

## TRANSPARÊNCIA E APLICABILIDADE NA INTELIGÊNCIA ARTIFICIAL: DESAFIOS E IMPLICAÇÕES PARA O ÂMBITO DA SAÚDE E DIREITO E O IMPACTO ÉTICO-JURÍDICO

### *TRANSPARENCY AND APPLICABILITY IN ARTIFICIAL INTELLIGENCE: CHALLENGES AND IMPLICATIONS FOR THE HEALTH AND LAW SCOPE AND THE ETHICAL-LEGAL IMPACT*

Santos, Amanda Ribeiro<sup>1</sup>  
Nascimento Junior, Francisco<sup>2</sup>

#### RESUMO

Este estudo explora a interseção fascinante e contemporânea entre a Inteligência Artificial (IA) e as áreas de Direito e Saúde. A IA possui o potencial para revolucionar o diagnóstico médico, aprimorando sua efetividade e precisão. No âmbito jurídico, ela pode transformar a maneira como o direito é exercido e interpretado, trazendo benefícios consideráveis em termos de eficiência e exatidão. O estudo propõe a aplicação da IA no Direito e na Saúde, para auxiliar profissionais do direito a lidar com questões complexas e em constante mudança, como processos envolvendo registros eletrônicos, casos de negligência médica, análise de provas periciais, entre outros. Contudo, também levanta questões relativas à ética, privacidade e responsabilidade. A metodologia utilizada inclui pesquisas bibliográficas, estudos *ex-post facto*, estudos de coorte, estudos de caso, pesquisas qualitativa e quantitativas e descritivas, e busca de jurisprudências. Assim, o estudo analisa essas questões em detalhes, ressaltando a importância de uma abordagem equilibrada e ordenada.

**Palavras-chave:** Inteligência Artificial; Direito; Saúde. Ética; Responsabilidade.

#### ABSTRACT

*This study explores the fascinating and contemporary intersection between Artificial Intelligence (AI) and the areas of Law and Health. AI has the potential to revolutionize medical diagnosis, enhancing its effectiveness and accuracy. In the legal field, it can transform the way law is practiced and interpreted, bringing considerable benefits in terms of efficiency and accuracy. The study proposes the application of AI in Law and Health, with the aim of assisting legal professionals to deal with complex and constantly changing issues, such as processes involving electronic records, cases of medical negligence, analysis of expert evidence, among others. However, it also raises questions related to ethics, privacy, and responsibility. The methodology used includes bibliographic research, ex-post facto studies, cohort studies, case studies, qualitative and quantitative descriptive research, and jurisprudence search. Thus, the study analyzes these issues in detail, highlighting the importance of a balanced and orderly approach.*

**Keywords:** Artificial Intelligence; Law; Health. Ethics; Responsibility.

<sup>1</sup> Acadêmica da Universidade Federal do Sul da Bahia. Especializando-se em Direito Constitucional, Direito Civil e Direito Processual Civil e Direito do Trabalho. Bolsista CNPq. <https://orcid.org/0009-0001-3212-1830>. <https://lattes.cnpq.br/6347807068102697>. E-mail: amanda.ribeiro@gfe.ufsb.edu.br

<sup>2</sup> Doutor em Educação - Universidade de São Paulo. Professor permanente do Programa de Pós Graduação em Ensino e Relações Étnico-Raciais (PPGER) da UFSB. <https://orcid.org/0000-0003-0587-8392>

## INTRODUÇÃO

A Inteligência Artificial (IA), uma das tecnologias mais inovadoras da atualidade, está começando a remodelar o mundo como o conhecemos, com implicações significativas para os campos da saúde e do direito. No campo da saúde, a IA, especialmente o subtipo de 'deep learning' (um algoritmo que utiliza redes neurais com múltiplas camadas para simular o pensamento humano), promete transformar a prática clínica, a administração de sistemas de saúde e a relação entre os pacientes e a rede de assistência nos próximos anos. Isso permitirá que os pacientes processem seus próprios dados para promover a saúde, impactando significativamente a prática clínica (Prado, 2002).

No campo jurídico, a IA está otimizando tarefas tradicionais, como análise de contratos e pesquisa jurídica, ao mesmo tempo que abre caminho para novas carreiras especializadas em proteção de dados, como inteligência artificial forense, direito digital e cibernético, e ética na inteligência. No entanto, a ausência de justificativa normativa afeta as decisões jurídicas. Na Saúde Digital, a IA está reformulando a estrutura dos serviços de saúde e dos sistemas de saúde nacionais, com grande potencial para melhorar a qualidade e reduzir os custos na assistência, especialmente no Sistema Único de Saúde (SUS).

Embora os benefícios sejam evidentes, o uso de aprendizado de máquina apresenta desafios significativos. Apesar de ter demonstrado resultados vantajosos, com análises preditivas e eficácia na elaboração de documentos, as correlações empíricas usadas por esses sistemas têm limitações em sua capacidade de esclarecimento. Isso prejudica as aplicações no campo legal, onde há uma demanda por justificativa normativa das decisões (Russell, S.J., Norvig, 1995). Portanto, enfatiza-se a necessidade de regulamentação ética em ambas as áreas para garantir que as decisões tomadas por sistemas de IA sejam transparentes e justificáveis (Allen, 2002; Wallach, 2004).

Do ponto de vista metodológico, este estudo segue uma abordagem qualitativa com base bibliográfica, ancorada em uma natureza interpretativa. Fundamenta-se na análise de artigos e referências bibliográficas relevantes ao tema central. A metodologia aplicada incluiu a exploração de recursos tanto em bibliotecas físicas quanto digitais, além da consulta a bancos de dados e periódicos especializados, com o objetivo de localizar materiais pertinentes.

Esta pesquisa envolve uma revisão bibliográfica integrativa, conduzida na Biblioteca Nacional de Medicina dos Estados Unidos da América (NLM), utilizando a base de dados

Medical Literature Analysis and Retrieval System Online (MEDLINE). Adicionalmente, foram feitas buscas na Scientific Electronic Library Online (SciELO) e nos sites da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), da Organização Mundial da Saúde (OMS) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Seguir-se-á uma metassíntese qualitativa dos resultados, isto é, uma interpretação sintetizada dos dados. Além disso, foram realizadas pesquisas descritivas, quantitativas e qualitativas, cujo objetivo é observar, documentar, analisar e associar fenômenos ou eventos, sem interferir no ambiente estudado, juntamente com os estudos exploratórios.

Entende-se, portanto, que o processo de investigação se configura como uma atividade científica fundamental que, por meio da indagação e reconstrução da realidade, nutre a atividade de ensino e a atualiza em relação à realidade. Da mesma forma, conecta o pensamento à ação, pois "nenhum problema intelectual pode existir sem antes ter sido um problema da vida cotidiana" (Minayo, 2001, p. 17). Segundo Minayo (1994), a objetivação não é alcançável quando se trabalha com dados predominantemente qualitativos, pois é inviável descrever a realidade com total precisão. Para a autora, a única forma de objetivação viável, nesse contexto, consiste na "rejeição da neutralidade", o que demanda atenção do pesquisador para "minimizar os juízos de valor" ao máximo. Nesse sentido, os métodos e as técnicas de coleta e tratamento dos dados ganham relevância.

Como a pesquisa bibliográfica tem sido um procedimento amplamente utilizado nos trabalhos de caráter exploratório-descritivo, reafirma-se a importância de definir e expor com clareza o método e os procedimentos metodológicos (modalidade de estudo, alcance específico, ferramentas de coleta de informações) que foram empregados na realização, especificando as fontes, de modo a expor as perspectivas que orientaram todo o processo de pesquisa e análise da proposta.

Na perspectiva deste artigo, pretende-se contribuir para a qualificação dos estudos que adotam a pesquisa bibliográfica como procedimento metodológico. Com base na pesquisa de Russell e Norvig (2010), Hengstler, Enkel e Duelli (2016), discutem a distinção entre dois principais grupos de Inteligência Artificial (IA): IA forte e IA fraca. A IA forte é uma hipótese que sugere a existência de sistemas com inteligência humana ou até mesmo superinteligência em todos os aspectos. Contudo, essas tecnologias ainda não são viáveis e permanecem no domínio da ficção científica. Por outro lado, a IA fraca refere-se às tecnologias atuais que são capazes de executar tarefas específicas, limitadas a certas capacidades humanas, como lógica, processamento de linguagem natural e reconhecimento de

padrões. Essas aplicações têm relevância comercial e são amplamente utilizadas. Em outras palavras, enquanto a IA forte representa um cenário futurista de inteligência artificial com capacidades sobre-humanas, a IA fraca é a nossa realidade atual, onde sistemas especializados desempenham funções específicas com base em algoritmos e dados disponíveis (Russell, S.J., Norvig, 1995).

Além disso, os computadores e a IA podem ser superiores aos humanos na tomada de decisões morais, uma vez que não são limitados pela racionalidade humana nem vulnerável ao lado emocional (Allen, 2002; Wallach, 2004). No entanto, os sistemas de inteligência artificial precisam ter sua excelência e proteção comprovadas, reconhecendo que atividades e assistências tradicionalmente realizadas por indivíduos estão sendo cada vez mais influenciadas e até executadas por sistemas automatizados. Isso questiona pressupostos fundamentais da regulamentação em saúde, do campo legal da medicina, dos direitos humanos e da ética.

Para o autor Allen, ele discute os métodos *top-down* (definição explícita de regras morais) e *bottom-up* (aprendizado a partir de exemplos) para construir sistemas morais. Allen também considera abordagens híbridas que combinam elementos de ambos os métodos. Wallach, por outro lado, concentra-se na ética e na regulação de máquinas autônomas. Ele explora métodos para ensinar robôs a fazer escolhas morais corretas e evitar ações prejudiciais. Ambos os autores enfatizam a importância de considerar questões éticas e legais no desenvolvimento e uso de sistemas de IA (Allen, 2002; Wallach, 2004)

A implementação do “direito à explicação” na área da saúde deve considerar as complexidades específicas da regulamentação da IA para uso clínico. Como esse direito agora faz parte da legislação brasileira, cabe aos órgãos reguladores definir sua amplitude e mecanismos para poder ser efetivamente aplicado. Além da atuação da Autoridade Nacional de Proteção de Dados, é necessária a intervenção de outros órgãos reguladores, como a Agência Nacional de Vigilância Sanitária e os conselhos de medicina.

Fica evidente que a regulamentação da IA é crucial nos setores da saúde e do direito. Toda intervenção afeta diretamente a saúde, portanto, a adoção dessas tecnologias emergentes deve ser incentivada paralelamente à criação de uma estrutura regulatória robusta que assegure seu uso em benefício da humanidade (Doherty *et al.*, 2016).

O artigo de Doherty *et al.* (2016), por exemplo, aborda a evolução na medicina, especialmente na oncologia. Os autores exploram as oportunidades oferecidas pela análise precisa de dados genômicos, facilitando escolhas racionais de tratamento adaptadas a pacientes individuais. Além disso, o artigo discute os desafios enfrentados no

desenvolvimento convencional de medicamentos e na aprovação regulatória. Embora o artigo não mencione especificamente a IA, podemos relacioná-la a algumas mudanças e questões éticas na medicina. A IA possibilita análises mais rápidas e precisas de grandes conjuntos de dados genômicos e clínicos. No entanto, surge a preocupação com a proteção dos dados dos pacientes e da própria pesquisa. Com a evolução da tecnologia, tornou-se muito fácil produzir evidências clínicas falsas, prejudicando a confiança pública na ciência e medicina e até mesmo os patrocinadores da pesquisa (Doherty et al., 2016).

A transparência é o princípio ético mais comumente encontrado nos códigos de diretrizes gerais para o uso da IA e é fundamental para a IA na saúde. Transparência implica que informações suficientes sobre as tecnologias de IA sejam documentadas antes da implantação, facilitando a consulta pública e o entendimento sobre seu funcionamento no mundo real. Espera-se que os sistemas sejam compreensíveis e explicáveis para desenvolvedores, profissionais de saúde, pacientes, usuários e reguladores, conforme a capacidade de compreensão de cada grupo e até mesmo de cada indivíduo.

A transparência dos algoritmos é necessária para que outros princípios-chave do uso da IA na saúde sejam eficazes: a proteção da autonomia humana, para garantir que as pessoas mantenham o controle sobre os sistemas de saúde e as decisões médicas; os requisitos regulatórios de segurança e eficácia, que asseguram que a IA não cause danos às pessoas e promova o bem-estar; a prestação de contas no uso de tecnologias de IA; a busca pela equidade, que promova a inclusão social e garanta que os algoritmos não reproduzam preconceitos e discriminação. A aplicação de todos esses princípios exige a transparência dos sistemas de IA.

Atualmente, o principal mecanismo para expressar a transparência algorítmica tem sido o “direito à explicação” sobre decisões automatizadas, considerado um elemento fundamental na regulamentação de algoritmos. Esse conceito vem se consolidando desde a elaboração do Regulamento Geral sobre a Proteção de Dados da União Europeia (GDPR), em vigor desde maio de 2018. O possuidor das informações deve ter “o privilégio de conseguir a interferência humana, de manifestar a sua perspectiva, de conseguir um esclarecimento sobre a resolução adotada na sequência desta avaliação e de contestar a decisão”. Ou seja, além de receber uma explicação compreensível, cria-se o direito à oportunidade de ser ouvido, de questionar e pedir revisão da decisão automatizada. Isso é o que tem sido denominado como “processo algorítmico adequado” e representa a expressão normativa do princípio da transparência de algoritmos, central na regulamentação dos sistemas de IA.

Desde a publicação do Regulamento Geral sobre a Proteção de Dados (GDPR), e

mesmo antes de sua entrada em vigor, têm-se travado intensos debates acerca da existência e do alcance do ‘direito à explicação’ nas decisões automatizadas. Assim como um algoritmo de Inteligência Artificial pode empregar uma infinidade de variáveis para atingir um resultado específico, a representação matemática complexa é frequentemente incompreensível para os seres humanos, é por essa razão que os algoritmos são comumente chamados de “caixa preta”. Tanto a LGPD – Lei Federal no 13.709/2018, quanto o GDPR se aplicam a qualquer empresa ou pessoa que trate dados pessoais dentro, claro, de suas respectivas jurisdições. A LGPD em território brasileiro, o GDPR no âmbito da União Europeia (Brasil, 1940).

Em linhas gerais, há quem defenda a viabilidade e o escopo do “direito à explicação” apenas no que diz respeito à funcionalidade geral do sistema, ao invés de decisões específicas e circunstâncias individuais. Por outro lado, existe a percepção de que a explicação também deve abranger decisões específicas, com a transparência restrita apenas pela natureza inerentemente “caixa-preta” dos algoritmos (Lebovitz (2021)

Desse modo, a relevância do “direito à explicação” é proporcionar aos pacientes a oportunidade de entender a lógica por trás das decisões automatizadas que afetam as ações tomadas em seus cuidados de saúde. Essa preocupação deve ser cada vez mais prevalente em várias situações clínicas. Por exemplo, já existem algoritmos de aprendizado profundo capazes de estabelecer critérios para transplantes de órgãos, como alocação, correspondência entre doador e receptor e previsão de sobrevida dos pacientes transplantados. É bastante provável que em breve esses algoritmos possam ser utilizados para esse propósito e que existam diferenças na ordem das filas de transplantes em comparação com as definidas por critérios clínicos feitos exclusivamente por humanos (Brasil, 1940).

Conforme discutido por Cervo e Bervian (2003, p. 67), esses estudos podem adotar várias formas: estudos descritivos; pesquisa de opinião; pesquisa de motivação; estudo de caso; e pesquisa documental. Andrade (2003, p. 124) observa que os estudos descritivos são frequentemente solicitados por empresas, instituições educacionais e partidos políticos, visando identificar um padrão específico na coleta de dados.

Os estudos explicativos, os últimos tipos que abordaremos neste trabalho acadêmico, apresentarão um caráter mais complexo em comparação aos outros. Além de observar, documentar, analisar e associar fenômenos, buscaremos aprofundar nosso entendimento dos temas "inteligência artificial no direito" e "saúde", identificando elementos determinantes para esclarecer as razões das coisas. No entanto, esses estudos estão mais propensos a cometer erros (Prado, 2002).

De acordo com Gil (2002, p. 42), pode-se afirmar que o conhecimento científico está

baseado nos resultados fornecidos pelos estudos explicativos. Isso não implica, no entanto, que as pesquisas exploratórias e descritivas sejam menos importantes, pois quase sempre representam uma fase inicial crucial para a obtenção de explicações científicas mais sólidas. Em outras palavras, a garantia do 'direito à transparência' na saúde exige mecanismos que tornem a IA compreensível e o reconhecimento das limitações na aplicação dos algoritmos.

Perceba que o equilíbrio ideal dependerá não apenas das competências tecnológicas dos sistemas inteligentes, mas também das especificidades de cada campo de aplicação e das exigências legais. É benéfico surgir uma literatura acadêmica brasileira sobre o assunto, em diálogo com o estado da arte da pesquisa mundial, abordando problemas no cenário brasileiro e dentro das condições do ordenamento jurídico nacional. É importante não perder de vista que muitos desses desafios também estão presentes em outros países, tanto na área jurídica quanto no setor de saúde (Hoffmann-Riem, 2021).

Assim, a análise do impacto ético-jurídico da inteligência artificial no direito e no setor de saúde é um tema de grande relevância e complexidade. A aplicabilidade dessas tecnologias traz consigo desafios significativos, mas também oportunidades para melhorar a eficiência e a eficácia em ambos os campos. É essencial que continuemos a explorar este tópico, a fim de garantir que a implementação da inteligência artificial seja feita de uma maneira que respeite os direitos e as necessidades das pessoas.

## **1. OS DESAFIOS E IMPLICAÇÕES PARA A SAÚDE E A PRÁTICA JURÍDICA: O IMPACTO DOS SISTEMAS INTELIGENTES**

A evolução da sociedade e as adaptações da lei ao longo do tempo têm moldado o que conhecemos hoje como direito e saúde. A progressão tecnológica é um fator crucial nessa transformação, tornando essencial que nos adaptemos e evoluamos com essas mudanças (Lebovitz, 2024).

A definição de normas morais para orientar a evolução da inteligência artificial (IA) é vital para direcionar seu uso para propósitos socialmente benéficos. Contudo, isso acarreta certos perigos, como a determinação do nível adequado de uso da IA. No judiciário brasileiro, por exemplo, a implementação do Processo Judicial Eletrônico (PJe), substituindo os processos físicos convencionais, tem impactado positivamente, aliviando a carga do judiciário e agilizando determinados processos.

Na área da saúde, a introdução de novas tecnologias redefine padrões éticos e jurídicos, assegurando a privacidade e proteção das informações. No entanto, a aplicação da

IA na saúde pode resultar em orientações equivocadas, violação de dados pessoais e disseminação de desinformação, conforme alertado pela Organização Mundial da Saúde (OMS). Ferramentas de modelo de linguagem (LLMs - *Large Language Models*), geradas por inteligência artificial, apresentam riscos para a saúde pública devido à rápida difusão e uso experimental sem mecanismos de controle adequados. Tais medidas devem incluir valores como transparência, inclusão, engajamento público, supervisão especializada e avaliação rigorosa (Allen, 2002; Wallach, 2004).

Os novos sistemas de inteligência artificial podem auxiliar em decisões clínicas e melhorar diagnósticos em locais com recursos limitados. No entanto, respostas geradas por Modelos de Linguagem de Grande Escala (LLMs) podem ser imprecisas ou equivocadas, especialmente em questões de saúde. A Organização Mundial da Saúde (OMS) adverte que a adoção precipitada de sistemas não testados pode resultar em erros, prejudicar pacientes e abalar a confiança na IA. É essencial que essas preocupações sejam baseadas em evidências sólidas, com análises de riscos e benefícios realizados antes da implementação generalizada em cuidados de saúde, seja por pacientes, prestadores de serviços ou legisladores.

Por outro lado, sistemas de suporte à decisão clínica, dispositivos vestíveis (*wearable devices*) e a crescente capacidade de armazenar e processar informações de pacientes (big data) já são uma realidade em muitos países. A IA processa esses dados utilizando algoritmos que se aprimoram automaticamente (*self-learning*), fornecendo diagnósticos cada vez mais precisos (Zhao, 2019). O processamento de grandes volumes de dados de saúde melhora a compreensão das causas, diagnóstico e tratamento de problemas de saúde, tanto ao nível individual quanto coletivo. Essa capacidade permite novas iniciativas de promoção, prevenção e restauração da saúde, exigindo uma reorganização dos sistemas de saúde (Zhao, 2019).

Um exemplo atual é a identificação de uma nova classe de antibióticos. Cientistas do Instituto de Tecnologia de Massachusetts (MIT) e da Universidade de Harvard, nos EUA, fizeram uma descoberta significativa ao utilizarem um algoritmo de inteligência artificial para identificar uma nova molécula antibiótica. Utilizando aprendizado profundo (*deep learning*) para prever a toxicidade, eles analisaram virtualmente diversas substâncias químicas. Esse método levou à identificação de antibióticos com atividade potente contra patógenos resistentes a múltiplos medicamentos (Zhao, 2019). Os cientistas testaram os compostos contra o *Staphylococcus aureus* resistente à meticilina (SARM) e descobriram que os novos compostos podem matar essa bactéria resistente. Além disso, prever quais subestruturas moleculares são responsáveis pela atividade antibacteriana foi uma inovação do estudo.

A partir disso, novos modelos de aprendizado profundo foram desenvolvidos para antecipar a toxicidade dos compostos em células humanas. Combinando essas informações com previsões de atividade antimicrobiana, os pesquisadores identificaram compostos eficazes contra microrganismos com efeitos adversos mínimos no corpo humano. Essa pesquisa evidencia a importância da inteligência artificial na descoberta de novos fármacos (Zhao, 2019).

Apesar dos avanços da IA, o papel do médico é essencial. O *know-what* (saber como) fornecido pelo computador deve ser complementado pelo *know-why* (saber por quê) discutido entre médico e paciente. Isso demanda um cuidado contínuo com a excelência na formação dos médicos e o reconhecimento de que o profissional de saúde é um agente terapêutico essencial (Lebovitz, 2024).

A partir dessa perspectiva, é importante complementar o estudo deste artigo discorrendo sobre o *machine learning* (aprendizado de máquina). Isso porque, hierarquicamente, o *deep learning* (aprendizado profundo) é uma extensão do *machine learning*. Ambos são tipos de aprendizado de máquina, porém o *deep learning* opera de maneira mais intensa. O *machine learning* é uma aplicação da inteligência artificial (*artificial intelligence*) que cria algoritmos permitindo que sistemas aprendam sem intervenção humana. Assim, não é necessário instruir o programa sobre o que fazer. À medida que é alimentado com novos dados e entradas automáticas, ele aprende e toma decisões autonomamente. O *deep learning*, por sua vez, é um subconjunto do *machine learning*, focando na construção de redes neurais artificiais (*artificial neural networks*) que simulam o cérebro humano. Isso permite que o algoritmo aprenda e se adapte a novas situações (Lebovitz, 2024).

No entanto, o *machine learning* precisa de treinamento para funcionar adequadamente em diferentes situações. No caso do *deep learning*, fornecemos apenas informações básicas, permitindo que a rede neural se construa e evolua automaticamente. Ou seja, basta fornecer dados para que o sistema utilize essas informações para melhorar seu desempenho, tornando-se indispensável na área da saúde, especialmente em casos de vida ou morte.

Ambos, *deep learning e machine learning*, têm capacidades de aprendizado, mas o *deep learning* realiza análises muito mais complexas e profundas. Ele também consegue reconhecer e executar várias tarefas simultaneamente, com base nos dados que coleta. Especialistas também ponderam que a inteligência artificial (IA) ainda precisa conquistar a confiança da comunidade médica para ser vista como uma técnica segura. "É uma tarefa na qual tanto pesquisadores, técnicos, quanto o sistema legislativo continuam no início. Um dos primeiros desafios a ser superado é a qualidade da informação que alimenta os algoritmos. O

setor de saúde produz uma quantidade enorme de dados, porém, poucas dessas informações são organizadas de maneira consolidada. Dados incompletos e inconsistentes geram o que chamamos de algoritmos ruins " (Lebovitz, 2024).

A expansão do uso de sistemas inteligentes tem o potencial de revolucionar a prática jurídica, não só introduzindo novas questões, mas também automatizando tarefas repetitivas. Isso exige uma transformação no perfil dos profissionais jurídicos, que precisam estar preparados para lidar com novas tecnologias (Lebovitz, 2024).

No campo do Direito, a adaptação envolve tanto o uso de ferramentas de IA pelos profissionais quanto a construção de sistemas computacionais que realizam atividades compatíveis com a lei. A expansão da IA demandará profissionais habilitados para lidar com mudanças tecnológicas e atuar em equipes interdisciplinares, empregando as habilidades de juristas, cientistas da computação e outros especialistas para desenvolver sistemas inteligentes que produzam resultados positivos e respeitem os direitos e interesses legais. Por exemplo, os direitos à explicação e à revisão de decisões de sistemas de IA estão necessariamente vinculados e precisam ser compreendidos conjuntamente. No Brasil, a estruturação desses direitos ainda requer regulamentação e futura formulação doutrinária e jurisprudencial, assim como em outras nações, seguindo o padrão europeu (Allen, 2002; Wallach, 2004).

A transparência algorítmica é fundamental para garantir o direito à explicação. Isso significa que a transparência em um sistema de IA deve se concentrar principalmente no processo, permitindo que as pessoas entendam como os algoritmos são desenvolvidos e implantados em termos gerais. Às vezes, pode incluir elementos sobre previsões ou decisões específicas, mas geralmente não envolve a divulgação de códigos ou bases de dados. Portanto, algum grau de opacidade é inevitável, seja pela complexidade dos sistemas ou por segredo empresarial ou de Estado, que visa proteger segredos de negócios ou dados sensíveis do usuário. Há também opacidade decorrente do "analfabetismo técnico" dos usuários.

Porém, ressalto que é necessária uma explicação concisa para abordar alguns aspectos essenciais. Primeiramente, destacar os principais fatores de decisão para previsões de IA, preferencialmente por ordem de relevância. Em seguida, esclarecer os fatores determinantes que afetam significativamente os resultados. Por último, discutir casos semelhantes que podem levar a resultados divergentes (Allen, 2002; Wallach, 2004).

A IA explicável está em constante evolução. Muitas pesquisas técnicas estão em andamento por corporações, entidades de padronização, organizações não governamentais e instituições governamentais, para desenvolver sistemas de IA que expliquem suas projeções. Um exemplo é o LIME (*Local Interpretable Model-Agnostic Explanations*), que cria modelos

substitutos simples para esclarecer previsões de modelos complexos, ajudando as pessoas a entenderem como um parecer foi estabelecido. Outro exemplo é o SHAP (*SHapley Additive exPlanations*), que usa a teoria dos jogos para atribuir valores a cada característica de uma previsão, ajudando a entender a contribuição de cada variável para o resultado. Essas técnicas visam tornar a IA mais transparente e confiável, especialmente em áreas críticas como saúde e finanças.

Desenvolver sistemas que ofereçam explicações é um processo complexo e caro. Isso se aplica tanto aos sistemas projetados para fornecer explicações desde o início (explicabilidade inerente) quanto aos que fornecem explicações após a decisão algorítmica (explicabilidade post hoc). Por essa razão, as pesquisas focam em modelos explicáveis em áreas de alto risco, como a assistência à saúde. Considerar o equilíbrio entre a capacidade de explicação e a precisão é essencial. Para que um sistema de IA seja explicável, muitas vezes é necessário reduzir as variáveis a um conjunto pequeno o suficiente para ser compreendido por humanos. No entanto, isso pode tornar impraticável a aplicação de alguns sistemas em questões complexas.

Para ilustrar essa ideia, Platão, em seus estudos com São Tomás de Aquino, defendia que a legislação humana não deveria contradizer as leis divinas e naturais, destacando a necessidade de harmonia entre essas esferas (Pedrosa, 2002). É crucial reconhecer e considerar os desafios para desenvolver uma IA explicável na área da saúde ao criar mecanismos regulatórios. Esses mecanismos devem levar em conta os limites da explicabilidade e garantir o direito à explicação e à revisão das decisões automatizadas na assistência médica, além de assegurar a ética e os direitos humanos (Pedrosa, 2002).

Contudo, um direito à explicação mais extensa, baseada na máxima transparência, pode ser incompatível com o uso de sistemas automatizados que buscam alta precisão. Os métodos atualmente acessíveis para explicabilidade conseguem fornecer descrições abrangentes de como o sistema de IA opera, em geral, mas são limitados ou inseguros para decisões específicas. Na realidade, as explicações podem ser extremamente valiosas em processos gerais de IA, como construção de modelo e inspeção, mas raramente são esclarecedoras em relação a resultados específicos fornecidos pelos algoritmos.

A falta de clareza atual deve continuar, pelo menos temporariamente. Em certa medida, a opacidade é comum na prática clínica. A medicina tradicionalmente emprega procedimentos com mecanismos não completamente entendidos, mas aplicados extensivamente devido aos efeitos comprovados, como muitos medicamentos. A barreira para desenvolver uma IA elucidativa na saúde devem ser identificadas e cuidadosamente avaliadas

ao elaborar mecanismos regulatórios que considerem os limites da explicabilidade e a extensão do direito à explicação e à revisão de decisões automatizadas no cuidado à saúde.

Dados do Conselho Nacional de Justiça (CNJ) indicam alguma convergência em torno dos princípios de transparência (clareza para o usuário de que está interagindo com um sistema artificial), explicabilidade (informações que permitam ao usuário entender os critérios para tomada de decisão), não discriminação (evitar vieses que possam ofender direitos fundamentais), não maleficência (sistemas de IA não podem prejudicar humanos), responsabilidade e privacidade/proteção de dados. No entanto, há divergências sobre o significado e a implementação desses princípios.

Há também o risco de sobre utilização de sistemas inteligentes em situações que podem ter impacto negativo nas pessoas. Por outro lado, o medo dessa probabilidade pode levar à subutilização da inteligência artificial, impedindo que se aproveitem de seus benefícios. Outro impasse é que os princípios são formulados de cima para baixo, com pretensão de universalidade, tornando seu conteúdo genérico e abstrato, dificultando sua aplicabilidade (Allen, 2002; Wallach, 2004).

A definição de regras gerais, em vez de uma regulamentação rígida, favorece o desenvolvimento tecnológico e permite que a sociedade se familiarize com os sistemas inteligentes antes de decidir como regulá-los. Um dos grandes desafios na regulamentação da IA é identificar não só os princípios aplicáveis, mas também os momentos em que esses princípios devem ser implementados por regras jurídicas, assim como os instrumentos legais mais adequados para essa regulamentação (Hoffmann-Riem, 2021).

A inteligência artificial pode revolucionar diversos segmentos da sociedade, incluindo saúde e direito. Contudo, é essencial que essas transformações sejam guiadas por princípios éticos robustos e regulamentações apropriadas. A adoção de uma estratégia de baixo para cima, considerando as particularidades de cada setor, pode assegurar que a IA beneficie a sociedade em geral, protegendo simultaneamente direitos e interesses individuais e coletivos.

Para concluir, a efetiva implementação desses princípios e regulamentações demandará um esforço persistente de todos os participantes, desde criadores de sistemas de IA até profissionais do direito e formuladores de políticas. Somente por meio de uma abordagem cooperativa e reflexiva, podemos garantir que a inteligência artificial seja utilizada de maneira ética e consciente.

## CONCLUSÃO

A Inteligência Artificial (IA) é mais do que uma inovação tecnológica revolucionária; é um marco na trajetória humana, com impactos significativos na economia, política e direito. A implementação dessas tecnologias emergentes deve ser feita de maneira cuidadosa, consciente e ética para atenuar os efeitos prejudiciais de seu uso inadequado.

Os princípios fundamentais na aplicação da IA, como a transparência e a explicabilidade, são essenciais para entender a lógica subjacente às decisões automatizadas. No entanto, a natureza da “caixa-preta” dos algoritmos pode dificultar para as pessoas entenderem completamente como as decisões são tomadas (Lebovitz, 2024). A IA também pode levar a violações de dados e à disseminação de desinformação, especialmente no campo da saúde. Portanto, a regulamentação da IA, incluindo iniciativas como a Lei Geral de Proteção de Dados (LGPD) no Brasil, é crucial. Ela deve abordar questões de equidade e inclusão para garantir o acesso igualitário aos benefícios da IA (Pedrosa, 2002).

A adoção de uma estratégia de baixo para cima, que considera as particularidades de cada setor, pode ser uma via promissora para garantir que a IA seja empregada para beneficiar a sociedade em geral, protegendo simultaneamente os direitos e interesses individuais e coletivos. O impacto da IA no trabalho, nas interações sociais (incluindo cuidados de saúde), na privacidade, na justiça e na segurança (incluindo iniciativas de paz e guerra) é enorme. O impacto social e ético da IA abrange muitos domínios, como os sistemas de classificação de máquinas, que levantam questões sobre privacidade e preconceitos, e veículos autônomos, que levantam questões sobre segurança e responsabilidade (Hoffmann-Riem, 2021).

Diante do exposto, a implementação efetiva desses princípios e regulamentações exigirá um esforço contínuo de todos os participantes, desde os criadores de sistemas de IA até os profissionais do direito e os responsáveis pela formulação de políticas. Somente por meio de uma abordagem cooperativa e reflexiva, podemos garantir que a IA seja utilizada de maneira ética e consciente.

## REFERÊNCIAS

ANDRADE, M. M. de. **Introdução à metodologia do trabalho científico**. 6. ed. São Paulo: Atlas, 2003

ALMEIDA, Haian de Assis Lopes; DE OLIVEIRA, Tamar Ramos. CRIMES VIRTUAIS: O AVANÇO DOS CRIMES ELETRÔNICOS E A EVOLUÇÃO DAS LEIS ESPECÍFICAS NO BRASIL. **Revista Ibero-Americana de Humanidades**, Ciências e Educação, v. 8, n. 11,

p. 277-294, 2022.

ALEXANDRINO, José. **Direitos Fundamentais: Introdução Geral**. Cascais: Editora Princípia, 2011.

ALEXANDER, R. D. **The biology of moral systems**. New York: Aldine de Gruyter, 1987.

ALLEN, C.; SMIT, I.; WALLACH, W. **Artificial morality: top-down, bottom-up, and hybrid approaches**. *Ethics and Information Technology*, v. 7, n. 3, p. 149-155, 2005.

BECCARIA, Cessare. **Dos Delitos e das Penas**. São Paulo-SP: ed.Martins Fontes, 1997.

BONINI, Luci Mendes et al. **Crimes Cibernéticos**. *Diálogos Interdisciplinares*, v. 7, n. 3, p. 223-236, 2018.

BRASIL. Código Penal. **Decreto-Lei nº 2.848, de 7 de dezembro de 1940**. Disponível em:<[http://www.planalto.gov.br/ccivil\\_03/decreto-lei/Del2848compilado.htm](http://www.planalto.gov.br/ccivil_03/decreto-lei/Del2848compilado.htm)> Acesso em: 19 jan. 2024.

BRASIL. **Constituição Federal de 1988**. Promulgada em 5 de outubro de 1988. Disponível em <[http://www.planalto.gov.br/ccivil\\_03/constituicao/constituicaocompilado.htm](http://www.planalto.gov.br/ccivil_03/constituicao/constituicaocompilado.htm)> Acesso em 21 fev. 2024.

BRASIL. **Lei Ordinária nº 12.735, de 30 de novembro de 2012**. Altera o Decreto-Lei no 2.848, de 7 de dezembro de 1940 - Código Penal, o Decreto-Lei no 1.001, de 21 de outubro de 1969 -Código Penal Militar, e a Lei no 7.716, de 5 de janeiro de 1989, para tipificar condutas realizadas mediante uso de sistema eletrônico, digital ou similares, que sejam praticadas contra sistemas informatizados e similares; e dá outras providências. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2011-2014/2012/Lei/L12735.htm](http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12735.htm)>. Acesso em: 25 de abril. 2024.

BRASIL. Lei Ordinária nº 12.737, de 30 de novembro de 2012. **Dispõe sobre a tipificação criminal de delitos informáticos**; altera o Decreto-Lei no 2.848, de 7 de dezembro de 1940 - Código Penal; e dá outras providências. Disponível em:<[http://www.planalto.gov.br/ccivil\\_03/\\_Ato2011-2014/2012/Lei/L12737.htm](http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12737.htm)>. Acesso em: 25 de abril. 2024.

BRASIL. Lei Ordinária nº 12.965, de 23 de abril de 2014. **Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil**. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2014/lei/l12965.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm)>. Acesso em: 26 de abril. 2024.

CALGAROTO, Cleber. **O direito à privacidade na internet: panorama, responsabilização civil e inovações do marco civil da internet (Lei nº 12.965/2014)**. Direito-Unisul Virtual, 2021.

CAZAROTI, Tatiane Martins Barros; PINHEIRO, Eduardo Fernandes. **Crimes Cibernéticos**.

**TCC-Direito, 2021.**

CERVO, A. L.; BERVIAN, P. A. **Metodologia científica**. 5. ed. São Paulo: Prentice Hall, 2003.

CRUZ, Diego; RODRIGUES, Juliana. Crimes cibernéticos e a falsa sensação de impunidade. **Revista Científica Eletrônica do Curso de Direito**, 2018.

DOHERTY, M. et al 2016. **Precision medicine and oncology: an overview of the opportunities presented by next-generation sequencing and big data and the challenges posed to conventional drug development and regulatory approval pathway**. *Annals of Oncology* 27: 1644–1646, 2016, DOI :10.1093/annonc/mdw165

DORNELAS, Natália Alves. **A Resposta Estatal Quanto Aos Crimes Cibernéticos: Uma Análise Direcionada Às Leis Nº 12.735/2012 E 12.737/2012**. Repositório de Trabalhos de Conclusão de Curso, 2019.

GATTI, B. A. **Estudos quantitativos em educação**. Educação e Pesquisa, São Paulo, SP, v. 30, n. 1, p. 11-30, jan, 2004.

GIL, A. C. **Método e técnicas de pesquisa social**. São Paulo, SP: Atlas. 1999.

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 5.ed.São Paulo: Atlas,2010

GIL, Antonio Carlos. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GIL, A. C. **Como elaborar projetos de pesquisa**. Atlas, 2008.

GIL, A. C. **Métodos e técnicas de pesquisa social**. Atlas, 1999.

GOMES, Walyson Milhomem; MEDRADO, Lucas Cavalcante. **CRIMES CIBERNÉTICOS UMA PONDERAÇÃO SOBRE A LEI 14.155 DE 2021 APLICÁVEL AO CRIME DE ESTELIONATO VIRTUAL**. **Revista Ibero-Americana de Humanidades**, Ciências e Educação, v. 9, n. 9, p. 1870-1894, 2023.

HABERMAS, Jürgen. **Um Ensaio sobre a Constituição da Europa**. Tradução Marian Toldy e Teresa Toldy. Lisboa: Edições 70 Lda, 2012

HERNANDEZ, Erika Fernanda Tangerino; DE TOLEDO, Nathália Karina Abucci. Crimes cibernéticos: seus efeitos revolucionários diante de uma legislação em constante evolução. **Revista Jurídica da UniFil**, v. 17, n. 17, p. 72-84, 2021.

HOFFMANN-RIEM, Wolfgang. **Teoria Geral do Direito Digital, Transformação Digital Desafios para o Direito**, Editora Forense, Rio de Janeiro, 2021, pp. 11-13

KRIEGUER, André Lemuel Ferreira; CERON, Antonio Luciano Bairros; MARCONDES, Aldair. **A Acelerada Evolução Social E Tecnológica Global Como Viabilizadores De Crimes Cibernéticos, Frente Ao Lento Desenvolvimento De Freios Legais Para Sua**

**Contenção.** Ponto de Vista Jurídico, p. 128-143, 2021.

LEBOVITZ, S.; LEVINA, N.; LIFSHITZ-ASSAF, H. **Is AI Ground Truth Really ‘True’? The Dangers of Training and Evaluating AI Tools Based on Experts’ Know-What.** Management Information Systems Quarterly, v. 45, n. 3b, p. 1501-1525, 2021. Disponível em: <<https://doi.10.25300/MISQ/2021/16564>>. Acesso em: 12 fev. 2024

LIMA, Yasmin Victoria et al. **Direito digital: aplicação nos crimes cibernéticos.** Anais da Semana de Pesquisa Jurídica, v. 1, p. 42-42, 2022.

MARCO, civil da internet entra em vigor. **Cultura Digital.** Disponível em: <http://culturadigital.br/marcocivil/2014/06/23/marco-civil-da-internet-entra-em-vigor/>. Acesso em: 26 fev. 2024.

MINAYO, M. C. S. Ciência, técnica e arte: o desafio da pesquisa social. In: MINAYO, M. C. S. (Org.). **Pesquisa social: teoria, método e criatividade.** Petrópolis, RJ: Vozes, 2001. p. 09-29.

MINAYO, M.C. de S. **O desafio do conhecimento: pesquisa qualitativa em saúde.** São Paulo-Rio de Janeiro, HUCITEC-ABRASCO, 1992

MPF, Ministério Público Federal. **Combate aos Crimes Cibernéticos.** Disponível em: <<http://www.mpf.mp.br/atuacao-tematica/ccr2/coordenacao/comissoes-e-grupos-detrabalho/combate-crimes-cirberneticos>>. Acesso em: 28 fev. 2024.

ONU News. **OMS aponta três riscos do uso da inteligência artificial na saúde.** Disponível em: <https://news.un.org/pt/story/2023/05/1814472>. Acesso em: 21 maio 2024.

RUSSELL, S.J., NORVIG, P., 1995. **Artificial Intelligence: A Modern Approach.** Prentice Hall,. Englewood Cliffs, New Jersey

RIBEIRO, José Mendes. **Saúde digital um sistema de saúde para o século XXI.** Lisboa: Fundação Francisco Manuel dos Santos, 2019.

SARLET, G; CALDEIRA, C. **O consentimento informado e a proteção de dados pessoais de saúde na internet: uma análise das experiências legislativas de Portugal e do Brasil para a proteção integral da pessoa humana.** Civilistica.com, a. 8. n. 1. 2019. Disponível:<http://civilistica.com/wp-content/uploads/2019/04/Sarlet-e-Caldeira-civilistica-com-a.8.n.1.2019.pdf>.

SARLET, W. Ingo. **Direitos Fundamentais e Direito Privado: algumas considerações em torno da vinculação dos particulares aos direitos fundamentais.** B. Cient. ESMPU, Brasília, a. 4 – n. 16, p. 193-259, jul./set. 2005. Disponível em: <https://repositorio.pucrs.br/dspace/handle/10923/11331>

SEER UFRGS. **Ao fazer intervir as novas tecnologias na relação.** Disponível em: <https://seer.ufrgs.br/index.php/ppgdir/article/view/121131>. Acesso em: 21 maio 2024.

PRADO, Luiz Regis. **Curso de Direito Penal Brasileiro**. 3 edição. São Paulo-SP: ed. Revista dos Tribunais, 2002.

PEDROSA, Ronaldo Leite. **Direito em História**. 4 edição. Nova Friburgo-RJ: ed. Imagem Virtual, 2002.

ZAFFARONI, Eugênio Raúl; PIERANGELI, José Henrique. Manual de Direito Penal Brasileiro. 4 edição. São Paulo-SP: ed. **Revista dos Tribunais**, 2002.

ZHAO, Q. et al. **MotifMap**: Integrative analysis of motif occurrences in human regulatory regions. *Cell Systems*, v. 9, n. 3, p. 269-278.e4, 2019. Disponível em: [https://www.cell.com/cell-systems/fulltext/S2405-4712\(19\)30312-6](https://www.cell.com/cell-systems/fulltext/S2405-4712(19)30312-6). Acesso em: 12 fev. 2024.