

# Inteligência artificial em saúde e implicações bioéticas: uma revisão sistemática

Mariele Abadia Elias<sup>1</sup>, Luciana Arruda Faversani<sup>1</sup>, Josiane Aparecida Vieira Moreira<sup>1</sup>, Anelise Viapiana Masieiro<sup>2</sup>, Natalia Veronez da Cunha Bellinati<sup>1</sup>

1. Universidade do Planalto Catarinense, Lages/SC, Brasil. 2. University of Iowa, Iowa City, United States of America.

## Resumo

A presença de inteligência artificial na saúde vem crescendo, ajudando em diagnósticos e tomadas de decisão, mas suas implicações geram dúvidas relacionadas à ética. Esta revisão sistemática, baseada nas diretrizes Prisma, identificou os usos de inteligência artificial na saúde e suas implicações bioéticas. Foi realizada busca nas bases de dados Science Direct e Scopus usando os descritores “*artificial intelligence*”, “*bioethics*” e “*health*”. Trabalhos em inglês, publicados entre 2017 e 2021 foram considerados, resultando em 102 artigos. Após aplicação dos critérios estabelecidos, 11 foram selecionados. Os estudos discutiram os princípios bioéticos da beneficência, não maleficência, autonomia e justiça, adicionando o elemento explicabilidade. Inteligência artificial mostrou correlação com imprevisibilidade, previsibilidade, confiança, papel do médico, desenvolvimento de sistemas, privacidade, segurança de dados, e aspectos sociais e financeiros. Desenvolvedores, profissionais da saúde e pacientes devem maximizar os benefícios e limitar os riscos das ferramentas que usam essa tecnologia.

**Palavras-chave:** Atenção à saúde. Aprendizado de máquina. Big data. Ética.

## Resumen

### Inteligencia artificial en salud y sus implicaciones bioéticas: una revisión sistemática

El uso de la inteligencia artificial en salud va en aumento por facilitar el diagnóstico y la toma de decisiones, pero sus implicaciones plantean dudas relacionadas con la ética. Esta revisión sistemática desde las directrices Prisma identificó los usos de la inteligencia artificial en salud y sus implicaciones bioéticas. Las búsquedas se realizaron en Science Direct y Scopus utilizando los descriptores “*artificial intelligence*”, “*bioethics*” y “*health*”. De los trabajos en inglés publicados entre 2017 y 2021, se obtuvo 102 artículos. Aplicados los criterios, quedaron 11. Los estudios abordaron los principios bioéticos de beneficencia, no maleficencia, autonomía y justicia, añadiendo el elemento explicabilidad. La inteligencia artificial se correlacionó con la imprevisibilidad, previsibilidad, confianza, papel de los médicos, desarrollo de sistemas, privacidad, seguridad de los datos y aspectos financieros y sociales. Los desarrolladores, los profesionales sanitarios y los pacientes deben maximizar los beneficios y limitar los riesgos que involucra esta tecnología.

**Palabras clave:** Atención a la salud. Aprendizaje automático. Macrodatos. Ética.

## Abstract

### Artificial intelligence in health and bioethical implications: a systematic review

The presence of artificial intelligence in healthcare is growing, helping in diagnosis and decision making. However, its application raises doubts, mostly related to ethics. This study aimed to identify its uses in health and its bioethical implications from a systematic literature review using the PRISMA guidelines. The ScienceDirect and Scopus databases were searched, using the descriptors “*artificial intelligence*,” “*bioethics*” and “*health*.” Works in English, published between 2017 and 2021 were considered, resulting in 102 articles found and, after applying the established criteria, 11 were selected. The studies reported on the bioethical principles of beneficence, non-maleficence, autonomy and justice, adding an element, explainability. Relationships were found between artificial intelligence in health and unpredictability, predictability, trust, physicians’ role, systems development, privacy, data security, financial and social aspects. Developers, healthcare professionals and patients must maximize the benefits and limit the risks of tools that use this technology.

**Keywords:** Health care. Machine learning. Big data. Ethics.

Declararam não haver conflito de interesse.

A inteligência artificial (IA) está sendo cada vez mais adotada em diferentes áreas. O termo em si é difícil de definir, uma vez que esse fenômeno depende de diferentes fatores<sup>1</sup>. Apesar desta dificuldade e de seus diversos conceitos, entende-se que a IA está associada a máquinas e computadores no intuito de ajudar a humanidade a resolver problemas e facilitar processos de trabalho<sup>2</sup>.

Os sistemas de IA trabalham utilizando algoritmos complexos e grandes conjuntos de dados que geram conclusões, substituindo o raciocínio humano pela análise de rotina<sup>3</sup>. Para alcançar a inteligência em nível humano, a IA precisa de orientação como modelo de realidade<sup>4</sup>. *Machine learning* (ML), traduzido como aprendizagem de máquina, é um sistema de IA que pode aprender com modelos e eventualmente se tornar autônomo, tomando decisões e gerando conclusões antes consideradas como restritas à competência da mente humana<sup>3</sup>.

A expansão da computação para armazenar, gerenciar, acessar e processar dados por meio de uma rede de servidores remotos – como o armazenamento em nuvem – levou à expansão de aplicativos de IA para o campo da saúde<sup>4</sup>. A IA e a ML têm o potencial de revolucionar a prestação de serviços nessa área<sup>3</sup>. Essas tecnologias podem ser utilizadas na informática em saúde, a qual descreve aquisição, armazenamento, recuperação e uso de informações para melhorar o atendimento ao paciente por meio de interações com o sistema<sup>4</sup>. Ferramentas de *big data*, ou seja, o armazenamento e a análise de dados volumosos, como os utilizados na área da saúde, também podem ser usados em associação com IAs<sup>5</sup>.

A IA tem melhorado o diagnóstico clínico e o desempenho na tomada de decisão em vários domínios do trabalho médico<sup>6</sup>. Essas ferramentas podem ajudar a adaptar os programas de saúde pública, garantindo que informações relevantes estejam disponíveis para políticas sólidas e tomada de decisão<sup>4</sup>. Atualmente, o diagnóstico automatizado por imagem é sem dúvida o domínio mais bem-sucedido do uso de IA na área médica. Muitas especialidades médicas, incluindo a radiologia, oftalmologia, dermatologia e patologia, dependem de diagnósticos baseados em imagem<sup>6</sup>.

No entanto, os pacientes reconhecem e estão começando a enfrentar as muitas questões levantadas pelos aplicativos de IA na área da saúde<sup>7</sup>. Essas demandas são legais, comerciais, sociais e, principalmente, éticas<sup>8</sup>. Projetar esses sistemas e utilizá-los não é um desafio meramente técnico, pois requer atenção aos princípios bioéticos<sup>3</sup>.

A bioética enfoca a relação entre os seres vivos e, quando a IA surge, os seres humanos devem se engajar eticamente com algo que não é natural por si só, ou seja, com sua própria criação<sup>2</sup>. São crescentes as preocupações sobre a potencial perda de controle na relação humano-IA, como até que ponto a IA pode ou deve apoiar decisões médicas ou mesmo tomá-las<sup>1</sup>.

Apesar de ser uma tecnologia-chave atualmente, em muitos casos será necessário entender como a decisão de uma máquina foi tomada e avaliar a explicação para tal escolha<sup>1</sup>. À medida que a IA avança, os marcos bioéticos precisam ser adaptados para incluir os problemas que esses sistemas podem apresentar, assim como o desenvolvimento dessas tecnologias precisa ser adaptado para incorporar princípios bioéticos<sup>3</sup>. Assim, esta revisão sistemática da literatura buscou identificar as aplicações da inteligência artificial em saúde e suas implicações bioéticas.

## Método

Esta revisão sistemática seguiu o Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA Statement)<sup>9</sup> e utilizou a estratégia População, Intervenção, Comparação, Desfecho (PICO). Esse método é utilizado na prática baseada em evidências e é recomendado para estruturar a busca bibliográfica de evidências para revisões<sup>10</sup>. A Tabela 1 descreve a estratégia PICO utilizada nesta revisão para responder à questão: quais são as aplicações da inteligência artificial em saúde e suas implicações bioéticas?

**Tabela 1.** Descrição da estratégia PICO

<b>P</b>	indivíduos, pessoas
<b>I</b>	o uso da inteligência artificial na saúde
<b>C</b>	não aplicável
<b>O</b>	implicações bioéticas

A busca pelos artigos foi realizada manualmente nas bases de dados ScienceDirect e Scopus em outubro de 2021. Foram utilizados os descritores em ciências da saúde (DeCS/MeSH) “inteligência artificial”, “bioética” e “saúde”, inter-relacionados pelo operador booleano “and”.

Os critérios de inclusão estabelecidos consistiram em artigos científicos publicados entre 2017 e 2021, em inglês e com acesso online. Livros, capítulos de livros, teses, dissertações, trabalhos

apresentados e publicados em eventos, artigos de revisão e notas editoriais foram excluídos.

Os títulos e resumos dos artigos encontrados foram lidos em uma primeira seleção. Os artigos selecionados foram lidos na íntegra, observando-se rigorosamente os critérios de inclusão e exclusão, verificação de duplicidade e adequação ao tema da pesquisa. Os artigos selecionados para revisão foram sistematizados em um quadro para a análise dos resultados. As etapas de seleção e leitura na íntegra dos artigos foram realizadas por três revisores independentes.

## Resultados

No total, foram identificados 102 artigos nas bases de dados seguindo os critérios de busca. Destes, 72 foram encontrados na base de dados ScienceDirect e 30 na Scopus. Após a leitura dos títulos e resumos, 86 artigos foram deletados por não abordarem o tema proposto. Um artigo foi excluído por se tratar de duplicata. Assim,

selecionaram-se 15 artigos potencialmente elegíveis, dos quais 11 foram mantidos após leitura na íntegra. A Figura 1 ilustra o processo de seleção.

Do total selecionado para inclusão na revisão, sete (63,64%) publicações foram acessadas na ScienceDirect e quatro (36,36%) na Scopus. Quanto ao ano de publicação, sete (63,64%) dos artigos foram publicados em 2021, e um (9,09%) artigo foi publicado a cada ano no restante do período estabelecido, ou seja, entre 2017 e 2020. Na análise da origem dos estudos, a maioria (45,45%) foi desenvolvida na América do Norte ou em colaboração com instituições daquele continente.

Quanto ao tipo de estudo, a maioria é transversal (45,45%), mas também foram encontrados estudos prospectivos (9,09%), estudos de caso (9,09%) e descritivos (9,09%), estudos exploratórios (9,09%), ensaios clínicos randomizados (9,09%) e estudo de abordagem multidimensional (9,09%).

Com base nos 11 trabalhos selecionados, elaborou-se uma tabela para análise dos objetivos, tipo de estudo, principais resultados e conclusões (Tabela 2).

**Figura 1.** Fluxograma do processo de seleção dos estudos para revisão sistemática

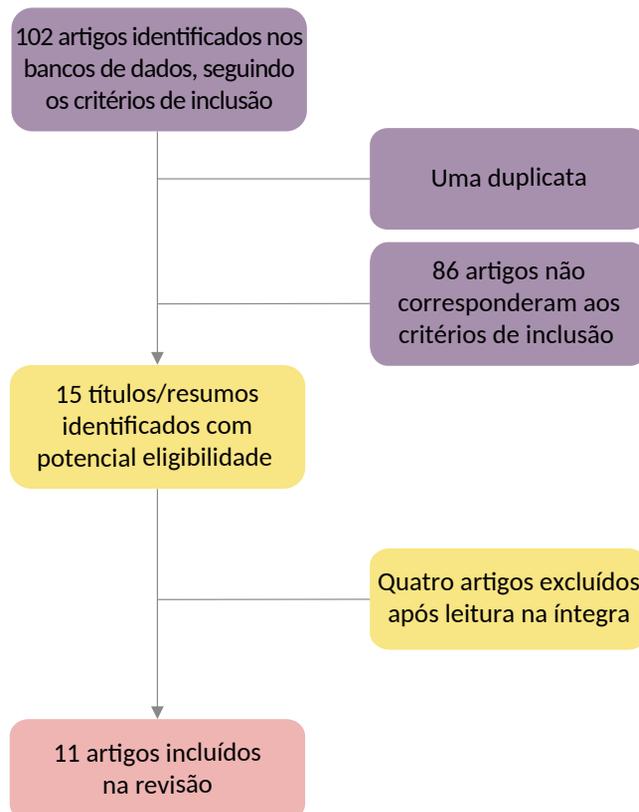


Tabela 2. Análise das principais informações apresentadas nas publicações: objetivo principal, tipo de estudo, principais resultados e conclusões

Autores; ano	Objetivo principal	Tipo de estudo	Principais resultados	Conclusões
Fernandes e colaboradores; 2019 <sup>11</sup>	Fornecer resultados que possam subsidiar o estudo da pesquisa médica no processo de tomada de decisão em bioética clínica, particularmente nos casos de eutanásia.	Transversal	Dados processados por métodos de seleção de características foram usados para criar modelos capazes de prever a decisão de eutanásia usando ML e um rastreador ocular. Experimentos estatísticos mostraram que o modelo preditivo resultante do algoritmo <i>perceptron</i> multicamadas (MLP) levou a melhor desempenho. Resultados interessantes (normas e regras) para a tomada de decisão bioética foram extraídos de simulações com modelos MLP. Alguns participantes tomaram uma decisão racional, respeitando o código de ética dos profissionais de enfermagem e o código penal brasileiro, o qual considera eutanásia como homicídio. Outros consideraram o aspecto emocional, vinculando essa decisão ao sofrimento do paciente.	O bom desempenho apresentado pelo modelo preditivo demonstra que a abordagem de investigação proposta pode ser utilizada para testar hipóteses científicas relacionadas à atenção visual e à tomada de decisão, verificando em que medida a visão é um fator determinante na tomada de decisão, particularmente na bioética clínica, quando se trata de questões de terminalidade da vida.
Silva, Lehoux, Hagemeister; 2018 <sup>12</sup>	Avaliar se uma inovação se qualifica como inovação responsável em saúde utilizando uma ferramenta desenvolvida em três etapas: triagem, avaliação e classificação; e discutir os aspectos políticos do uso da ferramenta.	Delphi (prospectivo)	O instrumento de triagem e avaliação da inovação responsável em saúde desenvolvido foi julgado por especialistas e, após a segunda rodada de comentários sobre o tema, chegou-se a um consenso para 16 das 20 questões quanto à importância, clareza e adequação da estrutura do instrumento. A sustentabilidade dos sistemas de saúde é prejudicada pela forma atual como as inovações em saúde são concebidas e colocadas no mercado. Houve consenso sobre a maioria dos critérios, atributos e escalas do instrumento. O uso futuro do instrumento pode contribuir para o desenvolvimento de inovações que proporcionem maior valor social.	O desenvolvimento desta ferramenta ajudará a preencher uma importante lacuna de conhecimentos e políticas, esclarecendo as decisões tomadas numa fase inicial pelas partes interessadas na inovação, tais como investidores, criadores de tecnologia, agências de financiamento de pesquisa e decisores políticos.
Lysaght e colaboradores; 2019 <sup>13</sup>	Abordar e analisar a estrutura ética para <i>big data</i> em saúde e pesquisa para demonstrar como a tomada de decisão pode ser baseada nela para o desenvolvimento e implementação de sistemas de suporte assistidos por IA em saúde de forma ética e responsável <sup>5</sup> .	Estudo de caso	Sistemas de apoio à decisão clínica (CDSS) são programas que geram informações em saúde. Os que usam ML e IA são complexos, razão pela qual médicos devem ser treinados para melhorar suas habilidades de tomada de decisão clínica ao usar esse tipo de recurso. Os algoritmos de IA do CDSS podem reforçar preconceitos sociais, mas também podem trazer benefícios, como sistemas de saúde pública mais eficientes. A inclusão e análise dos dados acontece de forma contínua, contribuindo com informações sobre práticas futuras adequadas. Alimentar o CDSS com dados de pacientes pode causar conflitos sobre o duplo papel dos médicos – cuidado e pesquisa. Os valores substantivos elencados são integridade profissional e equidade; e os valores processuais são transparência e prestação de contas. O CDSS assistido por IA deve ser explicável. A decisão final deve ser tomada pelo profissional de saúde e deve-se considerar que pode haver julgamentos morais que o programa não é capaz de fazer. O estudo de caso aborda a utilização desses programas em uma unidade de terapia intensiva. Apesar de trazer benefícios financeiros, essa ferramenta poderia gerar questões éticas como economia sobre saúde, desconfiância nas recomendações e preocupação com responsabilidade.	Dado o aumento dos custos dos cuidados em saúde, o desenvolvimento e a implementação da assistência baseada em IA na tomada de decisões clínicas talvez sejam inevitáveis. Os valores de integridade e responsabilidade profissionais desempenharão um papel mais proeminente ao nível dos cuidados prestados aos doentes, ao passo que os valores de justiça e o potencial de dano para o grupo devem ser equilibrados com imperativos de benefício público a nível social. Os imperativos de benefício público a nível social devem ser utilizados para o equilíbrio. A transparência afeta tanto a confiança na profissão médica como nos sistemas de saúde.

continua...

Tabela 2. Continuação

Autores; ano	Objetivo principal	Tipo de estudo	Principais resultados	Conclusões
Cawthorne, RobbIns-Van Wynsberghe; 2020 <sup>14</sup>	Criar uma estrutura ética a ser aplicada durante o desenho, o desenvolvimento, a implementação e a avaliação de drones em saúde pública.	Descritivo	<p>A hierarquia de valores utilizada consiste em princípios éticos, valores humanos, padrões e requisitos de <i>design</i>.</p> <p>A criação da estrutura considerou os quatro princípios bioéticos, além de um quinto princípio da ética em IA: a explicabilidade.</p> <p>A benevolência no campo da saúde pode ser traduzida em valores de bem-estar humano (e animal não humano), empregos e habilidades humanas e sustentabilidade ambiental.</p> <p>A não maleficência engloba privacidade, segurança, proteção, tranquilidade, empregos, habilidades humanas e sustentabilidade ambiental.</p> <p>A autonomia inclui o livre-arbítrio, os valores humanos, a responsabilidade e a confiança.</p> <p>A justiça inclui a distribuição equitativa de benefícios e indenizações.</p> <p>A adoção de drones na saúde pode levar à redução da infraestrutura de saúde local, reduzindo a assistência pessoal. No entanto, eles também podem conectar pessoas em lugares remotos a serviços modernos.</p> <p>A explicabilidade lida com a facilidade com que os sistemas podem ser compreendidos.</p> <p>O uso de uma estrutura ética é especialmente útil para aqueles com experiência limitada em ética tecnológica.</p>	<p>Os princípios éticos são abstratos e necessitam de maior contextualização e especificação para reflexão.</p> <p>A criação desta estrutura ética reforça o valor da integração da ética na prática e serve de modelo para o desenho e desenvolvimento em domínios de drones e não drones.</p> <p>A estrutura ajudou a identificar e refinar potenciais benefícios e mitigar riscos.</p>
Antes e colaboradores; 2021 <sup>15</sup>	Desenvolver uma nova medida e avaliar a abertura e a extensão das preocupações e dos benefícios percebidos em relação às tecnologias de saúde baseadas em IA em uma amostra de adultos nos Estados Unidos.	Transversal	<p>Os participantes estavam moderadamente abertos a tecnologias de saúde baseadas em IA, mas houve variação dependendo do tipo de aplicação.</p> <p>A confiança no sistema de saúde e na tecnologia foi o correlato mais forte e consistente de abertura, preocupação e percepção de benefício.</p> <p>Os participantes mais velhos eram menos abertos às tecnologias e os homens eram mais abertos do que as mulheres.</p> <p>O emprego de período integral foi associado a uma maior abertura e a uma menor preocupação.</p> <p>As duas tecnologias que fizeram previsões sobre doenças graves – risco de ataque cardíaco e probabilidade de sobrevivência ao câncer – foram as mais bem avaliadas.</p>	<p>A abertura dos participantes parece tênue, sugerindo que estratégias de promoção precoce e experimentos com novas tecnologias de IA podem influenciar fortemente as opiniões sobre o assunto.</p> <p>Podem ser necessário abordar a confiança no direcionamento da aceitação dessas inovações na área da saúde.</p>
Battie e colaboradores; 2021 <sup>16</sup>	Entender as melhores práticas no compartilhamento de dados de pacientes em instituições de saúde.	Exploratório	<p>Um grupo de trabalho identificou cinco amplos domínios de atividades importantes para a colaboração usando dados de pacientes: privacidade, consentimento informado, padronização de elementos de dados, contratos com fornecedores e avaliação de dados.</p> <p>Foram apresentados os métodos e a compreensão ética dos marcos legais comumente utilizados para esses fins, bem como o desenho do fluxo de dados que pode ajudar a informar como as permissões são criadas e revogadas.</p> <p>Uma descrição da cuidadosa preparação e anotação de conjuntos de dados é necessária quando se discute o anonimato e a desidentificação no zelo pela privacidade, apontando tecnicamente as dificuldades.</p> <p>O volume de dados necessários para preparação do algoritmo de IA é muito alto e, portanto, a preparação desses dados deve ocorrer de forma segura e compartilhável com seus proprietários (pacientes).</p>	<p>Criar uma relação de compartilhamento de dados envolve complexidade ética e tecnologia da informação.</p> <p>O anonimato e a privacidade dos pacientes mantêm a confiança e protegem as entidades que buscam compartilhar dados com segurança.</p>

continua...



Tabela 2. Continuação

Autores; ano	Objetivo principal	Tipo de estudo	Principais resultados	Conclusões
Green e colaboradores; 2021 <sup>17</sup>	Desenvolver ferramentas integradas em sistemas digitais de saúde para apoiar a tomada de decisão compartilhada e otimizar a preparação para o tratamento da doença renal crônica.	Ensaio clínico randomizado	Usando as ferramentas, 243 (24%) de 1.032 pacientes em quatro clínicas de nefrologia foram identificados com alto risco de evolução para insuficiência renal em dois anos. Especialistas em transição renal inscreveram 117 (48%) pacientes de alto risco até o final do primeiro ano de pesquisa. Enfermeiros usaram o aplicativo em 100% dos pacientes para documentar 287 etapas de planejamento da terapia renal substitutiva. Todos os especialistas em transição renal (100%) avaliaram a facilidade de uso e utilidade do instrumento, concordando ou concordando totalmente com todos os itens.	Enfermeiros relataram que as ferramentas desenvolvidas facilitaram a identificação de pacientes que necessitam de apoio e suas atividades de navegação. Houve, ainda, a rápida identificação de pacientes que precisam de tomada de decisão compartilhada e informada e sua preparação para tratamentos de substituição renal.
Martinho, Kroesen, Chorus; 2021 <sup>18</sup>	Obter informações sobre padrões de raciocínio e opiniões morais sobre IA em saúde de pessoas envolvidas na prática médica.	Transversal	Com base nas questões dos médicos sobre ética em torno da IA em saúde, quatro perspectivas principais foram identificadas: 1. A IA é uma ferramenta útil: permite que os médicos façam o que são treinados para fazer. 2. Regras e regulamentos são cruciais: as empresas privadas se preocupam apenas com dinheiro. 3. Ética é suficiente: as empresas privadas são confiáveis. 4. Ferramentas de IA explicáveis: o aprendizado é necessário e inevitável. Todas as perspectivas consideram que os médicos devem participar do processo de criação de tecnologias de IA em saúde, contribuindo para a explicabilidade. Médicos estão mais preocupados com o papel das grandes empresas na área da saúde e menos atentos ou preocupados com questões como equidade, preconceito e desigualdades em saúde.	Cada perspectiva fornece <i>insights</i> valiosos e muitas vezes contrastantes sobre questões éticas que devem ser operacionalizadas e consideradas no desenho e desenvolvimento de IA em saúde.
Shen e colaboradores; 2021 <sup>19</sup>	Analisar pesquisas de ressonância magnética (RM) altamente portáteis em ambientes internacionais remotos e com recursos limitados para criar orientação ética e legal em um cenário global complexo.	Transversal	É necessário assegurar que as comunidades locais sejam parceiras no empreendimento de pesquisa e garantir o valor social local do estudo. Os estudos de campo sobre RM precisam se responsabilizar pela segurança dos participantes e de todos ao seu redor. É importante prestar atenção às normas de privacidade e segurança de dados (locais e internacionais). É necessário identificar se a amostra na qual o modelo de IA foi treinado era diversa para que as previsões sejam mais precisas, considerando uma diversidade de fatores. Os resultados dos exames precisam ser divulgados aos participantes de forma esclarecedora e, em se tratando de achados incidentais, uma questão desafiadora é como fornecer suporte clínico e encaminhamento nessas comunidades remotas.	<i>Scanners</i> de ressonância magnética mais acessíveis e portáteis oferecem oportunidades para abarcar as necessidades de pesquisa não atendidas e as iniquidades em saúde em ambientes internacionais remotos e com recursos limitados. As comunidades locais devem ser parceiras contínuas na cocriação do conhecimento. A pesquisa deve produzir valor local para justificar os riscos e minimizar a possibilidade de abuso.

continua...

Tabela 2. Continuação

Autores; ano	Objetivo principal	Tipo de estudo	Principais resultados	Conclusões
Spiegel, Barker, Kistnasamy; 2021 <sup>20</sup>	<p>Descrever e avaliar a aplicação da IA no desenvolvimento de diagnósticos auxiliados por computador para apoiar a adjudicação mais eficiente de reivindicações para ex-mineradores de ouro com doença pulmonar ocupacional na África Austral.</p> <p>Transversal</p>		<p>Os resultados foram correlacionados com os princípios bioéticos. Beneficência: a IA poderia fornecer um julgamento mais consistente do que vários profissionais com diferentes níveis de habilidade. Não maleficência: manter a privacidade e a segurança dos dados e evitar a infiltração de determinadas tendências nos sistemas de tomada de decisão. Autonomia: a IA pode levar a perdas na qualificação profissional, à medida que os profissionais passam a confiar mais na tecnologia. Protocolos devem ser estabelecidos e os profissionais, treinados para identificar falsos negativos ou falsos positivos. Justiça: quando a demanda do mercado é fraca por investimentos em IA e as instituições públicas não respondem ao seu uso, o fracasso em lidar com a tecnologia pode ser uma forma de manter as desigualdades. Isso destaca o uso oportuno de inovações para beneficiar aqueles que precisam.</p>	<p>Os esforços para superar os desafios técnicos na aplicação da IA devem ser seguidos desde o início para garantir seu uso ético.</p>
Stahl e colaboradores; 2021 <sup>21</sup>	<p>Capturar teoricamente e medir empiricamente os benefícios e as desvantagens da IA para o progresso humano, para além dos princípios do aprendizado de máquina, contrabalançar os benefícios técnicos e econômicos da IA e seus aspectos legais, sociais e éticos.</p> <p>Abordagem multidimensional</p>		<p>O discurso ético da IA é discutido em três correntes:</p> <ol style="list-style-type: none"> <li>1. Problemas relacionados à aplicação do ML: é difícil prever até que ponto os dados serão usados para a finalidade determinada – já que um perfil pessoal leva a uma classificação para fins diferentes da inicial.</li> <li>2. Questões sociais e políticas emergentes em uma sociedade digital: esses sistemas exigem acesso a grandes quantidades de dados para fins de treinamento e validação, o que gera desconfiada relacionada a autonomia das máquinas, substituição de humanos por máquinas, injustiça na distribuição de custos e benefícios e controle de dados, além das consequências.</li> <li>3. Questões metafísicas sobre a natureza da realidade e da humanidade: estas dizem respeito a quais máquinas devem ser autorizadas a decidir autonomamente. Consequências econômicas, emprego, justiça, liberdade, contato humano, autonomia individual, desigualdade, integridade, propriedade, uso militar, assimetria de poder, responsabilidade e sustentabilidade se enquadraram nessa categoria.</li> </ol>	<p>Atualmente, não há consenso entre as diversas abordagens de governança e segurança da informação. A legislação de direitos humanos pode resolver muitas questões sociais e éticas. A complexidade dos contextos e cenários é um fator que pluraliza as formas de abordá-los e permite que estudiosos ou profissionais mantenham uma visão geral. Atenção ao uso atual da IA e ML e, até certo ponto, aos sistemas sociotécnicos mais amplos.</p>



## Discussão

Embora sejam uma grande promessa para a saúde, a ética e os direitos humanos das novas tecnologias que usam IA devem estar no centro de sua concepção, implantação e uso<sup>8</sup>. A incorporação de princípios bioéticos no desenvolvimento dessas tecnologias pode ajudar a proteger os direitos dos pacientes, minimizar riscos, estabelecer responsabilidades e instituir métricas robustas para estudar sua efetividade e seus benefícios<sup>3</sup>.

Desse ponto de vista, alguns estudos selecionados nesta revisão estabeleceram uma relação direta entre seus resultados e os princípios bioéticos da beneficência, não maleficência, autonomia e justiça<sup>14,20</sup>. Outro princípio considerado na ética da IA é a explicabilidade, que trata do entendimento das ferramentas que utilizam essa tecnologia<sup>13,14,18</sup>. A elaboração desses sistemas deve considerar os princípios da beneficência, manutenção de valores, responsabilidade e transparência – o que compreende a explicabilidade<sup>2</sup>.

A explicabilidade é fundamental, pois a confiança necessária e duradoura pode ser construída a partir dela para promover a aceitação da IA com futuros usuários<sup>1</sup>. Em um estudo, a confiança no sistema de saúde e na tecnologia foi o correlato mais forte e consistente de abertura, preocupação e percepção de benefício<sup>15</sup>. Em outra pesquisa, os participantes se sentiram desconfortáveis em confiar apenas nas recomendações feitas por uma ferramenta de IA, sem serem capazes de avaliar diretamente seu raciocínio<sup>7</sup>. Embora os sistemas sejam importantes no processo de tomada de decisão, os especialistas devem ser capazes de entender e refazer esse processo<sup>1</sup>.

Mesmo especialistas em desenvolvimento de IA não conseguem determinar como entradas são transformadas em saídas, ou seja, como um perfil pessoal gera uma decisão<sup>21</sup>. A imprevisibilidade no ML se torna, assim, a chave para a discussão, pois é um método para automatizar a análise de dados usando algoritmos que identificam iterativamente padrões nos dados e aprendem com eles<sup>4</sup>. As consequências são impossíveis de mensurar quando a máquina é programada para aprender sozinha<sup>21</sup>.

O que as máquinas devem poder decidir de forma autônoma e quem é responsável pela decisão? Ainda não há respostas para essas perguntas. Por enquanto,

médicos devem participar e entender a IA e, em seguida, decidir de forma autônoma, embora com base na recomendação da ferramenta<sup>21</sup>.

Assim como a imprevisibilidade, a previsão é outro ponto relevante discutido na IA. Um instrumento capaz de prever o risco e priorizar pacientes para o tratamento da doença renal crônica auxiliou na rápida identificação de indivíduos com necessidade de substituição renal<sup>17</sup>. Outro estudo avaliou a reação e decisão de enfermeiros sobre a prática da eutanásia em pacientes terminais utilizando a tecnologia de rastreamento ocular (*eye-tracking*), trazendo uma divisão entre decisões racionais e emocionais, demonstrando a eficácia do modelo de predição<sup>11</sup>. Mesmo quando a IA indica tratamentos, profissionais da saúde e pacientes podem fazer julgamentos morais que o programa não é capaz de produzir<sup>13</sup>.

Apesar das indicações da IA, a decisão final em relação à saúde deve ser tomada por um profissional<sup>13</sup>. Estudo realizado com pacientes atendidos em uma unidade básica de saúde constatou que eles acreditam que seus médicos devem protegê-los dos danos decorrentes de erros da IA, de modo que a decisão final deve caber ao médico e à seguradora de saúde<sup>7</sup>. Nesse sentido e considerando a má prática, as seguradoras devem ser claras quanto à cobertura quando as decisões são tomadas por sistemas de IA, mesmo que parcialmente<sup>6</sup>.

É provável que a IA no contexto da saúde dê aos médicos mais tempo para outras tarefas, como estabelecer contato direto com seus pacientes<sup>6</sup>. No entanto, a IA pode causar conflitos sobre o papel dos médicos, que devem prestar cuidados enquanto alimentam o sistema com dados do paciente para fins de suporte e pesquisa clínica<sup>13</sup>. À medida que a IA é implementada para tarefas clínicas específicas, os papéis dos profissionais da saúde continuarão a evoluir, uma vez que vários módulos de IA serão incorporados ao cuidado<sup>6</sup>.

A integridade profissional é um dos três principais valores éticos da *big data* na tomada de decisão, uma vez que engloba a responsabilidade pelo paciente, tornando a formação médica central<sup>13</sup>. Médicos precisarão se adaptar às suas novas funções como integradores e intérpretes de informações e como apoiadores dos pacientes, e o sistema de educação médica terá que lhes prover as ferramentas e os métodos necessários<sup>6</sup>.

As ferramentas de IA são úteis, mas os médicos não foram treinados para compreendê-las e, ao mesmo tempo, sua participação ativa no seu desenvolvimento é vital<sup>18</sup>.

Para enfrentar os desafios, pesquisadores de IA e clínicos devem trabalhar juntos para priorizar e desenvolver aplicativos que atendam às necessidades clínicas cruciais<sup>6</sup>. Dado o risco potencial de causar danos, o envolvimento de bioeticistas no desenho dessas tecnologias é necessário<sup>3</sup>. Colaborações multidisciplinares e multissetoriais serão necessárias para facilitar o desenvolvimento e a implantação de aplicativos de IA na área médica<sup>6</sup>.

Vários conjuntos de dados devem ser manipulados para desenvolver ferramentas de IA, o que significa que essa tecnologia só pode avançar com o uso de *big data*<sup>5</sup>. Para desenvolver e treinar algoritmos, os dados dos registros de saúde precisam ser acessados, e os pacientes nem sempre estão cientes de que tais informações estão sendo compartilhadas<sup>16</sup>. Esse dilema requer um estudo aprofundado dessas tecnologias emergentes e a evolução dos princípios bioéticos de privacidade e confidencialidade do paciente<sup>3</sup>.

Discussões em torno da ética dos registros eletrônicos em saúde e da IA têm se concentrado principalmente na privacidade, na confidencialidade, na segurança de dados, no consentimento informado e na propriedade de dados. No entanto, a relevância de cada um varia dependendo das diferenças culturais, da alfabetização, da relação paciente-profissional, da infraestrutura disponível e da regulamentação de cada país<sup>4</sup>.

Apesar das muitas discussões sobre a ética do *big data* em variados contextos, pouca orientação está disponível sobre quais valores estão em jogo e como as decisões devem ser tomadas em um ambiente de saúde e pesquisa cada vez mais complexo<sup>5</sup>. O anonimato e a privacidade do paciente são essenciais para criar confiança, além da necessidade de listar níveis de permissão nas conexões entre instituição de saúde, titular dos dados (paciente) e instituições com fins lucrativos relacionadas (planos de saúde, provedores e parceiros de tratamento/exame)<sup>16</sup>.

Várias questões éticas complexas surgem quando se considera o uso de *big data* e, portanto, um arcabouço ético para abordá-las e orientar ações é importante<sup>5</sup>. Da mesma forma, esse planejamento

ético deve ser aplicado em outros projetos que envolvam IA na área da saúde, como uma estrutura a ser utilizada em todas as etapas da obtenção de drones em saúde pública, envolvendo princípios éticos, valores humanos, padrões e requisitos de projeto<sup>14</sup>.

O papel das empresas de *big data* e IA, segmento voltado para o lucro, é questionado diante das implicações éticas na organização da saúde<sup>18</sup>. A sustentabilidade dos sistemas de saúde é atualmente prejudicada pela forma como as inovações em saúde são concebidas e colocadas no mercado<sup>12</sup>. Os aspectos financeiros e o poder concentrado nos detentores de dados e nas tecnologias parecem ser mais relevantes do que aqueles relacionados a equidade, preconceito e desigualdades<sup>18</sup>.

Quem controlará ou lucrará com a aplicação da IA ainda precisa ser determinado, portanto, a prioridade deve ser o equilíbrio entre as salvaguardas regulatórias e as forças de mercado para garantir que os pacientes se beneficiem<sup>6</sup>. A criação de regulamentos é um ponto fundamental na evolução desse tema<sup>18</sup>.

Apesar de ter imenso potencial para corrigir erros humanos e melhorar a prestação de cuidados, aplicativos de IA baseados em ML podem agravar vieses<sup>3</sup>. Pacientes se preocupam com a possibilidade de as ferramentas de IA reforçarem os vieses existentes, o que pode ocorrer se um conjunto de dados de aprendizagem for tendencioso ou os desenvolvedores incorporarem involuntariamente seu próprio viés em um algoritmo<sup>7</sup>, o que já é uma possibilidade com algoritmos de CDSS<sup>13</sup>.

A IA pode ser transformadora de aspectos sociais para a saúde pública em países com poucos recursos, pois à medida que esses locais se tornam mais conectados e criam dados de maior qualidade, a capacidade das ferramentas de IA de enfrentar os desafios de saúde provavelmente aumentará<sup>4</sup>.

A portabilidade é outro aspecto importante quando se discute a disponibilidade de IA em locais remotos, por exemplo, *scanners* de ressonância magnética mais acessíveis e portáteis oferecem oportunidades para abordar necessidades de saúde não atendidas e iniquidades em ambientes remotos e com recursos limitados<sup>19</sup>. A infraestrutura requerida nesses ambientes demanda investimentos substanciais para ser implementada, nem sempre tendo o acesso necessário para transmitir grandes conjuntos de dados para sistemas

em nuvem<sup>4</sup>. Nesse caso, tornam-se relevantes as regulamentações de privacidade e segurança de dados (locais e internacionais) e a forma de fornecer suporte e encaminhamento clínico nessas comunidades localizadas em locais remotos<sup>19</sup>.

No estudo da IA no diagnóstico assistido por computador, os pedidos de indenização para trabalhadores de minas com câncer de pulmão ocupacional foram acelerados<sup>20</sup>. Mesmo com baixa demanda por investimentos públicos e privados e alto risco de desqualificação da ferramenta na esfera jurídica, essa abordagem proporciona um benefício social comprovado quando protocolos éticos e monitoramento fazem parte do processo desde o início.

Duas tecnologias de IA preditiva foram melhor avaliadas entre idosos e indivíduos com menor letramento digital, demonstrando que o alívio para doenças graves, como infarto e câncer, casos envolvendo urgência, vulnerabilidade e risco, pode ser aceito sob uma premissa de confiabilidade<sup>15</sup>.

Embora pacientes se apresentem geralmente entusiasmados com a possibilidade de a IA melhorar seus cuidados, eles também estão preocupados com as possibilidades de segurança e supervisão<sup>7,15</sup>. Preocupações de que as ferramentas de IA possam aumentar os custos de saúde e que tais custos possam ser repassados aos pacientes também são levantadas<sup>7</sup>. No entanto, como as decisões médicas são tomadas por sistemas automatizados, o custo-efetividade e o racionamento de despesas podem ser priorizados<sup>3</sup>.

O uso de IA na tomada de decisão clínica é provavelmente inevitável, dados os custos crescentes da assistência à saúde<sup>13</sup>. Estudos prospectivos poderão melhor identificar as fragilidades dos modelos de IA em ambientes clínicos heterogêneos e indicar maneiras de integrá-los aos fluxos de trabalhos clínicos atuais<sup>6</sup>. No entanto, os esforços para superar os desafios técnicos na implantação da IA devem seguir, desde o início, aspectos éticos<sup>20</sup>.

Corroborando a literatura, evidenciam-se benefícios e riscos no uso da IA na saúde. A fim de maximizar os benefícios para o interesse público e limitar as ameaças, a Organização Mundial da Saúde propõe seis princípios para o uso da IA que se concentram nos tópicos abordados nesta

discussão, a saber: 1) proteger a autonomia humana; 2) promover o bem-estar, a segurança e o interesse público; 3) assegurar a transparência, a explicabilidade e a inteligibilidade; 4) promover a responsabilidade e a prestação de contas; 5) garantir a inclusão e a equidade; e 6) promover tecnologias responsivas e sustentáveis<sup>8</sup>.

## Considerações finais

Entender a IA na área da saúde requer compreender seus passos: alimentação de dados, desenvolvimento de algoritmos e saída de decisões. Nesse processo, os quatro princípios bioéticos, somados aos da explicabilidade e aos da legislação de direitos humanos, são os construtos garantidores da ética.

A garantia do anonimato e do consentimento para o uso de dados, sem fins comerciais de maior relevância na implementação da IA, permanece diversificada e juridicamente regionalizada nos procedimentos de pesquisa, diagnóstico e tratamento. A transparência é, portanto, a principal base da IA em saúde e as perguntas são mais amplas do que as respostas.

Entender como as decisões foram programadas, tirando-as da caixa preta da tecnologia incompreensível, é o aspecto central. Quando a máquina aprende sozinha, o médico pode ser responsabilizado pelas decisões? As regulamentações precisam evoluir, tanto na definição de responsabilidades quanto no entendimento e na confiança na autonomia dessas decisões geradas pela IA. Permitir que máquinas decidam por conta própria, com base em seu próprio aprendizado, é um conceito em evolução, uma vez que nem mesmo especialistas nesse desenvolvimento podem prever suas consequências, tanto em termos de assertividade e confiabilidade quanto na manutenção do uso inicialmente definido.

O que podemos afirmar é que estamos em um caminho sem volta, em que todas as questões levantadas neste artigo precisarão ser reguladas considerando acesso, privacidade, questões sociais e de justiça, ainda sendo tocadas por interesses de controle, poder e lucratividade.

*Esta pesquisa não recebeu subsídios específicos de agências de fomento dos setores público, comercial ou sem fins lucrativos.*

## Referências

1. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov* [Internet]. 2019 [acesso 5 nov 2021];9(4):e1312. DOI: 10.1002/widm.1312
2. Tai MCT. The impact of artificial intelligence on human society and bioethics. *Tzu Chi Med J* [Internet]. 2020 [acesso 5 nov 2021];32(4):339-43. DOI: 10.4103/tcmj.tcmj\_71\_20
3. Nabi J. How bioethics can shape artificial intelligence and machine learning. *Hastings Cent Rep* [Internet]. 2018 [acesso 5 nov 2021];48(5):10-3. DOI: 10.1002/hast.895
4. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Heal* [Internet]. 2018 [acesso 5 out 2021];3(4):e000798. DOI: 10.1136/bmjgh-2018-000798
5. Xafis V, Schaefer GO, Labude MK, Brassington I, Ballantyne A, Lim HY *et al.* An ethics framework for big data in health and research. *Asian Bioeth Rev* [Internet]. 2019 [acesso 5 nov 2021];11(3):227-254. DOI: 10.1007/s41649-019-00099-x
6. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng* [Internet]. 2018 [acesso 5 nov 2021];2(10):719-731. DOI: 10.1038/s41551-018-0305-z
7. Richardson JP, Smith C, Curtis S, *et al.* Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ Digit Med* [Internet]. 2021 [acesso 5 nov 2021];4(1):140. DOI: 10.1038/s41746-021-00509-1
8. World Health Organization. Ethics and governance of artificial intelligence for health [Internet]. Geneva: WHO Guidance; 2021 [acesso 4 nov 2021]. Disponível: <https://bit.ly/3Ry4M8P>
9. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* [Internet]. 2009 [acesso 5 set 2021];339(7716):332-6. DOI: 10.1136/bmj.b2535
10. Santos CMDC, Pimenta CADM, Nobre MRC. A estratégia PICO para a construção da pergunta de pesquisa e busca de evidências. *Rev Latinam Enferm* [Internet]. 2007 [acesso 5 set 2021];15(3):508-11. DOI: 10.1590/S0104-11692007000300023
11. Fernandes DL, Siqueira-Batista R, Gomes AP, Souza CR, Costa IT, Cardoso FSL *et al.* Investigation of the visual attention role in clinical bioethics decision-making using machine learning algorithms. *Procedia Comput Sci* [Internet]. 2017 [acesso 11 out 2022];108:1165-74. DOI: 10.1016/j.procs.2017.05.032
12. Silva HP, Lehoux P, Hagemester N. Developing a tool to assess responsibility in health innovation: results from an international delphi study. *Heal Policy Technol* [Internet]. 2018 [acesso 11 out 2022];7(4):388-96. DOI: 10.1016/j.hlpt.2018.10.007
13. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-assisted decision-making in healthcare. *Asian Bioeth Rev* [Internet]. 2019 [acesso 11 out 2022];11(3):299-314. DOI: 10.1007/s41649-019-00096-0
14. Cawthorne D, Robbins-van Wynsberghe A. An ethical framework for the design, development, implementation, and assessment of drones used in public healthcare. *Sci Eng Ethics* [Internet]. 2020 [acesso 11 out 2022];26(5):2867-91. DOI: 10.1007/s11948-020-00233-1
15. Antes AL, Burrous S, Sisk BA, Schuelke MJ, Keune JD, DuBois JM. Exploring perceptions of healthcare technologies enabled by artificial intelligence: an online, scenario-based survey. *BMC Med Inform Decis Mak* [Internet]. 2021 [acesso 11 out 2022];21(1):221. DOI: 10.1186/s12911-021-01586-8
16. Batlle JC, Dreyer K, Allen B, Cook T, Roth CJ, Kitts AB *et al.* Data sharing of imaging in an evolving health care world: report of the ACR Data Sharing Workgroup, part 1: data ethics of privacy, consent, and anonymization. *J Am Coll Radiol* [Internet]. 2021 [acesso 11 out 2022];18(12):1646-54. DOI: 10.1016/j.jacr.2021.07.014
17. Green JA, Ephraim PL, Hill-Briggs F, Browne T, Strigo TS, Hauer CL *et al.* Integrated digital health system tools to support decision making and treatment preparation in CKD: the PREPARE NOW Study. *Kidney Med* [Internet]. 2021 [acesso 11 out 2022];3(4):565-75.e1. DOI: 10.1016/j.xkme.2021.03.009

18. Martinho A, Kroesen M, Chorus C. A healthy debate: exploring the views of medical doctors on the ethics of artificial intelligence. *Artif Intell Med* [Internet]. 2021 [acesso 11 out 2022];121(October):102190. DOI: 10.1016/j.artmed.2021.102190
19. Shen FX, Wolf SM, Bhavnani S, Deoni S, Elison JT, Fair D *et al.* Emerging ethical issues raised by highly portable MRI research in remote and resource-limited international settings. *Neuroimage* [Internet]. 2021 [acesso 11 out 2022];238:118210. DOI: 10.1016/j.neuroimage.2021.118210
20. Spiegel JM, Ehrlich R, Yassi A, Riera F, Wilkinson J, Lockhart K *et al.* Using artificial intelligence for high-volume identification of silicosis and tuberculosis a bio-ethics approach. *Ann Glob Heal* [Internet]. 2021 [acesso 11 out 2022];87(1):58. DOI: 10.5334/aogh.3206
21. Stahl BC, Andreou A, Brey P, Hatzakis T, Kirichenko A, Macnish K *et al.* Artificial intelligence for human flourishing: beyond principles for machine learning. *J Bus Res* [Internet]. 2021 [acesso 11 out 2022];124:374-88. DOI: 10.1016/j.jbusres.2020.11.030

Mariele Abadia Elias – Graduada – marieleabadia@uniplaclages.edu.br

 0000-0001-5724-8400

Luciana Arruda Faversoni – Graduada – luarrudaf@uniplaclages.edu.br

 0000-0002-5495-1122

Josiane Aparecida Vieira Moreira – Graduada – josiane@uniplaclages.edu.br

 0000-0003-2683-5172

Anelise Viapiana Masiero – Doutora – anemasiero@uniplaclages.edu.br

 0000-0003-3203-5876

Natalia Veronez da Cunha – Doutora – nat\_cunha@uniplaclages.edu.br

 0000-0002-8522-5836

#### Correspondência

Mariele Abadia Elias – Av. Marechal Castelo Branco, 170, bloco 2, Universitário CEP 88509-900. Lages/SC, Brasil.

#### Participação das autoras

Mariele Abadia Elias, Luciana Arruda Faversoni e Josiane Aparecida Vieira Moreira foram responsáveis pela conceituação, pelo método, pela pesquisa, pela redação e pela elaboração do artigo original. Anelise Viapiana Masiero e Natalia Veronez da Cunha Bellinati foram responsáveis pela elaboração, revisão e edição do artigo.

**Recebido:** 7.4.2023

**Revisado:** 6.7.2023

**Aprovado:** 13.10.2023