Avanços e riscos da inteligência artificial na atenção à saúde[*]

Avances y riesgos de la inteligencia artificial en la atención sanitaria

Advances and risks for artificial intelligence in healthcare

Marcelo Lamy
Doutor, PPGD em Direito da Saúde (UNISANTA)

Klauss Carvalho de Malta
Mestrando, PPGD em Direito da Saúde (UNISANTA)

RESUMO: Contextualização: A inteligência artificial (IA) tem experimentado um progresso notável, tornando-se uma ferramenta indispensável para resolver uma ampla gama de desafios tecnológicos e econômicos. **Problema**: Este artigo explora o estado atual da IA, com foco especial nas técnicas de aprendizado de máquina, incluindo redes neurais artificiais, discute o estado atual dessas áreas, os desafios que elas enfrentam e as oportunidades de pesquisa que se apresentam; além disso, destaca preocupações relacionadas aos impactos sociais e questões éticas da IA. **Objetivos**: apresentar problemas e questões relacionados à IA, com a avaliação do impacto na sociedade e apresentação de possíveis formas para mitigar os riscos da IA. **Métodos**: ancorado em textos científicos e documentos fez-se revisão crítico-narrativa. **Resultados**: A IA é uma tecnologia promissora com o potencial de gerar benefícios significativos para a sociedade. No entanto, a IA também apresenta riscos e desafios que precisam ser considerados. Os estudos sobre os impactos sociais e questões éticas da IA são importantes para garantir que essa tecnologia seja desenvolvida e usada de forma responsável. **Conclusões**: A IA é uma tecnologia com potencial de impacto positivo, mas também apresenta riscos que precisam ser mitigados. É importante que os estudos sobre os impactos sociais e questões éticas da IA sejam realizados de forma contínua e que enfrentem os seguintes aspectos: (1) técnicos: a) preconceito e discriminação; b) falta de transparência; c) falta de controle; (2) sociais: d) perda de empregos; e) desigualdade social; f) impactos ambientais; (3) éticos: g) autonomia; h) responsabilidade; i) privacidade.

PALAVRAS-CHAVE: Discriminação Social; Equidade; Autonomia Pessoal; Responsabilidade Civil; Políticas de Controle Social.

RESUMEN: ***Contextualización***: *La Inteligencia Artificial (IA) ha experimentado un progreso notable, convirtiéndose en una herramienta indispensable para resolver una amplia gama de desafíos tecnológicos y económicos.* ***Problema***: *este artículo explora el estado actual de la IA, con especial atención a las técnicas de aprendizaje automático, incluidas las redes neuronales artificiales, analiza el estado actual de estas áreas, los retos a los que se enfrentan y las oportunidades de investigación que se presentan; además, destaca las preocupaciones relacionadas con las repercusiones sociales y las cuestiones éticas de la IA.* ***Objetivos***: *presentar los problemas y cuestiones relacionados con la IA, evaluar su impacto en la sociedad y presentar posibles formas de mitigar los riesgos de la IA.* ***Método***: *se realizó una revisión crítico-narrativa a partir de textos y documentos científicos.* ***Resultados***: *La IA es una tecnología prometedora con potencial para generar importantes beneficios para la sociedad. Sin embargo, la IA también presenta riesgos y retos que deben tenerse en cuenta. Los estudios sobre las repercusiones sociales y las cuestiones éticas de la IA son importantes para garantizar que esta tecnología se desarrolle y utilice de forma responsable.* ***Conclusiones***: *la IA es una tecnología con potencial para tener un impacto positivo, pero también presenta riesgos que hay que mitigar. Es importante que los estudios sobre los impactos sociales y las cuestiones éticas de la IA se lleven a cabo de forma continuada y que aborden los siguientes aspectos: (1) técnicos: a) prejuicios y discriminación; b) falta de transparencia; c) falta de control; (2) sociales: d) pérdida de puestos de trabajo; e) desigualdad social; f) impactos medioambientales; (3) éticos: g) autonomía; h) responsabilidad; i) privacidad.*

PALABRAS CLAVE: *Discriminación Social; Equidad; Autonomía Personal; Responsabilidad Civil; Políticas de Control Social.*

[*] Esse trabalho foi apresentado originalmente no V Congresso Internacional de Direito da Saúde, realizado em 19, 20 e 21 de outubro de 2023 na Universidade Santa Cecília (UNISANTA). Em função da recomendação de publicação da Comissão Científica do Congresso, fez-se a presente versão.

***ABSTRACT**: **Context**: Artificial Intelligence (AI) has experienced remarkable progress, becoming an indispensable tool for solving a wide range of technological and economic challenges. **Problem**: This article explores the current state of AI, with a special focus on machine learning techniques, including artificial neural networks, discusses the current state of these areas, the challenges they face and the research opportunities that present themselves; in addition, it highlights concerns related to the social impacts and ethical issues of AI. **Objectives**: to present problems and issues related to AI, assessing the impact on society and presenting possible ways to mitigate the risks of AI. **Methods**: A critical-narrative review was carried out based on scientific texts and documents. **Results**: AI is a promising technology with the potential to generate significant benefits for society. However, AI also presents risks and challenges that need to be considered. Studies into the social impacts and ethical issues of AI are important to ensure that this technology is developed and used responsibly. **Conclusions**: AI is a technology with the potential for positive impact, but it also presents risks that need to be mitigated. It is important that studies into the social impacts and ethical issues of AI are carried out on an ongoing basis and that they tackle the following aspects: (1) technical: a) prejudice and discrimination; b) lack of transparency; c) lack of control; (2) social: d) loss of jobs; e) social inequality; f) environmental impacts; (3) ethical: g) autonomy; h) responsibility; i) privacy.*

***KEYWORDS**: Social Discrimination; Equity; Personal Autonomy; Damage Liability; Social Control Policies.*

Introdução

Nos últimos anos, a ciência testemunhou um notável avanço na aplicação da inteligência artificial (IA) na área da saúde, marcando uma revolução na medicina e nos cuidados de saúde em todo o mundo.

Esta transformação é impulsionada pelo progresso vertiginoso das tecnologias de informação, das quais a IA emerge como uma força motriz. Ao contrário dos primórdios da IA, em que as máquinas eram restritas a tarefas repetitivas e comandos específicos, a IA moderna abrange uma gama impressionante de funções cognitivas, englobando aspectos daquilo que tradicionalmente denominamos como inteligência humana.

Swaminathan (OMS, 2021, p.5) reconhece o potencial transformador da IA na saúde, afirmando que ela representa uma promissora oportunidade para elevar os padrões de cuidados médicos em escala global.

A IA não apenas promete aprimorar a rapidez e precisão dos diagnósticos e triagens de doenças, mas também oferece suporte nos serviços clínicos e fortalece pesquisas médicas, contribuindo ainda para avanços significativos no desenvolvimento de medicamentos.

Contudo, o uso crescente da IA na saúde não é desprovido de desafios. A coleta e o tratamento de dados sigilosos de pacientes levantam preocupações éticas e de privacidade, demandando medidas rigorosas de proteção. Além disso, o entusiasmo em torno do potencial da IA pode conduzir a abordagens excessivamente tecnológicas para problemas complexos de saúde, desafiando a busca pela equidade no acesso às inovações médicas.

Questões de viés racial e preconceito também surgem quando dados de baixa qualidade ou inadequados são utilizados na IA.

Este artigo explora a ascensão da IA na medicina, destacando seus impactos positivos e desafios éticos. Examina, ainda, a evolução das redes neurais profundas e suas aplicações na saúde, assim como a importância de dados de alta qualidade e a seleção criteriosa de algoritmos.

Aborda também a necessidade de uma distribuição equitativa das tecnologias de saúde baseadas em IA e a importância de manter sistemas atualizados e eficazes. Neste contexto, analisa-se como a IA está redefinindo os diagnósticos, tratamentos e pesquisas médicas, com um foco crítico nas questões éticas e sociais que envolvem esse avanço tecnológico.

No tocante à legislação, a iminente revolução da inteligência artificial (IA) na área da saúde está redefinindo paradigmas, com a implementação da IA, em especial a abordagem de aprendizagem profunda (*deep learning*), desempenhando um papel proeminente. A supervisão da IA na saúde tornou-se crucial, demandando a promoção de inovações tecnológicas e a criação de um arcabouço regulatório sólido para garantir seu uso em prol do bem-estar humano.

O Projeto de Lei (PL) n. 2.338/2023 no Brasil, em consonância com a Lei Geral de Proteção de Dados (LGPD), propõe uma abordagem principiológica abrangente, visando regulamentar a IA na saúde. Esses princípios englobam questões de privacidade, ética, transparência e responsabilidade, refletindo os debates internacionais em andamento.

O PL também trata também de sistemas de alto risco, dos *sandboxes* (ambiente de teste utilizado por programadores e desenvolvedores para testar de modo isolado os novos programas, aplicativos e plataformas com segurança, sem o risco de interferirem ou danificarem qualquer ambiente, sem ramificações no mundo real) e da supervisão exercida pela Autoridade Nacional de Proteção de Dados (ANPD), representando um avanço significativo na regulação da IA no país, com ênfase no desenvolvimento responsável dessa tecnologia.

Valendo-se da coleta de textos científicos (realizada nos portais Scielo, na Biblioteca Virtual da Saúde e no portal da Organização Mundial da Saúde) e de documentos (levantados no portal de legislação do Planalto brasileiro), fez-se revisão crítico-narrativa de viés qualitativo (Lamy, 2020, pp. 337-340).

1 Conceitos técnicos estruturantes

Antes de ingressar no âmbito específico da saúde, é relevante estabelecer alguns conceitos técnicos estruturantes que permitem uma discussão mais aprofunda sobre a temática da presente investigação.

1.1 Aprendizado de máquina

O objetivo do aprendizado de máquina (*machine learning*) é treinar programas por meio de exemplos. Com uma quantidade substancial de exemplos (*input*) a máquina constrói hipóteses (*output*), derivadas desses dados (Mitchell, 1997).

O aprendizado de máquina é fruto de uma área da ciência da computação que se concentra no desenvolvimento de algoritmos (sequência de passos lógicos e estruturados para a realização de uma tarefa) que permitem esse aprendizado e se adaptam a partir disso. Os algoritmos de aprendizado de máquina são projetados para encontrar padrões nos dados e usá-los para fazer previsões, sugerir ou tomar decisões.

A inferência indutiva é um dos principais métodos utilizados no aprendizado de máquina. Trata-se de um processo de generalização de padrões observados em um conjunto de dados. A precisão da inferência indutiva depende da quantidade e da qualidade dos dados. Mais dados e dados mais precisos levam a generalizações mais precisas.

Existem três categorias principais de aprendizado de máquina: supervisionado, não supervisionado e por reforço (Honda *et al.*, 2017).

No aprendizado supervisionado – abordagem mais comumente empregada (Sas, 2023) – o algoritmo é fornecido com um conjunto de dados de treinamento que inclui exemplos de objetos e suas classificações desejadas. O objetivo do algoritmo é permitir o processo

automatizado de identificar objetos e suas classificações de exemplos que ainda não foram rotulados.

No aprendizado não supervisionado, os exemplos são apresentados ao algoritmo sem serem rotulados. O algoritmo, então, analisa as semelhanças entre os atributos dos exemplos e os agrupa com base nessas semelhanças. A partir dessa análise, o algoritmo tenta identificar se existem agrupamentos ou *clusters* naturais entre os exemplos. Após a identificação dos agrupamentos, é comum conduzir uma análise adicional para compreender o significado de cada grupo dentro do contexto do problema em questão.

No aprendizado por reforço, o algoritmo não recebe a resposta correta, mas é orientado por sinais de reforço, recompensa ou punição. O algoritmo formula hipóteses com base nos exemplos e avalia se essas hipóteses foram bem-sucedidas ou não com base nos sinais de reforço. O aprendizado por reforço encontra aplicação significativa em campos como jogos e robótica, sendo amplamente utilizado nesses contextos (Honda *et al*, 2017).

Resolver problemas por meio do aprendizado de máquina nem sempre é uma tarefa simples e requer determinados requisitos. Primeiramente, é essencial dispor de um conjunto sólido de exemplos, frequentemente exigindo a construção e a manutenção contínua desse conjunto, pois os dados nem sempre são ideais. É crucial empregar técnicas para aprimorar a qualidade dos dados, quando necessário.

Além disso, vale destacar que nem todos os algoritmos de aprendizado de máquina são adequados para resolver todos os tipos de problemas. Portanto, é importante realizar uma seleção criteriosa dos algoritmos apropriados para cada problema em questão.

Após o treinamento, é crucial avaliar se o algoritmo está efetivamente resolvendo o problema e com que grau de precisão. Por fim, é importante ressaltar que os sistemas de aprendizado de máquina devem ser atualizados regularmente, uma vez que mudanças nos dados podem afetar o desempenho e a eficácia do sistema.

O aprendizado da máquina depende também de dois grandes desafios: (a) fazer com que as máquinas entendam a linguagem humana da mesma forma que os humanos a compreendem (processamento de linguagem natural, em inglês: *natural language processing* – NLP), (b) permitir que as máquinas realizem tarefas que normalmente requerem a visão humana, como o reconhecimento de objetos, detecção de movimentos, análise de expressões faciais (visão computacional – em inglês: *computer vision*).

1.2 Redes neurais artificiais

As redes neurais artificiais (RNA) – modelos matemáticos inspirados na estrutura das redes neurais biológicas – permitem um aprendizado de máquina para solucionar concomitante diversos problemas. Em razão disso, permitem um aprendizado profundo (*deep learning*).

O *perceptron*, um dos modelos mais básicos de aprendizado em redes neurais, foi concebido por Frank Rosenblatt em 1957 para resolver problemas de separação linear simples. Trata-se de um modelo de aprendizado simples composto por uma única camada de neurônios McCulloch-Pitts - MCP (Mcculloch et al, 1943).

Esse modelo utiliza uma regra de aprendizado baseada na correção de erros, ou seja, na diferença entre a resposta desejada (inserida pelo homem) e a resposta da rede (decorrente da aplicação do algoritmo).

O processo envolve duas fases: treinamento e teste. Durante a fase de treinamento, a rede é apresentada a um conjunto de dados rotulados, ou seja, um conjunto de dados que contém exemplos de entrada e saída esperados. Os parâmetros da rede, como os pesos, são ajustados a cada novo exemplo, de modo que a rede aprenda a gerar a resposta esperada para cada entrada (Rosenblatt,1957). Após o ajuste dos parâmetros, a rede é avaliada com um conjunto de dados diferente, não rotulado. O desempenho da rede é avaliado com base na sua capacidade de gerar respostas corretas para os exemplos de teste.

Redes neurais *perceptron* simples, como o *perceptron* original, só podem resolver problemas lineares. Para problemas mais complexos, são necessárias redes neurais com mais de uma camada de neurônios. Essas redes são chamadas de redes neurais multicamadas (*multi layer perceptron* – MLP, do inglês). O método mais comum para treinar MLPs é o algoritmo *backpropagation*, que ajusta os pesos das conexões entre os neurônios de forma a minimizar o erro entre a saída da rede e o valor desejado (Rumelhart; Hinton; Williams, 1986).

Outras técnicas, como as máquinas de vetores de suporte - *support vectors machine* - SVM, do inglês (Cortes et al., 1995) – passaram a demonstrar resultados superiores às MLPs em várias aplicações, como o reconhecimento de imagens. No entanto, o cenário mudou (Bengio et al., 2015) com o surgimento das redes neurais profundas (*deep neural network* – DNN, do inglês).

As DNNs são MLPs com múltiplas camadas ocultas, o que lhes permite aprender representações mais complexas de dados. As DNNs revolucionaram o campo do Aprendizado de Máquina, alcançando resultados superiores em uma ampla gama de tarefas, incluindo reconhecimento de imagens, Processamento de Linguagem Natural (PLN) e tradução automática.

As redes neurais profundas (DNNs) foram inspiradas na capacidade do cérebro humano de detectar características locais e realizar seleção de informações relevantes. Atualmente, elas são a técnica de aprendizado de máquina mais eficaz para a resolução de problemas.

1.3 Inteligência artificial

Partindo do conceito de que inteligência artificial (IA) é a habilidade de um ente não natural avaliar informações recebidas e extrair inferências a partir dessas informações, Boldissera (2023) afirma que a inteligência artificial pode ser classificada em três categorias: IA especializada, IA geral e IA superinteligente.

A IA especializada, também conhecida como IA limitada ou fraca, é moldada por algoritmos altamente especializados em domínios específicos. Esses sistemas armazenam grandes volumes de dados e aplicam algoritmos para executar tarefas complexas, mas limitadas ao escopo para o qual foram desenvolvidos. Nessa categoria, que se encontra muito desenvolvida, estão incluídos os sistemas especialistas e os sistemas de recomendação (Granatyr, 2017).

A IA geral, ou IA forte, é estruturada por algoritmos com capacidades semelhantes às humanas em uma ampla variedade de tarefas, como as que se almejam no processamento da linguagem natural e na visão computacional. Segundo Boldissera (2023), esse é o estágio que a ciência concentra seus esforços atuais.

A IA superinteligente depende de algoritmos que superem significativamente os seres humanos em praticamente todas as tarefas. Ainda não existem sistemas de IA com esse nível de capacidade e permanece incerto se futuramente serão desenvolvidos sistemas mais inteligentes do que os humanos (Trueman, 2013).

As soluções de sucesso usando IA têm em comum a capacidade de resolver problemas complexos de forma eficiente e eficaz. Essas soluções geralmente são baseadas em técnicas de aprendizado de máquina, que permitem que os sistemas aprendam com dados e se adaptem a novas situações.

2 Inteligência artificial na saúde

Atualmente testemunha-se uma nova era de transformação industrial, impulsionada pelo avanço de tecnologias da informação, incluindo a inteligência artificial (IA). Nesse contexto, as máquinas não se limitam mais a tarefas repetitivas, mas também desempenham funções cognitivas, envolvendo o uso do que tradicionalmente consideramos como inteligência.

Swaminathan (OMS, 2021, p.5) afirmou no primeiro relatório global para orientação e uso da IA: a inteligência artificial (IA) representa uma promissora oportunidade para aprimorar os cuidados de saúde em escala global; ela tem o potencial de aprimorar a rapidez e a precisão dos diagnósticos e a triagem de doenças, prestar apoio nos serviços clínicos e fortalecer as investigações em saúde, além de contribuir para o avanço na pesquisa de medicamentos.

A IA é uma tecnologia cada vez mais explorada na área da saúde. O avanço da medicina do futuro é foco de pesquisa em diversos países e o Brasil tem avançado na implantação de políticas públicas de saúde que utilizam IA (Lemes, 2020, p.166).

Segundo Torres (2023), a IA tem o potencial de ser uma ferramenta poderosa para o combate às doenças tropicais, pois a detecção precoce de surtos é uma das aplicações da IA na saúde pública. Ela pode ser utilizada para analisar grandes conjuntos de dados de pacientes, incluindo dados clínicos, dados ambientais e dados de mídia social. Isso pode ajudar a identificar padrões e tendências que podem indicar um surto em andamento.

Em áreas com recursos e infraestrutura limitados, essa tecnologia pode ser uma solução viável para melhorar a vigilância de doenças. Pode ser utilizada para automatizar tarefas, como a classificação de casos suspeitos, o rastreamento de contatos e a análise de dados epidemiológicos. Isso pode liberar recursos humanos para outras atividades, como o atendimento médico e a educação em saúde (OMS, 2021, p.6).

Além da detecção precoce de surtos, as redes neurais podem melhorar o diagnóstico e o tratamento de doenças tropicais, auxiliar no desenvolvimento de novos medicamentos e vacinas e ampliar as respostas a emergências de saúde pública (SBMT, 2023).

A automação vem assumindo certo protagonismo na área de diagnóstico, com avanços notáveis. Atualmente, existem sistemas de diagnóstico automático que com precisão e acurácia excepcionais, ocasionalmente superando até mesmo os diagnósticos realizados por profissionais de saúde. Um exemplo notável é o robô desenvolvido pela empresa iFlytek, que foi aprovado no exame nacional de licenciamento de médicos da China, demonstrando a capacidade dessas tecnologias em oferecer diagnósticos confiáveis e de alta qualidade (Costa, 2019).

A evolução da IA está revolucionando a medicina, aprimorando diagnósticos e acelerando a pesquisa médica. Com bastante otimismo, Swaminathan (OMS, 2021, p. 5)

destaca seu potencial global na melhoria da saúde. No entanto, seu uso suscita questões éticas e de segurança de dados.

Apesar dos avanços, o uso da IA na saúde levanta preocupações éticas, legais, comerciais e sociais transnacionais. Muitas dessas preocupações não são exclusivas da IA. O uso de software e computação na área da saúde tem desafiado desenvolvedores, governos e prestadores de serviços de saúde por meio século, e a IA apresenta desafios éticos adicionais e novos que vão além da competência dos reguladores tradicionais e dos participantes nos sistemas de saúde.

Desafios éticos devem ser adequadamente abordados se a IA for amplamente utilizada para melhorar a saúde humana, notadamente os de preservar a autonomia humana e de garantir acesso equitativo a essas tecnologias (OMS, 2021, p.25).

O otimismo desenfreado em relação aos benefícios potenciais da IA não pode agravar a distribuição desigual de acesso a tecnologias de saúde dentro e entre países ricos e de baixa renda. A divisão digital poderia agravar o acesso desigual a tecnologias de saúde por geografia, gênero, idade ou disponibilidade de dispositivos, se os países não tomarem medidas apropriadas. Ademais, o uso inadequado da IA também pode perpetuar ou agravar preconceitos (OPAS, 2021).

O uso de dados limitados, de baixa qualidade e não representativos na IA pode aprofundar preconceitos e disparidades na assistência médica. Inferências tendenciosas, análises de dados enganosas e aplicativos e ferramentas de saúde mal projetados podem ser prejudiciais. Algoritmos preditivos baseados em dados inadequados ou inapropriados podem resultar em viés racial ou étnico significativo (Leme, 2020, p. 117).

2.1 IA na medicina

A IA na medicina tem como principal foco a criação de programas que possam realizar diagnósticos e oferecer recomendações terapêuticas.

Coiera (2014) explicou há tempos que programas ou sistemas especialistas (SE), baseados em modelos cognitivos das doenças, por suas conexões com os dados do paciente e os sintomas clínicos, consideram a tomada de decisão médica uma decorrência de um raciocínio controlável.

Os SE podem ser utilizados para uma variedade de aplicações médicas, incluindo diagnóstico, prognóstico e planejamento terapêutico. Eles têm o potencial de melhorar a qualidade do cuidado médico, tornando-o mais preciso, eficiente e acessível. (Widman,1998)

Os softwares utilizados na área da medicina, muitos deles dotados de inteligência artificial, são frutos de uma trajetória evolutiva que remonta aos primórdios do desenvolvimento dos sistemas especialistas (SE). Um marco importante nesse processo foi o MYCIN, um sistema desenvolvido na Universidade de Stanford nos em 1970 que se destacou por sua capacidade de realizar diagnósticos e recomendar tratamentos médicos com base em informações até mesmo incertas ou incompletas.

A inteligência artificial (IA) e a robótica têm se tornado ferramentas cada vez mais importantes na medicina. No entanto, a adoção desses recursos por parte dos profissionais de saúde ainda enfrenta obstáculos razoáveis, como a resistência dos médicos a perder o controle do processo de tomada de decisão, a desconfiança na própria capacidade das máquinas e a dificuldade de aprendizado dos sistemas.

Um exemplo de sistema médico de apoio à decisão (*sequence-aware diffusion model for longitudinal medical image generation* - SADM) é o DXplain, desenvolvido no Laboratório de ciências informáticas do Hospital Geral de Massachusetts, que auxilia no diagnóstico por meio da análise de dados clínicos. Ele gera uma lista dos diagnósticos mais prováveis e justifica cada um deles, quando necessário. Outro exemplo de SADM é o software Risco Cirúrgico MED, que avalia o risco cardiológico e pulmonar do paciente antes da cirurgia (Silva, 2013).

A IA e a robótica são utilizadas para melhorar o diagnóstico e o tratamento de doenças, bem como para reduzir as infecções hospitalares. O uso de robôs na medicina começou em 1985, com o PUMA 560, desenvolvido para guiar agulhas durante uma biópsia cerebral (Claudio, 2020).

O Hospital Albert Einstein, por exemplo, adotou a tecnologia robótica em 2008. Desde então, uma equipe de médicos especializados em cirurgias robóticas tem utilizado o Vinci Surgical System, um robô cirurgião desenvolvido pela Intuitive Surgical (Claudio, 2020).

Os avanços da IA e da robótica na medicina ilustram como essas tecnologias têm o potencial de melhorar a qualidade do atendimento e a segurança dos pacientes.

A tecnologia emergente da IA está redefinindo paradigmas e é vista como uma das inovações mais significativas no campo da saúde. A implementação da IA na área médica, particularmente a abordagem de aprendizagem profunda (*deep learning*), é especialmente notável (Obermeyer & Emanuel, 2016). Espera-se que tenha um efeito notável nos próximos anos no atendimento médico, na administração de sistemas de saúde e na interação entre os pacientes e os serviços de saúde, capacitando os pacientes a gerenciarem seus próprios dados e melhorar sua saúde (Topol, 2019).

3 Desafios da regulação da IA na saúde

Diferentemente de produtos que seguem a regulamentação tradicional, como medicamentos e equipamentos médicos, os algoritmos estão sempre se desenvolvendo e despertam muitos novos problemas (Rajkomar et al., 2019).

Até agora, leis e regulamentos compreensivos de todos os novos problemas não se produziram, está avançada apenas a área da privacidade de dados, como se fez no Brasil com a LGPD (Lei Geral de Proteção de Dados), embora existam diversas propostas de regulação mais compreensivas, como é a proposta brasileira alinhavada pelo projeto de lei n. 2.338 de 2023 (Jobin; Ienca; Vayena, 2019). No âmbito internacional, o foco das discussões tem desaguado no estabelecimento de princípios regulatórios essenciais (Richman, 2018), como os de garantir as revisões de decisões automatizadas.

O projeto de lei (PL) n. 2.338/2023 apresenta um arcabouço principiológico (artigos 2 e 3) que trata da implementação e do uso de sistemas de inteligência artificial no Brasil:

> "a) Valorização da dignidade humana; b) Respeito aos direitos fundamentais e princípios democráticos; c) Liberdade para o desenvolvimento da individualidade; d) Proteção ao meio ambiente e incentivo à sustentabilidade; e) Igualdade, não discriminação, diversidade e respeito aos direitos trabalhistas; f) Estímulo ao progresso tecnológico e à inovação; g) Incentivo à livre iniciativa, concorrência leal e proteção ao consumidor; h) Proteção da privacidade, segurança de dados e controle de informações pessoais; i) Incentivo à pesquisa e desenvolvimento para promover a inovação nos setores produtivos e no setor público; j) Acesso à informação e educação sobre sistemas de inteligência artificial e suas aplicações; k) Crescimento inclusivo, desenvolvimento sustentável e bem-estar geral; l) Autodeterminação e liberdade de escolha; m) Participação ativa do ser humano no ciclo da inteligência artificial e supervisão humana eficaz; n) Luta contra a discriminação; o) Justiça, equidade e inclusão social; p) Transparência, clareza, compreensibilidade e auditabilidade; q) Confiança e robustez dos sistemas de inteligência

> artificial e segurança da informação; r) Garantia de devido processo legal, contestação e contraditório; s) Rastreabilidade das decisões ao longo do ciclo de vida dos sistemas de IA para responsabilização; t) Responsabilidade, prestação de contas e reparação total de danos; u) Responsabilização pelos sistemas de IA, responsabilidade por suas ações e reparação integral por eventuais prejuízos; x) Prevenção, precaução e atenuação de riscos sistêmicos decorrentes de usos intencionais ou não intencionais, bem como de efeitos não previstos dos sistemas de inteligência artificial; y) Não prejudicar e manter a proporção entre os métodos empregados e as finalidades legítimas dos sistemas de IA." (Senado Federal, 2023).

Apresentado pelo presidente do Senado, Rodrigo Pacheco, em 3 de maio de 2023, propõe a regulamentação do uso da inteligência artificial no Brasil. Esse PL está baseado no trabalho de uma comissão de juristas e especialistas que estudou o tema e realizou mais de 70 audiências públicas (STJ, 2023).

O Projeto segue o modelo regulatório do *AI Act*, aprovado pelo Parlamento Europeu em maio de 2023. Este regramento europeu preconiza a necessidade de supervisão humana, segurança, transparência e sustentabilidade no desenvolvimento e aplicação de tecnologias de IA (UE, 2023).

A Autoridade Nacional de Proteção de Dados (ANPD) publicou uma análise preliminar do Projeto de Lei n. 2.338/2023. O documento apresenta os pontos de convergência e conflito entre o PL e a LGPD, reforça o posicionamento da ANPD de fomento à inovação em IA, desde que feita de forma responsável (ANPD, 2023).

O PL e a LGPD compartilham o objetivo de proteger os direitos individuais e coletivos dos titulares de dados pessoais, em especial no que diz respeito à privacidade, à segurança e à transparência no tratamento de dados (ANPD, 2023). Além disso, tanto o PL quanto a LGPD estabelecem que o controlador é responsável pelo tratamento de dados pessoais e deve adotar medidas técnicas e administrativas para garantir a segurança e a proteção dos dados. Os dispositivos normativos do PL conferem ao titular de dados pessoais uma série de direitos, como o direito de acesso, de retificação, de eliminação, de portabilidade e de oposição ao tratamento de dados.

Além disso, o PL atribui para uma autoridade competente definida pelo Executivo (ANPD) a competência para fiscalizar e aplicar sanções administrativas em relação ao tratamento de dados pessoais, e demais atributos da inteligência artificial, nos termos do art. 32 (Senado Federal, 2023).

O PL estabelece a responsabilidade civil objetiva para os sistemas de alto risco, que são aqueles que apresentam um risco significativo para os direitos e liberdades fundamentais dos titulares de dados pessoais (arts. 16, 17 e 18 do PL 2.338/2023). O que constitui avanço significativo com relação à LGPD, que não se manifesta sobre esta matéria (BRASIL, 2023).

Merece destaque também a regulamentação da utilização de *sandboxes* pelo PL, dos ambientes controlados para o desenvolvimento e teste de novas tecnologias. Tópica que a LGPD também não regulou. (ANTT, 2023)

A ANPD recomenda que os pontos de conflito entre o PL e a LGPD sejam sanados, principalmente aqueles que dizem respeito às atribuições legais da ANPD. A Autoridade também destaca a importância de que o PL detalhe questões relativas à proteção de dados pessoais em *sandboxes*, em especial em sistemas de alto risco (ANPD 2023).

Na análise preliminar da ANPD, esse PL é um passo importante para a regulação da inteligência artificial no Brasil: "O PL nº 2338/2023, ao estabelecer princípios, diretrizes e instrumentos de governança para o desenvolvimento e uso responsável da inteligência artificial, representa um avanço significativo na regulação dessa tecnologia no Brasil" (ANPD, 2023).

Considerações finais

A IA é uma tecnologia cujo uso tem um significativo potencial positivo, mas que também apresenta uma série de riscos que precisam ser mitigados.

O uso ético e responsável depende do enfrentamento prático e regulatório dos seguintes aspectos, que sintetizamos:

(1) Aspectos técnicos:

a. Preconceito e discriminação (assegurar que o uso da tecnóloga não gere novas discriminações, por exemplo, restringindo a usabilidade das análises ancoradas em amostragens de agrupamentos não-diversificados);

b. Falta de transparência e de inteligibilidade (assegurar a publicação do desenho da inteligência utilizada, e que essa publicação seja compreensível para todos os destinatários, ou seja, desenvolvedores, provedores das informações, profissionais da saúde, gestores da saúde e usuários);

c. Falta de controle (verificar a qualidade dos *inputs* e dos métodos de análise empreendidos; observar se as respostas da tecnologia são adequadas e apropriadas às expectativas constitutivas; assegurar que as sugestões da máquina sejam reveladas apenas se passarem nos filtros da segurança, da precisão e da eficácia para usos definidos, fixando pontos de supervisão humana);

(2) Aspectos sociais:

a. Perda de empregos (verificar e controlar os efeitos adversos da IA que transcendem a questão médica, como os de dimensão trabalhista);

b. Desigualdade social (assegurar a disponibilização da tecnologia para todos os públicos);

c. Impactos ambientais (verificar e controlar os efeitos adversos da IA que transcendem a questão médica, como os de dimensão ambiental);

(3) Aspectos éticos:

a. Autonomia (assegurar que a decisão seja apenas sugerida pela máquina e sempre tomada pelo profissional da saúde em conjunto com o paciente; garantir que as pessoas que decidirão, em cada caso, recebam as informações que foram consideradas na sugestão e o método que pautou a análise e a sugestão da máquina);

b. Responsabilidade (assegurar que a responsabilidade pelos eventuais danos continue a ser humana e construir modelos de responsabilização adequados a realidade da IA, nos quais se fixe a responsabilidade dos provedores da tecnologia, dos sistemas de atenção que utilizam a tecnologia, dos profissionais que se valem das tecnologias);

c. Privacidade (assegurar que a utilização dos dados seja autorizada, restrita ao autorizado e anonimizada, assegurando o sigilo dos dados).

Referências

AGÊNCIA NACIONAL DE TRANSPORTE TERRESTRE (ANTT). Ambiente regulatório experimental. Brasil. 2023. disponível em: <https:// www.antt.gov.br/sandboxregulatorio>. Acesso em: 12 de outubro de 2023.

BENGIO, Yoshua; LECUN, Yann; HINTON, Geoffrey. Aprendizado Profundo. **Nature,** Reino Unido, 521:436-444, 2015.

BRASIL. ANPD publica análise preliminar do Projeto de Lei nº 2338/2023, que dispõe sobre o uso da Inteligência Artificial: ANPD. Brasília, Disponível em: https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-analise-preliminar-do-projeto-de-lei-no-2338-2023-que-dispoe-sobre-o-uso-da-inteligencia-artificial. Acesso em: 11 out. 2023.

BRASIL. Projeto que regula IA é apresentado ao Senado após trabalho da comissão liderada pelo ministro Cueva: STJ. Brasília, Disponível em: https://www.stj.jus.br/sites/portalp/Paginas/Comunicacao/Noticias/2023/04052023-Projeto-que-regula-IA-e-apresentado-ao-Senado-apos-trabalho-da-comissao-liderada-pelo-ministro-Cueva.aspx Acesso em: 11 out. 2023.

BRASIL. Senado Federal. Projeto de Lei nº 2338, de 2023. Tramitação conjunta com os projetos de lei PL 5.051 e 5.691, de 2019; PL 21, de 2020; PL 872, de 2021; PL 3.592, de 2023, que dispõe sobre o uso da Inteligência Artificial. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233 Acesso em: 11 out. 2023.

CLAUDIO, Luiz. Resumo sobre a história e o futuro da cirurgia robótica. **Comunidade Sanar**. 2020. Disponível em: < https://www.sanarmed.com/ia-o-primeiro-medico-robo-do-mundo-ja-atende-pacientes-colunistas>. Acesso em: 25 de setembro de 2023.

COIERA, Eurico. Inteligência Artificial na Medicina. **Informática Médica,** 4(1):27-46, jul-ago 1998.

CORTES, Corinna, VAPNIK, Vladimir. Redes vetoriais de apoio. **Mach Learn**, Berlin, 20: 273–297, 1995.

COSTA, Ana Carolina. IA: O primeiro médico robô do mundo já atende pacientes. **Comunidade Sanar**. 2019. Disponível em: < https://www.sanarmed.com/ia-o-primeiro-medico-robo-do-mundo-ja-atende-pacientes-colunistas>. Acesso em: 25 de setembro de 2023.

GRANATYR, Jones. IA Forte x IA Fraca. **Expert Academy**. 2017. Disponível em: <https://iaexpert.academy/2017/01/17/ia-forte-x-ia-fraca/>. Acesso em: 25 de setembro de 2023.

HONDA, Hugo et al. Os três tipos de aprendizagem de máquina. **UnB**. Brasília. 2017. Disponível em: <https://lamfo-unb.github.io/2017/07/27/tres-tipos-am/>. Acesso em: 25 de setembro de 2023.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. O cenário global das diretrizes de ética da IA. Nature Machine Intelligence. London, v. 1, n. 9, p. 389-399, set. 2019.

LAMY, Marcelo. Metodologia da Pesquisa: técnicas de investigação, argumentação e redação. 2ª ed. rev. atual. e ampl. São Paulo: Matrioska Editora, 2020.

LEMES, Marcelle Martins; Lemos, Amanda Nunes Lopes Espiñeira. O uso da inteligência artificial em saúde pela Administração Pública brasileira. Brasília: Cadernos Ibero-Americanos de Direito Sanitário, 9(3):166-182, jul-set 2020.

MCCULLOCH, Warren; PITTS, Water. Um cálculo lógico das ideias imanentes na atividade nervosa. Bulletin of Mathematical Biophysics, p.117-135, 1943.

MITCHELL, Thomas Michael. Machine Learning. New York, NY: McGraw-Hill, 1997.

OBERMEYER, Ziad; EMANUEL, Ezekiel. Prevendo o Futuro — Big Data, Aprendizado de Máquina e Medicina Clínica. New England Journal of Medicine. Waltham, v. 375, n. 13, p. 1216-1219, set. 2016.

Organização Mundial da Saúde (OMS). Ética e Governança da Inteligência Artificial para a Saúde. Disponível em: https://ppl-ai-file-upload.s3.amazonaws.com/web/direct-files/4747645/63aab737-4a7a-4d22-82a1-d64b2007403f/OMS.pdf. Acesso em: 12 out. 2023.

Organização Pan-Americana da Saúde (OMS). Ethics and governance of artificial intelligence for health, 2021. Disponível em: https://www.who.int/publications/i/item/9789240029200. Acesso em: 27 de setembro de 2023.

Organização Pan-Americana da Saúde (OPAS). Núcleo de Informação e Coordenação do Ponto BR. Martínez AL, Ortiz NO, Senne F (ed.). Medição da saúde digital: recomendações metodológicas e estudos de caso. São Paulo: Comitê Gestor da Internet no Brasil. 2019.

RAJKOMAR, Alvin. et al. Aprendizado de Máquina na Medicina. New England Journal of Medicine. Waltham, v. 380, n. 14, p. 1347-1358, abr. 2019.

RICHMAN, Barak. Regulamentação Sanitária para a Era Digital - Corrigindo a Incompatibilidade. New England Journal of Medicine. Waltham, v. 379, n. 18, p. 1694-1695, 2018.

ROSENBLA\TT, Frank. O Perceptron - Um autômato que percebe e reconhece. Relatório 85-460-1. EUA. Laboratório Aeronáutico Cornell. 1957.

RUMELHART, David Everett; HINTON, Geoffrey Everest; WILLIAMS, Ronald. Aprendendo representações por erros de retropropagação, *Nature,* Londres*,* 323:533-536, 1986.

SILVA, Cleyton Cézar Souto; VIANNA, Rodrigo Pinheiro De Toledo; MORAES, Ronei Marcos De. Sistema de Apoio a Decisão: a Segurança Alimentar e o Modelo em Rede Neural. Revista Brasileira de Ciências da Saúde. 2013; 16(1): 74-89

TOPOL, Eric. Medicina de alto desempenho: a convergência da inteligência humana e artificial. Nature Medicine. Reino Unido, v. 25, n. 1, p. 44-56, 2019.

TRUEMAN, Charlotte, et al. OpenAI alerta para riscos da IA superinteligente. Computerworld Brasil, Rio de Janeiro, 14(12):12-14, jul 2013. Disponível em: <https://itforum.com.br/computerworld/openai-riscos-ia-superinteligente/>. Acesso em: 25 de setembro de 2023.

União Europeia (EU). Lei de IA da UE: primeiro regulamento sobre inteligência artificial. 2023. Disponível em: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence. Acesso em: 12 de outubro de 2023.

WIDMAN, Lawrence. Sistemas Especialistas em Medicina. Informática Médica 5(1):2-5, set-out 1998.

ZHANG, Aston. et al. Mergulhe no Deep Learning. Versão 2.0. Londres, Inglaterra: Cambridge University Press, 2022.