prejuízo à privacidade e à autonomia individual (Spector-bagdady et al., 2020, p. 2072-2073).

Em certos casos, o uso de dados sem consentimento é aceito quando há interesse público relevante. A regulação admite que a privacidade possa ser limitada, desde que isso seja necessário, proporcional e não existam alternativas menos invasivas para atingir o fim coletivo (Townend, 2022, p. 78-79).

Em certas situações, como o desenvolvimento de IA voltada à saúde pública, o uso de dados sem consentimento pode ser justificável. Quando há potencial para salvar vidas, espera-se que autoridades compartilhem essas informações de forma ética, em benefício coletivo (Cohen et al., 2018, p. 2).

Para justificar o uso de dados sem consentimento, é preciso que haja um objetivo legítimo e bem definido de interesse público, como a resposta a ameaças à saúde coletiva. Essa exceção só se sustenta quando for proporcional, necessária e houver transparência, sem interesses comerciais

envolvidos. A regra continua sendo a proteção da privacidade, sua flexibilização só se justifica quando o compartilhamento gerar benefícios concretos à saúde pública que não seriam viáveis por outros meios (Price II et al., 2019, p. 40-41).

O aprendizado federado é uma técnica de IA que permite treinar modelos sem transferir os dados para um servidor central. Os dados permanecem nos dispositivos locais, o que aumenta a segurança e protege a privacidade. Isso viabiliza o uso de grandes volumes de dados distribuídos sem comprometer a confidencialidade das informações (Yang et al., 2019, p. 2-3).

Ao distribuir o treinamento entre diferentes fontes, o aprendizado federado reduz riscos de falhas centralizadas e melhora a segurança dos dados. Isso fortalece a privacidade, facilita o cumprimento das leis de proteção de dados e ainda contribui para criar modelos mais eficazes e generalizáveis (Li et al., 2020, p. 2).

Apesar de desafios como a diversidade dos dados médicos e a dificuldade de rastrear responsabilidades, o aprendizado federado tem se mostrado promissor na saúde. Ele permite a construção de modelos robustos sem centralizar dados sensíveis, preservando a privacidade e favorecendo diagnósticos mais precisos, inclusive para doenças raras (Rieke et al., 2020, p. 2-3).

Centros de dados em saúde são estruturas digitais criadas para permitir o compartilhamento seguro de informações sensíveis. Eles funcionam como pontos de acesso regulado, organizando dados de múltiplas fontes e garantindo conformidade com normas de governança e proteção de dados (Alvarez-Romero et al., 2023, p. 2).

Um exemplo de infraestrutura regulatória voltada à proteção e compartilhamento de dados em saúde é o Espaço Europeu de Dados de Saúde (EHDS), proposto em 2022 pela Comissão Europeia. A iniciativa visa garantir que cidadãos europeus tenham maior controle sobre seus dados de saúde, ao mesmo tempo em que promove o acesso seguro e padronizado a esses dados por instituições públicas, pesquisadores e desenvolvedores de tecnologias em saúde, respeitando os princípios do GDPR (COMISSÃO EUROPEIA, 2022).

O EHDS propõe a criação de um mercado único para tecnologias digitais em saúde, como IA e dispositivos médicos, unificando regras e garantindo que

os titulares mantenham controle sobre seus dados. Segundo a OMS, uma governança de dados eficaz é essencial para que a regulação da IA na saúde seja ética, segura e voltada ao interesse público (COMISSÃO EUROPEIA, 2022, p. 1; OMS, 2021, p. 81).

Leis de proteção de dados devem prever o uso responsável de informações de saúde, garantindo o consentimento informado, fiscalização por órgãos autônomos e transparência no tratamento por entidades públicas e privadas. O uso desses dados deve sempre priorizar o interesse público.

## 3.4 Síntese reflexiva

A análise das bases normativas propostas pela Organização Mundial da Saúde evidencia a crescente preocupação internacional com os impactos éticos, jurídicos e sociais da inteligência artificial no campo da saúde. Ao reconhecer tanto o potencial transformador dessas tecnologias quanto os riscos associados à sua execução indiscriminada, a OMS propõe um marco de atuação pautado na proteção da dignidade humana, na promoção da equidade e no fortalecimento da autonomia dos sujeitos.

O conjunto de princípios identificados, que inclui transparência, explicabilidade, justiça distributiva, *accountability* e proteção de dados, constitui um alicerce normativo indispensável para orientar a concepção e a aplicação de sistemas de IA em contextos assistenciais. Tais diretrizes não apenas contribuem para mitigar eventuais danos, como também estimulam a criação de soluções tecnológicas mais alinhadas às necessidades reais das populações e às exigências dos sistemas de saúde contemporâneos.

A normatização internacional, nesse sentido, representa um esforço de antecipação regulatória frente à velocidade com que essas tecnologias vêm sendo desenvolvidas. Trata-se de construir uma governança da IA que seja, ao mesmo tempo, promotora da inovação e garantidora dos direitos fundamentais, assegurando que os avanços técnicos não se sobreponham à centralidade do ser humano no cuidado em saúde.

Para tanto se observou a síntese normativa na consubstancialização dos seguintes princípios normativos segundo as dimensões éticas, jurídicas e institucionais:

**Quadro 3** – Síntese do conjunto de princípios identificados nas bases normativas na saúde.

| PRINCÍPIO | DIMENSÃO | REFERÊNCIA |
|---|---|---|
| **Dignidade da pessoa humana** | Ética e Jurídica | Valor fundante que garante respeito à integridade, autonomia e singularidade do ser humano. |
| **Autonomia do paciente** | Ética e Jurídica | Direito do indivíduo de tomar decisões livres e conscientes sobre sua saúde. |
| **Equidade** | Ética | Correção de desigualdades no acesso aos benefícios e recursos das tecnologias em saúde. |
| **Justiça distributiva** | Ética e Institucional | Distribuição justa dos benefícios da IA entre diferentes grupos sociais. |
| **Transparência algorítmica** | Ética e Jurídica | Clareza sobre o funcionamento e critérios de decisão dos sistemas automatizados. |
| **Explicabilidade** | Ética | Dever de tornar as decisões da IA compreensíveis e justificáveis para humanos. |
| **Accountability (prestação de contas)** | Jurídica e Institucional | Atribuição clara de deveres e responsabilidades pelo uso da IA. |
| **Não maleficência** | Ética | Evitar danos intencionais ou previsíveis ao paciente. |
| **Beneficência** | Ética | Dever de promover o bem-estar e maximizar os benefícios para o paciente. |
| **Consentimento informado** | Ética e Jurídica | Garantia de decisão consciente, baseada em informação clara sobre riscos e limites da IA. |
| **Privacidade e proteção de dados** | Jurídica | Garantia legal do sigilo e controle sobre dados pessoais sensíveis. |
| **Sustentabilidade sistêmica** | Ética e Institucional | Uso da IA que preserve a integridade dos sistemas de saúde e evite sobrecargas estruturais. |

**Fonte**: Elaborado pelo autor

Nos próximos capítulos, aprofundar-se-á a análise sobre os referenciais regulatórios específicos vinculados à atuação da OMS, com foco nos critérios de confiabilidade tecnológica, nos mecanismos de transparência algorítmica e nos regimes de responsabilidade jurídica aplicáveis às soluções inteligentes na saúde.

# 4 REFERENCIAIS PARA O MARCO REGULATÓRIO DA IA

A atuação normativa da Organização Mundial da Saúde na inteligência artificial reflete um esforço sistemático de estruturar diretrizes regulatórias que orientem o uso ético, seguro e eficaz dessas tecnologias nos sistemas de saúde. Reconhecendo a complexidade envolvida na integração entre algoritmos inteligentes e decisões clínicas, a OMS propõe parâmetros técnicos e institucionais que buscam compatibilizar os avanços tecnológicos com os direitos fundamentais dos pacientes e com os princípios que regem a saúde pública global.

Neste contexto, a construção de um marco regulatório específico para a inteligência artificial em saúde exige a consideração de múltiplos vetores, entre os quais se destacam: a necessidade de garantir a confiabilidade dos sistemas automatizados; a clareza nos processos de tomada de decisão algorítmica; a responsabilização de agentes públicos e privados diante de eventuais danos; e a proteção robusta dos dados sensíveis gerados no ambiente assistencial.

Este capítulo tem por finalidade analisar os principais referenciais estabelecidos pela OMS com vistas à regulação da inteligência artificial no setor da saúde. Serão discutidos os critérios técnicos para validação de soluções baseadas em IA, os instrumentos voltados à explicabilidade das decisões automatizadas, bem como os fundamentos jurídicos e institucionais que sustentam a noção de responsabilidade nos sistemas inteligentes. Trata-se de identificar os elementos essenciais para a construção de uma governança internacional da IA em saúde, capaz de assegurar não apenas inovação, mas também legitimidade, equidade e proteção ao paciente.

## 4.1 Garantia de confiabilidade nas aplicações com IA

A confiança nos sistemas de IA em saúde depende, antes de tudo, da garantia de segurança. Esses requisitos são essenciais, pois refletem a

natureza crítica dessas tecnologias. Quando utilizados como “sistemas de informação como dispositivos médicos”, os modelos de *Deep Learnig* passam a exigir avaliação e aprovação por órgãos reguladores, assegurando que atendam aos padrões exigidos para proteger a saúde dos pacientes.

### 4.1.1 Soluções digitais na Saúde com função de dispositivo médico

A definição harmonizada de dispositivo médico foi estabelecida em 2005 pela Força-Tarefa de Harmonização Global (GHTF) e desde então vem sendo adotada por diversas agências reguladoras ao redor do mundo. Segundo essa definição, dispositivos médicos incluem instrumentos, máquinas, implantes, reagentes para uso in vitro, softwares e outros materiais destinados a finalidades médicas específicas, como diagnóstico, prevenção, tratamento de doenças, apoio à vida e controle da concepção.

A definição também abrange dispositivos usados para fornecer informações médicas por meio de análise de amostras do corpo humano. Para ser enquadrado nessa categoria, o dispositivo não pode exercer sua principal ação no organismo por meios farmacológicos, imunológicos ou metabólicos, embora possa ser auxiliado por esses mecanismos. Apesar de complexa, essa definição tornou-se a base para a regulação de tecnologias médicas em diversas jurisdições e continua sendo um marco fundamental na delimitação do escopo regulatório dos dispositivos, incluindo aqueles baseados em inteligência artificial (Aronson et al., 2020, p. 84).

Criado em 2011 por autoridades reguladoras da Austrália, Canadá, China, União Europeia, Japão e Estados Unidos, o IMDRF dá continuidade ao trabalho da GHTF com foco na harmonização internacional das normas sobre dispositivos médicos. O fórum busca acelerar a convergência regulatória global e já conta com a participação de países como Reino Unido, Coreia do Sul e Singapura, além da OMS e agências da Suíça e Argentina como observadoras (IMDRF, 2020, p. 3).

Desde 2013, o IMDRF tem desenvolvido diretrizes específicas para regular softwares com finalidade médica, conhecidos como SaMD (Software as

a Medical Device). Segundo a definição adotada, trata-se de qualquer software que realize funções médicas por conta própria, sem depender de um dispositivo de hardware. Isso inclui aplicações voltadas ao diagnóstico, monitoramento ou tratamento, desde que operem de forma independente, mesmo quando integradas a outras tecnologias (IMDRF, 2013, p. 6).

A regulação de softwares como dispositivos médicos (SaMD) tornou-se uma peça central do cenário regulatório, à medida que aplicações digitais ganham papel estratégico em diagnósticos, tratamentos e decisões clínicas. O modelo proposto pelo IMDRF organiza os SaMD segundo uma classificação de risco, definida por dois critérios combinados: (1) o grau de importância da informação fornecida para a decisão clínica e (2) a gravidade da condição de saúde a que se destinam. A partir dessa combinação, são estabelecidas quatro categorias (I a IV), que indicam níveis crescentes de impacto potencial na saúde individual ou coletiva. Essa categorização orienta a aplicação proporcional dos requisitos regulatórios (IMDRF, 2014, p. 10-14).

O padrão internacional ISO 14971, atualmente em sua terceira edição (2019), é a principal referência para gerenciamento de riscos em dispositivos médicos, incluindo softwares médicos (SaMD). Ele estabelece princípios e procedimentos para identificar e controlar riscos associados ao uso desses produtos. Parte da premissa de que todo dispositivo envolve riscos inerentes, definidos pela combinação de dois fatores: a probabilidade de ocorrência de um dano e a gravidade de suas consequências. Esses riscos podem afetar pacientes, usuários, terceiros, propriedades ou o meio ambiente, e sua gestão é essencial para garantir a segurança e a eficácia dos dispositivos ao longo de seu ciclo de vida (ISO, 2019).

Segundo a ISO 14971:2019, o gerenciamento de riscos em dispositivos médicos começa com a identificação dos perigos e das sequências de eventos que podem levar à exposição a esses riscos. Avaliam-se, então, a probabilidade e a gravidade dos danos, definindo a estimativa de risco. Essa estimativa é comparada com critérios de aceitabilidade, e, se necessário, aplicam-se medidas de controle até que o risco residual seja aceitável. Caso contrário, o projeto deve ser revisto. Esse processo, chamado de análise de risco, abrange todas as etapas do ciclo de vida do produto e também se aplica a softwares médicos com IA (ISO, 2019; Odaibo, 2021, p. 3).

Fabricantes de SaMD devem garantir que o software seja desenvolvido e documentado de forma compatível com seu risco e uso previsto. Também é necessário monitorar o desempenho após a comercialização, com detecção de falhas e análise do uso real. Mudanças no software devem ser cuidadosamente controladas, já que podem afetar diretamente a prática clínica (IMDRF, 2014, p. 21).

A norma ISO 13485:2016 é a principal referência internacional para sistemas de gestão da qualidade (QMS) voltados a dispositivos médicos. Ela estabelece requisitos específicos para todas as etapas do ciclo de vida do produto, desde o projeto até a manutenção e desativação, e é adotada como padrão regulatório em diversas jurisdições. Diferente da ISO 9001, a ISO 13485 foi desenvolvida com foco exclusivo na segurança e no desempenho de dispositivos médicos, incluindo softwares médicos, assegurando conformidade regulatória em cada fase da produção (ISO, 2016, p. 1-2).

Segundo o IMDRF, o QMS para SaMD deve incluir: (a) estrutura organizacional com governança e responsabilização; (b) processos de suporte como planejamento, gestão de riscos e melhoria contínua; e (c) etapas do ciclo de vida do software, do desenvolvimento à desativação. Tudo deve ser proporcional ao porte da organização e voltado à segurança e eficácia do produto (IMDRF, 2015, p. 8-9).

A avaliação clínica é parte essencial da regulação de softwares médicos. Segundo o IMDRF, trata-se de um processo contínuo que verifica se o SaMD tem associação clínica válida com a condição de saúde visada, se o software processa corretamente os dados de entrada e se sua aplicação gera benefícios clínicos reais. Essa abordagem integra os princípios de gestão de riscos e qualidade, com o objetivo de garantir que o software seja clinicamente relevante e seguro para o uso previsto (IMDRF, 2017, p. 4-10). Além disso, ela é um processo contínuo voltado à verificação da segurança, eficácia e relevância clínica do produto. Conforme o IMDRF, essa avaliação se apoia em três pilares: associação clínica válida, que estabelece se a saída do software está alinhada com a condição de saúde visada; validação analítica, que comprova a precisão técnica na geração dos resultados; e validação clínica, que verifica se o uso do software resulta em benefícios clínicos concretos. O processo deve incluir dados pré e pós-comercialização, utilizando evidências

científicas, ensaios clínicos e dados do mundo real para garantir a melhoria contínua do SaMD ao longo de seu ciclo de vida (IMDRF, 2017, p. 9-19).

A estrutura de avaliação clínica do IMDRF, com foco em evidências e revisão contínua, converge com princípios reconhecidos para a inteligência artificial na saúde. Ambas destacam a importância do monitoramento permanente e da atualização dos algoritmos para garantir segurança ao longo do tempo (Feng et al., 2022, p. 1-2).

### 4.1.2 Tecnologias médicas com Aprendizado de Máquina integrado

Em 29 de janeiro de 2025, o IMDRF publicou o documento "*Good Machine Learning Practice for Medical Device Development: Guiding Principles*", consolidando princípios orientadores para o desenvolvimento de dispositivos médicos que incorporam aprendizado de máquina (IMDRF, 2025).

O documento "*Good Machine Learning Practice for Medical Device Development: Guiding Principles*" (IMDRF/AIML WG/N88 FINAL: 2025), publicado pelo *International Medical Device Regulators Forum* (IMDRF), apresenta um conjunto de 10 princípios orientadores (*guiding principles*) que visam assegurar o desenvolvimento seguro, eficaz e de alta qualidade de dispositivos médicos baseados em aprendizado de máquina (*Machine Learning-enabled Medical Devices* - MLMD). Abaixo, os principais elementos dessas diretrizes: (IMDRF, 2025):

1. Clareza no uso pretendido e abordagem multidisciplinar: O dispositivo deve ter seu uso pretendido claramente definido, incluindo o contexto clínico e os riscos/benefícios para o paciente. Especialistas de diferentes áreas devem participar de todas as fases do ciclo de vida do produto;

2. Boas práticas de engenharia de software e segurança: O projeto deve seguir princípios robustos de engenharia de software, segurança cibernética, gestão da qualidade, rastreabilidade e integridade dos dados. A infraestrutura para implantação, monitoramento e manutenção do modelo deve ser bem planejada;

3. Conjuntos de dados representativos: Os dados usados no treinamento e teste devem refletir as características da população-alvo (idade, sexo, etnia, localização etc.), a fim de garantir a generalização dos resultados;

4. Separação entre dados de treinamento e de teste: Os conjuntos de treinamento e de teste devem ser independentes, com controle sobre possíveis fontes de dependência (como local de coleta ou características dos pacientes);

5. Padrões de referência adequados: Os padrões de referência usados para avaliar o modelo devem ser clinicamente relevantes, bem definidos e adequados ao contexto de uso do dispositivo;

6. Escolha e design do modelo: O modelo deve ser projetado com base nos dados disponíveis e no uso pretendido, com atenção para mitigar riscos como *overfitting* e degradação de desempenho;

7. Avaliação considerando a interação humano-IA: A avaliação do dispositivo deve incluir a interação entre humanos e IA no ambiente clínico, considerando fatores como habilidades do usuário, compreensão dos resultados e risco de uso indevido;

8. Testes em condições clínicas relevantes: O desempenho do dispositivo deve ser testado em condições reais de uso clínico, com planos metodológicos e estatísticos sólidos;

9. Informação clara e essencial ao usuário: O fabricante deve fornecer aos usuários informações claras sobre: a) uso pretendido; b) benefícios e riscos; c) dados usados no treinamento/teste; d) limitações do modelo; e) atualizações do dispositivo.

10. Monitoramento contínuo e re-treinamento: Modelos implantados devem passar por monitoramento contínuo de desempenho. O re-treinamento deve ser controlado para evitar viés, *overfitting* e degradação, especialmente com a introdução de novos dados (*dataset drift*) (IMDRF, 2025, p. 6-9).

Esses princípios têm como objetivo guiar a regulamentação internacional e harmonizar boas práticas, promovendo a inovação responsável na integração de IA no setor da saúde.

### 4.1.3 Modelo regulatório dos Estados Unidos

A FDA segue o paradigma regulatório do IMDRF para softwares como dispositivos médicos (SaMD), mas adota uma classificação própria de risco em três níveis: classe I (baixo risco), classe II (risco intermediário) e classe III (alto risco). Há três vias de aprovação para SaMD: (a) a autorização 510(k), que reconhece a equivalência substancial a um produto já aprovado; (b) a via De Novo, usada para inovações de menor risco sem precedentes legais; e (c) a aprovação pré-comercialização (PMA), exigida para dispositivos de alto risco, com necessidade de evidências clínicas robustas (FDA, 2022, p. 2-3).

Nos Estados Unidos, a maioria dos dispositivos médicos com base em inteligência artificial tem sido aprovada por meio do mecanismo simplificado de notificação 510(k), reservado para produtos considerados substancialmente equivalentes a outros já aprovados. Isso indica que avaliações regulatórias mais completas, como as exigidas para a via De Novo ou a aprovação pré-comercialização (PMA), ainda são pouco aplicadas a esse tipo de tecnologia (Benjamens et al., 2020, p. 2).

Sem uma regulação específica para IA adaptativa, a FDA adotou a regra do “algoritmo travado”, permitindo apenas versões fixas dos softwares. Isso impede atualizações após a aprovação, limitando a utilidade clínica e aumentando o risco de perda de desempenho com o tempo (Babic et al., 2019, p. 2-4).

A FDA entendeu que a regulação atual de SaMD não atende aos dispositivos com inteligência artificial, que mudam com o tempo, assim propôs uma estrutura baseada no ciclo de vida total, permitindo que fabricantes apresentem, na fase pré-comercial, um plano de mudanças com duas partes: as alterações previstas (pré-especificações) e a forma segura de aplicá-las (protocolo de mudança de algoritmo). A proposta busca equilibrar inovação com segurança, por meio de monitoramento contínuo.

A proposta da FDA levou, em 2021, à publicação de um plano de ação para SaMD com IA/ML. O documento destaca a criação de regras específicas, baseadas em boas práticas (GMLP), que envolvem qualidade de dados, validação de modelos e interpretabilidade. Também propõe transparência com

os usuários e ações para mitigar vieses, além de testes-piloto com monitoramento em tempo real do desempenho dos dispositivos (FDA, 2022, p. 8-10).

A 21st Century Cures Act, sancionada em 2016, reposicionou o papel da FDA no ecossistema regulatório norte-americano, ao incentivar a inovação tecnológica na saúde. Com foco em acelerar o ingresso de novos produtos no mercado, especialmente softwares médicos como os sistemas de suporte à decisão clínica (CDSS), a lei flexibilizou requisitos regulatórios tradicionais, permitindo o uso de evidências alternativas. Essa reconfiguração reforçou a transição da FDA de um papel centrado na segurança para uma atuação voltada à promoção da inovação (Lievevrouw et al., 2022, p. 563).

A partir da Lei de Curas, a FDA estabeleceu diretrizes para diferenciar softwares clínicos que são regulados como dispositivos médicos daqueles que não são. Em setembro de 2022, foram definidos quatro critérios que, se atendidos de forma cumulativa, excluem o software da categoria regulatória de dispositivo. São eles: o sistema deve apenas sugerir alternativas ao profissional, sem impor condutas; deve apresentar informações médicas usualmente compartilhadas entre profissionais; não pode processar sinais, imagens ou padrões clínicos; e deve permitir que o usuário compreenda, de forma autônoma, como as sugestões foram geradas (FDA, 2022, p. 6).

Desta forma, a entidade ao não distinguir softwares baseados em IA ou aprendizado de máquina, pode excluir da regulação sistemas amplamente usados na saúde, como os de suporte à decisão clínica baseados em modelos opacos. Esses algoritmos, muitas vezes incompreensíveis para humanos, impedem que profissionais revisem os fundamentos das recomendações emitidas. Assim, torna-se inviável atender ao princípio da interpretabilidade, o que deveria impedir sua dispensa do controle regulatório (Gerke, 2023, p. 97).

Nos EUA, softwares que se declaram voltados ao bem-estar, sem fins diagnósticos ou terapêuticos, podem ficar fora do escopo da regulação como dispositivos médicos. Segundo a 21st Century Cures Act, essa exclusão se aplica, por exemplo, a sistemas de IA que apenas incentivam hábitos saudáveis, o que permite que escapem da supervisão da FDA (Minssen et al., 2020, p. 4).

Tem ganhado força nos Estados Unidos uma proposta que pode inspirar modelos regulatórios internacionais: a substituição da análise de produtos pela avaliação de sistemas. Diante da complexidade dos SaMD baseados em IA/ML, torna-se insuficiente examinar apenas o dispositivo. É preciso considerar como ele interage com médicos, pacientes, outros dispositivos e os dados usados. Essa “visão de sistema” permitiria compreender melhor os riscos e garantir a eficácia, especialmente frente a algoritmos adaptativos. No entanto, adotar tal abordagem exigiria mudanças profundas nas normas e na própria estrutura das agências reguladoras (Gerke et al., 2020, p. 2).

### 4.1.4 Estrutura normativa europeia para a IA

Na União Europeia, o marco normativo que regula softwares como dispositivos médicos foi consolidado com o Regulamento (UE) 2017/745, adotado em abril de 2017 e aplicável a partir de maio de 2021. Este regulamento substituiu as antigas diretivas e passou a estabelecer um padrão unificado e mais rigoroso, com foco na segurança, desempenho e rastreabilidade dos dispositivos médicos, incluindo os digitais (UE, 2017, p. 1).

O *Medical Device Regulation* (MDR), em vigor na União Europeia desde maio de 2021, substituiu a antiga Diretiva de Dispositivos Médicos (MDD) com o objetivo de atualizar o arcabouço normativo frente ao avanço tecnológico. Embora preserve a estrutura e princípios da MDD, o MDR introduz requisitos mais rigorosos para avaliação de desempenho, vigilância pós-mercado e classificação de risco, o que amplia o controle sobre softwares médicos, incluindo os baseados em IA (Niemiec, 2022, p. 2).

O regulamento europeu vigente para dispositivos médicos estabelece critérios obrigatórios de segurança, desempenho e documentação técnica para qualquer tecnologia de saúde, inclusive softwares. A conformidade é verificada por meio de um processo que autoriza o uso da marca CE (*Conformité Européenne*) - símbolo que indica que um produto atende aos requisitos legais da União Europeia. Também foi implementado o sistema UDI, que aprimora a

rastreabilidade dos produtos. Na avaliação de riscos, aplica-se a norma ISO 14971, harmonizada com as diretrizes do IMDRF (UE, 2022, p. 28).

O MDR exige que fabricantes comprovem que os riscos dos dispositivos são aceitáveis frente aos benefícios. Para isso, adota uma escala de risco crescente: classes I, IIa, IIb e III. Dispositivos da classe I podem ser autodeclarados; já os de risco superior exigem avaliação por “organismos notificados”, empresas independentes credenciadas pelos Estados-membros e não pela EMA, que não regula diretamente esses produtos.

No caso dos softwares como dispositivos médicos (SaMD), o MDR prevê regras específicas (Regra 11, Anexo VIII), que elevam o nível de risco conforme a gravidade do impacto clínico: a) softwares usados para decisões clínicas - classe IIa; b) se houver risco de dano grave ou cirurgia - classe IIb; c) se houver risco de morte ou dano irreversível - classe III; d) monitoramento fisiológico geral - classe IIa; e) monitoramento de parâmetros vitais com risco imediato - classe IIb; f) outros softwares - classe I.

Com isso, muitos softwares que antes eram considerados de baixo risco passaram a categorias mais elevadas. Espera-se que a maioria dos sistemas de inteligência artificial se enquadre nas classes IIa ou IIb, exigindo certificação por organismos notificados (Minssen et al., 2020, p. 8-14).

A escassez de infraestrutura para operacionalizar o MDR tem sido apontada como uma das principais fragilidades da regulação europeia de dispositivos médicos. O número reduzido de organismos notificados aptos a realizar avaliações vem gerando atrasos na certificação de novos produtos. Isso pode comprometer o acesso a inovações na saúde, criando um paradoxo entre o aumento da exigência regulatória e a capacidade real de implantação, sobretudo em relação a tecnologias baseadas em IA (Gilbert et al., 2023, p. 3-4).

O MDR dedica um anexo exclusivo à avaliação clínica, exigida para todos os dispositivos, independentemente da classe de risco. O processo é composto por etapas como: elaboração do plano de avaliação, identificação e análise crítica dos dados clínicos disponíveis, e, se necessário, realização de investigações clínicas. Para dispositivos de classe III e implantáveis, esses estudos são obrigatórios. A norma também permite que fabricantes aleguem

equivalência com produtos já certificados, embora isso tenha se tornado mais restritivo sob o novo regulamento (Niemiec, 2022, p. 3).

O Medical Device Coordination Group (MDCG), criado pelo MDR para apoiar a aplicação uniforme do regulamento, publicou diretrizes específicas sobre avaliação clínica e de desempenho de softwares médicos. Essas orientações detalham os critérios metodológicos para demonstrar associação clínica válida, desempenho técnico e desempenho clínico dos softwares, alinhando-se aos princípios internacionais do IMDRF. O documento ainda destaca que a avaliação deve ser contínua e proporcional ao risco, considerando evidências clínicas, dados do mundo real e atualizações após a entrada no mercado (MDCG, 2020, p. 9-15).

Ela, baseadas no modelo do IMDRF, definem critérios para a avaliação clínica de softwares médicos, focando em associação clínica, validação analítica e clínica. Já o MDR exige plano de acompanhamento clínico pós-comercialização, mas ainda não regula de forma específica os sistemas de inteligência artificial. Falta um protocolo claro para mudanças em algoritmos adaptativos, o que representa um ponto crítico no modelo europeu atual (Gilbert et al., 2021, p. 4).

Portanto, para resolver essa questão foi adotado o Regulamento da União Europeia sobre Inteligência Artificial (Regulamento (UE) 2024/1689), publicado em 12 de julho de 2024, que estabelece o primeiro quadro jurídico específico para sistemas de IA na UE, incluindo dispositivos médicos baseados em inteligência artificial (CUATRECASAS, 2024), ele classifica os sistemas de IA conforme o nível de risco. Dispositivos médicos que incorporam IA são geralmente considerados de alto risco, exigindo: a) avaliação de conformidade por organismos notificados, que devem demonstrar competência nos termos dos regulamentos MDR e IVDR; b) sistemas de gestão de qualidade específicos para IA, incluindo governança e gestão de dados para conjuntos de dados de treino, manutenção de registos, concessão transparente, supervisão humana, precisão e cibersegurança; c) monitoramento contínuo pós-comercialização, com obrigações para os responsáveis pela implantação de sistemas de IA de risco elevado, como hospitais ou profissionais de saúde, incluindo garantir uma utilização correta, atribuir uma supervisão humana,

realizar a monitorização e a vigilância, manter registos do sistema e realizar avaliações do impacto na proteção de dados (CUATRECASAS, 2024).

Embora o regulamento tenha sido publicado em 2024, sua aplicação foi faseada. Algumas disposições, como as relativas às práticas de IA proibidas, tornaram-se aplicáveis em 2 de fevereiro de 2025, enquanto outras em 2 de agosto de 2026 (UE, 2024).

Além disso, a Agência Europeia de Medicamentos (EMA, 2025) estabeleceu, em fevereiro de 2025, um procedimento regular para que fabricantes de determinados dispositivos médicos de alto risco solicitem aconselhamento científico sobre suas estratégias de desenvolvimento clínico. Isso visa apoiar a inovação e promover o acesso mais rápido dos pacientes a dispositivos mais seguros e eficazes.

## 4.2 Clareza nos processos e decisões algorítmicas

A transparência é amplamente reconhecida como um princípio ético fundamental na aplicação da inteligência artificial à saúde. Mais do que divulgar algoritmos, ela envolve tornar compreensíveis as informações que embasam decisões geradas por sistemas de IA, garantindo que profissionais e pacientes compreendam suas implicações. Essa abordagem é vista como essencial para promover confiança pública e permitir responsabilização, sendo considerada uma das bases de um modelo ético para a IA na sociedade (Floridi et al., 2018, p. 700), ela envolve não apenas a explicação dos sistemas de IA, mas também o entendimento sobre quem os desenvolve, como são geridos e sob quais regras operam. Documentação técnica e avaliações de impacto são formas práticas de tornar visíveis essas informações e reforçar a responsabilidade sobre o uso da IA (Mittelstadt, 2023, p. 4).

Em IA, a transparência pode significar apenas compreender a lógica geral do algoritmo, sem acesso direto aos dados de treinamento ou às previsões específicas (OCDE, 2019, p. 92), também pode envolver o acesso aos elementos que explicam como o modelo toma decisões, como seus parâmetros e dados de treinamento (Lipton, 2016, p. 3-4), intrinsecamente

vinculado às noções de compreensão da lógica operacional dos sistemas de IA/ML.

### 4.2.1 Entre a transparência operacional e a clareza de decisão

Na IA, interpretabilidade e explicabilidade são termos próximos e muitas vezes usados como sinônimos, já que ainda não há definições consolidadas ou distinções claras entre eles (Mittelstadt, 2023, p. 3). Interpretabilidade é o quanto uma decisão de IA pode ser compreendida por humanos, permitindo reconhecer sua causa de forma inteligível (Miller, 2019, p. 3), ela é interpretável quando um ser humano entende por que ela chegou a certo resultado. Já nos modelos de caixa preta, essa relação entre entrada e saída é obscura e difícil de explicar (Burrell, 2016, p. 2-3).

Um modelo é transparente quando é simples ou pode ser dividido em partes que fazem sentido para o usuário, como entradas, parâmetros e cálculos (Lipton, 2016, p. 4), explicação é o modo pelo qual uma pessoa entende a razão de uma decisão tomada por um sistema de IA (Miller, 2019, p. 3), pode ocorrer após a decisão do modelo, mesmo em sistemas opacos, servindo para justificar o resultado sem revelar seu funcionamento interno (Lipton, 2016, p. 10).

A explicabilidade não depende da interpretabilidade do modelo. Mesmo sistemas opacos podem ser explicados por meios externos, razão pela qual o termo se aplica tanto a modelos transparentes quanto a caixas-pretas (Amann et al., 2022, p. 2) e apesar da proximidade entre os termos, na saúde é útil distinguir IA interpretável, cujos modelos são transparentes, de IA explicável, que busca tornar compreensíveis os resultados de modelos opacos por meio de explicações externas (Babic et al., 2021, p. 284).

Modelos considerados interpretáveis são aqueles baseados em estruturas transparentes, como árvores de decisão ou funções lineares, compreensíveis por profissionais da área. Porém, com o aumento dos parâmetros, essa clareza se perde, tornando o modelo difícil de entender como um todo (Molnar, 2022, p. 23-24). Na IA explicável, parte-se de um modelo

opaco, cujo funcionamento não é acessível, e cria-se um segundo modelo interpretável para simular suas decisões. Esse modelo auxiliar serve apenas para explicar os resultados, não para prever, e costuma simplificar os dados para torná-los compreensíveis (Babic et al., 2021, p. 284), desta forma, IA interpretável já é explicável, pois sua estrutura permite compreensão direta. Por outro lado, a IA explicável só se torna interpretável por meio de técnicas pós-decisão, que aproximam suas saídas de modelos compreensíveis (Lipton, 2016, p. 3-4).

### 4.2.2 Explicabilidade como direito fundamental na era da IA

A chamada ***explicabilidade algorítmica*** no contexto das decisões automatizadas, refere-se à prerrogativa conferida aos indivíduos de compreender os fundamentos e a lógica subjacente às decisões tomadas por sistemas de inteligência artificial que impactam diretamente sua esfera pessoal, social ou jurídica. Embora seja um conceito relativamente recente, sua importância vem ganhando solidez, sobretudo a partir das discussões que marcaram a formulação do Regulamento Geral sobre a Proteção de Dados da União Europeia (RGPD).

O referido regulamento estabelece um conjunto de garantias voltadas às pessoas afetadas por decisões inteiramente automatizadas. De acordo com o RGPD, os titulares de dados possuem o direito de ser informados sempre que forem submetidos a decisões automatizadas que possam produzir efeitos jurídicos relevantes ou interferências significativas em sua vida. Esse direito inclui o acesso a informações pertinentes sobre a lógica empregada no processamento automatizado, bem como sobre o alcance e as implicações potenciais dessa atividade (conforme disposto nos artigos 13º, 14º e 15º).

Além disso, o regulamento europeu assegura aos titulares o direito de não se submeterem exclusivamente a decisões baseadas em tratamento automatizado, inclusive aquelas fundamentadas em técnicas de profiling (automação de definição de perfis), salvo em hipóteses excepcionais, como quando tais decisões forem necessárias à execução de um contrato ou

autorizadas mediante consentimento expresso do titular. Nessas circunstâncias, o regulamento impõe salvaguardas adicionais, permitindo ao indivíduo requerer a intervenção humana, expressar seu ponto de vista e contestar a decisão (art. 22.º).

O considerando 71 do RGPD orienta que o titular dos dados tem direito à intervenção humana, a se manifestar, obter explicação e contestar decisões automatizadas. Essa proteção dá origem ao chamado “devido processo algorítmico” (Kaminski, 2019, p. 198).

Desde o RGPD, se discute se existe um direito à explicação em decisões automatizadas ou se ele depende da interpretação de outras garantias, como o direito de contestar e obter informações sobre o processo decisório (Bygrave, 2020, p. 523-524).

A complexidade matemática dos modelos de IA, especialmente os de caixa-preta, torna suas decisões difíceis ou impossíveis de compreender por humanos. Isso alimenta o debate jurídico sobre a garantia da explicabilidade, dividido entre os que defendem explicações apenas sobre o funcionamento geral dos sistemas e os que defendem explicações específicas para cada decisão automatizada (Wachter et al., 2017, p. 2-4).

Outra corrente sustenta que a garantia à explicabilidade deve abranger também decisões individuais, desde que a opacidade técnica do modelo não impeça totalmente a transparência. Essa visão defende que a explicação deve permitir ao titular exercer seus direitos sob o RGPD e proteger sua autonomia e dignidade conforme os princípios dos direitos humanos (Selbst et al., 2017, p. 235), e existe o entendimento de que diferentes públicos devem receber diferentes níveis de informação. Indivíduos recebem explicações acessíveis, enquanto reguladores podem acessar dados técnicos, como o código-fonte (Kaminski, 2019, p. 210).

O art. 20 da LGPD garante a revisão de decisões automatizadas, mas não prevê o direito à intervenção humana nem a recusa à decisão automática. Já a explicabilidade, embora não expresso, pode ser extraído da interpretação do artigo e de normas constitucionais e consumeristas (Monteiro et al, 2021, p. 120-124). A legislação brasileira garante a quem for afetado por decisões automatizadas o acesso a informações objetivas sobre os critérios adotados,

sendo a explicação condição essencial para assegurar a transparência do processo.

### 4.2.3 Instrumentos de explicabilidade aplicados à IA na Saúde

A explicabilidade, tanto em modelos transparentes quanto em sistemas opacos, é vista como essencial na saúde, onde compreender as decisões da IA é crucial para garantir autonomia e confiança (Herzog, 2022, p. 2-3). Essa exigência se traduz na necessidade de máxima transparência sobre o funcionamento dos sistemas, de modo que tanto profissionais quanto pacientes possam entender os critérios adotados pelas tecnologias utilizadas (Watson et al., 2019, p. 1).

Nesse contexto, a confiança dos profissionais de saúde desempenha um papel decisivo. Quando esses profissionais compreendem como um sistema de IA opera e sobre quais bases formula suas recomendações, tendem a aceitá-lo como uma ferramenta útil na prática clínica, o que favorece sua integração segura ao cuidado (Morley et al., 2020, p. 2).

Por outro lado, a transparência não deve ser restrita aos especialistas. Para que a autonomia dos pacientes seja respeitada, é fundamental que eles também compreendam, ao menos minimamente, os critérios utilizados por sistemas automatizados em decisões que os afetem diretamente (Amann et al., 2020, p. 6). Esse cuidado se torna ainda mais necessário em contextos sensíveis, como os transplantes, nos quais a IA já vem sendo aplicada para auxiliar na alocação de órgãos, avaliação de compatibilidade e estimativas de sobrevida (Khorsandi et al., 2021, p. 2942).

No entanto, permitir o uso ético da IA em decisões críticas exige mais do que compreensão básica. É necessário garantir que tais sistemas não atuem de forma autônoma, sem supervisão humana. Explicações claras sobre os resultados gerados são fundamentais não apenas para justificar decisões, mas também para fortalecer o controle, promover a responsabilidade e ampliar a confiança dos usuários. Como diferentes usuários possuem diferentes objetivos e níveis de conhecimento, construir explicações que sejam simultaneamente

úteis e acessíveis é um dos grandes desafios da área. É nesse cenário que se insere o campo da XAI (eXplainable Artificial Intelligence), que busca desenvolver estratégias para tornar compreensíveis até mesmo os modelos mais complexos (Gunning et al., 2019).

Ainda assim, a explicação não precisa abranger todo o processo decisório da IA. A literatura aponta que a explicabilidade é mais necessária quando o sistema substitui diretamente o julgamento humano. Nesses casos, é essencial que ele possa responder, de maneira compreensível, a perguntas relevantes sobre o porquê de uma determinada decisão ter sido tomada (OCDE, 2019, p. 94). Para isso, recomenda-se que as explicações evidenciem os fatores mais relevantes considerados, aqueles que foram determinantes para o resultado final, e os motivos pelos quais casos semelhantes podem ter gerado decisões distintas. Essa abordagem facilita a avaliação do impacto específico de cada variável no resultado obtido (Doshi-velez et al., 2019, p. 4).

Uma das maneiras mais diretas de garantir explicabilidade é recorrer a algoritmos intrinsicamente interpretáveis, como regressão linear, regressão logística e árvores de decisão. Esses modelos utilizam estruturas matemáticas simples, como somas ponderadas ou regras condicionais, que permitem ao usuário visualizar com clareza como cada variável de entrada influenciou a previsão final (Molnar, 2022, p. 38-85). No entanto, como muitos modelos aplicados na saúde são opacos por natureza, é necessário recorrer a técnicas de explicação pós-hoc[22]. Nesse sentido, destacam-se ferramentas como o

[22] A expressão “explicação pós-hoc” (do latim post hoc, que significa “depois disso”) refere-se a métodos aplicados depois que um modelo de inteligência artificial já foi treinado e está operando, com o objetivo de interpretar ou explicar decisões que, por natureza, são opacas ou de difícil compreensão. Muitos modelos usados na área da saúde, como redes neurais profundas, *random forests* e modelos de aprendizado por reforço, são considerados “caixas-pretas”: eles fornecem resultados precisos, mas não revelam claramente como chegaram àquela conclusão. No contexto clínico, isso representa um desafio ético e jurídico: como validar ou confiar em uma decisão automatizada (como um diagnóstico ou recomendação de tratamento) se não compreendemos os critérios que a embasam? As técnicas pós-hoc são soluções que não alteram o funcionamento interno do modelo original, mas criam interpretações externas que ajudam a entender por que determinada decisão foi tomada. Elas atuam como tradutores de modelos complexos. Exemplos clássicos: LIME (Local Interpretable Model-Agnostic Explanations) cria um modelo explicável mais simples (como uma regressão linear) em torno de uma previsão específica do modelo complexo. Assim, é possível entender quais variáveis influenciaram aquela decisão individual, mesmo que o modelo geral seja opaco. SHAP (SHapley Additive exPlanations), baseado em teoria dos jogos, calcula o peso (importância) de cada variável na formação de uma previsão, distribuindo de forma justa o “mérito” da decisão entre os fatores envolvidos.

LIME e o SHAP, que utilizam modelos simplificados para simular e tornar inteligíveis as previsões dos sistemas mais complexos (Ribeiro et al., 2016).

O SHAP, especificamente, é uma das abordagens mais utilizadas para mensurar a contribuição individual de cada variável. Ele se baseia nos valores de Shapley, conceito originado da teoria dos jogos, que calcula a média da influência de cada fator considerando todas as combinações possíveis. Dessa forma, oferece uma estimativa justa e transparente do peso de cada entrada na decisão final (Lundberg et al., 2017, p. 2-4).

Além disso, no caso das redes neurais, um dos modelos mais complexos e difíceis de interpretar, há técnicas específicas voltadas à sua explicação, como os mapas de saliência, que realçam os pixels mais relevantes em classificações de imagem (Molnar, 2022, p. 213-215). Essas técnicas fazem parte de um amplo conjunto de métodos voltados à explicabilidade em diferentes algoritmos, reforçando o compromisso com a transparência e a confiança no uso da IA na saúde.

### 4.2.4 A Complexidade da explicação algorítmica na prática clínica

A exigência de explicações como base da transparência na inteligência artificial, especialmente na área da saúde, foi por muito tempo considerada um consenso. No entanto, essa ideia passou a ser questionada nos últimos anos diante das limitações práticas e conceituais da chamada IA explicável, com impactos diretos sobre o modo como se pensa a regulação no setor (Amann et al., 2022, p. 2).

Uma das principais tensões gira em torno da escolha entre desempenho e interpretabilidade. Modelos altamente precisos, como os de *Deep Learning*, oferecem excelentes resultados em tarefas complexas, mas sua lógica interna é extremamente difícil de ser compreendida. Por outro lado, modelos mais simples, embora mais fáceis de explicar, frequentemente apresentam desempenho inferior. Nesse sentido, a exigência de explicabilidade total pode restringir o uso de sistemas de IA de alto desempenho em contextos críticos da saúde (London, 2019, p. 15-17).

Apesar disso, há evidências de que, em cenários com dados bem estruturados, como ocorre frequentemente na saúde, modelos transparentes e interpretáveis podem alcançar desempenho similar ao dos modelos opacos. Por essa razão, defende-se que, sempre que possível, essas alternativas mais compreensíveis sejam priorizadas, sobretudo em decisões de alto impacto para os pacientes (Rudin, 2019, p. 2-3).

Ainda assim, as explicações atualmente oferecidas por sistemas de IA tendem a ser mais úteis para auditorias e análises coletivas do que para decisões clínicas individuais. Além disso, há o risco de que os usuários atribuam um grau excessivo de confiança a essas explicações, mesmo quando elas não são totalmente precisas, o que pode, ironicamente, aumentar a ocorrência de erros em vez de mitigá-los (Ghassemi et al., 2021, p. 746-748).

Outro ponto crítico diz respeito à equidade. Técnicas explicativas de IA podem apresentar desempenho desigual entre diferentes grupos populacionais. Isso é especialmente preocupante na saúde, onde a precisão e a justiça das decisões automatizadas devem ser garantidas para todos. Quando explicações favorecem apenas certos perfis, reforçam-se desigualdades já existentes no acesso e na qualidade do cuidado (Balagopalan et al., 2022, p. 2-3).

Além disso, explicações geradas por sistemas menos precisos e possivelmente falhos exigem uma abordagem cautelosa. Na área da saúde, onde decisões impactam diretamente a vida das pessoas, confiar em justificativas frágeis ou técnicas de baixa acurácia pode representar um risco ético e clínico relevante (Ghassemi et al., 2021, p. 748-749).

Em razão do seu caráter aproximado, técnico e, muitas vezes, opaco, as explicações atualmente disponíveis não oferecem suporte suficiente para exigências regulatórias robustas no campo da saúde. Elas são frequentemente limitadas, insinceras ou inadequadas para fundamentar decisões éticas ou jurídicas de maneira satisfatória (Babic et al., 2023, p. 858-859).

Como alternativa, uma proposta regulatória emergente sugere que o uso de modelos opacos seja condicionado à comparação com modelos interpretáveis. Caso uma solução transparente apresente desempenho equivalente, a adoção de sistemas do tipo "caixa-preta" poderia ser vetada, especialmente em decisões de alto risco. Abordagens como essa ganham força

e devem ser consideradas pelas autoridades responsáveis pela regulação da IA em saúde (Rudin, 2019, p. 8-9).

### 4.2.5 Registros técnicos como instrumentos de prestação de contas

Na regulação da inteligência artificial aplicada à saúde, a exigência de transparência é um dos pilares fundamentais. Para tanto, torna-se indispensável a documentação completa de todo o ciclo de vida do sistema, abrangendo desde os métodos empregados até os dados utilizados e as decisões tomadas durante o desenvolvimento. Esse tipo de registro detalhado permite não apenas a supervisão eficaz por parte das autoridades reguladoras, mas também fortalece a confiança social no uso dessas tecnologias (OMS, 2023, p. 8-10).

A segurança e a eficácia dos sistemas de IA dependem diretamente dessa transparência. Nesse sentido, iniciativas como o TRIPOD-AI e o PROBAST-AI surgem como ferramentas valiosas para padronizar os relatórios científicos e avaliar o risco de viés nos modelos clínicos. Esses instrumentos não apenas orientam pesquisadores, como também fornecem subsídios técnicos consistentes para os reguladores no processo de validação e controle desses sistemas (Collins et al., 2015, p. 1-7).

Complementando essas iniciativas, a proposta MINIMAR estabelece um conjunto de padrões mínimos para a documentação de sistemas de IA na área da saúde. O foco recai sobre a necessidade de explicitar dados essenciais, como a população usada no treinamento, a arquitetura do modelo e sua finalidade de uso. Ao exigir esses elementos, o modelo busca promover maior transparência, reduzir vieses e favorecer a justiça algorítmica no desenvolvimento e aplicação dessas tecnologias (Hernandez-boussard et al., 2020, p. 2012-2014).

## 4.3 *Accountability* em sistemas de IA

Na área da saúde, a responsabilização não apenas representa um princípio ético fundamental, mas também funciona como uma forma indireta, porém eficaz, de regulação da inteligência artificial. Isso se deve ao fato de que normas legais moldam comportamentos de profissionais e instituições, influenciando de maneira relevante os desfechos clínicos, mesmo sem atuar diretamente sobre os algoritmos (Gostin, 2003, p. 2).

Nesse contexto, a responsabilização assume um papel normativo que deve ser operacionalizado especialmente por meio dos mecanismos próprios da responsabilidade civil, funcionando como ferramenta regulatória essencial para garantir o uso ético da IA em saúde.

### 4.3.1 Fundamentos jurídicos da responsabilidade civil aplicada à IA

A responsabilidade civil é um instrumento jurídico que reflete como diferentes culturas lidam com riscos, danos e a necessidade de compensação. Ela impõe deveres a indivíduos e organizações, funcionando como um mecanismo de controle social e jurídico diante de prejuízos causados por condutas ou produtos (Engel et al., 2009, p. 1).

No campo do direito, a responsabilidade civil é regida por normas do direito privado e aplica-se sempre que uma pessoa física ou jurídica causa dano a outrem. Seu propósito principal é restaurar o equilíbrio violado, seja ele patrimonial ou moral. Embora os sistemas jurídicos variem, principalmente entre os modelos de *common law* e *civil law*, ambos compartilham princípios fundamentais, como a responsabilidade subjetiva, que exige a comprovação de culpa, e a responsabilidade objetiva, que prescinde dessa comprovação, bastando o nexo causal entre conduta e dano (Bussani et al., 2022, p. 7-9).

Na aplicação da inteligência artificial à saúde, a responsabilidade tende a seguir modelos já utilizados para dispositivos médicos. Ao serem equiparados

a produtos de interesse à saúde, os sistemas de IA passam a ser regulados sob a lógica da responsabilidade civil que se aplica a tais bens (Minssen et al., 2020, p. 621-623).

A responsabilidade objetiva dos fabricantes, por exemplo, foi historicamente consolidada como uma forma de proteger os consumidores contra produtos defeituosos. Essa doutrina baseia-se na premissa de que quem coloca um produto no mercado deve responder por danos decorrentes de seu uso, mesmo sem culpa, já que está em melhor posição para prever riscos e absorver os custos relacionados a acidentes (Bussani et al., 2022, p. 221).

Entretanto, os modelos de IA que aprendem autonomamente apresentam desafios inéditos. Seus criadores nem sempre conseguem prever todas as decisões futuras da máquina, o que gera uma lacuna no regime tradicional de responsabilidade: em muitos casos, não há um agente claramente identificável que possa ser responsabilizado por eventos inesperados (Matthias, 2004, p. 176).

Diante disso, uma proposta da Oxford Internet Institute (OII), da Universidade de Oxford, é repartir a responsabilidade entre os diversos agentes envolvidos, tanto humanos quanto artificiais. Ainda que não haja intenção direta, essa abordagem reconhece a natureza coletiva e complexa dos sistemas de IA, nos quais danos podem decorrer de interações entre múltiplos elementos (Floridi et al., 2018, p. 694).

Tradicionalmente, a doutrina consolidada em decisões da Suprema Corte de diversos estados americanos do “intermediário instruído” (learned intermediary doctrine) transferia ao médico a responsabilidade pelas decisões clínicas, isentando o fabricante do dispositivo. Contudo, essa lógica perde força diante de sistemas de IA cada vez mais autônomos e opacos, nos quais se torna difícil identificar com precisão a origem da decisão, enfraquecendo a atribuição clara de responsabilidade (Sullivan et al., 2019, p. 162-163).

Além disso, a responsabilidade civil na saúde costuma distinguir os deveres de médicos e instituições. Enquanto os médicos são responsáveis por agir com diligência, sem garantia de resultado, os hospitais assumem responsabilidade por falhas sistêmicas. Com a introdução de sistemas opacos

de IA, essa divisão tradicional torna-se imprecisa, exigindo novas formas de distribuição de deveres legais (Price, 2018, p. 390).

Para que a IA seja amplamente adotada na prática clínica, é essencial que as responsabilidades jurídicas e éticas dos diversos atores estejam bem definidas. A incerteza quanto a quem deve responder por erros ou falhas limita a confiança nas tecnologias e dificulta sua incorporação segura aos cuidados de saúde (Vayena et al., 2018, p. 4).

Essa incerteza levanta uma questão prática: médicos poderão ser responsabilizados por seguirem recomendações da IA, ou por decidirem ignorá-las? O debate se intensifica à medida que diretrizes clínicas passam a incorporar algoritmos como referência, o que pode tornar a não observância dessas recomendações passível de ser interpretada como negligência (Schönberger, 2019, p. 192).

A atribuição de responsabilidade legal no uso da IA em saúde dependerá, portanto, de escolhas normativas conscientes. Os sistemas jurídicos precisarão decidir quais comportamentos incentivar ou coibir, além de definir com clareza qual será o novo padrão de cuidado a ser exigido dos profissionais da saúde diante da consolidação dessas ferramentas (OMS, 2021, p. 76).

No caso das instituições, a responsabilidade objetiva dos hospitais pelos danos causados durante o atendimento permanece. Com o uso crescente da IA, aumenta-se a expectativa de que essas entidades adotem cuidados adicionais na seleção, elaboração e monitoramento desses sistemas, além de promover o treinamento adequado dos profissionais. No entanto, muitas vezes os hospitais acabam confiando nas validações fornecidas por fabricantes ou agências reguladoras, já que nem sempre dispõem da capacidade técnica para avaliar as ferramentas por si mesmos (Price et al., 2022, p. 11).

Essa confiança, contudo, pode se mostrar problemática, especialmente quando se reconhece que a imprevisibilidade do comportamento da IA fragiliza os modelos tradicionais de responsabilização. Se os atos do sistema forem considerados fora do controle humano direto, torna-se ainda mais difícil responsabilizar os desenvolvedores de forma justa e juridicamente sustentável (Sullivan et al., 2019, p. 163).

Por fim, a discussão sobre responsabilidade também atinge os próprios reguladores. Nos Estados Unidos, há indícios de que dispositivos aprovados por meio do processo padrão da FDA poderiam ser resguardados contra ações judiciais estaduais sobre segurança ou eficácia. Em contrapartida, produtos autorizados por vias menos rigorosas, como o processo 510(k), continuariam sujeitos a litígios civis, o que ressalta a complexidade do cenário regulatório (Price et al., 2022, p. 19).

### 4.3.2 Definição de responsáveis pelo uso da IA na área da Saúde

Na Europa, observa-se uma tendência consolidada de qualificar softwares baseados em inteligência artificial como dispositivos médicos sempre que possuem uma finalidade diagnóstica ou terapêutica declarada pelo fabricante. Essa classificação não é apenas formal: ela traz implicações jurídicas importantes, especialmente no campo da responsabilidade civil, pois esses sistemas passam a ser equiparados a produtos de interesse à saúde, sujeitos à regulação técnica e jurídica aplicável a tais bens (Minssen et al., 2020, p. 621).

Nesse contexto, aplica-se a lógica da responsabilidade objetiva por produtos, segundo a qual os fabricantes devem arcar com os danos decorrentes do uso dos bens que colocam no mercado, independentemente de culpa. Tal regime é justificado pelo princípio de que quem lucra com os riscos do consumo deve também assumir os custos dos danos, o que protege os usuários e incentiva práticas mais seguras (Bussani et al., 2022, p. 221).

Contudo, quando se trata de sistemas de IA com capacidade de aprendizado autônomo, a lógica tradicional da responsabilidade enfrenta sérios desafios. A imprevisibilidade de suas ações dificulta a atribuição direta de responsabilidade aos fabricantes, pois nem mesmo seus criadores conseguem antever todas as decisões que os sistemas podem tomar, gerando uma lacuna jurídica ainda não plenamente resolvida (Matthias, 2004, p. 176).

Uma resposta a esse impasse é a adoção da ideia de responsabilidade moral distribuída. Essa abordagem reconhece que, em sistemas complexos

como os de IA, os danos muitas vezes resultam da interação entre humanos e máquinas, sem que haja necessariamente uma intenção individual identificável. Com isso, torna-se possível responsabilizar coletivamente os envolvidos, ainda que nenhum deles tenha agido com dolo ou negligência direta (Floridi et al., 2018, p. 693).

Nos Estados Unidos, a doutrina do "intermediário instruído" ainda exerce forte influência. Por essa perspectiva, o médico é considerado o responsável final pelas decisões clínicas, mesmo quando utiliza dispositivos fabricados por terceiros, isentando o fabricante da responsabilidade por eventuais danos. No entanto, essa lógica torna-se frágil frente à crescente autonomia e opacidade dos sistemas de IA, que dificultam ao profissional compreender ou avaliar as recomendações fornecidas (Sullivan et al., p. 162).

No contexto do direito civil brasileiro e de sistemas jurídico-privatistas similares, a responsabilidade civil na prática médica distingue entre profissionais e instituições: os médicos respondem com base na culpa (negligência, imprudência ou imperícia), enquanto as instituições de saúde, como hospitais, estão sujeitas à responsabilidade objetiva por falhas estruturais ou defeitos na prestação do serviço. Contudo, com a introdução de sistemas opacos de inteligência artificial, os chamados modelos de caixa-preta, essa distinção se torna menos clara, exigindo revisão dos critérios normativos vigentes (Price, 2018, p. 390).

A ausência de regras claras sobre quem deve ser responsabilizado em caso de falhas atribuíveis à IA compromete a segurança jurídica e limita sua adoção clínica. A incerteza jurídica pode frear a incorporação dessas tecnologias nos sistemas de saúde, especialmente quando não há diretrizes bem definidas sobre responsabilidade (Vayena et al., 2018, p. 3).

Essa dificuldade é acentuada quando se considera que os médicos, além de poderem ser responsabilizados por seguir decisões da IA que causem danos, também correm o risco de serem acusados de negligência por ignorar recomendações automatizadas formalizadas em diretrizes clínicas. O dilema coloca o profissional em uma situação de vulnerabilidade jurídica sem precedentes (Schönberger, 2019, p. 192).

A definição da responsabilidade jurídica na aplicação da IA à saúde dependerá, portanto, de escolhas normativas estratégicas. Caberá aos

legisladores e reguladores estabelecer que condutas serão incentivadas ou desestimuladas, bem como qual será o novo padrão de cuidado exigido dos profissionais diante do uso consolidado da IA na prática clínica (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2021, p. 76).

As instituições hospitalares, por sua vez, seguem sendo responsabilizadas objetivamente por danos ocorridos durante o atendimento. Com a entrada da IA nos ambientes clínicos, espera-se dessas organizações o dever de cuidado na escolha, supervisão e correta integração dessas tecnologias, além de fornecer capacitação adequada aos profissionais que as utilizam. No entanto, muitas vezes os hospitais dependem das validações fornecidas por desenvolvedores e reguladores, pois não dispõem de meios técnicos próprios para avaliar a complexidade desses sistemas (Price et al., 2022, p. 11).

A autonomia e imprevisibilidade da IA também dificultam a responsabilização de seus desenvolvedores. Quando as ações da tecnologia fogem ao controle humano direto, fragiliza-se a aplicação da responsabilidade objetiva por produto, o que pode deixar os pacientes sem reparação adequada diante de falhas clínicas (Sullivan et al., 2019, p. 163).

A aplicação da doutrina da preempção federal a dispositivos médicos baseados em inteligência artificial tem sido debatida por juristas e pesquisadores de prestígio nos EUA, como I. Glenn Cohen, Nicolas Terry e Barbara Evans, além de instituições como a Harvard Law School, o Petrie-Flom Center, o Journal of Law and the Biosciences e o Brookings Institution. O foco das discussões está na compatibilidade entre os processos regulatórios tradicionais da FDA (como o PMA) e as características adaptativas, opacas e dinâmicas dos sistemas de IA, que desafiam os marcos clássicos de responsabilidade civil e segurança do paciente. (Price et al., 2022, p. 20).

### 4.3.3 Instrumentos viáveis de imputação de responsabilidade

Propostas doutrinárias vêm sendo desenvolvidas para enfrentar os desafios da atribuição de responsabilidade no uso de sistemas de inteligência

artificial, especialmente no campo da saúde, onde os modelos tradicionais baseados na culpa mostram-se insuficientes para lidar com a complexidade dessas tecnologias. Entre as alternativas sugeridas estão a adaptação do padrão de cuidado exigido dos profissionais de saúde, a criação de regimes de responsabilidade compartilhada entre os diversos agentes envolvidos e, até mesmo, a concessão de personalidade jurídica limitada a sistemas autônomos (Sullivan et al., 2019, p. 164).

Uma das abordagens mais discutidas é a divisão de responsabilidades entre todos os participantes do ciclo de vida da IA, desenvolvedores, fabricantes, gestores hospitalares e profissionais clínicos, evitando concentrar toda a responsabilização em um único agente. Essa estrutura permitiria oferecer compensação às vítimas mesmo em casos nos quais a causa exata do erro não pode ser claramente identificada (Vladeck, 2014, p. 129).

Diante da crescente dificuldade em rastrear falhas específicas em sistemas opacos, propõe-se ainda a aplicação do regime de responsabilidade solidária, que permite à vítima acionar qualquer dos envolvidos, cabendo a estes, posteriormente, discutir a repartição interna dos encargos. Outra possibilidade discutida é a concessão de personalidade jurídica restrita à IA, com a atribuição de certos deveres legais e contratuais, ainda que sem reconhecê-la como sujeito pleno de direitos (Allain, 2013, p. 1074).

Como a inteligência artificial, por ora, não pode ser tratada como sujeito jurídico autônomo pleno, recai sobre as empresas desenvolvedoras e fornecedoras o dever de assumir os ônus legais decorrentes do uso de suas soluções, inclusive por meio de cláusulas contratuais específicas. No entanto, essa solução não resolve o núcleo do problema, principalmente quando há envolvimento clínico. Em muitos casos, transfere-se ao profissional de saúde, especialmente quando a IA está incorporada a diretrizes clínicas, a responsabilidade por avaliar e aplicar as decisões do sistema (Price, 2018, p. 421).

Nesse cenário, a responsabilização do médico pelo uso de sistemas de suporte à decisão clínica baseados em IA/ML dependerá da sua adesão ao padrão de cuidado vigente. A jurisprudência tende a proteger o profissional que atua de acordo com esse padrão, mesmo em casos de erro. Por outro lado, o médico que ignora uma recomendação correta ou adota uma sugestão errada,

ambas em desacordo com o padrão aceito, poderá ser responsabilizado em caso de dano. À medida que os sistemas de IA forem formalmente integrados às diretrizes clínicas, suas recomendações passarão a influenciar diretamente a definição do padrão de cuidado exigido (Price II et al., 2019, p. 1766).

Essa abordagem, no entanto, pode ter um efeito desestimulante sobre o uso da IA na prática clínica. Enquanto essas tecnologias não forem incorporadas de maneira oficial e validada aos protocolos terapêuticos reconhecidos, persistirá entre os profissionais o receio de infringir o padrão de cuidado. Essa incerteza, motivada por questões éticas ou pelo medo de responsabilização, tende a fazer com que os médicos evitem seguir recomendações algorítmicas, ainda que potencialmente benéficas, o que acaba por limitar a expansão da IA na medicina. Para mudar esse panorama, é fundamental que entidades profissionais e autoridades reguladoras incorporem esses sistemas em diretrizes clínicas robustas, validadas e submetidas a constante auditoria (OMS, 2021, p. 79).

Considerando as limitações estruturais dos modelos tradicionais de responsabilidade civil para lidar com os danos provocados pela IA na saúde, autores têm defendido a adoção de um sistema de compensação sem culpa. Esse modelo, já implementado em países como Nova Zelândia, Dinamarca e Suécia, permite que pacientes recebam indenizações por eventos adversos inesperados sem a necessidade de provar culpa ou entrar com ações judiciais. A compensação seria administrada por meio de fundos públicos e privados, aliviando as vítimas do ônus de identificar o responsável em sistemas altamente complexos e de difícil rastreabilidade (Holm et al., 2021, p. 184).

### 4.3.4 Novas abordagens para a responsabilidade civil diante da IA

Atualmente, tramitam na União Europeia duas propostas legislativas com o objetivo de adaptar as normas de responsabilidade civil aos desafios impostos pela inteligência artificial. Essas iniciativas buscam responder às dificuldades específicas relacionadas à opacidade, autonomia e complexidade dos sistemas de IA, que tornam a prova da culpa e do nexo causal

especialmente problemática. A primeira proposta, apresentada pela Comissão Europeia, visa modificar as regras de responsabilidade extracontratual para garantir uma proteção mais eficaz às vítimas e proporcionar maior segurança jurídica às empresas, especialmente em situações transfronteiriças (UE, 2022c, p. 1-3).

Em paralelo, a segunda proposta legislativa propõe a revisão da Diretiva 85/374/EEC sobre responsabilidade por produtos defeituosos, com o intuito de atualizá-la diante dos riscos específicos associados às tecnologias digitais, como softwares autônomos e sistemas de IA. A proposta reforça a responsabilidade objetiva dos fabricantes, inclusive de software e serviços digitais integrados, reduz o ônus da prova para os consumidores em casos complexos e amplia a cobertura de compensações, incluindo danos materiais e prejuízos decorrentes da perda ou corrupção de dados (UE, 2022b, p. 1-2).

A chamada Diretiva de Responsabilidade da IA (AI Liability Directive - AILD), por sua vez, concentra-se nas ações baseadas em culpa. Reconhecendo as barreiras probatórias impostas pela natureza técnica dos sistemas de IA, o texto propõe três medidas para mitigar essas dificuldades: (a) harmonização das obrigações de documentação por parte dos fabricantes; (b) presunção de causalidade sempre que normas de segurança forem violadas; e (c) inversão parcial do ônus da prova, quando a única forma de demonstrar o dano estiver nos mecanismos internos do sistema. Ainda, o texto prevê uma revisão futura para analisar a viabilidade de harmonizar regras de responsabilidade objetiva e de exigir seguros obrigatórios, com foco na proteção dos direitos fundamentais e prevenção de lacunas jurídicas (UE, 2022a, p. 8-9).

Por sua vez, a proposta de revisão da Diretiva de Responsabilidade por Produtos de 1985 aprofunda a adaptação às tecnologias digitais. Embora preserve a lógica da responsabilidade objetiva que dispensa a comprovação de culpa, ela amplia as definições de “produto” e “defeito” para incluir riscos decorrentes de falhas de atualização, vulnerabilidades cibernéticas e desconexões tecnológicas. A noção de dano também é ampliada, passando a incluir prejuízos à saúde psicológica e à perda de dados. Para proteger os consumidores diante da complexidade técnica de sistemas como a IA de caixa-

preta, a proposta admite presunções de causalidade que facilitam a responsabilização em litígios de alta complexidade (De Luca, 2025, p. 4-6).

De forma geral, ambas as propostas legislativas europeias sinalizam um aumento do risco regulatório para fabricantes e fornecedores de produtos baseados em IA e *Machine Learning*, ao mesmo tempo em que ampliam a proteção ao consumidor. Contudo, por se tratarem de diretivas, essas normas ainda precisarão ser transpostas ao ordenamento jurídico dos Estados-membros. Apesar dos avanços, persistem preocupações quanto à responsabilização por danos causados por sistemas médicos de IA classificados como “caixa-preta”, especialmente nos casos em que os sistemas operam conforme os requisitos legais, mas ainda assim geram efeitos lesivos. Nessas situações, tanto os desenvolvedores quanto os profissionais da saúde podem ser isentados de responsabilidade, deixando as vítimas sem acesso à reparação, o que revela lacunas jurídicas que ainda precisam ser enfrentadas (Duffourc et al., 2023, p. 2-4).

## 4.4 Síntese reflexiva

A análise dos referenciais normativos propostos pela Organização Mundial da Saúde para o uso da inteligência artificial na saúde revela a importância de se estabelecer um arcabouço regulatório robusto, capaz de garantir que os benefícios dessas tecnologias sejam efetivamente convertidos em avanços para os sistemas assistenciais, sem comprometer os direitos fundamentais dos indivíduos. A confiabilidade dos sistemas, a explicabilidade dos algoritmos e a definição clara de responsabilidades jurídicas constituem eixos estruturantes desse modelo regulatório emergente.

A OMS propõe uma abordagem multidimensional, que reconhece a especificidade da saúde e demanda a adoção de mecanismos de governança adaptativos, transparentes e centrados no ser humano. Ao enfatizar a responsabilidade compartilhada entre desenvolvedores, profissionais, instituições e gestores públicos, a organização sinaliza para a necessidade de uma regulação colaborativa, que combine rigor técnico com sensibilidade ética.

A construção de uma governança global para a inteligência artificial em saúde passa, necessariamente, pelo fortalecimento da *accountability*, pela padronização dos critérios de segurança, e pela promoção de ambientes regulatórios que favoreçam tanto a inovação quanto a proteção dos usuários. Nesse sentido, os referenciais da OMS funcionam como guias para a criação de políticas públicas orientadas pela equidade, pela justiça distributiva e pela preservação da confiança no sistema de saúde.

**Quadro 4** – Dimensões da abordagem multidimensional da OMS para a governança da IA em saúde.

| DIMENSÃO | REFERÊNCIA |
|---|---|
| Centralidade no ser humano | Adoção de tecnologias que respeitem a dignidade, os direitos e as necessidades dos pacientes. |
| Responsabilidade compartilhada | Distribuição de deveres entre desenvolvedores, profissionais, instituições e poder público. |
| Transparência | Clareza nos processos, critérios e decisões produzidas por sistemas de IA em saúde. |
| Segurança | Estabelecimento de padrões técnicos rigorosos para minimizar riscos aos usuários. |
| Accountability | Prestação de contas e mecanismos de responsabilização por eventuais falhas ou abusos. |
| Governança adaptativa | Estrutura normativa flexível e responsiva às rápidas transformações tecnológicas. |
| Colaboração multissetorial | Participação integrada de atores públicos, privados, acadêmicos e da sociedade civil. |
| Promoção da inovação ética | Estímulo ao desenvolvimento tecnológico alinhado a valores éticos e ao bem comum. |
| Equidade e inclusão | Garantia de acesso justo às tecnologias, com atenção às populações vulnerabilizadas. |
| Regulação global articulada | Construção de padrões normativos internacionais coerentes e interoperáveis. |

**Fonte**: Elaborado pelo autor

No capítulo seguinte, serão abordadas as novas tendências e estruturas regulatórias em desenvolvimento, com ênfase nas iniciativas jurídicas voltadas ao controle, à testagem e à normatização de tecnologias emergentes, como forma de consolidar uma inteligência artificial compatível com os direitos fundamentais e os princípios de um sistema de saúde universal.

# 5 RUMOS DA INTELIGÊNCIA ARTIFICIAL

A rápida expansão da inteligência artificial nas últimas décadas tem exigido, de forma progressiva, uma reformulação das estratégias regulatórias tradicionais. À medida que essas tecnologias adquirem maior capacidade decisória, adaptativa e autônoma, torna-se cada vez mais evidente a necessidade de estruturas normativas mais responsivas, capazes de acompanhar o ritmo das inovações sem comprometer a proteção dos direitos fundamentais.

Nessa conjuntura, observa-se um movimento internacional de transição entre diretrizes éticas de caráter geral e normas jurídicas específicas voltadas à regulação das tecnologias emergentes. A crescente sofisticação dos modelos algorítmicos e sua inserção em contextos sensíveis, como a saúde, impõem desafios inéditos para os formuladores de políticas públicas, exigindo o desenvolvimento de novos mecanismos de testagem, validação e responsabilização (arquivamento de informações[23]).

Este capítulo tem como objetivo analisar os rumos contemporâneos da regulação da inteligência artificial, com foco nas transformações institucionais e jurídicas que vêm sendo promovidas para lidar com os riscos e oportunidades associados ao uso dessas tecnologias. Serão examinadas, em especial, as novas estruturas de testagem jurídica, como as *sandboxes* regulatórias, bem como a consolidação de regimes legais mais rígidos voltados à governança de sistemas inteligentes, buscando oferecer uma leitura crítica e atualizada dos

[23] Arquivos de log (ou logs) são registros automáticos e contínuos de eventos, processos e decisões executados por um sistema computacional, incluindo, no caso da inteligência artificial, as etapas internas envolvidas no processamento de dados, no funcionamento do modelo e na geração das respostas. No contexto de sistemas de IA na saúde, os arquivos de log podem incluir, por exemplo: a) dados de entrada recebidos (ex: sintomas ou exames analisados); b) versão do modelo ou algoritmo usado; c) parâmetros de decisão ativados; d) Tempo e data da operação; e) resultado gerado pelo sistema (ex: sugestão de diagnóstico); f) erros ou falhas identificadas durante o processo; g) quem ou qual sistema acessou ou modificou o modelo. Para que servem? **Auditoria**: Permitem rastrear e verificar como uma decisão automatizada foi tomada, elemento fundamental em ambientes regulados como a saúde. **Responsabilização**: Em caso de erro clínico, ajudam a identificar se houve falha humana, técnica ou algorítmica, sendo, portanto, instrumentos-chave para imputar responsabilidade civil ou institucional; **Explicabilidade**: Podem ser usados junto com técnicas pós-hoc para reconstruir o raciocínio do sistema, mesmo quando ele é opaco; **Segurança**: auxiliam na detecção de invasões, manipulações indevidas ou mau uso do sistema.

caminhos possíveis para uma inteligência artificial ética, segura e humanamente orientada.

## 5.1 Entre diretrizes gerais e normas específicas

A tendência observada nas políticas de proteção de dados pessoais parece se repetir na regulação da inteligência artificial em sentido mais amplo. A União Europeia adota uma abordagem legislativa horizontal e uniforme, como demonstra a proposta do AI Act, que estabelece regras aplicáveis a todos os Estados-membros. Já os Estados Unidos mantêm uma estratégia setorial, sem a previsão de uma legislação federal geral sobre IA, focando em normas específicas emitidas por agências reguladoras e ações executivas. Exemplo dessa abordagem é o Decreto Executivo de 2023, que estabelece princípios orientadores como segurança, transparência, equidade e proteção de direitos civis, ao mesmo tempo em que atribui responsabilidades regulatórias a diversas agências federais (EUA, 2023).

Apesar das diferenças estruturais entre os modelos europeu e norte-americano, observa-se uma aproximação no que se refere à regulação da IA aplicada à saúde. Ambos reconhecem a necessidade de normativos específicos para esse campo, dadas suas implicações éticas, clínicas e de segurança. Na União Europeia, o AI Act se articula com o Regulamento de Dispositivos Médicos (MDR); nos Estados Unidos, a Food and Drug Administration (FDA) lidera a regulamentação por meio de uma abordagem setorial. Ainda que partam de arquiteturas distintas, os dois modelos convergem em relação à necessidade de lidar com temas como aprendizado contínuo, viés algorítmico e exigência de transparência em aplicações médicas de IA (Vokinger et al., 2021, p. 738-739).

O movimento regulatório europeu teve início em fevereiro de 2020, com a publicação do Livro Branco sobre a Inteligência Artificial: Uma abordagem europeia para a excelência e a confiança, pela Comissão Europeia. O documento estabelece uma estratégia de dupla abordagem, voltada à

promoção simultânea da excelência na inovação e da regulação responsável. Entre suas diretrizes centrais estão: a necessidade de criar um ecossistema de confiança, com foco em segurança, transparência e explicabilidade; o desenvolvimento de uma estrutura regulatória baseada no risco, que impõe exigências proporcionais ao impacto dos sistemas de IA; e o reforço dos direitos fundamentais, como privacidade, não discriminação e dignidade humana. O Livro Branco também delineia medidas para apoiar pesquisa, qualificação profissional e cooperação internacional, revelando a intenção estratégica da União Europeia de posicionar-se como referência ética e normativa global na regulação da IA (UNIÃO EUROPEIA, 2020, p. 1-3).

Embora tenha destacado a saúde como área prioritária, o Livro Branco limitou-se a apresentar princípios gerais, sem avançar na formulação de regras específicas para o setor. Essa lacuna regulatória abriu espaço para desafios importantes, especialmente no que se refere à compatibilização entre inovação e segurança clínica (Cohen et al., 2020, p. e377).

A consolidação do marco normativo europeu ocorreu com a apresentação do AI Act, em 2021, aprovado em 2024. Trata-se do primeiro regulamento abrangente sobre IA no mundo. A norma adota uma abordagem baseada em risco, classificando os sistemas de IA em quatro categorias: risco inaceitável (proibidos), alto risco (submetidos a avaliação de conformidade), risco limitado (com exigências mínimas de transparência) e risco mínimo (sem obrigações legais, com adesão voluntária) (UNIÃO EUROPEIA, 2024, p. 1-3).

O AI Act organiza os sistemas de alto risco em dois eixos distintos: de um lado, produtos já cobertos por normas de segurança específicas, como dispositivos médicos, e de outro, aplicações consideradas sensíveis, como reconhecimento biométrico, educação, relações de trabalho e justiça. Essas aplicações exigem registro em um banco de dados europeu e devem obedecer a regras técnicas rigorosas antes de sua liberação no mercado (Madiega, 2024, p. 3-4).

Para assegurar a aplicação harmonizada do regulamento, o AI Act institui um modelo de governança multinível, com a criação do Comitê Europeu de inteligência artificial, responsável pela coordenação supranacional, e a designação de autoridades nacionais para fiscalização e supervisão em cada país-membro. Apesar de seu alcance, o regulamento não trata com

profundidade das especificidades da IA na saúde, especialmente no que diz respeito a aspectos clínicos, éticos e de desempenho técnico. Por essa razão, espera-se que os dispositivos regulatórios já existentes, como o MDR, continuem a atuar como referência setorial, ao passo que a Agência Europeia de Medicamentos (EMA) sinaliza a necessidade de regulamentações complementares voltadas exclusivamente à IA médica (UNIÃO EUROPEIA, 2024, p. 1-3).

Comparando os modelos, observa-se que a União Europeia adota uma abordagem normativa horizontal e preventiva, com o AI Act estabelecendo regras uniformes e articuladas a regulamentos setoriais, como o dos dispositivos médicos. Já os Estados Unidos seguem uma lógica setorial, com regulação fragmentada conduzida por agências como a FDA. Apesar dessas diferenças estruturais, ambos convergem na área da saúde ao reconhecerem a necessidade de normas específicas para aplicações médicas de IA, diante de seus impactos sobre direitos fundamentais. Essa convergência demonstra um consenso internacional sobre a importância de tratar a IA em saúde com rigor técnico, transparência e segurança jurídica.

## 5.2 Estruturas de testagem jurídica para tecnologias emergentes

Inspiradas inicialmente pelo setor financeiro britânico, as *sandboxes* regulatórias surgem como ambientes controlados para testar tecnologias inovadoras com regras mais flexíveis e sob supervisão estatal. No contexto do AI Act, esse modelo foi incorporado como uma estratégia para equilibrar inovação e segurança, permitindo que sistemas de IA sejam avaliados em condições reais, com acompanhamento regulatório. O objetivo é facilitar o acesso ao mercado e, ao mesmo tempo, aprimorar a capacidade dos reguladores de compreender e acompanhar o desenvolvimento tecnológico (Buocz et al., 2023, p. 359-360).

Além da União Europeia, muitos países também adotaram *sandboxes* regulatórios, especialmente no setor financeiro. A expansão dessa ferramenta

em nível global reflete seu potencial para testar inovações com controle de riscos. Segundo relatório do Banco Mundial, até 2020 havia 73 *sandboxes* em funcionamento em 57 países, demonstrando o crescente interesse por modelos regulatórios mais dinâmicos e experimentais (BANCO MUNDIAL, 2020, p. IX-X).

No âmbito específico do AI Act, as *sandboxes* regulatórias são propostas com dois objetivos centrais: permitir a testagem supervisionada de soluções inovadoras e contribuir para a formulação de normas mais adequadas a partir da experiência prática. Essa combinação entre experimentação e gestão de risco beneficia tanto o setor privado quanto os órgãos reguladores, ao promover um ambiente colaborativo para inovação responsável (Madiega, 2022, p. 2).

O regulamento europeu prevê ainda exceções específicas para o uso de dados pessoais em *sandboxes* regulatórios, quando esses testes envolvem domínios sensíveis como saúde pública, meio ambiente, segurança pública e justiça penal. Essas autorizações excepcionais visam garantir que a experimentação seja feita com segurança, e que os dados sejam utilizados conforme regras definidas pelas autoridades nacionais competentes ou pela Autoridade Europeia para a Proteção de Dados (UNIÃO EUROPEIA, 2024, art. 53-54).

A experiência recente da FDA, nos Estados Unidos, durante a pandemia de COVID-19, evidenciou o valor de abordagens regulatórias mais ágeis e adaptativas. As autorizações emergenciais concedidas à época funcionaram, na prática, como uma espécie de sandbox, permitindo a rápida incorporação de inovações em saúde pública. Esse precedente reforça a viabilidade das *sandboxes* regulatórias como ferramenta útil também no setor da saúde (Sherkow, 2022, p. 359-361).

Nesse cenário, cresce a expectativa de que as *sandboxes* se consolidem como instrumentos relevantes para avaliação de tecnologias emergentes na área da saúde digital. Elas podem contribuir para adaptar os processos regulatórios às rápidas transformações do setor, promovendo um ambiente de inovação mais seguro e eficiente (Leckenby et al., 2021, p. 867).

Particularmente no caso da IA aplicada à saúde, as *sandboxes* regulatórias podem desempenhar papel decisivo. Esses ambientes oferecem a

possibilidade de compreender melhor os sistemas em desenvolvimento, testar critérios de segurança e eficácia antes de sua ampla utilização, e envolver os desenvolvedores em processos mais transparentes e responsáveis (OMS, 2023, p. 37).

Apesar de seu potencial, persistem importantes incertezas quanto à aplicação das *sandboxes* no setor da saúde. Ainda não há clareza sobre quais autoridades seriam competentes para regulamentar e conduzir os testes, como garantir igualdade de oportunidades entre empresas participantes e não participantes, nem como lidar com os riscos jurídicos e de responsabilização decorrentes de falhas durante os períodos de experimentação (Buocz et al., 2023, p. 370-371).

Dessa forma, as *sandboxes* regulatórias representam uma inovação institucional promissora para enfrentar os desafios regulatórios impostos pela inteligência artificial, especialmente na área da saúde. Ao oferecer um ambiente controlado para experimentação supervisionada, elas permitem testar soluções tecnológicas de maneira segura, promovendo aprendizado regulatório contínuo e redução de incertezas jurídicas. No entanto, sua efetividade dependerá da definição clara de competências institucionais, da adoção de critérios equitativos para participação e da estruturação de mecanismos de responsabilização adequados. O êxito dessas iniciativas exigirá, portanto, não apenas flexibilidade regulatória, mas também firmeza no compromisso com os direitos fundamentais e a proteção da saúde pública.

## 5.3 Rumos da IA na realidade brasileira

No contexto normativo brasileiro, destaca-se o Projeto de Lei nº 2.338/2023, atualmente em tramitação na Câmara dos Deputados, como a principal iniciativa voltada à regulação da inteligência artificial. Embora inspirado em modelos estrangeiros como o AI Act europeu, o texto brasileiro incorpora diretrizes próprias, classificando como sistemas de alto risco, entre outros, aqueles voltados a aplicações na área da saúde, especialmente quando envolvem riscos relevantes à integridade física e mental das pessoas. Para

esses casos, o projeto prevê uma série de mecanismos de controle e fiscalização, como a exigência de supervisão humana, testes de segurança, explicabilidade, mitigação de vieses, comunicação obrigatória de incidentes graves e a criação de um Sistema Nacional de Regulação e Governança da IA, com atuação de autoridades setoriais (BRASIL, 2025, p. 9).

Dessa forma, o projeto prevê uma série de mecanismos de controle e fiscalização. Entre eles, destaca-se a obrigatoriedade de supervisão humana nos sistemas de IA de alto risco, como forma de prevenir ou minimizar impactos negativos sobre os direitos e liberdades dos indivíduos. Essa supervisão deve permitir que operadores humanos compreendam, intervenham e priorizem riscos, exceto nos casos em que sua implementação seja comprovadamente inviável, hipótese em que devem ser adotadas medidas alternativas eficazes (BRASIL, 2025, p. 7).

Outro ponto de destaque é a exigência de medidas técnicas de governança e documentação, previstas no art. 18. Essas medidas incluem: testes de confiabilidade e segurança, mitigação de vieses discriminatórios, rastreabilidade das decisões e explicabilidade dos resultados. Tanto desenvolvedores quanto aplicadores têm obrigações nesse processo, o que fortalece a estrutura regulatória preventiva (BRASIL, 2025, p. 13).

Adicionalmente, o projeto estabelece a obrigatoriedade de comunicação de incidentes graves (art. 42), como falhas que afetem a integridade física de pessoas ou provoquem violações significativas a direitos fundamentais. Nesses casos, os agentes responsáveis devem notificar a autoridade setorial competente, que poderá determinar providências corretivas e medidas de mitigação (BRASIL, 2025, p. 22).

A proposta também prevê a criação de uma base de dados pública de sistemas de IA de alto risco, coordenada pela autoridade competente em parceria com as autoridades setoriais. Essa base deve disponibilizar, de forma acessível ao público, os documentos relativos às avaliações de impacto algorítmico, respeitando os limites legais de sigilo e proteção de dados (Ibidem).

Por fim, o PL institui o Sistema Nacional de Regulação e Governança da Inteligência Artificial (SIA), composto por diversos órgãos como a ANPD, autoridades setoriais, comitês técnicos e especialistas. Esse sistema visa

coordenar e harmonizar as ações regulatórias, garantindo a eficácia da fiscalização e a compatibilidade entre normas gerais e específicas, inclusive no setor da saúde (BRASIL, 2025, p. 23).

Antes do PL 2.338/2023, o Projeto de Lei nº 21/2020, de autoria do deputado Eduardo Bismarck (PDT-CE), foi apresentado em fevereiro de 2020 na Câmara dos Deputados. Esse PL estabelecia fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil, destacando valores como liberdade, igualdade, pluralismo e direitos digitais (BRASIL, Câmara, 2020). O projeto chegou a ser aprovado pela Câmara em setembro de 2021, mas não avançou no Senado, sendo arquivado em dezembro de 2024 (BRASIL, 2024). Apesar de não ter apresentado um modelo regulatório operacionalizável, o PL 21/2020 foi fundamental para pautar o debate nacional sobre IA, influenciando a elaboração posterior do PL 2.338/2023 (BRASIL, Câmara, 2020; BRASIL, Senado, 2024).

Além dos projetos de lei, o Brasil dispõe de marcos legais já vigentes que dialogam diretamente com a temática, como o Marco Civil da Internet (Lei nº 12.965/2014), que estabelece princípios fundamentais para o uso da internet, como liberdade de expressão, privacidade, proteção de dados pessoais e neutralidade de rede (BRASIL, 2014); a Lei Geral de Proteção de Dados (Lei nº 13.709/2018), que dispõe sobre o tratamento de dados pessoais, definindo princípios como finalidade, minimização, segurança, transparência e direitos dos titulares (BRASIL, 2018); e a Estratégia Brasileira de Inteligência Artificial (EBIA), instituída em abril de 2021 pelo Ministério da Ciência, Tecnologia e Inovações, que estabelece diretrizes para o desenvolvimento ético, seguro e confiável da IA no Brasil (BRASIL, 2021). Esses instrumentos, ainda que não sejam específicos para a regulação da IA na saúde, fornecem elementos jurídicos e políticos fundamentais para a construção de um modelo normativo nacional voltado à proteção dos direitos fundamentais.

### 5.3.1 Inteligência Artificial na Proposta de Reforma do Código Civil.

Recentemente, o anteprojeto de Reforma do Código Civil brasileiro, formulado por comissão de juristas presidida pelo ministro Luís Felipe Salomão e apresentado pelo senador Rodrigo Pacheco, foi protocolado no Senado em 31 de janeiro de 2025 como Projeto de Lei nº 4/2025 (BRASIL, 2025). Pela primeira vez, inclui um Livro VI – Direito Civil Digital, com dispositivos específicos sobre patrimônio digital, assinatura eletrônica, responsabilidade por danos online, proteção de dados e inteligência artificial (BRASIL, 2025, p. 202).

No tocante à responsabilidade civil, ele introduz sanções específicas e impõe novos critérios operacionais a plataformas digitais e sistemas de IA, incluindo obrigações de diligência, moderação de conteúdo, avaliação periódica de riscos, canais de reparação e transparência nos termos de uso (BRASIL, 2025, p. 208-210).

Ainda que inova ao reconhecer formalmente a importância da explicabilidade e da identificação robusta dos sistemas de IA, a proposta permanece com escopo geral, sem estabelecer critérios específicos para avaliação clínica em saúde, nem detalhamento sobre governança algorítmica, gestão de risco, logs ou responsabilidade solidária, elementos fundamentais à luz das diretrizes da OMS ou do AI Act.

Assim, embora represente avanço no reconhecimento normativo da IA no sistema jurídico, o anteprojeto ainda não preenche totalmente as lacunas de governança clínica e segurança jurídica essenciais para a regulação setorial da saúde digital.

### 5.3.2 Plano Brasileiro de Inteligência Artificial (PBIA)

Em julho de 2024, durante a 5ª Conferência Nacional de Ciência, Tecnologia e Inovação (5ª CNCTI), realizada nos dias 30, 31 de julho e 1º de agosto, foi lançado o Plano Brasileiro de Inteligência Artificial (PBIA) 2024–

2028. A entrega oficial ocorreu ao Presidente da República Luiz Inácio Lula da Silva, em cerimônia promovida pelo MCTI, com apoio institucional do CGEE (BRASIL, 2024). O plano é considerado um marco estratégico, pois prevê investimentos de R$ 23 bilhões ao longo de quatro anos, com o objetivo de tornar o Brasil uma referência global em inteligência artificial (BRASIL, 2024).

O PBIA também enfatiza a importância da soberania digital, incluindo o desenvolvimento de modelos de linguagem em português, como o “SoberanIA”, um LLM totalmente nacional, desenvolvido pelo Piauí em parceria com o MCTI, e a consolidação da Nuvem Soberana, uma infraestrutura de dados estratégicos a ser operada por Serpro e Dataprev com investimento de R$ 1 bilhão (BRASIL, Agência GOV, 2025).

Todavia, apesar de sua abrangência orçamentária e institucional, o PBIA não oferece detalhamento específico para IA na saúde, como critérios clínico-regulatórios, mecanismos de supervisão humana ou requisitos de explicabilidade em diagnósticos automatizados.

### 5.3.3 Observatório Brasileiro de Inteligência Artificial (OBIA)

Em consonância com o Plano Brasileiro de Inteligência Artificial (PBIA), foi criado em setembro de 2024 o Observatório Brasileiro de Inteligência Artificial (OBIA), que integra o Eixo 5 do PBIA. O observatório está sob coordenação do NIC.br/CGI.br, recebendo apoio institucional do CGI.br, e tem como missão monitorar, compilar e prover informações confiáveis para orientar políticas e diretrizes relacionadas à IA (NIC.br, 2024).

O Observatório Brasileiro de Inteligência Artificial (OBIA) oferece dados de adoção da IA no Brasil, produz relatórios públicos e realiza monitoramento contínuo, atuando como elemento estratégico para validar políticas regulatórias, avaliar riscos socioeconômicos e acompanhar impactos da IA na saúde, um papel fundamental para a governança clínica responsável (NIC.br, 2024). No entanto, ainda que o OBIA seja um mecanismo promissor para a coleta de evidências, ele atua como instrumento informativo e não como órgão decisório ou regulador, incumbência que continua nas mãos do Estado e de

agências setoriais. Por isso, reforça-se, uma vez mais, a importância de complementar esta estrutura informacional com regulação normativa específica, responsabilização jurídica e supervisão técnica contínua, conforme proposto nesta pesquisa.

## 5.4 Síntese reflexiva.

O Capítulo 5 apresentou uma análise crítica sobre os rumos regulatórios da inteligência artificial (IA), discutindo a transição entre diretrizes principiológicas e instrumentos normativos vinculantes. Foi evidenciado que, diante do avanço acelerado das tecnologias emergentes, os sistemas jurídicos precisam desenvolver mecanismos de regulação que conciliem inovação com segurança, especialmente por meio de estruturas flexíveis como as sandboxes regulatórias e os modelos de testagem controlada.

Destacou-se também que a consolidação de normas específicas para a IA, como exemplificado pelo AI Act europeu e pelas propostas brasileiras em discussão, exige um equilíbrio contínuo entre proporcionalidade regulatória e proteção de direitos fundamentais, sobretudo em aplicações de alto impacto, como a saúde. A crescente importância das estruturas normativas experimentais demonstra que a regulação não pode se furtar à adaptabilidade, devendo ser responsiva, iterativa e tecnicamente embasada.

A inserção da realidade brasileira no debate normativo ampliou a compreensão da complexidade nacional na implementação de modelos regulatórios para IA. O contexto jurídico e institucional brasileiro apresenta desafios particulares, como assimetrias tecnológicas, desigualdades regionais e a necessidade de compatibilização entre o desenvolvimento da IA e os marcos constitucionais de proteção à dignidade humana.

Nesse sentido, iniciativas como a Estratégia Brasileira de IA, o Observatório Brasileiro de IA (OBIA) e o Projeto de Lei 2338/2023 apontam para um cenário promissor, embora ainda em consolidação normativa.

Conclui-se, portanto, que o futuro da regulação da inteligência artificial no setor da saúde, tanto em nível global quanto nacional, dependerá da capacidade de articular princípios ético-jurídicos com ferramentas regulatórias eficazes. Essa articulação é essencial para assegurar uma governança algorítmica que promova inovação responsável, equidade no acesso e proteção efetiva dos direitos fundamentais.

**Quadro 5** - Estrutura conceitual das sandboxes regulatórias aplicadas à inteligência artificial.

| ELEMENTO | REFERÊNCIA |
|---|---|
| Definição | Ambiente regulado de experimentação para testar tecnologias inovadoras com supervisão estatal. |
| Objetivo | Conciliar inovação tecnológica com proteção jurídica e institucional. |
| Formato | Espaço controlado e temporário, com regras flexibilizadas e avaliação contínua. |
| Supervisão | Monitoramento por órgãos reguladores para garantir segurança e legalidade. |
| Participantes | Startups, empresas, agências reguladoras, instituições públicas e operadores do direito. |
| Benefício regulatório | Estrutura normativa flexível e responsiva às rápidas transformações tecnológicas. |
| Exemplo internacional | AI Act da União Europeia prevê sandboxes para testar IA de alto risco antes da liberação ampla. |
| Aplicação na saúde | Teste de sistemas de IA diagnóstica, suporte clínico e gestão, sem comprometer a segurança do paciente. |
| Flexibilidade normativa | Adaptação de exigências legais para fomentar aprendizado regulatório e inovação segura. |
| Finalidade estratégica | Aprimorar marcos regulatórios por meio da prática, ajustando normas à realidade tecnológica. |

**Fonte**: Elaborado pelo autor

Por outro lado, a construção de marcos regulatórios nacionais e supranacionais evidencia uma tendência à institucionalização da inteligência artificial como objeto de direito, o que demanda a definição de limites, deveres e garantias no uso dessas ferramentas. Essa nova fase regulatória deverá ser marcada pela busca de equilíbrio entre os potenciais da IA e os valores fundantes do Estado Democrático de Direito, com ênfase na proteção da dignidade humana, na equidade e na responsabilização efetiva dos agentes envolvidos.

Com base nessa trajetória, o tópico seguinte apresentará as considerações finais da presente dissertação, com o intuito de sintetizar os principais achados da pesquisa, destacar suas contribuições teóricas e normativas, e apontar perspectivas futuras para o desenvolvimento de uma

inteligência artificial alinhada aos referenciais da saúde, justiça social e direitos fundamentais.

## Considerações finais

A presente dissertação se propôs a investigar, sob uma perspectiva normativa e interdisciplinar, a superestrutura principiológica que deve orientar a aplicação da inteligência artificial no campo da saúde, com especial atenção às diretrizes propostas pela Organização Mundial da Saúde. A partir dessa delimitação temática, buscou-se compreender não apenas os fundamentos técnicos e conceituais da inteligência artificial, mas sobretudo os desafios éticos, jurídicos e sociais implicados em sua crescente inserção da saúde.

No primeiro momento, foi possível demonstrar que a inteligência artificial constitui um campo plural, composto por abordagens que vão desde estruturas simbólicas baseadas em regras até arquiteturas conexionistas inspiradas no funcionamento das redes neurais humanas. A distinção entre IA estreita e IA geral, bem como entre os paradigmas supervisionados, não supervisionados, semi-supervisionados e por reforço, evidencia o grau de sofisticação e diversidade das tecnologias em uso, tornando indispensável a formulação de critérios regulatórios que considerem tal complexidade.

Em seguida, foram analisadas as implicações da inteligência artificial na saúde, identificando-se transformações relevantes nas formas de diagnóstico, prognóstico, tratamento e gestão assistencial. A aplicação de algoritmos preditivos, redes neurais profundas e modelos fundacionais têm impactado significativamente as rotinas clínicas, exigindo uma governança normativa que assegure a compatibilidade entre inovação tecnológica e a centralidade da dignidade humana.

No tocante à proposta de uma superestrutura principiológica, observou-se que, embora existam iniciativas normativas internacionais consistentes — como o AI Act da União Europeia, os princípios da OCDE, as diretrizes éticas da OMS e da Comissão Europeia, nenhuma delas, de modo isolado, abarca de forma plena a totalidade dos elementos estruturantes aqui delineados. A superestrutura proposta se diferencia por articular princípios éticos fundamentais (como autonomia, explicabilidade e justiça) com critérios normativos específicos voltados à realidade da saúde e, ainda, com

instrumentos de viabilidade regulatória que consideram as assimetrias institucionais de países em desenvolvimento, como o Brasil.

Sua aplicação prática deve ocorrer por meio de três dimensões: (i) a incorporação explícita dos princípios fundamentais nos marcos legais e regulatórios nacionais e internacionais; (ii) a institucionalização de mecanismos de supervisão e responsabilidade contínua, como auditorias algorítmicas e relatórios de impacto ético; e (iii) a criação de ambientes regulatórios experimentais, como sandboxes supervisionadas, que possibilitem o desenvolvimento e o teste de sistemas de IA em condições controladas, assegurando sua conformidade ética e sua responsividade prática. Além disso, recomenda-se que a superestrutura seja incorporada como diretriz obrigatória em políticas públicas de saúde, servindo como referencial para aquisições públicas, certificações técnicas e programas de capacitação profissional.

A dissertação também demonstrou que o desafio da regulação da IA na saúde não se limita à adoção de diretrizes abstratas, mas exige a construção de modelos normativos operacionais, capazes de incidir sobre realidades práticas e complexas.

A emergência de estruturas de testagem jurídica, como as *sandboxes* regulatórias, representa uma inovação importante nesse campo, permitindo que tecnologias emergentes sejam avaliadas em ambientes supervisionados, de forma a promover o aprendizado regulatório e evitar a produção de danos irreversíveis. A experiência de diferentes jurisdições, como União Europeia, Estados Unidos e Brasil, revela que há um movimento global em curso rumo à formalização de marcos legais específicos para a inteligência artificial, especialmente em setores sensíveis como a saúde.

O estudo realizado permite concluir que o uso da inteligência artificial na saúde deve ser guiado por um novo modelo de governança tecnológica, fundado na articulação entre princípios éticos, normas jurídicas e mecanismos de controle institucional. Essa governança deve ser orientada não apenas pela mitigação de riscos, mas também pela promoção ativa do bem-estar coletivo, da equidade no acesso aos cuidados e da preservação da confiança nas instituições de saúde.

A pesquisa também contribuiu para o debate acadêmico ao propor uma leitura integradora entre os elementos técnicos da IA e os fundamentos

normativos da proteção à saúde, demonstrando que é possível, e necessário, construir pontes entre os campos da ciência da computação, da bioética, do direito e da saúde pública. O reconhecimento da inteligência artificial como tema jurídico de alta densidade normativa exige, portanto, o envolvimento de múltiplos atores institucionais e sociais na construção de soluções legítimas, transparentes e efetivamente democráticas.

Como perspectivas futuras, destaca-se a necessidade de aprofundar o debate sobre a soberania dos dados em saúde, o fortalecimento da supervisão humana sobre decisões automatizadas e a prática concreta de princípios de justiça algorítmica. Além disso, o contínuo monitoramento dos impactos da IA sobre os sistemas de saúde, sobretudo em contextos de vulnerabilização social, será fundamental para garantir que o avanço tecnológico se traduza em efetiva melhoria das condições de vida e respeito aos direitos fundamentais.

Em termos práticos a superestrutura proposta a ser observada é a seguinte:

**Quadro 6** – A Superestrutura principiológico-normativa proposta para a IA na saúde

| EIXO | COMPONENTE | DEFINIÇÃO |
|---|---|---|
| Cuidados técnicos | Definição do desafio clínico | Delimitar claramente o problema e o impacto da IA na prática assistencial |
| | Qualidade dos dados | Coleta, curadoria e tratamento adequados dos dados para assegurar validade |
| | Avaliação de desempenho | Métricas apropriadas e testes independentes são fundamentais na validação dos modelo |
| | Transparência técnica | Reduzir a opacidade algorítmica e tornar o sistema compreensível para o usuário |
| | Atualização contínua | Garantir supervisão e melhorias constantes dos modelos em uso clínico |
| Diretrizes éticas (OMS) | Proteger a autonomia humana | Assegurar a decisão informada do paciente, mesmo com suporte da IA |
| | Promover o bem-estar, a segurança humana e o interesse público | Garantir a Beneficência e evitar a maleficência |
| | Garantia de transparência, explicabilidade e inteligibilidade. | Tornar os processos decisórios da IA compreensíveis, explicáveis e com transparências algorítmica. |
| | Promover responsabilidade e prestação | Responsabilizar agentes |

| | | |
|---|---|---|
| | de contas | por falhas ou decisões danosas |
| | Garantir inclusão e equidade | Garantir acesso inclusivo e equitativo às tecnologias em saúde |
| | Promover uma IA que seja responsiva e sustentável | Uso ético da IA que preserve os recursos e a integridade institucional |
| Governança de dados | Proteção de dados sensíveis | Garantia da privacidade e controle do paciente sobre seus dados |
| | LGPD e regulamentações internacionais | Aplicação das normas nacionais e alinhamento com o AI Act e HIPAA |
| | Pseudonimização e desidentificação | Técnicas que reduzem riscos de reidentificação e violação de privacidade |
| Diretrizes para dispositivos médicos | Avaliação proporcional ao risco | Classificação dos dispositivos conforme impacto clínico e exigência regulatória |
| | Boas práticas de engenharia | Qualidade de software, segurança cibernética e rastreabilidade são essenciais |
| | Inclusão de dados representativos | Dados de treinamento devem refletir a diversidade da população atendida |
| Diretrizes para boas práticas algorítmicas | Explicabilidade como direito | O paciente tem o direito de entender o raciocínio algorítmico por trás da decisão |
| | Registros técnicos | Manutenção documental detalhada dos processos e parâmetros do sistema |
| | Supervisão humana | Reforço da supervisão profissional sobre as decisões automatizadas |
| Diretrizes para Accountability (responsabilidade e prestação de contas) | Responsabilidade civil | Marco jurídico para imputar deveres e reparações por danos oriundos da IA |
| | Identificação dos agentes responsáveis | Definição clara de quem responde pelo uso do sistema em saúde |
| | Mecanismos de imputação | Procedimentos para atribuir responsabilidade objetiva ou subjetiva |
| Sandboxes regulatórias | Ambiente de testes supervisionado | Espaço controlado para desenvolvimento e monitoramento de tecnologias inovadoras |
| | Flexibilização normativa | Aplicação de regras temporariamente ajustadas sob supervisão estatal |
| | Finalidade normativa | Promover aprendizado regulatório e proteger os usuários em fase experimental |

**Fonte**: Elaborado pelo autor.

Conclui-se, assim, que a construção de um marco regulatório robusto da inteligência artificial na saúde depende da adoção de fundamentos ético-jurídicos sólidos, da articulação com instrumentos regulatórios adaptativos e da concretização de mecanismos institucionais que assegurem o respeito à autonomia dos pacientes e à justiça nos sistemas de cuidado. A proposta apresentada nesta dissertação visa oferecer não apenas um diagnóstico crítico, mas também uma contribuição propositiva à governança da IA em saúde, promovendo um modelo de desenvolvimento tecnológico alinhado à dignidade humana, à equidade e à sustentabilidade sistêmica.

## Referências

ACOSTA, Julián N. et al. IA biomédica multimodal. Nature Medicine, v. 28, p. 1773-1784, 2022. Disponível em: https://www.nature.com/articles/s41591-022-01981-2. Acesso em: 05 de abril de 2025.

AGÊNCIA EUROPEIA DE MEDICAMENTOS (EMA). EMA estabelece procedimento regular para aconselhamento científico sobre certos dispositivos médicos de alto risco. EMA, 10 de fevereiro de 2025. Disponível em: https://www.ema.europa.eu/en/news/ema-establishes-regular-procedure-scientific-advice-certain-high-risk-medical-devices. Acesso em: 11 de abril de 2025.

AGÊNCIA GOV. MCTI e governo do Piauí firmam parceria para avançar em projetos de IA. Brasília, 21 maio 2025. Disponível em: https://agenciagov.ebc.com.br/noticias/202505/mcti-e-piaui-firmam-parceria-para-projetos-em-ia. Acesso em: 21 de maio de 2025.

AHMED, Tanbir; AZIZ, Md Momin Al; MOHAMMED, Noman. Desidentificação de prontuários eletrônicos usando redes neurais. Scientific Reports, v. 10, n. 18600, 2020. DOI: 10.1038/s41598-020-75544-1. Disponível em: https://www.nature.com/articles/s41598-020-75544-1. Acesso em: 09 de abril de 2025.

ALAMI, Hassane et al. inteligência artificial na saúde: construindo as bases para uma inovação responsável, sustentável e inclusiva nos países de baixa e média renda. Globalization and Health, v. 16, n. 52, 2020. DOI: https://doi.org/10.1186/s12992-020-00584-1. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7315549/pdf/12992_2020_Article_584.pdf Acesso em: 8 de abril de 2025.

ALLAIN, Jessica S. De Jeopardy à Icterícia: implicações da responsabilidade médica do Dr. Watson e outros sistemas de inteligência artificial. Louisiana Law Review, v. 73, n. 4, p. 1051-1077, 2013. Disponível em: https://digitalcommons.law.lsu.edu/lalrev/vol73/iss4/7/. Acesso em: 12 de abril de 2025.

ALVAREZ-ROMERO, Celia et al. Centros de dados em saúde: uma análise das características de governança existentes. Health Research Policy and Systems, v. 21, n. 70, 2023. DOI: 10.1186/s12961-023-01026-1. Disponível em: https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-023-01026-1. Acesso em: 09 de abril de 2025.

AMANN, Julia et al. Explicabilidade para inteligência artificial na saúde: uma perspectiva multidisciplinar. BMC Medical Informatics and Decision Making, v. 20, n. 310, 2020. Disponível em: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6. Acesso em: 11 de abril de 2025.

AMANN, Julia et al. Explicar ou não explicar? Explicabilidade da inteligência artificial em sistemas de apoio à decisão clínica. PLOS Digital Health, v. 1, n. 2, e0000016, 2022. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC9931364/pdf/pdig.0000016.pdf. Acesso em: 10 de abril de 2025.

ARONSON, Jeffrey K.; HENEGHAN, Carl; FERNER, Robin E. Dispositivos médicos: definição, classificação e implicações regulatórias. Drug Safety, v. 43, p. 83-93, 2020. DOI: 10.1007/s40264-019-00878-3. Disponível em: https://link.springer.com/article/10.1007/s40264-019-00878-3. Acesso em: 10 abril de 2025.

BABIC, Boris et al. Cuidado com as explicações da IA na saúde: os benefícios da IA explicável não são o que parecem. Science, 2021. Pre-print. Disponível em: https://borisbabic.com/research/beware_march2023.pdf. Acesso em: 7 de abril de 2025.

BABIC, Boris et al. Algoritmos em bloqueio regulatório na medicina. Science Manuscript (versão pré-publicação), 2019. Disponível em: https://philsci-archive.pitt.edu/18796/1/science_manuscript_revised.docx.pdf. Acesso em: 09 de abril de 2025.

BABIC, Boris; COHEN, I. Glenn. A isca e troca da explicabilidade algorítmica. Minnesota Law Review, v. 108, p. 857-884, 2023. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4541487. Acesso em: 12 de abril de 2025.

BALAGOPALAN, Aparna et al. A estrada para a explicabilidade é pavimentada com viés: medindo a equidade das explicações. Massachusetts Institute of Technology, 2022. Disponível em: https://arxiv.org/pdf/2205.03295. Acesso em: 12 de abril de 2025.

BANCO MUNDIAL. Experiências globais de sandboxes regulatórios. Nota Fintech n.º 8. Finance, Competitiveness & Innovation Global Practice. Washington, Capital: Grupo Banco Mundial, 2020. Disponível em: https://documents1.worldbank.org/curated/en/912001605241080935/pdf/Global-Experiences-from-Regulatory-Sandboxes.pdf. Acesso em: 12 de abril de 2025.

BATES, David W. Como regular os algoritmos de integridade da IA em evolução. Nature Medicine, [S.l.], v. 29, n. 1, p. 26, jan. 2023. DOI: 10.1038/s41591-022-02165-8. Disponível em: https://www.nature.com/articles/s41591-022-02165-8. Acesso em: 7 de abril de 2025.

BATES, David W. et al. Big Data na área da saúde: usando analytics para identificar e gerenciar pacientes de alto risco e alto custo. Health Affairs, v. 33, n. 7, p. 1123-1131, 2014. Disponível em: https://sci-hub.se/10.1377/hlthaff.2014.0041. Acesso em: 05 de abril de
2025.

BEAM, Andrew L.; KOHANE, Isaac S. Big Data e aprendizado de máquina na área da saúde. JAMA, [online], published online March 12, 2018. p. E1-E2. Disponível em: https://sci-hub.se/10.1001/jama.2017.18391. Acesso em: 06 de abril de 2025.

BEAUCHAMP, Tom L. O modelo dos quatro princípios na ética da atenção à saúde. In: BEAUCHAMP, Tom L. Standing on Principles: Collected Essays. New York: Oxford University Press, 2010. p. 30-49. Disponível em: https://bioethics.pitt.edu/sites/default/files/Stott%2C%20Chapter%203%20Beauchamp.pdf. Acesso em: 7 de abril de 2025.

BERNIER, Alexander et al. O panorama internacional da governança de dados. Journal of Law and the Biosciences, v. 9, p. 1-45, 2022. Disponível em: https://pubmed.ncbi.nlm.nih.gov/35382430/. Acesso em: 8 de abril de 2025.

BISHOP, Christopher M. Reconhecimento de padrões e aprendizado de máquina. New York: Springer, 2006. Disponível em: http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf Acesso em: 05 de abril de 2025.

BOBBIO, Norberto. Norberto Bobbio: Uma Era de Direitos sem Fundações. In: BACCELLI, Luca. Iris, Firenze University Press, v. II, n. 3, p. 405-426, out. 2010. Disponível em: https://www.academia.edu/5903821/Norberto_Bobbio_An_Age_of_Rights_without_Foundations_1_L_B?auto=download. Acesso em: 6 abr. 2025.

BÖCKENFÖRDE, Ernst-Wolfgang. Direitos Fundamentais: Teoria e Interpretação. In: KÜNKLER, Mirjam; STEIN, Tine (ed.). Teoria Constitucional e Política: Escritos Selecionados. Oxford: Oxford University Press, 2017. p. 266-289. Disponível em: https://api.pageplace.de/preview/DT0400.9780191024573_A30389640/preview-9780191024573_A30389640.pdf. Acesso em: 06 de abril de 2025.

BOHR, Adam; MEMARZADEH, Kaveh. A ascensão da inteligência artificial em aplicações de saúde. In: inteligência artificial na Saúde. 1. ed. Elsevier, 2020. p. 25-56. Disponível em: https://www.sciencedirect.com/science/article/pii/B9780128184387000022. Acesso em: 05 de abril de 2025.

BOTELHO, Marcos César; CAMARGO, Elimei Paleari do Amaral. A aplicação da Lei Geral de Proteção de Dados na saúde. Revista de Direito Sanitário, São Paulo, v. 21, e-0021, 2021. Disponível em: https://doi.org/10.11606/issn.2316-9044.rdisan.2021.168023. Acesso em: 07 de abril de 2025.

BRASIL. Câmara dos Deputados. Câmara começa a discutir projeto que regulamenta o uso da inteligência artificial no Brasil. Portal da Câmara dos Deputados, Brasília, 2023. Disponível em: https://www.camara.leg.br/noticias/1140392-camara-comeca-a-discutir-projeto-que-regulamenta-a-inteligencia-artificial-no-brasil/. Acesso em: 09 de abril de 2025.

BRASIL. Câmara dos Deputados. Projeto de Lei nº 21, de 2020: estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil e dá outras providências. Apresentado pelo Deputado Eduardo Bismarck. Brasília, DF, 4 fev. 2020. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=1853928&filename=PL%2021/2020. Acesso em: 7 de abril de 2025.

BRASIL. Câmara dos Deputados. Projeto de Lei nº 2.338, de 2023: dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, DF: Câmara dos Deputados, 2025. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2868197&filename=PL%202338/2023. Acesso em: 7 de abril de 2025.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Diário Oficial da União: seção 1, Brasília, DF, 24 abr. 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 5 de abril de 2025.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Lei Geral de Proteção de Dados Pessoais - LGPD). Diário Oficial da União: seção 1, Brasília, DF, 15 ago. 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 07 de abril de 2025.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. Estratégia Brasileira de Inteligência Artificial – EBIA. Brasília: MCTI, 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf. Acesso em: 04 de abril de 2025.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. Plano Brasileiro de Inteligência Artificial (PBIA) 2024–2028: Tecnologia e Inovação. Brasília: MCTI/CGEE, jul. 2024. Disponível em: https://www.gov.br/lncc/pt-br/assuntos/noticias/ultimas-noticias-1/plano-brasileiro-de-inteligencia-artificial-pbia-2024-2028. Acesso em: 01 de maio de 2025.

BRASIL. Senado Federal. Projeto de Lei nº 4, de 2025: dispões sobre a atualização da Lei nº 10.406/2002 (Código Civil) e da legislação correlata. Brasília, DF, 31 jan. 2025. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/166998. Acesso em: 05 de abril de 2025.

BRASIL. Senado Federal. Senado aprova regulamentação da inteligência artificial; texto vai à Câmara. Brasília, 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: 07 abril de 2025.

BROWNSTEIN, John S. et al. Avanços na inteligência artificial para vigilância de doenças infecciosas. New England Journal of Medicine, v. 388, n. 17, p. 1597-1607, 2023. Disponível em: https://www.binasss.sa.cr/bibliotecas/bhm/abr23/23.pdf. Acesso em: 05 de abril de 2025.

BUOCZ, Thomas; PFOTENHAUER, Sebastian; EISENBERGER, Iris. As sandboxes regulatórias no AI Act: conciliando inovação e segurança? Law, Innovation and Technology, v. 15, n. 2, p. 357-389, 2023. Disponível em: https://www.tandfonline.com/doi/epdf/10.1080/17579961.2023.2245678?needAccess=true. Acesso em: 12 abr. 2025.

BURRELL, Jenna. Como a máquina 'pensa': Entendendo a opacidade em algoritmos de aprendizado de máquina. Big Data & Society, v. 3, n. 1, p. 1-12, 2016. Disponível em: https://moodle2.units.it/pluginfile.php/560396/mod_resource/content/1/Bussani_Sebok_Infantino_Product_Liability.pdf. Acesso em 06 de abril de 2025.

BUSSANI, Mauro; SEBOK, Anthony J.; INFANTINO, Marta. Perspectivas de Direito Comum e Direito Civil sobre o Direito de Responsabilidade Civil. New York: Oxford University Press, 2022.

BYGRAVE, Lee A. Artigo 22 - Tomada de decisão individual automatizada, incluindo a definição de perfis. In: KUNER, Christopher; BYGRAVE, Lee A.; DOCKSEY, Christopher (org.). O Regulamento Geral de Proteção de Dados da UE (RGPD). Oxford: Oxford University Press, 2020. Disponível em:

https://academic.oup.com/book/41324/chapter/352298561?login=true. Acesso em: 12 de abril de 2025.

CEBRIAN, Fabiana Faraco; FREITAS, Cinthia Obladen de Almendra. Possibilidades e desafios da implementação da inteligência artificial generativa na transparência algorítmica. In: PEIXOTO, Fabiano Hartmann; PEREIRA, João Sergio dos Santos Soares; CALADO, Vinicius de Negreiros (Orgs.). Dr.IA – inteligência artificial generativa, modelos de linguagem e argumentação jurídica [recurso eletrônico on-line]. Belo Horizonte: Skema Business School, 2023. v. 1. 45p . Disponível em: https://site.conpedi.org.br/publicacoes/s5y6p2k5/q6eh7k97/FyK49r7HGGJ918n3.pdf Acesso em: 07 de junho de 2025.

CALADO, Vinícius de Negreiros; LAMY, Marcelo. Teleconsulta médica: os limites éticos e o risco de negligência informacional. Cadernos Ibero-Americanos de Direito Sanitário. 2020 jul./set.; 9(3): 89-122. Disponível em: http://dx.doi.org/10.17566/ciads.v9i3.713. Acesso em: 07 de junho de 2025.

CALLAWAY, Ewen. Todo o universo das proteínas. Nature, v. 608, p. 15-16, 2022. Disponível em: https://www.nature.com/articles/d41586-022-02083-2. Acesso em: 05 de abril de 2025.

CHAPELLE, Olivier; SCHÖLKOPF, Bernhard; ZIEN, Alexander (Orgs.). Aprendizado semi-supervisionado. Cambridge, Massachusetts: MIT Press, 2010.

CHAR, Danton S.; ABRÀMOFF, Michael D.; FEUDTNER, Chris. Identificando considerações éticas para aplicações de Machine Learning na saúde. *The American Journal of Bioethics*, v. 20, n. 11, p. 7-17, 2020. DOI: 10.1080/15265161.2020.1819469. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7737650/pdf/nihms-1648280.pdf. Acesso em: 7 abr. 2025.

CHATTERJEE, Sheshadri. A privacidade de dados é um direito fundamental na Índia? Uma análise e recomendações sob a perspectiva jurídica e de políticas públicas. International Journal of Law and Management, v. 61, n. 1, p. 170-190, 2019. DOI: 10.1108/IJLMA-01-2018-0013. Disponível em: https://www-emerald-com.ez67.periodicos.capes.gov.br/insight/content/doi/10.1108/ijlma-01-2018-0013/full/pdf?title=is-data-privacy-a-fundamental-right-in-india-an-analysis-and-recommendations-from-policy-and-legal-perspective. Acesso em: 09 de abril de 2025.

CHEN, Hongming et al. A ascensão do aprendizado profundo na descoberta de medicamentos. Drug Discovery Today, v. 23, n. 6, p. 1241-1250, 2018. Disponível em: https://www.sciencedirect.com/science/article/pii/S1359644617303598?via%3Dihub. Acesso em: 05 de abril de 2025.

CHEN, Jonathan H. et al. Relevância decadente dos dados clínicos para decisões futuras em conjuntos de pedidos clínicos de pacientes internados orientados por dados. International Journal of Medical Informatics, v. 102, p. 71-79, 2017. Disponível em: https://www.sciencedirect.com/science/article/pii/S138650561730059X?via%3Dihub. Acesso em: 06 de abril de 2025.

CHEN, Po-Hsuan Cameron; LIU, Yun; PENG, Lily. Como desenvolver modelos de aprendizado de máquina para a área da saúde. Nature Materials, v. 18, p. 410-414, 2019. Disponível em: https://www.nature.com/articles/s41563-019-0345-0. Acesso em: 06 de abril de 2025.

CHEVRIER, Raphaël et al. Uso e compreensão de anonimização e desidentificação na literatura biomédica: revisão de escopo. Journal of Medical Internet Research, v. 21, n. 5, e13484, 2019. DOI: 10.2196/13484. Disponível em: https://www.jmir.org/2019/5/e13484/PDF. Acesso em: 09 abril de 2025.

CHIKWETU, Lucy et al. A desidentificação de dados de dispositivos vestíveis nos dá uma falsa sensação de segurança? Uma revisão sistemática. The Lancet Digital Health, v. 5, n. 4, p. e239-e247, abr. 2023. DOI: 10.1016/S2589-7500(22)00234-5. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC10040444/pdf/nihms-1883636.pdf. Acesso em: 09 de abril de 2025.

COHEN, I. Glenn; MELLO, Michelle M. A HIPAA e a proteção de informações de saúde no século XXI. JAMA - Journal of the American Medical Association, [s.l.], publicado online em 24 maio 2018. p. 2. DOI: https://doi.org/10.1001/jama.2018.5630. Disponível em: https://law.stanford.edu/wp-content/uploads/2018/05/jama_Cohen_Mello_2018_vp_180053.pdf. Acesso em: 8 de abril de 2025.

COHEN, I. Glenn; EVGENIOU, Theodoros; GERKE, Sara; MINSSEN, Timo. A estratégia europeia de inteligência artificial: implicações e desafios para a saúde digital. The Lancet Digital Health, v. 2, p. e376-e379, jul. 2020. Disponível em: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(20)30112-6/fulltext. Acesso em: 12 de abril de 2025.

COLLINS, Gary S. et al. Relatório transparente de um modelo de previsão multivariável para prognóstico ou diagnóstico individual (TRIPOD): a Declaração TRIPOD. BMC Medicine, v. 13, n. 1, p. 1-10, 2015. DOI: 10.1186/s12916-014-0241-z. Disponível em: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-014-0241-z. Acesso em: 12 de abril de 2025.

COMISSÃO EUROPEIA. Proposta de Regulamento do Parlamento Europeu e do Conselho relativo ao Espaço Europeu de Dados de Saúde. Estrasburgo: Comissão Europeia, 3 maio 2022. COM(2022) 197 final. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52022PC0197. Acesso em: 09 de abril de 2025.

COMITÊ DE DIREITOS ECONÔMICOS, SOCIAIS E CULTURAIS (CESCR). Comentário Geral nº 14: O direito ao mais alto padrão possível de saúde (artigo 12 do PIDESC). Adotado na 22ª Sessão, em 11 de agosto de 2000. Documento E/C.12/2000/4. Disponível em: https://www.ohchr.org/sites/default/files/Documents/Issues/Women/WRGS/Health/GC14.pdf. Acesso em: 06 de abril de 2025.

CUATRECASAS. Impacto do Regulamento da IA nos dispositivos médicos. Cuatrecasas, 2024. Disponível em: https://www.cuatrecasas.com/pt/portugal/farmaceutico-saude/art/impacto-do-regulamento-da-ia-nos-dispositivos-medicos?utm_source=chatgpt.com. Acesso em: 11 abr. 2025.

DAVIS, Sharon E.; WALSH, Colin G.; MATHENY, Michael E. Perguntas abertas e lacunas de pesquisa para monitorar e atualizar ferramentas habilitadas para IA em ambientes clínicos. Frontiers in Digital Health, v. 4, p. 1-9, 2022. Disponível em: https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2022.958284/full. Acesso em: 6 de abril de 2025.

DOSHI-VELEZ, Finale et al. Responsabilidade da IA perante a lei: o papel da explicação. Harvard University: Berkman Klein Center Working Group, 2019. Disponível em: https://arxiv.org/pdf/1711.01134. Acesso em: 12 de abr de 2025.

DE LUCA, Stefano. Revisão da Diretiva de Responsabilidade por Produtos Defeituosos. Serviço de Pesquisa do Parlamento Europeu (EPRS), PE 739.341, fev. 2025. Disponível em: https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739341/EPRS_BRI(2023)739341_EN.pdf. Acesso em: 12 de abril de 2025.

DUBBER, Markus D.; PASQUALE, Frank; DAS, Sunit (org.). O Manual Oxford de Ética da IA. New York: Oxford University Press, 2020. ISBN 978-0-19-006739-7. Disponível em: https://dokumen.pub/qdownload/the-oxford-handbook-of-ethics-of-ai-9780190067397-019006739x.html. Acesso em: 7 de abril de 2025.

DUFFOURC, Mindy Nunez; GERKE, Sara. As propostas de diretivas da UE sobre responsabilidade por IA deixam lacunas preocupantes com impacto na IA médica. npj Digital Medicine, v. 6, art. 77, 2023. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC10133336/. Acesso em: 12 de abril de 2025.

EMANUEL, Ezekiel J.; WACHTER, Robert M. inteligência artificial na área da saúde: o valor corresponderá ao Exagero? JAMA, v. 321, n. 23, p. 2281-2282, 2019. Disponível em: https://sci-hub.se/downloads/2019-07-17/7b/10.1001@jama.2019.4914.pdf. Acesso em: 06 de abril de 2025.

ENGEL, David; MCCANN, Michael W. (org.). Linhas de falha: o direito da responsabilidade civil como prática cultural. Stanford: Stanford University Press, 2009. Disponívem em: https://go.gale.com/ps/i.do?p=AONE&u=anon%7Ec6e2f63c&id=GALE%7CA251860345&v=2.1&it=r&aty=open-web-entry. Acesso em: 12 de abril de 2025.

ESTADOS UNIDOS. FDA. Software de apoio à decisão clínica: guia para a indústria e para a equipe da Food and Drug Administration. Silver Spring, MD: Departamento de Saúde e Serviços Humanos dos Estados Unidos, 2022. Disponível em: https://www.fda.gov/media/109618/download. Acesso em: 7 de abril de 2025.

ESTADOS UNIDOS. Food and Drug Administration. Proposta de estrutura regulatória para modificações em software baseado em IA/ML como dispositivo médico. Washington, Capital, 2022. Disponível em: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf. Acesso em: 10 de abril de 2025.

ESTADOS UNIDOS. Casa Branca. Decreto Executivo sobre o desenvolvimento e uso seguro, protegido e confiável da inteligência artificial. Washington, Capital, 30 de outubro de 2023. Disponível em: https://www.acc.com/sites/default/files/2024-08/2023.10.30-Executive-Order-on-the-Safe--Secure--and-Trustworthy-Development-and-Use-of-Artificial-Intelligence-_-The-White-House.pdf. Acesso em: 12 de abril de 2025.

ESTEVA, Andre et al. Um guia para aprendizado profundo na área da saúde. Nature Medicine, v. 25, n. 1, p. 24-29, 2019. Disponível em: https://www.nature.com/articles/s41591-018-0316-z. Acesso em: 05 de abril de 2025.

EUA. Guia sobre privacidade e segurança das informações eletrônicas de saúde. Versão 2.0, abril de 2015. Washington, Capital: Escritório do Coordenador Nacional para Tecnologia da Informação em Saúde, 2015. p. 5-13. Disponível em: https://www.healthit.gov/sites/default/files/pdf/privacy/privacy-and-security-guide.pdf. Acesso em: 09 de abril de 2025.

FENG, Jiayun et al. Melhoria da qualidade da inteligência artificial clínica: rumo ao monitoramento contínuo e atualização de algoritmos de IA na área da saúde. npj Digital Medicine, v. 5, n. 66, p. 1-9, 2022. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC9156743/pdf/41746_2022_Article_611.pdf. Acesso em: 06 de abril de 2025.

FINLAYSON, Samuel G. et al. A mudança do clínico e do conjunto de dados na inteligência artificial. The New England Journal of Medicine, v. 385, n. 3, p. 283, 2021. Disponível em: https://www.nejm.org/doi/pdf/10.1056/NEJMc2104626. Acesso em: 06 de abril de 2025.

FIRDAUS, Muhammad. Uma revisão da Lei de Proteção de Dados Pessoais na Indonésia. Programa Interdisciplinar de Segurança da Informação, Pukyong National University, 2022. Disponível em: https://www.pkc.pknu.ac.kr. Acesso em: 09 de abril de 2025.

FJELLAND, Ragnar. Por que a inteligência artificial geral não será realizada. Humanities and Social Sciences Communications, [s.l.], v. 7, n. 10, p. 1-9, 2020. Springer Nature. Disponível em: https://doi.org/10.1057/s41599-020-0494-4. Acesso em: 5 abr. 2025.

FLORIDI, Luciano et al. Uma estrutura ética para uma boa sociedade de IA: oportunidades, riscos, princípios e recomendações. Minds and Machines, [S.l.], v. 28, p. 689-707, 2018. Disponível em: https://link.springer.com/article/10.1007/s11023-018-9482-5. DOI: 10.1007/s11023-018-9482-5. Acesso em: 7 de abril de 2025.

FOLLAND, Sherman; GOODMAN, Allen C.; STANO, Miron. A economia da saúde e cuidados de saúde. 7. ed. Boston: Prentice Hall, 2013. Cap. 20 Regulamentação Governamental: Principais Mecanismos Regulatórios, p. 407-432. Disponível em: https://dl.icdst.org/pdfs/files3/f19422deae07b90088024d54f312c5c2.pdf. Acesso em: 06 de abril de 2025.

FUTOMA, Joseph et al. O mito da generalização na pesquisa clínica e aprendizado de máquina na área da saúde. The Lancet Digital Health, v. 2, n. 9, p. e489-e492, 2020. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7444947/pdf/main.pdf. Acesso em: 06 de abril de 2025.

GENESERETH, Michael R.; NILSSON, Nils J. Fundamentos lógicos da inteligência artificial. San Mateo: Morgan Kaufmann, 2012. Disponível em: https://kryten.mm.rpi.edu/COURSES/LOGAIS02/nonmonotonic.pdf. Acesso em: 04 de abril de 2025.

GERKE, Sara; MINSSEN, Timo; COHEN, Glenn. Desafios éticos e legais da inteligência artificial na saúde. In: TOPOL, Eric. inteligência artificial na Saúde. Amsterdam: Elsevier, 2020. p. 295-322. DOI: 10.1016/B978-0-12-818438-7.00012-5. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7332220/pdf/main.pdf. Acesso em: 09 de abril de 2025.

GERKE, Sara et al. A necessidade de uma visão sistêmica para regular o software baseado em inteligência artificial/aprendizado de máquina como dispositivo médico. NPJ Digital Medicine, [S.l.], v. 3, n. 53, p. 1-5, 2020. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7138819/pdf/41746_2020_Article_262.pdf. Acesso em: 10 abr. 2025.

GERKE, Sara. "Rótulos de informações nutricionais" para dispositivos médicos baseados em inteligência artificial/aprendizado de máquina - a necessidade urgente de padrões de rotulagem. The George Washington Law Review, [S.l.], v. 91, n. 1, p. 79-161, 2023. Disponível em: https://www.gwlr.org/wp-content/uploads/2023/03/91-Geo.-Wash.-L.-Rev.-79-2023.pdf. Acesso em: 10 abr. 2025.

GHASSEMI, Marzyeh; OAKDEN-RAYNER, Luke; BEAM, Andrew L. A falsa esperança das abordagens atuais de explicabilidade em inteligência artificial na saúde. The Lancet Digital Health, v. 3, p. e745-e750, nov. 2021. Disponível em: https://pubmed.ncbi.nlm.nih.gov/34711379/. Acesso em: 12 de abril de 2025.

GIANFRANCESCO, Milena A. et al. Vieses potenciais em algoritmos de aprendizado de máquina usando dados de registros eletrônicos de saúde. JAMA Internal Medicine, v. 178, n. 11, p. 1544-1547, 2018. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6347576/. Acesso em: 06 de abril de 2025.

GILBERT, Stephen et al. Protocolos de alteração algorítmica na regulação de dispositivos médicos baseados em aprendizado de máquina adaptativo. Journal of Medical Internet Research, v. 23, n. 10, e30545, 2021. DOI: 10.2196/30545. Disponível em: https://www.jmir.org/2021/10/e30545/PDF. Acesso em: 7 de abril de 2025.

GIOVANOLA, Benedetta; TIRIBELLI, Simona. Para além do viés e da discriminação: redefinindo o princípio ético da equidade nos algoritmos de aprendizado de máquina na saúde. AI & Society, v. 38, p. 549-563, 2023. DOI: https://doi.org/10.1007/s00146-022-01455-6. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC9123626/pdf/146_2022_Article_1455.pdf Acesso em: 8 de abril de 2025.

GODDARD, Kate; ROUDSARI, Abdul; WYATT, Jeremy C. Viés de automação: uma revisão sistemática de frequência, mediadores de efeito e mitigadores. Journal of the American Medical Informatics Association, v. 19, n. 1, p. 121-127, 2012. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC3240751/pdf/amiajnl-2011-000089.pdf. Acesso em: 06 de abril de 2025.

GOERTZEL, Ben. Inteligência Artificial Geral: Conceito, Estado da Arte e Perspectivas Futuras. Journal of Artificial General Intelligence, v. 5, n. 1, p. 1-46, 2014. Disponível em: https://www.researchgate.net/publication/271390398_Artificial_General_Intelligence_Concept_State_of_the_Art_and_Future_Prospects/fulltext/5b161b57aca272d43b7e95b8/Artificial-General-Intelligence-Concept-State-of-the-Art-and-Future-Prospects.pdf?origin=publication_detail&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uRG93bmxvYWQiLCJwcmV2aW91c1BhZ2UiOiJwdWJsaWNhdGlvbiJ9fQ. Acesso em:  05 de abril de 2025.

GOMES, Bruna; ASHLEY, Euan A. Inteligência artificial em medicina molecular. New England Journal of Medicine, v. 388, n. 26, p. 2456-2465, 2023. Disponível em: https://www.binasss.sa.cr/bibliotecas/bhm/jun23/54.pdf. Acesso em: 05 de abril de 2025.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Aprendizado profundo. Cambridge: MIT Press, 2016. Disponível em: https://virtualmmx.ddns.net/gbooks/DeepLearning.pdf. Acesso em: 5 de abril de 2025.

GOSTIN, Lawrence O. Quando o terrorismo ameaça a saúde: até que ponto as limitações aos direitos humanos são justificadas? The Journal of Law, Medicine & Ethics, v. 31, n. 4, p. 524-538, 2003. Disponível em: file:///H:/1%20Documentos%20Importantes/Projetos/9.%20Projetos%202023/3%20Mestrado/2025/3.%20disserta%C3%A7%C3%A3o/refer%C3%AAncias%20bibliogr%C3%A1ficas/(GOSTIN).pdf. Acesso em: 06 de abril de 2025.

GREENLEAF, Graham. Leis Globais de Privacidade de Dados 2023: 162 Leis Nacionais e 20 Projetos de Lei. UNSW Law & Justice Research Series, [s.l.], n. 48, 2023. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4426146. Acesso em: 7 de abril de 2025.

GROTE, Thomas; BERENS, Philipp. Sobre a ética da tomada de decisão algorítmica na saúde. Journal of Medical Ethics, v. 46, n. 3, p. 205-211, 2020. Disponível em: https://pubmed.ncbi.nlm.nih.gov/31748206/. Acesso em: 7 de abril de 2025.

GSTREIN, Oskar J.; HALEEM, Noman; ZWITTER, Andrej. Regulamento de IA de uso geral e a Lei de IA da União Europeia. Internet Policy Review, v. 13, n. 3, 2024. DOI: https://doi.org/10.14763/2024.3.1790. Disponível em: https://policyreview.info/pdf/policyreview-2024-3-1790.pdf?utm_source=chatgpt.com. Acesso em: 9 de abril de 2025.

GUNNING, David et al. Inteligência Artificial Explicável (XAI). Science Robotics, v. 4, eaay7120, 2019. Disponível em: https://www.science.org/doi/epdf/10.1126/scirobotics.aay7120. Acesso em: 12 de abril de 2025.

HAENLEIN, Michael; KAPLAN, Andreas. Uma breve história da inteligência artificial: sobre o passado, presente e futuro da inteligência artificial. California Management Review, v. 61, n. 4, p. 5-14, 2019. Disponível em: https://www.researchgate.net/profile/Michael-Haenlein/publication/334539401_A_Brief_History_of_Artificial_Intelligence_On_the_Past_Present_and_Future_of_Artificial_Intelligence/links/60cd82e5a6fdcc01d482dc23/A-Brief-History-of-Artificial-Intelligence-On-the-Past-Present-and-Future-of-Artificial-Intelligence.pdf?origin=publication_detail&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uRG93bmxvYWQiLCJwcmV2aW91c1BhZ2UiOiJwdWJsaWNhdGlvbiJ9fQ&__cf_chl_tk=8Tu8H2iTUYx8NZwlMvSbRYEHkvOwb0T_jQyLp0PJqoA-1743821431-1.0.1.1-XqQkLhyP1gAX3pnh_zBEIBfwYZOxvU6WrSBcFkNJVEk. Acesso em: 04 de abril de 2025.

HARRISON, James H. Jr. et al. Introdução à inteligência artificial e aprendizado de máquina para patologia. Archives of Pathology & Laboratory Medicine, v. 145, n. 10, p. 1228-1254, 2021. Disponível em: https://allen.silverchair-cdn.com/allen/content_public/journal/aplm/145/10/10.5858_arpa.2020-0541-cp/2/i1543-2165-145-10-1228.pdf?Expires=1746921095&Signature=AlJTg0TtZ7OTBYABu~~Ys4JcjsfTuR2SN5AH6re793r0G3L4h8tB6EDSlH26T0OghRdekM9ovVmhqyEpP1xWdwl4hhN7KBMkoK4ZtimIq2FfyFYtloBB-Pq7d1X5BW1MuXKaps1Wc1GCmBzfLZlp~KcrmDBR8yu7w5aWCMl5FnHRAf13BgHeu8Jg1okBqXFDKArVdHQv2wb-

00P0u5Ulcr2gROQE1wNQ3pyr1wacpwKsloKEBTUWDZdIq3OXprOJQuJNeXnhY1tU~E1r9ieObWYzS1vBYHEGhMjBGIoYgbM9zKdOaIPDPIOsAWUHFzOL9MzCZNVOHgizuifemlbQcA__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA. Acesso em: 05 de abril de 2025.

HASSABIS, D. et al. Inteligência Artificial Inspirada na Neurociência. Neuron, v. 95, n. 2, p. 245-258, 2017. Disponível em: https://www.cell.com/action/showPdf?pii=S0896-6273%2817%2930509-3. Acesso em: 04 de abril de 2025.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. Os Elementos do Aprendizado Estatístico: Mineração de Dados, Inferência e Previsão. 2. ed. New York: Springer, 2009. Disponível em: https://www.sas.upenn.edu/~fdiebold/NoHesitations/BookAdvanced.pdf Acesso em: 05 de abril de 2025.

HATHERLEY, Joshua James. Dados sobre o diálogo: por que é improvável que a inteligência artificial humanize a medicina. 2023. Tese (Doutorado em Filosofia) - Monash University, Melbourne, 2023. Disponível em: https://philpapers.org/archive/HATDOD.pdf. Acesso em: 08 de abril de 2025.

HAUG, Charlotte J.; DRAZEN, Jeffrey M. Inteligência Artificial e Aprendizado de Máquina em Medicina Clínica. *The New England Journal of Medicine*, v. 388, n. 13, p. 1201-1208, 2023. Disponível em: https://www.uio.no/studier/emner/matnat/fys/HON1000/h23/begrenset/nejmra2302038.pdf. Acesso em: 06 de abril de 2025.

HERNANDEZ-BOUSSARD, Tina et al. MINIMAR (Minimum Information for Medical AI Reporting): desenvolvendo padrões de relato para inteligência artificial na saúde. Journal of the American Medical Informatics Association, v. 27, n. 12, p. 2011-2015, 2020. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7727333/pdf/ocaa088.pdf. Acesso em: 12 de abril de 2025.

HERZOG, Christian. Sobre a utilidade ética e epistemológica da IA explicável na medicina. Philosophy & Technology, v. 35, art. 50, 2022. Disponível em: https://link.springer.com/article/10.1007/s13347-022-00546-y. Acesso em: 12 de abril de 2025.

HILBERT, Martin; LÓPEZ, Priscila. A capacidade tecnológica mundial de armazenar, comunicar e computar informações. Science, v. 332, n. 6025, p. 60-65, 2011. Disponível em: https://www-science-org.ez67.periodicos.capes.gov.br/doi/epdf/10.1126/science.1200970 Acesso em: 04 de abril de 2025.

HINTON, Geoffrey. *Deep Learning* - Uma tecnologia com potencial para transformar os cuidados de saúde. JAMA, v. 320, n. 11, p. 1101-1102, 2018. Disponível em: https://sci-hub.se/downloads/2019-01-07/f4/hinton2018.pdf. Acesso em: 05 de abril de 2025.

HOI, Steven C. H. et al. Aprendizagem online: uma pesquisa abrangente. Neurocomputing, v. 459, p. 249-289, 2021. Disponível em: https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?params=/context/sis_research/article/7962/&path_info=OnlineLearning_Survey_sv.pdf. Acesso em: 5 abr. 2025.

HOLM, Søren; STANTON, Catherine; BARTLETT, Benjamin. Um novo argumento para compensação sem culpa na saúde: a introdução de sistemas de inteligência artificial.

Health Care Analysis, v. 29, p. 171-188, 2021. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC8321978/. Acesso em: 12 de abril de 2025.

HOOFNAGLE, C. J.; VAN BERLO, W.; ZARIN, M. R.; GOURNAY, B. de. O Regulamento Geral sobre a Proteção de Dados da União Europeia: o que é e o que significa. Information & Communications Technology Law, v. 28, n. 1, p. 65-98, 2019. DOI: https://doi.org/10.1080/13600834.2019.1573501. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3254511. Acesso em: 8 de abril de 2025.

HOSNY, Ahmed; AERTS, Hugo J. W. L. inteligência artificial para a saúde global: tecnologias socialmente responsáveis podem ajudar a enfrentar desigualdades no cuidado. Science, v. 366, n. 6468, p. 955-956, 2019. DOI: https://doi.org/10.1126/science.aay5189. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7790340/pdf/nihms-1657312.pdf Acesso em: 8 de abril de 2025.

HOVY, Eduard H. Geração automatizada de discurso usando relações de estrutura de discurso. Inteligência artificial, v. 63, p. 341-385, 1993. Disponível em: https://www.sciencedirect.com/science/article/pii/0004370293900213?via%3Dihub. Acesso em: 04 de abril de 2025.

HUNT, Paul; BACKMAN, Gunilla. Sistemas de saúde e o direito ao mais alto padrão de saúde possível. Health and Human Rights, vol. 10, n. 1, p. 81-92, 2008. Disponível em: https://repository.essex.ac.uk/4410/1/106.pdf. Acesso em: 06 de abril de 2025.

ÍNDIA. Lei de Proteção de Dados Pessoais Digitais, 2023. Digital Personal Data Protection Act, n. 113, de 2023. Publicado no Official Gazette. Disponível em: https://prsindia.org/files/bills_acts/bills_parliament/2023/Digital%20Personal%20Data%20Protection%20Bill,%202023.pdf. Acesso em: 09 de abril de 2025.

INTERNATIONAL MEDICAL DEVICE REGULATORS FORUM. Software como Dispositivo Médico (SaMD): Aplicação do Sistema de Gestão de Qualidade. IMDRF/SaMD WG/N23 FINAL:2015, 2 out. 2015. Disponível em: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-151002-samd-qms.pdf. Acesso em: 10 de abril de 2025.

INTERNATIONAL MEDICAL DEVICE REGULATORS FORUM. Software como Dispositivo Médico (SaMD): Avaliação Clínica. IMDRF/SaMD WG/N41FINAL:2017, 21 de setembro de 2017. Disponível em: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-170921-samd-n41-clinical-evaluation_1.pdf. Acesso em: 10 de abril de 2025

INTERNATIONAL MEDICAL DEVICE REGULATORS FORUM. Plano Estratégico do IMDRF 2021-2025. IMDRF/MC/N39FINAL:2020 (Edição 2), 25 de setembro de 2020. Disponível em: https://www.imdrf.org/sites/default/files/docs/imdrf/final/procedural/imdrf-proc-n39-strategic-plan-200925.pdf. Acesso em: 09 de abril de 2025.

INTERNATIONAL MEDICAL DEVICE REGULATORS FORUM. Software como dispositivo médico (SaMD): principais definições. IMDRF/SaMD WG/N10FINAL:2013. 9 de dezembro de 2013. Disponível em: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf. Acesso em: 09 de abril de 2025.

INTERNATIONAL MEDICAL DEVICE REGULATORS FORUM. Software como dispositivo médico (SaMD): possível estrutura para categorização de risco e considerações correspondentes. IMDRF/SaMD WG/N12FINAL:2014, 18 de setembro de 2014. Disponível em: https://www.imdrf.org/documents/software-medical-device-samd-possible-framework-risk. Acesso em: 09 de abril de 2025.

INTERNATIONAL MEDICAL DEVICE REGULATORS FORUM (IMDRF). Boas práticas de aprendizado de máquina para desenvolvimento de dispositivos médicos: princípios orientadores. IMDRF/AIML WG/N88 FINAL:2025. 27 de janeiro de 2025. Disponível em: https://www.imdrf.org/documents/machine-learning-enabled-medical-devices-key-terms-and-definitions. Acesso em: 10 de abril de 2025.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 13485:2016 - Dispositivos médicos - Sistemas de gestão da qualidade - Requisitos para fins regulamentares. 3. ed. Geneva: ISO, 2016. Disponível em: https://cdn.standards.iteh.ai/samples/59752/2580a43a872a4bbf9919167741ef13fa/ISO-13485-2016.pdf . Acesso em: 09 de abril de 2025.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 14971:2019 - Dispositivos médicos - Aplicação da gestão de riscos a dispositivos médicos. 3. ed. Geneva: ISO, 2019. Disponível em: https://modirfa.com/wp-content/uploads/2023/01/ISO100152019-ENGLISH.pdf?utm_source=chatgpt.com. Acesso em: 10 de abril de 2025.

IQBAL, Jeffrey David; BILLER-ANDORNO, Nikola. A lacuna regulatória na saúde digital e caminhos alternativos para preenchê-la. Health Policy and Technology, [S.l.], v. 11, p. 100663, 2022. Disponível em: https://www.sciencedirect.com/science/article/pii/S2211883722000703?via%3Dihub. Acesso em: 7 de abril de 2025.

JANSSEN, Marijn et al. Governança de dados: organizando dados para uma inteligência artificial confiável. Government Information Quarterly, v. 37, n. 3, p. 101493, 2020. DOI: https://doi.org/10.1016/j.giq.2020.101493. Disponível em: https://www.sciencedirect.com/science/article/pii/S0740624X20302719?ref=pdf_download&fr=RR-9&rr=92d599b3ee83bcd7 Acesso em: 8 de abril de 2025.

JIANG, Fei et al. Inteligência artificial na área da saúde: passado, presente e futuro. AVC e Neurologia Vascular, v. 2, n. 4, p. e000101, 2017. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC5829945/pdf/svn-2017-000101.pdf. Acesso em: 05 de abril de 2025.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. Inteligência artificial: o cenário global das diretrizes éticas. Nature Machine Intelligence, [S.l.], v. 1, p. 389-399, 2019. Disponível em: https://arxiv.org/pdf/1906.11668. Acesso em: 7 de abril de 2025.

JORDAN, M. I.; MITCHELL, T. M. Aprendizado de máquina: tendências, perspectivas e perspectivas. Science, v. 349, n. 6245, p. 255-260, 2015. Disponível em: https://www.cs.cmu.edu/~tom/pubs/Science-ML-2015.pdf. Acesso em: 04 de abril de 2025.

JUMPER, John et al. Previsão da estrutura de proteínas altamente precisa com o AlphaFold. Nature, v. 596, p. 583-589, 2021. Disponível em: https://www.nature.com/articles/s41586-021-03819-2. Acesso em 05 de abril de 2025.

KAMINSKI, Margot E. O direito à explicação, explicado. Berkeley Technology Law Journal, v. 34, n. 1, p. 189-220, 2019. Disponível em: https://doi.org/10.15779/Z38TD9N83. Acesso em: 11 de abril de 2025.

KEANE, Pearse A.; TOPOL, Eric J. De olho na IA e no diagnóstico autônomo. npj Digital Medicine, v. 1, n. 40, p. 1-3, 2018. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6550235/pdf/41746_2018_Article_48.pdf. Acesso em: 06 de abril de 2025.

KELLY, Christopher J. et al. Principais desafios para fornecer impacto clínico com inteligência artificial. BMC Medicine, v. 17, n. 195, p. 1-9, 2019. Disponívell em: https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1426-2. Acesso em: 06 de abril de 2025.

KHORSANDI, Shirin E. et al. Inteligência artificial no transplante hepático. Transplantation Proceedings, v. 53, p. 2939-2944, 2021. Disponível em: https://www.sciencedirect.com/science/article/pii/S0041134521007417?via%3Dihub. Acesso em: 11 de abril de 2025.

KILKENNY, Monique F.; ROBINSON, Kerin M. Qualidade dos dados: "entrada e saída de lixo ". Health Information Management Journal, v. 47, n. 3, p. 103-105, 2018. Disponível em: https://journals.sagepub.com/doi/10.1177/1833358318774357. Acesso em: 06 de abril de 2025.

KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. Classificação ImageNet com redes neurais convolucionais profundas. Communications of the ACM, v. 60, n. 6, p. 84-90, 2017. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf. Acesso em: 05 de abril de 2025.

KUNER, Christopher; BYGRAVE, Lee A.; DOCKSEY, Christopher (org.). O Regulamento Geral sobre a Proteção de Dados da UE: um comentário - atualização 2021 de artigos selecionados. Oxford: Oxford University Press, 2021. p. 45-46. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3839645. Acesso em: 09 de abril de 2025.

KURZWEIL, Ray. A singularidade está próxima: quando os humanos transcendem a biologia. New York: Viking, 2005. Disponível em: https://paisdospuntocero.wordpress.com/wp-content/uploads/2018/04/book-kurzweil-singularity-is-near-1.pdf. Acesso em: 04 de abril de 2025.

KUSHIDA, Clete A. et al. Estratégias para desidentificação e anonimização de dados de prontuários eletrônicos de saúde para uso em estudos de pesquisa multicêntricos. Medical Care, v. 50, supl., p. S82-S101, 2012. DOI: 10.1097/MLR.0b013e3182585355. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6502465/pdf/nihms-374697.pdf. Acesso em: 09 abril de 2025.

LAUPACIS, Andreas; SACKETT, David L.; ROBERTS, Robin S. Uma avaliação de medidas clinicamente úteis das consequências do tratamento. The New England Journal of Medicine, v. 318, n. 26, p. 1728-1733, 1988. Disponível em: https://sci-hub.se/10.1056/nejm198806303182605. Acesso em: 06 de abril de 2025.

LEARY, Virginia A. O Direito à Saúde no Direito Internacional dos Direitos Humanos. Health and Human Rights, v. 1, n. 1, p. 24-56, 1994. Disponível em: http://www.jstor.org/stable/4065261. Acesso em: 06 de abril de 2025.

LECHTERMAN, Theodore M. O conceito de accountability na ética e governança da IA. In: BULLOCK, Justin B. et al. (org.). O Manual Oxford de Governança de IA. Oxford: Oxford University Press, 2022. p. 164-182. Disponível em: https://ora.ox.ac.uk/objects/uuid:c954516e-9139-47ed-941e-978a82f7d84e/files/rgb19f7562. Acesso em: 08 de abril de 2025.

LECKENBY, Emily; DAWOUD, Dalia; BOUVY, Jacoline; JÓNSSON, Páll. A abordagem sandbox e seu potencial de uso na avaliação de tecnologias em saúde: uma revisão da literatura. Economia da Saúde Aplicada e Política de Saúde, v. 19, p. 857-869, 2021. Disponível em: https://pubmed.ncbi.nlm.nih.gov/34254275/. Acesso em: 12 de abril de 2025.

LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Aprendizado profundo. Nature, [s. l.], v. 521, p. 436-444, 2015. Disponível em: https://www.cs.toronto.edu/~hinton/absps/NatureDeepReview.pdf. Acesso em: 5 abr. 2025.

LEE, Cecilia S.; LEE, Aaron Y. Aplicações de aprendizado contínuo de aprendizado de máquina na prática clínica. The Lancet Digital Health, v. 2, n. 6, p. e279-e281, 2020. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC8259323/pdf/nihms-1701649.pdf. Acesso em: 06 de abril de 2025.

LEVITINA, Carolina. Humanos na tomada de decisão automatizada sob a Lei RGPD e IA. In: DE MIGUEL ASENSIO, Pedro Alberto (coord.). A regulação da inteligência artificial e seu impacto no campo da saúde. Madrid: Reus, 2022. p. 231-232. Disponível em: https://recyt.fecyt.es/index.php/cidob/article/view/112020/82014. Acesso em: 09 de abril de 2025.

LI, Tian et al. Aprendizado federado: desafios, métodos e direções futuras. Foundations and Trends® in *Machine Learning*, 2020. arXiv:1908.07873. Disponível em: https://arxiv.org/pdf/1908.07873. Acesso em: 09 de abril de 2025.

LIEVEVROUW, Elisa; MARELLI, Luca; VAN HOYWEGHEN, Ine. O processo de padronização da FDA para tecnologias médicas de saúde digital: coprodução de inovação tecnológica e organizacional. BioSocieties, [s.l.], v. 17, p. 549-576, 2022. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC8116827/pdf/41292_2021_Article_232.pdf. Acesso em: 10 de abril de 2025.

LIPTON, Zachary C. O mito da interpretabilidade dos modelos. Workshop sobre Interpretabilidade Humana em Aprendizado de Máquina (ICML), Nova York, 2016. Disponível em: https://arxiv.org/abs/1606.03490. Acesso em: 11 de abril de 2025.

LITJENS, Geert; KOOI, Thijs; BEJNORDI, Babak Ehteshami; SETIO, Arnaud Arindra Adiyoso; CIOMPI, Francesco; GHAFOORIAN, Mohsen; VAN DER LAAK, Jeroen A. W. M.; VAN GINNEKEN, Bram; SÁNCHEZ, Clara I. Uma pesquisa sobre aprendizado profundo em medicina
análise de imagem. Análise de imagens médicas, v. 42, p. 60 88, dezembro. 2017. https://doi.org/10.1016/j.media.2017.07.005.

LIU, Hongfang et al. Inteligência artificial na área da saúde: a esperança, o exagero, a promessa, o perigo. Washington, Capital: Academia Nacional de Medicina; Gabinete de Responsabilidade do Governo dos Estados Unidos, 2019. (GAO-20-215SP).

Disponível em: https://www.gao.gov/assets/gao-20-215sp.pdf. Acesso em: 06 de abril de 2025.

LIU, Xiaoxuan et al. A auditoria algorítmica médica. The Lancet Digital Health, v. 4, p. e384-e397, 2022. DOI: https://doi.org/10.1016/S2589-7500(22)00003-6. Disponível em: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00003-6/fulltext. Acesso em: 07 de abril de 2025.

LONDON, Alex John. Inteligência artificial e decisões médicas em caixa-preta: acurácia versus explicabilidade. Hastings Center Report, v. 49, n. 1, p. 15-21, 2019. Disponível em: https://pubmed.ncbi.nlm.nih.gov/30790315/. Acesso em: 12 de abril de 2025.

LUNDBERG, Scott M.; LEE, Su-In. Uma abordagem unificada para interpretar previsões de modelos. In: Advances in Neural Information Processing Systems 30 (NIPS 2017), p. 4765-4774. Disponível em: https://dl.acm.org/doi/pdf/10.5555/3295222.3295230. Acesso em: 12 de abril de 2025.

MADIEGA, Tambiama. Lei de inteligência artificial e sandboxes regulatórios. Serviço de Pesquisa do Parlamento Europeu - EPRS, PE 733.544, jun. 2022. Disponível em: https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/733544/EPRS_BRI(2022)733544_EN.pdf. Acesso em: 12 de abril de 2025.

MADIEGA, Tambiama. Lei de inteligência artificial. Serviço de Pesquisa do Parlamento Europeu - EPRS, PE 698.792, set. 2024. Disponível em: https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf. Acesso em: 12 de abril de 2025.

MATTHIAS, Andreas. A lacuna de responsabilidade: atribuindo responsabilidade às ações de autômatos que aprendem. Ética e Tecnologia da Informação, v. 6, p. 175-183, 2004. Disponível em: https://link.springer.com/article/10.1007/s10676-004-3422-1. Acesso em: 12 de abril de 2025.

MAY, Thomas. Riscos sociogenéticos - Testes de ancestralidade genética, identidade de terceiros e proteção da privacidade. The New England Journal of Medicine, v. 379, n. 5, p. 410-412, 2018. DOI: 10.1056/NEJMp1805870. Disponível em: https://pubmed.ncbi.nlm.nih.gov/29924688/. Acesso em: 09 de abril de 2025.

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. Uma proposta para o projeto de pesquisa de verão de Dartmouth sobre inteligência artificial. Hanover: Dartmouth College, 31 ago. 1955. Disponível em: https://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html. Acesso em: 04 de abril de 2025.

MCCARTHY, John. Programas com bom senso. Universidade de Stanford, 1959. Disponível em: http://www-formal.stanford.edu/jmc/mcc59.pdf. Acesso em: 04 de abril de 2025.

MCCARTHY, John. O que é inteligência artificial? Universidade de Stanford, 2007. Disponível em: http://www-formal.stanford.edu/jmc/whatisai/whatisai.pdf. Acesso em: 04 de abril de 2025.

MCNAIR, Douglas; PRICE II, W. Nicholson. Inteligência artificial na assistência à saúde: direito, regulação e políticas públicas. In: MATHENY, Michael et al. (org.). Inteligência artificial na saúde: a esperança, a ilusão, a promessa, o risco. Washington,

Capital: Academia Nacional de Medicina, 2019. p. 181-213. Disponível em: https://repository.law.umich.edu/book_chapters/357/. Acesso em: 08 de abril de 2025.

MEDICAL DEVICE COORDINATION GROUP (MDCG). Orientações sobre avaliação clínica (MDR) / avaliação de desempenho (IVDR) de softwares de dispositivos médicos (MDSW). MDCG 2020-1, mar. 2020. Disponível em: https://health.ec.europa.eu/system/files/2020-09/md_mdcg_2020_1_guidance_clinic_eva_md_software_en_0.pdf. Acesso em: 11 de abril de 2025.

MILLER, Randolph A. Sistemas de apoio à decisão de diagnóstico médico: passado, presente e futuro. Journal of the American Medical Informatics Association, v. 1, n. 1, p. 8-27, 1994. Disponívem em: https://www.sciencedirect.com/science/article/pii/S0004370218305988?via%3Dihub. Acesso em: 05 de abril de 2025.

MINSKY, Marvin. Logical Versus Analogical or Symbolic Versus Connectionist or Neat Versus Scruffy. AI Magazine, v. 12, n. 2, p. 34-51, 1991. Disponível em: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/894. Acesso em: 05 abril de 2025.

MINSSEN, Timo et al. Respostas regulatórias ao aprendizado de máquina na medicina. Journal of Law and the Biosciences, v. 7, n. 1, p. 1-18, 2020. DOI: 10.1093/jlb/lsaa002. Acesso em: 07 de abril de 2025.

MINSSEN, Timo; MIMLER, Marc; MAK, Vivian. Quando um software autônomo é considerado um dispositivo médico na União Europeia? Medical Law Review, v. 28, n. 3, p. 615-624, 2020. Disponível em: https://academic.oup.com/medlaw/article/28/3/615/5865470?login=true. Acesso em: 12 de abril de 2025.

MITCHELL, Tom M. Aprendizado de máquina. New York: McGraw-Hill, 1997. Disponível em: https://www.cs.cmu.edu/~tom/files/MachineLearningTomMitchell.pdf. Acesso em: 05 abril de 2025.

MITTELSTADT, Brent. Interpretabilidade e transparência na inteligência artificial. SSRN Electronic Journal, 2023. Disponível em: https://ssrn.com/abstract=4317938. Acesso em: 07 de abril de 2025.

MOLNAR, Christoph. Aprendizado de máquina interpretável: um guia para tornar modelos de caixa-preta explicáveis. [S.l.]: Leanpub, 2022. Disponível em: https://originalstatic.aminer.cn/misc/pdf/Molnar-interpretable-machine-learning_compressed.pdf. Acesso em: 11 de abril de 2025.

MONTEIRO, Renato L.; MACHADO, Caio V.; SILVA, Leoncio. O direito à explicação na Lei Brasileira de Proteção de Dados. International Journal of Digital and Data Law, v. 7, p. 119-136, 2021. Disponível em: https://ojs.imodev.org/?journal=RIDDN&page=article&op=view&path%5B%5D=406&path%5B%5D=550. Acesso em: 12 de abril de 2025.

MORLEY, Jessica et al. A ética da IA na saúde: uma revisão de mapeamento. Social Science & Medicine, v. 260, 2020. Disponível em: https://www.sciencedirect.com/science/article/pii/S0277953620303919?via%3Dihub. Acesso em: 11 de abril de 2025.

MURDOCH, Blake. Privacidade e inteligência artificial: desafios para a proteção de informações de saúde em uma nova era. BMC Medical Ethics, v. 22, n. 122, 2021. DOI: 10.1186/s12910-021-00687-3. Disponível em: https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-021-00687-3. Acesso em: 09 de abril de 2025.

MURPHY, K. P. Aprendizado de máquina: uma perspectiva probabilística. Cambridge, MA: MIT Press, 2012. Disponível em: https://www.cs.ubc.ca/~murphyk/MLbook/pml-toc-1may12.pdf. Acesso em: 05 de abril de 2025.

NASCIMENTO, Tiago. RGPD e cibersegurança no setor público: os dados são o novo "petróleo" do século XXI. Portal do DPO, 2020. Disponível em: https://www.academia.edu/45392549/Os_dados_s%C3%A3o_o_novo_petr%C3%B3leo_do_s%C3%A9culo_XXI. Acesso em: 08 de abril de 2025.

NATURE. O potencial da IA para acelerar a descoberta de medicamentos precisa de uma verificação da realidade. Nature, v. 622, p. 217, 2023. Disponível em: https://www.nature.com/articles/d41586-023-03159-x. Acesso em: 05 de abril de 2025.

NEWELL, Allen; SIMON, Herbert A. Ciência da Computação como Investigação Empírica: Símbolos e Pesquisa. Comunicações do ACM, v. 19, n. 3, p. 113-126, 1976. Disponível em: https://doi.org/10.1145/360018.360022. Acesso em: 04 de abril de 2025.

NIC.br. Observatório Brasileiro de Inteligência Artificial é lançado. São Paulo, 4 set. 2024. Disponível em: https://www.nic.br/noticia/na-midia/observatorio-brasileiro-de-inteligencia-artificial-e-lancado/. Acesso em: 06 de abril de 2025.

NIEMIEC, Emilia. O Regulamento de Dispositivos Médicos da UE ajudará a melhorar a segurança e o desempenho dos dispositivos médicos de IA? Digital Health, [S.l.], v. 8, p. 1-8, 2022. Disponível em: https://doi.org/10.1177/20552076221089079. Acesso em: 10 de abril de 2025.

NOVELLI, Claudio; TADDEO, Mariarosaria; FLORIDI, Luciano. Responsabilidade em inteligência artificial: o que é e como funciona. AI & Society, v. 39, p. 1871-1882, 2024. Disponível em: https://link.springer.com/article/10.1007/s00146-023-01635-y. Acesso em: 08 de abril de 2025.

NTOUTSI, Eirini et al. Viés em sistemas de inteligência artificial orientados por dados - uma introdução geral. WIREs Data Mining and Knowledge Discovery, v. 10, n. 3, e1356, 2020. Disponível em: https://wires.onlinelibrary.wiley.com/doi/epdf/10.1002/widm.1356. Acesso em: 08 de abril de 2025.

OBERMEYER, Ziad; EMANUEL, Ezekiel J. Prevendo o futuro - *Big Data, Machine Learning* e medicina clínica. New England Journal of Medicine, v. 375, n. 13, p. 1216-1219, 2016. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC5070532/pdf/nihms821556.pdf. Acesso em: 05 de abril de 2025.

ODAIBO, Stephen G. Gestão de riscos em software como dispositivo médico com IA/ML (SaMD): sobre a ISO 14971 e normas relacionadas. RETINA-AI Health, Inc., 2021. Disponível em: https://arxiv.org/pdf/2109.07905. Acesso em: 10 de abril de 2025.

OMS. Considerações regulatórias sobre inteligência artificial para a saúde. Genebra: Organização Mundial da Saúde, 2023. Disponível em: https://reliefweb.int/report/world/regulatory-considerations-artificial-intelligence-health. Acesso em: 08 de abril de 2025.

OMS. Departamento do Alto Comissariado das Nações Unidas para os Direitos Humanos. O direito à saúde: Ficha informativa nº 31. Geneva: ONU, 2008. Disponível em: https://www.ohchr.org/sites/default/files/Documents/Publications/Factsheet31.pdf. Acesso em: 06 de abril de 2025.

OMS. Ética e governança da inteligência artificial para a saúde: orientação da OMS. Genebra: Organização Mundial da Saúde, 2021. Disponível em: https://apps.who.int/iris/handle/10665/341996. Acesso em: 07 de abril de 2025.

ONU. Comentário Geral nº 16 - Artigo 17: Direito à Privacidade. Comitê de Direitos Humanos da ONU. Adotado em 8 de abril de 1988. Disponível em: https://www.refworld.org/legal/general/hrc/1988/en/27539. Acesso em: 8 de abril de 2025.

ONU. Declaração Universal dos Direitos Humanos. Paris: Organização das Nações Unidas, 1948. Disponível em: https://www.un.org/en/about-us/universal-declaration-of-human-rights. Acesso em: 06 de abril de 2025.

ONU. Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais. Nova York: Assembleia Geral das Nações Unidas, 1966. Disponível em: https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights. Acesso em: 06 de abril de 2025.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). Inteligência artificial na sociedade. Paris: OECD Publishing, 2019. Disponível em: https://www.oecd.org/en/publications/artificial-intelligence-in-society_eedfee77-en.html. Acesso em: 07 de abril de 2025.

OWENS, John; CRIBB, Alan. “Meu Fitbit Acha que Posso Fazer Melhor!”: As tecnologias vestíveis de promoção da saúde apoiam a autonomia pessoal?”. Philosophy & Technology, [S.l.], v. 32, p. 23-38, 2019. DOI: 10.1007/s13347-017-0266-2. Disponível em: https://link.springer.com/article/10.1007/s13347-017-0266-2. Acesso em: 07 de abril de 2025.

PARIKH, Ravi B.; OBERMEYER, Ziad; NAVATHE, Amol S. Regulamentação da análise preditiva na medicina: os algoritmos devem atender aos padrões regulatórios de benefício clínico. Science, Washington, v. 363, n. 6429, p. 810-812, 2019. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6557272/. Acesso em: 07 de abril de 2025.

PARISI, German I. et al. Aprendizagem contínua ao longo da vida com redes neurais: uma revisão. Neural Networks, v. 113, p. 54-71, 2019. Disponível em: https://pure.hw.ac.uk/ws/portalfiles/portal/24019831/1_s2.0_S0893608019300231_main.pdf. Acesso em: o5 de abril de 2025.

PHANSALKAR, Shobha et al. Interações medicamentosas que devem ser não interruptivas para reduzir a fadiga de alerta em registros eletrônicos de saúde. Journal of the American Medical Informatics Association, v. 20, n. 3, p. 489-493, 2013. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC3628052/pdf/amiajnl-2012-001089.pdf. Acesso em: 06 de abril de 2025.

PLOTKIN, David. Administração de dados: um guia prático para a gestão eficaz de dados e governança de dados. 2. ed. Londres: Academic Press, 2021. Disponível em: https://dokumen.pub/qdownload/data-stewardship-an-actionable-guide-to-effective-data-management-and-data-governance-2nbsped-0128221321-9780128221327.html Acesso em: 08 de abril de 2025.

PRICE II, W. Nicholson; COHEN, I. Glenn. Privacidade na era dos megadados médicos. Nature Medicine, v. 25, n. 1, p. 37-43, 2019. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6376961/pdf/nihms-1005787.pdf. Acesso em: 08 de abril de 2025.

PRICE II, W. Nicholson; GERKE, Sara; COHEN, I. Glenn. Potencial responsabilidade de médicos pelo uso de inteligência artificial. JAMA - Journal of the American Medical Association, v. 322, n. 18, p. 1765-1766, 2019. Disponível em: https://repository.law.umich.edu/articles/2475/. Acesso em: 12 de abril de 2025.

PRICE, W. Nicholson II. Erros médicos e medicina de caixa-preta. In: COHEN, I. Glenn et al. (org.). *Big Data*, direito da saúde e bioética. Cambridge: Cambridge University Press, 2018. p. 390. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3147243. Acesso em: 12 de abril de 2025.

PRICE, W. Nicholson II; GERKE, Sara; COHEN, I. Glenn. Responsabilidade pelo uso de inteligência artificial na medicina. In: SOLAIMAN, Barry; COHEN, I. Glenn (org.). Manual de Pesquisa em Saúde, IA e Direito (a publicar). Edward Elgar Publishing, 2023. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4115538 Acesso em: 12 de abril de 2025.

PRIVACY INTERNATIONAL. Chaves para a proteção de dados: um guia para o engajamento político em proteção de dados. Londres: Privacy International, 2018. Disponível em: https://privacyinternational.org/sites/default/files/2018-09/Data%20Protection%20COMPLETE.pdf. Acesso em: 08 de abril de 2025.

RAJKOMAR, Alvin; DEAN, Jeffrey; KOHANE, Isaac. Aprendizado de máquina na medicina. The New England Journal of Medicine, v. 380, n. 14, p. 1347-1357, 2019. DOI: 10.1056/NEJMra1814259. Acesso em: 05 de abril de 2025.

RAJKOMAR, Alvin et al. Garantindo equidade na aprendizagem de máquina para promover justiça em saúde. Annals of Internal Medicine, v. 169, n. 12, p. 866-872, 2018. DOI: https://doi.org/10.7326/M18-1990. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6594166/pdf/nihms-1033618.pdf. Acesso em: 08 de abril de 2025.

RAJKOMAR, Alvin; OREN, Eyal; CHEN, Kai; DAI, Andrew M.; HAJAJ, Nissan; HARDT, Michaela; LIU, Peter J.; LIU, Xiaobing; MARCUS, Jake; SUN, Mimi; SUNDBERG, Patrik; YEE, Hector; ZHANG, Kun; ZHANG, Yi; FLORES, Gerardo; DUGGAN, Gavin E.; IRVINE, Jamie; LE, Quoc; LITSCH, Kurt; DEAN, Jeffrey. Aprendizado profundo escalável e preciso com registros eletrônicos de saúde. NPJ Digital Medicine , v. 1, n. 1, p. 18, 8 maio 2018. Disponível em: https://doi.org/10.1038/s4174601800291. Acesso em: 05 de abril de 2025

RAJPURKAR, Pranav; LUNGREN, Matthew P. O estado atual e futuro da interpretação de imagens médicas por IA. New England Journal of Medicine, v. 388, n. 21, p. 1981-1990, 2023. Disponível em:

https://strawman.com/member/uploads/objects/1b/b9492b8109bbe389c23c38a5a3a0874ab2ec11.pdf. Acesso em: 05 de abril de 2025.

RAWLS, John. Uma teoria da justiça. Tradução de Almiro Pisetta e Leonel Vallandro. 2. ed. revista. São Paulo: Martins Fontes, 2002. (Baseado na edição original: RAWLS, John. A Theory of Justice. Rev. ed. Cambridge, MA: Belknap Press of Harvard University Press, 1999). Disponível em: https://giuseppecapograssi.wordpress.com/wp-content/uploads/2014/08/rawls99.pdf. Acesso em: 08 de abril de 2025.

RENDA, Andrea; ENGLER, Alex. O que há em um nome? Acertando na definição de inteligência artificial na Lei de IA da UE. CEPS Explainer, Brussels, fev. 2023. Disponível em: https://www.ceps.eu/ceps-publications/whats-in-a-name/. Acesso em: 04 de abril de 2025.

RIBEIRO, Marco Tulio et al. T-Explainer: Um framework de explicabilidade independente de modelo baseado em gradientes. Preprint submetido ao IEEE, 2016. Disponível em: https://arxiv.org/pdf/2404.16495v1. Acesso em: 12 de abril de 2025.

RICHMAN, Barak. Regulamentação da saúde para a era digital. New England Journal of Medicine, [s.l.], v. 379, n. 18, p. 1694-1695, 1 nov. 2018. Disponível em: https://www.nejm.org/doi/full/10.1056/NEJMp1806848. Acesso em: 07 de abril de 2025.

RIEKE, Nicola et al. O futuro da saúde digital com aprendizado federado. NPJ Digital Medicine, v. 3, n. 119, 2020. DOI: 10.1038/s41746-020-00323-1. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7490367/pdf/41746_2020_Article_323.pdf. Acesso em: 09 de abril de 2025.

ROCHER, Luc; HENDRICKX, Julien M.; MONTJOYE, Yves-Alexandre de. Estimando o sucesso de reidentificações em conjuntos de dados incompletos com modelos generativos. Nature Communications, v. 10, n. 3069, 2019. DOI: 10.1038/s41467-019-10933-3. Disponível em: https://www.nature.com/articles/s41467-019-10933-3. Acesso em: 09 de abril de 2025.

ROSKI, J. et al. inteligência artificial na Saúde: A Esperança, o Exagero, a Promessa, o Perigo. Washington, Capital: Academia Nacional de Medicina. Gabinete de Responsabilidade do Governo dos Estados Unidos, 2019. (GAO-20-215SP). Disponível em: https://www.gao.gov/assets/gao-20-215sp.pdf. Acesso em: 05 de abril de 2025.

RUDIN, Cynthia. Pare de explicar modelos de aprendizado de máquina em caixa-preta para decisões de alto impacto e use modelos interpretáveis em vez disso. Duke University, 2019. Disponível em: https://arxiv.org/pdf/1811.10154. Acesso em: 12 de abril de 2025.

RUMELHART, David E.; MCCLELLAND, James L. (Ed.). Processamento distribuído paralelo: Explorations in the Microstructure of Cognition (Explorações na Microestrutura da Cognição). Volume 1: Foundations. Cambridge, MA: MIT Press, 1986. Disponível em: https://www.cs.utoronto.ca/~hinton/absps/naturebp.pdf. Acesso em: 05 de abril de 2025.

RUSSELL, Stuart J.; NORVIG, Peter. Inteligência artificial: Uma Abordagem Moderna. 3rd ed. Upper Saddle River, NJ: Pearson Education, 2010. Disponível em: http://repo.darmajaya.ac.id/5272/1/Artificial%20Intelligence-A%20Modern%20Approach%20%283rd%20Edition%29%20%28%20PDFDrive%20%29.pdf. Acesso em: 4 de abril de 2025.

SAHNI, Nikhil R.; CARRUS, Brandon. A inteligência artificial na prestação de cuidados de saúde nos EUA. New England Journal of Medicine, v. 389, n. 4, p. 348-358, 2023. Disponível em: https://statmodeling.stat.columbia.edu/wp-content/uploads/2024/01/nejmra2204673.pdf. Acesso em: 06 de abril de 2025.

SAITO, Takaya; REHMSMEIER, Marc. O gráfico de precisão-recuperação é mais informativo do que o gráfico ROC na avaliação de classificadores binários em conjuntos de dados desequilibrados. PLoS ONE, v. 10, n. 3, 2015. Disponível em: 10.1371/journal.pone.0118432. Acesso em 06 de abril de 2025.

SANTORO, Adam; LAMPINEN, Andrew; MATHEWSON, Kory; LILLICRAP, Timothy; RAPOSO, David. Comportamento simbólico na inteligência artificial. arXiv preprint arXiv:2102.03406, 2021. Disponível em: https://arxiv.org/abs/2102.03406. Acesso em: 04 de abril de 2025.

SCHÖNBERGER, Daniel. Inteligência artificial na saúde: uma análise crítica das implicações legais e éticas. International Journal of Law and Information Technology, v. 27, n. 2, p. 171-203, 2019. Disponível em: https://academic.oup.com/ijlit/article/27/2/171/5485669?rss=1&login=true. Acesso em: 12 de abril de 2025.

SCHWALBE, Nina; WAHL, Brian. Inteligência artificial e o futuro da saúde global. The Lancet, v. 395, p. 1579-1586, 2020. DOI: https://doi.org/10.1016/S0140-6736(20)30226-9. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7255280/pdf/main.pdf. Acesso em: 07 de abril de 2025.

SELBST, Andrew D.; POWLES, Julia. Informação significativa e o direito à explicação. International Data Privacy Law, v. 7, n. 4, p. 233-242, 2017. Disponível em: https://academic.oup.com/idpl/article/7/4/233/4762325?login=true. Acesso em: 12 de abril de 2025.

SENDAK, Mark P. et al. Um caminho para a tradução de produtos de aprendizagem automática na prestação de cuidados de saúde. EMJ Innovations, jan. 2020. Disponível em: https://www.emjreviews.com/wp-content/uploads/2020/01/A-Path-for-Translation-of-Machine-Learning.....pdf. Acesso em: 05 de abril de 2025.

SHAH, Pratik et al. Inteligência artificial e aprendizagem automática no desenvolvimento clínico: uma perspetiva translacional. NPJ Digital Medicine, v. 2, n. 69, p. 1-5, 2019. Disponível em: https://www.nature.com/articles/s41746-019-0148-3. Acesso em: 06 de abril de 2025.

SHERKOW, Jacob S. Áreas de testes regulatórias e saúde pública. University of Illinois Law Review, v. 2022, p. 357-408. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3792217. Acesso em: 12 de abril de 2025.

SMOLENSKY, Paul. Sobre o tratamento adequado do conexionismo. Ciências Comportamentais e Cerebrais, v. 11, n. 1, p. 1-74, 1988. Disponível em: https://home.csulb.edu/~cwallis/382/readings/482/smolensky.proper.treat.pdf. Acesso em: 05 de abril de 2025.

SPECTOR-BAGDADY, Kayte et al. Compartilhamento de dados e amostras biológicas em saúde com a indústria - uma abordagem prática orientada por princípios. The New

England Journal of Medicine, v. 382, n. 22, p. 2072-2075, 2020. DOI: 10.1056/NEJMp1915298. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7332220/pdf/main.pdf. Acesso em: 09 de abril de 2025.

SULLIVAN, Hannah R.; SCHWEIKART, Scott J. As doutrinas atuais de responsabilidade civil são adequadas para lidar com danos causados por IA? AMA Journal of Ethics, v. 21, n. 2, p. 160-166, 2019. Disponível em: https://www.semanticscholar.org/paper/Are-Current-Tort-Liability-Doctrines-Adequate-for-Sullivan-Schweikart/15cf1e421bd6bea554f6cfdb7f6591faf92aad3a. Acesso em: 12 de abril de 2025.

SURESH, Harini; GUTTAG, John. Uma estrutura para entender fontes de danos ao longo do ciclo de vida do aprendizado de máquina. In: Equidade e Acesso em Algoritmos, Mecanismos e Otimização (EAAMO ’21), 5-9 out. 2021, Nova York: ACM, 2021. p. 1-9. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3465416.3483305. Acesso em: 06 de abril de 2025.

SUTTON, Reed T. et al. Uma visão geral dos sistemas de suporte à decisão clínica: benefícios, riscos e estratégias para o sucesso. NPJ Digital Medicine, [S. l.], v. 3, n. 17, p. 1-10, 2020. Disponível em: https://doi.org/10.1038/s41746-020-0221-y. Acesso em: 05 de abril de 2025.

SZOLOVITS, Peter et al. Inteligência artificial, aprendizado de máquina e sistemas de saúde. Journal of Global Health, v. 8, n. 2, p. 2-3, 2018. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6199467/pdf/jogh-08-020303.pdf. Acesso em: 05 de abril de 2025.

THEODOS, Kim; SITTIG, Scott. Leis de privacidade da informação em saúde na era digital: a HIPAA não se aplica. Perspectives in Health Information Management, [S.l.], Winter 2021. p. 5-6. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC7883355/pdf/phim0018-0001l.pdf. Acesso em: 08 de abril de 2025.

TOPOL, Eric J. Medicina de alto desempenho: a convergência da inteligência humana e artificial. Nature Medicine, v. 25, n. 1, p. 44-56, 2019. Disponível em: https://www.sbra.be/sites/default/files/1._topol_j._e._high-performance_medicine_the_convergence_of_human_and_artificial_intelligence_nature_medicine_volume_25_january_2019.pdf. Acesso em 05 de abril de 2025.

TOWNEND, David. Privacidade. In: SEDERBERG, Audrey et al. (org.). Manual de Cambridge sobre inteligência artificial responsável: perspectivas interdisciplinares. Cambridge: Cambridge University Press, 2022. p. 73-80. DOI: https://doi.org/10.1017/9781108620024.010. Disponível em: https://www.cambridge.org/core/books/cambridge-handbook-of-health-research-regulation/privacy/E584F9EF39E9597F2F8A0F83ECBE34D6. Acesso em: 08 de abril de 2025.

TURING, A. M. Maquinário de computação e inteligência. Mind, v. 59, n. 236, p. 433-460, 1950. Disponível em: https://courses.cs.umbc.edu/471/papers/turing.pdf. Acesso em: 04 de abril de 2025.

UNESCO. Recomendação sobre a Ética da Inteligência Artificial. Paris: Organização das Nações Unidas para a Educação, Ciência e Cultura, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137. Acesso em: 8 de abril de 2025.

UNESCO. Relatório do IBC sobre *Big Data* e Saúde. Paris: Organização das Nações Unidas para a Educação, Ciência e Cultura. 2017. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000258724. Acesso em: 07 de abril de 2025.

UNESCO. Readiness Assessment Methodology: a tool of the Recommendation on the Ethics of Artificial Intelligence. Publicado em 28 de agosto de 2023. Disponível em: https://www.unesco.org/en/articles/readiness-assessment-methodology-tool-recommendation-ethics-artificial-intelligence. Acesso em: 01 de maio de 2025.

UNIÃO EUROPEIA. Comissão Europeia. Livro Branco sobre a inteligência artificial: uma abordagem europeia para a excelência e a confiança. Bruxelas, 19 fev. 2020. COM(2020) 65 final. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52020DC0065. Acesso em: 12 de abril de 2025.

UNIÃO EUROPEIA. Comissão Europeia. Proposta de Diretiva do Parlamento Europeu e do Conselho relativa à adaptação das regras de responsabilidade civil extracontratual à inteligência artificial (Diretiva Responsabilidade da IA). COM(2022) 496 final, 28 set. 2022. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52022PC0496. Acesso em: 12 de abril de 2025.

UNIÃO EUROPEIA. Comissão Europeia. Proposta de Diretiva do Parlamento Europeu e do Conselho relativa à responsabilidade por produtos defeituosos. Bruxelas, 28 set. 2022. COM(2022) 495 final. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52022PC0495. Acesso em: 12 de abril de 2025.

UNIÃO EUROPEIA. Comissão Europeia. Relatório de avaliação de impacto que acompanha a proposta de diretiva relativa à adaptação das regras de responsabilidade civil extracontratual à inteligência artificial. Bruxelas, 28 set. 2022. SWD(2022) 319 final. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52022SC0319. Acesso em: 12 de abril de 2025.

UNIÃO EUROPEIA. Grupo de Peritos de Alto Nível sobre Inteligência Artificial. Orientações éticas para uma IA de confiança. Bruxelas: Comissão Europeia, 2019. Disponível em: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Acesso em: 07 de abril de 2025.

UNIÃO EUROPEIA. Grupo de Especialistas de Alto Nível em Inteligência Artificial. Uma definição de IA: principais capacidades e disciplinas científicas. Bruxelas: Comissão Europeia, 2019. Disponível em: https://ec.europa.eu/futurium/en/ai-alliance-consultation. Acesso em: 04 de abril de 2025.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. Jornal Oficial da União Europeia, L 119/1, 4 maio 2016. p. 1. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679. Acesso em: 08 de abril de 2025.

UNIÃO EUROPEIA. Parlamento Europeu; Conselho da União Europeia. Regulamento (UE) 2017/745, de 5 de abril de 2017: relativo aos dispositivos médicos, que altera a Diretiva 2001/83/CE, o Regulamento (CE) nº 178/2002 e o Regulamento (CE) nº 1223/2009 e que revoga as Diretivas 90/385/CEE e 93/42/CEE do Conselho. Jornal Oficial da União Europeia, L 117, p. 1-175, 5 maio 2017. Disponível em: https://eur-

lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32017R0745. Acesso em: 07 de abril de 2025.

UNIÃO EUROPEIA. Comissão Europeia. Decisão de Execução (UE) 2022/757, de 11 de maio de 2022: que altera a Decisão de Execução (UE) 2021/1182 no que diz respeito às normas harmonizadas relativas aos sistemas de gestão da qualidade, à esterilização e à aplicação da gestão de risco aos dispositivos médicos. Jornal Oficial da União Europeia, L 138, p. 27-29, 17 maio 2022. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A32022D0757. Acesso em: 07 de abril de 2025.

UNIÃO EUROPEIA. Regulamento (UE) 2024/... do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial). Bruxelas: Parlamento Europeu, 2024. Disponível em: https://eur-lex.europa.eu. Acesso em: 07 de abril de 2025.

VALVERDE-ALBACETE, Francisco J.; PELÁEZ-MORENO, Carmen. Precisão de classificação de 100% é considerada prejudicial: O fator de transferência de informações normalizado explica o paradoxo da precisão. PLoS ONE, v. 9, n. 1, e84217, 2014. Disponível em: https://pubmed.ncbi.nlm.nih.gov/24427282/. Acesso em: 06 de abril de 2025.

VAMATHEVAN, Jessica et al. Aplicações de aprendizado de máquina na descoberta e desenvolvimento de medicamentos. Nature Reviews Drug Discovery, v. 18, n. 6, p. 463-477, 2019. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6552674/pdf/nihms-1029624.pdf. Acesso em: 05 de abril de 2025.

VAN DE VEN, Gido M. et al. Três tipos de aprendizagem incremental. Nature Machine Intelligence, v. 4, n. 12, p. 1185-1197, dez. 2022. Disponível em: https://doi.org/10.1038/s42256-022-00568-3. Acesso em: 05 de abril de 2025.

VASWANI, A. et al. Atenção é tudo que você precisa. In: Avanços em Sistemas de Processamento de Informação Neural, v. 30, 2017. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 05 de abril de 2025.

VAYENA, Effy; BLASIMME, Alessandro. Megadados biomédicos: novos modelos de controle de acesso, uso e governança. Bioethical Inquiry, v. 14, p. 501-513, 2017. DOI: https://doi.org/10.1007/s11673-017-9809-6. Disponível em: https://pubmed.ncbi.nlm.nih.gov/28983835/. Acesso em: 08 de abril de 2025.

VAYENA, Effy; BLASIMME, Alessandro; COHEN, I. Glenn. Machine Learning na medicina: enfrentando desafios éticos. PLOS Medicine, v. 15, n. 11, p. e1002689, 2018. DOI: https://doi.org/10.1371/journal.pmed.1002689. Disponível em: https://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1002689&type=printable. Acesso em: 07 de abril de 2025.

VICKERS, Andrew J.; ELKIN, Elena B. Análise de curva de decisão: um novo método para avaliar modelos de previsão. Medical Decision Making, v. 26, n. 6, p. 565-574, 2006. Disponível em: https://journals.sagepub.com/doi/pdf/10.1177/0272989X06295361?casa_token=yBJDmhOwEEoAAAAA:erjn8k7uPiSlEERMwr9lSXsiyqf91S-lYiZCc1RV9zl__v6g-AkqcXZ1oXwmDULFXdmyBN_x8mE. Acesso em: 06 de abril de 2024.

VLADECK, David C. Máquinas sem responsáveis: regras de responsabilidade civil e inteligência artificial. Washington Law Review, v. 89, n. 1, p. 117-150, 2014. Disponível em: https://digitalcommons.law.uw.edu/cgi/viewcontent.cgi?article=4800&context=wlr. Acesso em: 12 de abril de 2025.

VOKINGER, Kerstin N.; GASSER, Urs. Regulando a IA na medicina nos Estados Unidos e na Europa. *Nature Machine Intelligence*, v. 3, n. 9, p. 738-739, 2021. Disponível em: https://pubmed.ncbi.nlm.nih.gov/34604702/. Acesso em: 12 de abril de 2025.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Por que um direito à explicação de decisões automatizadas não existe no RGPD. SSRN Electronic Journal, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2903469. Acesso em: 12 de abril de 2025.

WANG, Yichuan; KUNG, LeeAnn; BYRD, Terry A. Análise de *Big Data*: entendendo suas capacidades e potenciais benefícios para organizações de saúde. Previsão Tecnológica e Mudança Social, v. 126, p. 3-13, 2018. Disponível em: https://www.ehidc.org/sites/default/files/resources/files/big%20data%20analytics.pdf. Acesso em: 05 de abril de 2025.

WARREN, Samuel D.; BRANDEIS, Louis D. O direito à privacidade. Harvard Law Review, v. 4, n. 5, p. 193-220, 1890. Disponível em: https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf. Acesso em: 08 de abril de 2025.

WATSON, David S. et al. Aplicações clínicas de algoritmos de aprendizado de máquina: para além da caixa-preta. Oxford Internet Institute, 2019. Preprint. Disponível em: https://philpapers.org/archive/WATCAO-2.pdf. Acesso em: 07 de abril de 2025.

WIENS, Jenna et al. Não causar danos: um roteiro para aprendizado de máquina responsável na área da saúde. Nature Medicine, v. 25, n. 9, p. 1337-1340, 2019. Disponível em: https://www.nature.com/articles/s41591-019-0548-6. Acesso em: 06 de abril de 2025.

WYNANTS, Laure et al. Três mitos sobre limites de risco para modelos de previsão. BMC Medicine, v. 17, n. 192, p. 1-7, 2019. Disponível em: . https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-019-1425-3. Acesso em: 06 de abril de 2025.

YANG, Qiang et al. Aprendizado federado: conceito e aplicações. Transações da ACM em Sistemas e Tecnologia Inteligentes, v. 10, n. 2, art. 12, p. 1-19, fev. 2019. DOI: 10.1145/3298981. Disponível em: https://arxiv.org/pdf/1902.04885. Acesso em: 09 de abril de 2025.

YU, KH., Beam, A.L. & Kohane, I.S. Inteligência artificial na saúde. Nat Biomed Eng 2, 719-731 (2018). Disponível em: https://www.proquest.com/docview/2389685397?sourcetype=Scholarly%20Journals. Acesso em: 05 de abril de 2025.

YU, Kun-Hsing; KOHANE, Isaac S. Enquadrando os desafios da inteligência artificial na medicina. BMJ Quality & Safety, v. 28, n. 3, p. 238-241, 2019. Disponível em: https://qualitysafety.bmj.com/content/qhc/28/3/238.full.pdf?casa_token=BBgOfAYJmJAAAAAA:fTAPOCVd4GKRtA70fO9YD2NbrVj2gzgkZQXxPLqnADcC_-sdB_37JUGQL09JSrcZrT0wt8SNrw. Acesso em: 06 de abril de 2025.

**APÊNDICE**

| | PROJETO/ARTIGO | PUBLICAÇÃO |
|---|---|---|
| A) | Centro de estudo e pesquisa em direito, Inteligência artificial e saúde (CEPEDIAS). | Esta proposta criar, desenvolver e ampliar um Centro de Estudo e Pesquisa em Direito e Inteligência Artificial aplicada à Saúde (CEPEDIAS) na Universidade Santa Cecília (UNISANTA), com a colaboração de diversos profissionais, docentes e estudiosos de diversas áreas de conhecimento e em especial, das áreas jurídica, saúde e biotecnologias. É objetivo principal do CEPEDIAS desenvolver estudos e pesquisas que buscam a integração de metodologias avançadas de IA para desenvolver soluções inovadoras para a saúde, analisando as possibilidades, legalidade e aplicabilidade jurídica dessas metodologias no mundo, e também, no ordenamento jurídico brasileiro. Esses estudos e pesquisas abordarão desafios críticos, como diagnóstico precoce e preciso de doenças complexas, planos de tratamento personalizados, monitoramento remoto de pacientes e otimização dos sistemas de gestão de saúde. Ao enfatizar os estudos e as pesquisas desenvolvidas pelo CEPEDIAS, pretende-se verificar os limites da utilização da IA, e os seus impactos no atendimento ao paciente e a gestão da saúde, melhorando a eficiência, a precisão e os resultados. O CEPEDIAS, além dos objetivos acima descritos, também priorizará a educação e a disseminação do conhecimento, treinando uma nova geração de profissionais em IA e saúde, além de promover a transferência de tecnologia para garantir que nossa pesquisa tenha um impacto social amplo. Situação: Em andamento Natureza: Projetos de desenvolvimento tecnológico. Alunos envolvidos: Mestrado acadêmico (1); Integrantes: Klauss Carvalho De Malta; Daniela Rezende de Oliveira; Marcelo Lamy (Responsável) |
| B) | A Crise do Estado de Bem-Estar Social - Impactos na Seguridade e Saúde Pública | Unisanta Law and Social Science. v.13, p.399 - 409, 2024. |

| | | |
|---|---|---|
| | no Brasil. | |
| C) | Advances and Risks for artificial intelligence in healthcare. | Revista de Direito da Saúde Global. v.2, p.153-180 - 180, 2024 doi:10.5281/zenodo.13119081 |
| D) | Avanços na saúde e justiça em atenção ao sistema penitenciário brasileiro: desafios, inovações e o caminho para a humanização. | Revista Direitos Humanos & Sociedade. v.7, p.91 - 10, 2024. |
| E) | Modelos de Análise de Políticas Públicas: uma revisão crítica. | Unisanta Law and Social Science. v.13, p.271 - 284, 2024. |
| F) | Impacto da Análise de *Big Data* na Área da Saúde | Anais do XIII Encontro Nacional de Pós-graduação (2024). v.8, p.310 - 315, 2024. Referências adicionais: Português. Meio de divulgação: Meio digital. Home page: [https://periodicos.unisanta.br/ENPG/article/view/2236/2229] |
| G) | Orientações da OMS sobre Ética e Governança da Inteligência Artificial. | Unisanta Law and Social Science. v.13, p.285 - 294, 2024. Referências adicionais: Português. Meio de divulgação: Meio digital. Home page: [https://ojs.unisanta.br/LSS/article/view/2507] [doi:10.5281/zenodo.14479603] |
| H) | Avanços e riscos da inteligência artificial na atenção à saúde. | Unisanta Law and Social Science. v.12, p.108 - 119, 2023. Palavras-chave: Discriminação Social, Equidade, Autonomia Pessoal, Responsabilidade Civil, Políticas de Controle Social. Áreas do conhecimento: Direito Referências adicionais: Português. Meio de divulgação: Meio digital. Home page: [https://periodicos.unisanta.br/index.php/lss/article/view/3785] [doi:10.5281/zenodo.13119180] |
| I) | Los retos de la Autonomía del paciente frente a la inteligencia Artificial. | CADERNOS IBERO-AMERICANOS DE DIREITO SANITÁRIO. , 2024. Palavras-chave: Autonomia Pessoal, bem-estar, aplicações clínicas da IA, avanços da saúde, *Big Data* Áreas do conhecimento: Direito da Saúde, *Big Data* Setores de atividade: Educação Referências adicionais: Espanhol. TRABALHO PREMIADO EM 3º LUGAR NA MODALIDADE - PREMIO A LAS MEJORES COMUNICACIONES ORALES |

| | | |
|---|---|---|
| | | MODALIDAD PRESENCIAL. Tendo em vista a aprovação e apresentação do trabalho no I Congreso Internacional de Derecho Sanitario, entramos em contato para convidá-lo(a) a publicar o inteiro teor de sua pesquisa, no formato de artigo científico, no periódico Cadernos Ibero-Americanos de Direito Sanitário, que é editado e publicado pelo Programa de Direito Sanitário da Fundação Oswaldo Cruz de Brasília (Prodisa/Fiocruz Brasília). O artigo trata dos desafios da autonomia do paciente diante da inteligência artificial. |
| J) | Regulação e desafios legais da inteligência artificial na medicina. | Unisanta Law and Social Science. , 2024. Palavras-chave: Inteligência Artificial na Medicina, Governança e Ética em IA, Proteção de Dados na Saúde, Transparência e Responsabilidade, Regulamentação da IA Áreas do conhecimento: Direito da Saúde Referências adicionais: Português. Este artigo explora os desafios e as necessidades de regulamentação da inteligência artificial (IA) na medicina, um campo onde a IA está transformando diagnósticos, tratamentos e monitoramento de pacientes. A revolução tecnológica traz avanços significativos, mas também levanta preocupações éticas e jurídicas, especialmente no que se refere à proteção de dados pessoais e à privacidade dos pacientes. O artigo enfatiza a importância de regulamentações como a GDPR europeia e a LGPD brasileira, que impõem padrões rigorosos de segurança e privacidade, sendo essenciais para evitar a exposição e o uso indevido de dados sensíveis em contextos de saúde. |
| K) | Federal Food and Drug Administration (FDA) Guidance on Cybersecurity in Medical Devices. | In: DIREITO DA SAÚDE COMPARADO EM TEMAS SENSÍVEIS E-HEALTH - ABORTO – EUTANÁSIA – AGÊNCIAS REGULADORAS - LICENÇA MÉDICA – AUTONOMIA DO MÉDICO - CYBERSECURITY - TRATADOS, ed.1. Nova York: Lawinter Editions, 2024, v.1, p. 7 - 136. Palavras-chave: FDA, CYBERSECURITY IN MEDICAL DEVICES, FEDERAL FOOD AND DRUG ADMINISTRATION Áreas do conhecimento: Direito Internacional Público Referências adicionais: Brasil/Português. Meio de divulgação: Meio digital, ISBN: 9783039270408, Home page: https://www.amazon.com/dp/3039270400 In the contemporary scenario, the intersection between technology and health is marked by advances extraordinary, providing innovations that positively impact medical treatment. However, this progress also brings with it considerable challenges, with cybersecurity emerging as a critical concern. This work aims to explore the intersection between |

| | | |
|---|---|---|
| | | cybersecurity and medical devices, anchored in an alarming event that highlights potential fatal consequences of this scenario. |
| L) | A violação do Princípio da autonomia do usuário do serviço de Saúde como consequência da disseminação de Fake news durante a pandemia do coronavírus. | In: Cultura, mídia e persuasão: discurso midiático durante a Pandemia do Coronavírus, ed.1. Curitiba: Bagai, 2020, v.1, p. 7 - 16. Palavras-chave: Cultura, Mídia, Pandemia do coronavírus Áreas do conhecimento: Direito Constitucional,Saúde Pública Setores de atividade: Educação Referências adicionais: Brasil/Português. Meio de divulgação: Meio digital, ISBN: 9786587204475, Home page: https://editorabagai.com.br/product/cultura-midia-e-persuasaodiscurso-midiatico-durante-a-pandemia-do-coronavirus/. |
| M) | Algumas aplicação e desafios da iA na saúde. | (Seminário,Apresentação de Trabalho) Palavras-chave: desafios, Inteligência Artificial, Saúde Áreas do conhecimento: Direito Internacional Público Setores de atividade: Educação Referências adicionais: Brasil/Português. Meio de divulgação: Impresso. |
| N) | Directrices para los protocolos de ensayos clínicos. | (Simpósio,Apresentação de Trabalho) Palavras-chave: Ensaios, Saúde Áreas do conhecimento: Direito Internacional Público Setores de atividade: Educação Referências adicionais: Brasil/Espanhol. Meio de divulgação: Impresso |
| O) | Inteligência Artificial na Saúde - Promovendo o Bem-Estar e a Segurança Humana. | Palavras-chave: Inteligência Artificial, Saúde, bem-estar, segurança, ética. Áreas do conhecimento: Direito da Saúde. Setores de atividade: Educação Referências adicionais: Brasil/Português. Meio de divulgação: Vários. Home page: https://periodicos.unisanta.br/index.php/lss/article/view/3785 |
| P) | Orientação da OMS sobre Ética e Governança. | (Congresso, Apresentação de Trabalho) Palavras-chave: Saúde, OMS, Orientações Áreas do conhecimento: Direito Constitucional Setores de atividade: Educação Referências adicionais: Brasil/Português. Meio de divulgação: Outro. Home page: https://ojs.unisanta.br/LSS/article/view/2507 |