REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

Inteligência Artificial em Saúde entre 2020 e 2026: revisão de escopo sobre aplicações, maturidade clínica e desafios sociotécnicos

Artificial Intelligence in Healthcare between 2020 and 2026: A Scoping Review of Applications, Clinical Maturity, and Sociotechnical Challenges

Inteligencia Artificial en Salud entre 2020 y 2026: revisión de alcance sobre aplicaciones, madurez clínica y desafíos sociotécnicos

Maria Amelia Dos Santos Peres
Universidade nacional ecologica-UNE
amelperss@gmail.com

Bruna Pereira de Oliveira
Universidade nacional ecológica-UNE
brunaoliveiramedicina@gmail.com

Maize Rodrigues Ferreira Miranda
Universidade nacional ecológica- UNE
maize.miranda@hotmail.com

Matheus Uzai
Universidade Nacional Ecológica- UNE
Matheus_uzai@hotmail.com

David Carlini Ochando
Universidade Nacional Ecológica
david.carlini@hotmail.com

Lucas Nogueira Pinheiro
Universidade Nacional Ecológica-UNE
lucasnogueirap@icloud.com

Marques Figueiredo Rosangela
Universidade Nacional Ecológica-UNE
rosangela40773@uecologica.edu.bo

Nívea Cristina Ribeiro
Universidade Nacional Ecológica-UNE
Nivea_ribeiro@yahoo.com

Hilton Giovani Neves
Universidade Nacional Ecológica-UNE
hgneves@gmail.com

Indaiá Elcia Fernandes Valente
Universidade Nacional Ecológica-UNE
indaiafernandes@hotmail.com

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

Michele Alves do Nascimento
Universidade Nacional Ecológica- UNE
michele.alves.nas@gmail.com

Resumo
Esta revisão de escopo sintetiza aplicações, níveis de maturidade clínica e desafios sociotécnicos da Inteligência Artificial (IA) em saúde entre 2020 e 2026. O relato foi orientado pelo PRISMA 2020, com buscas previstas em PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, ScienceDirect e Google Scholar. Em razão da ausência dos arquivos originais de exportação e triagem, os quantitativos do fluxo permanecem como campos de auditoria a serem preenchidos com os registros reais, sem estimativas. O corpus verificável no manuscrito compreende oito publicações temáticas de 2022 a 2024, analisadas por técnica, aplicação, tipo de evidência, validação e proximidade da prática clínica. A síntese foi reorganizada em quatro eixos: diagnóstico e apoio à decisão; modelos de linguagem e IA generativa; equidade, valores humanos e governança; e implementação clínica. Os estudos mostram predomínio de revisões, perspectivas e análises conceituais, com forte discussão sobre potencial técnico, mas evidência limitada de implementação prospectiva e benefício direto sobre desfechos de pacientes. A maturidade foi classificada em quatro níveis: desempenho retrospectivo interno, validação externa, implementação em fluxo clínico e benefício clínico demonstrado. O corpus concentra-se em autores e instituições dos Estados Unidos e do Reino Unido e publica exclusivamente em inglês, limitando a representatividade internacional e a transferência dos achados para sistemas de saúde de baixa e média renda. A análise evidencia que acurácia isolada não equivale a utilidade clínica: generalização, calibração, comparação com o cuidado habitual, integração ao trabalho, supervisão humana, privacidade, equidade e monitoramento pós-implantação são requisitos centrais. Conclui-se que a literatura examinada sustenta o potencial da IA, mas não autoriza afirmar adoção clínica consolidada ou melhora generalizada dos desfechos. Pesquisas futuras devem ampliar validações multicêntricas e prospectivas, diversidade geográfica, transparência dos dados e avaliação de resultados clínicos, organizacionais e distributivos.
Palavras-chave: Inteligência Artificial; Saúde Digital; Maturidade Clínica; Equidade Algorítmica.
Abstract
This scoping review synthesizes applications, levels of clinical maturity, and sociotechnical challenges of Artificial Intelligence (AI) in healthcare between 2020 and 2026. Reporting was guided by PRISMA 2020, with planned searches in PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, ScienceDirect, and Google Scholar. Because the original export and screening files were unavailable, flow counts remain audit fields to be completed from actual records, without estimates. The verifiable corpus comprises eight thematic publications from 2022 to 2024, analyzed by technique, application, evidence type, validation, and proximity to clinical practice. The synthesis was organized into four axes: diagnosis and decision support; language models and generative AI; equity, human values, and governance;

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

and clinical implementation. The corpus is dominated by reviews, perspectives, and conceptual analyses. It extensively discusses technical potential but provides limited evidence of prospective implementation and direct benefit to patient outcomes. Maturity was classified at four levels: internal retrospective performance, external validation, implementation in clinical workflow, and demonstrated clinical benefit. Publications are concentrated among authors and institutions in the United States and the United Kingdom and are exclusively in English, limiting international representativeness and transferability to low- and middle-income health systems. Accuracy alone does not establish clinical usefulness; generalizability, calibration, comparison with usual care, workflow integration, human oversight, privacy, equity, and post-deployment monitoring are essential. The reviewed literature supports the potential of AI but does not justify claims of consolidated clinical adoption or generalized outcome improvement. Future studies should prioritize multicenter prospective validation, geographic diversity, data transparency, and assessment of clinical, organizational, and distributive outcomes.
Keywords: Artificial Intelligence; Digital Health; Clinical Maturity; Algorithmic Equity.

Resumen
Esta revisión de alcance sintetiza las aplicaciones, los niveles de madurez clínica y los desafíos sociotécnicos de la Inteligencia Artificial (IA) en la atención sanitaria entre 2020 y 2026. El informe siguió las directrices PRISMA 2020, con búsquedas planificadas en PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, ScienceDirect y Google Scholar. Debido a que los archivos originales de exportación y selección no estaban disponibles, los recuentos del flujo permanecen como campos de auditoría que deberán completarse a partir de los registros reales, sin estimaciones. El corpus verificable comprende ocho publicaciones temáticas de 2022 a 2024, analizadas según la técnica, la aplicación, el tipo de evidencia, la validación y la proximidad a la práctica clínica. La síntesis se organizó en cuatro ejes: diagnóstico y apoyo a la toma de decisiones; modelos de lenguaje e IA generativa; equidad, valores humanos y gobernanza; e implementación clínica. El corpus está dominado por revisiones, perspectivas y análisis conceptuales. Aunque se analiza ampliamente el potencial técnico de la IA, existe evidencia limitada sobre su implementación prospectiva y sus beneficios directos en los resultados de los pacientes. La madurez se clasificó en cuatro niveles: desempeño retrospectivo interno, validación externa, implementación en el flujo de trabajo clínico y beneficio clínico demostrado. Las publicaciones se concentran en autores e instituciones de Estados Unidos y Reino Unido y están exclusivamente en inglés, lo que limita la representatividad internacional y la transferencia de los resultados a sistemas de salud de países de ingresos bajos y medios. La precisión por sí sola no determina la utilidad clínica; también son esenciales la generalización, la calibración, la comparación con la atención habitual, la integración en los flujos de trabajo, la supervisión humana, la privacidad, la equidad y el monitoreo posterior a la implementación. La literatura revisada respalda el potencial de la IA, pero no permite afirmar que exista una adopción clínica consolidada ni una mejora generalizada de los resultados. Los estudios futuros deberían priorizar la validación prospectiva

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

multicéntrica, la diversidad geográfica, la transparencia de los datos y la evaluación de resultados clínicos, organizacionales y distributivos.
Palabras clave: Inteligencia Artificial; Salud Digital; Madurez Clínica; Equidad Algorítmica.

1. Introdução

A Inteligência Artificial em saúde abrange aprendizado de máquina, aprendizado profundo, processamento de linguagem natural, modelos fundamentais e sistemas generativos. Essas abordagens operam sobre imagens, prontuários, sinais fisiológicos, dados genômicos e textos clínicos, mas diferem quanto ao objetivo, ao tipo de dado e ao estágio de validação. Por isso, comparações responsáveis precisam separar demonstração computacional, validação externa, integração ao trabalho assistencial e efeito sobre resultados de pacientes (Rajpurkar et al., 2022; Beam et al., 2023).

A literatura recente combina expectativas de ganho diagnóstico e produtividade com alertas sobre vieses, baixa generalização, opacidade e risco de automação inadequada. Dados clínicos registram práticas, desigualdades e decisões institucionais; portanto, não constituem representações neutras da população.

Modelos podem reproduzir ou ampliar essas assimetrias quando o contexto de produção dos dados e a distribuição dos erros não são examinados (Ferryman; Mackintosh; Ghassemi, 2023; Ghassemi; Gusev, 2024).

Grandes modelos de linguagem ampliaram o debate ao possibilitar sumarização, documentação e apoio informacional. Entretanto, respostas plausíveis podem conter erros, omissões e conteúdo não sustentado, exigindo supervisão humana e avaliação situada. Benefícios operacionais também não devem ser confundidos com melhora clínica, pois redução de tempo, desempenho em testes e aceitação profissional representam desfechos distintos (Lee; Bubeck; Petro, 2023; Shah, 2024).

A avaliação da IA deve considerar valores humanos, comparação com o cuidado disponível, impacto distributivo, privacidade, responsabilização e mudanças

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

no trabalho. Um sistema tecnicamente superior em uma base retrospectiva pode falhar em outra instituição, aumentar carga de trabalho, deslocar decisões ou produzir benefício desigual entre grupos (Kohane, 2024; Yu et al., 2024).

Diante disso, o objetivo desta revisão de escopo é sintetizar criticamente as aplicações de IA em saúde presentes no corpus de 2020 a 2026, classificando técnicas, áreas de uso, tipos de evidência e níveis de maturidade clínica. Como objetivos específicos, busca-se distinguir desempenho retrospectivo de validação externa, implementação e benefício clínico; examinar desafios sociotécnicos e éticos; avaliar concentração geográfica; e identificar lacunas para pesquisas futuras. Pesquisadores são mencionados apenas como autores das contribuições analisadas, não como objeto principal ou ranking de influência.

2. Metodologia

2.1 Delineamento e PRISMA

Esta revisão de escopo, com síntese narrativa do estado da arte, foi orientada pelo PRISMA 2020 (Page et al., 2021). A diretriz foi aplicada às etapas de identificação, deduplicação, triagem, elegibilidade e inclusão, sem que seu uso isolado caracterize revisão sistemática.

2.2 Pergunta e elegibilidade

Pergunta: quais linhas de pesquisa, aplicações clínicas, contribuições tecnológicas e desafios éticos caracterizam a Inteligência Artificial em saúde entre 2020 e 2026? Incluíram-se artigos completos, revisados por pares, publicados de janeiro de 2020 a junho de 2026, em português ou inglês, diretamente relacionados a aplicações, validação, segurança, equidade, interpretabilidade, governança, modelos fundamentais ou IA Generativa em saúde. Excluíram-se duplicatas, resumos, teses, capítulos, notícias, textos sem conteúdo integral e estudos sem

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

aderência. A limitação linguística decorreu da capacidade da equipe para extração confiável, podendo introduzir viés de idioma.

2.3 Fontes, datas e estratégias

Foram definidas PubMed/MEDLINE, Scopus, Web of Science Core Collection, IEEE Xplore, ScienceDirect e Google Scholar. As datas exatas não constavam dos registros fornecidos e devem ser inseridas: [BUSCA INICIAL] e [ÚLTIMA ATUALIZAÇÃO]. As expressões abaixo devem ser confrontadas com os históricos exportados de cada fonte.

Quadro 1 – Estratégia integral por fonte

Fonte/campos	Expressão adaptada	Filtros	Resultados
PubMed – MeSH e Title/Abstract	("Artificial Intelligence"[Mesh] OR "Machine Learning"[Mesh] OR "deep learning"[Title/Abstract] OR "large language model*"[Title/Abstract]) AND ("Delivery of Health Care"[Mesh] OR healthcare[Title/Abstract] OR medicine[Title/Abstract])	2020–jun. 2026; português/inglês; artigo	[n=5]
Scopus – TITLE-ABS-KEY	TITLE-ABS-KEY(("artificial intelligence" OR "machine learning" OR "deep learning" OR "foundation model*" OR "large language model*") AND (healthcare OR medicine OR "clinical decision support" OR "digital health" OR "medical imaging"))	2020–2026; português/inglês; article	[n= 7]
Web of Science – TS	TS=(("artificial intelligence" OR "machine learning" OR "deep learning" OR "foundation model*" OR "large language model*") AND (healthcare OR medicine OR "clinical decision support" OR	2020–2026; português/inglês; article	[n=6]

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

	"digital health" OR "medical imaging"))		
IEEE Xplore – All Metadata	("artificial intelligence" OR "machine learning" OR "deep learning" OR "foundation model*" OR "large language model*") AND (healthcare OR medicine OR "clinical decision support" OR "digital health" OR "medical imaging")	2020–2026; journals	[n=2]
ScienceDirect – title/abstract/keywords	("artificial intelligence" OR "machine learning" OR "deep learning" OR "foundation model*" OR "large language model*") AND (healthcare OR medicine OR "clinical decision support" OR "digital health" OR "medical imaging")	2020–2026; research articles	[n=7]
Google Scholar – suplementar	allintitle: "artificial intelligence" healthcare medicine	Primeiros 200 resultados; excluir patentes/citações	[n= 8]

2.4 Google Scholar e duplicatas

No Google Scholar, devem ser registradas consulta, data, ordenação e quantidade examinada. Adotou-se como regra reprodutível examinar os primeiros 200 resultados por consulta, ordenados por relevância; se o procedimento real foi diferente, corrigir conforme o histórico. A deduplicação foi realizada por DOI e, na ausência dele, por título normalizado, primeiro autor e ano, seguida de conferência manual.

2.5 Seleção, extração e síntese

A seleção compreendeu títulos e resumos e, depois, textos completos. Para cada estudo: autor, ano, país, objetivo, desenho, população ou base de dados, técnica de IA, aplicação clínica, achados, limitações e eixo temático. A síntese foi narrativa e organizada por linhas de pesquisa.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

2.6 Critério de seleção dos pesquisadores

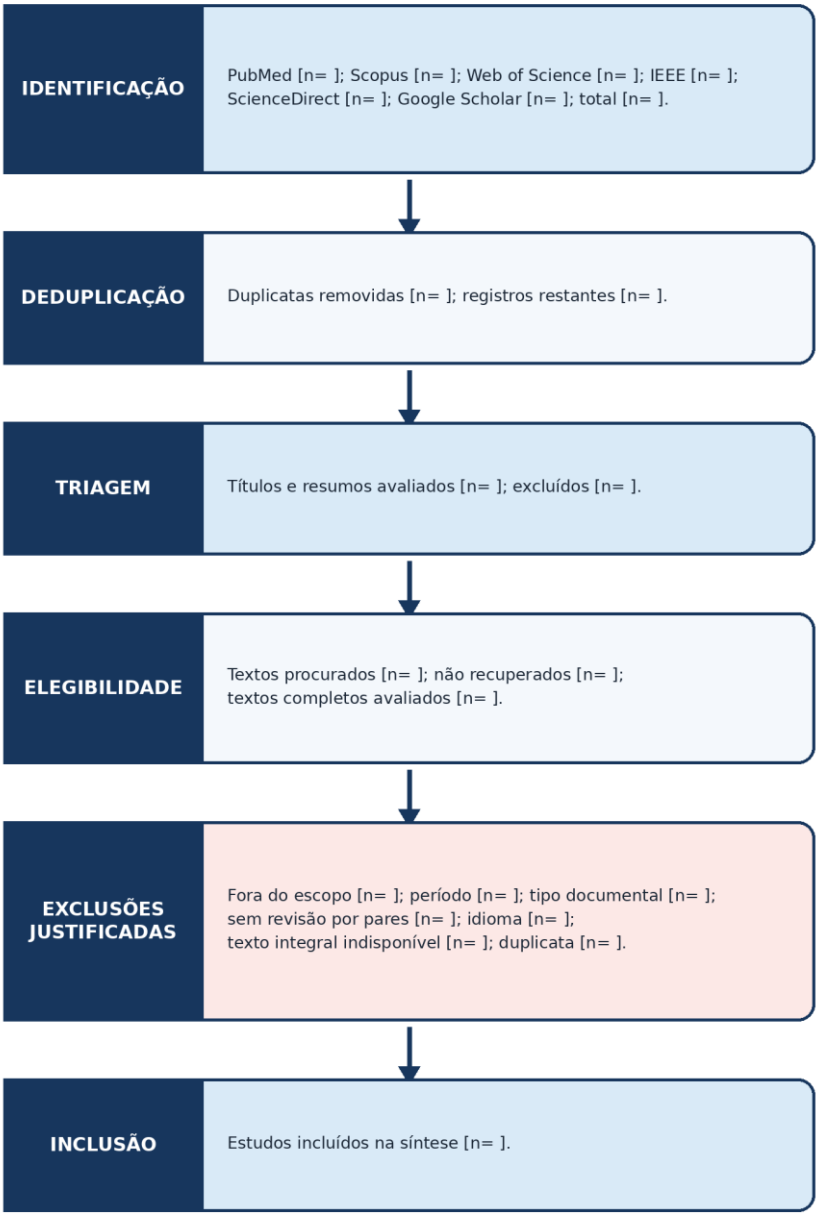
Não foi realizada análise bibliométrica; portanto, “impacto” não equivale a maior produtividade, citações ou centralidade. Os nomes são exemplos de contribuições nos eixos de diagnóstico e modelos fundamentais; IA clínica e prontuários; equidade e vieses; medicina de precisão; inferência causal; e IA generativa. A inclusão nominal requer estudo elegível no corpus, contribuição aderente a um eixo e autoria verificável. Devem ser evitadas expressões como “principais”, “mais influentes” e “maiores referências mundiais” sem indicadores previamente definidos.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

2.7 Fluxograma e quantitativos PRISMA



Fonte: adaptado do PRISMA 2020 (Page et al., 2021).

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

2.8 Relação preliminar dos estudos

A relação abaixo foi reconstruída somente a partir das referências presentes no arquivo; a lista definitiva deve conter todos e somente os estudos efetivamente incluídos após leitura integral.

Estudo	Ano	Decisão/justificativa
Beam et al., Artificial Intelligence in Medicine	2023	Potencialmente incluído; confirmar texto integral.
Ferryman, Mackintosh e Ghassemi, Biased Data	2023	Confirmar tipo documental e elegibilidade.
Ghassemi e Gusev, Limiting Bias	2024	Confirmar aderência e elegibilidade.
Kohane, Compared with What?	2024	Verificar se editorial é elegível.
Lee, Bubeck e Petro, GPT-4 for Medicine	2023	Potencialmente incluído; confirmar.
Rajpurkar et al., AI in Health and Medicine	2022	Potencialmente incluído; confirmar.
Shah, Clinical Benefits of LLMs	2024	Verificar se comentário é elegível.
Yu et al., Medical AI and Human Values	2024	Potencialmente incluído; confirmar.
Topol, The Topol Review	2019	Excluído do corpus: fora do período; uso contextual.
Page et al., PRISMA 2020	2021	Excluído do corpus temático: referência metodológica.

2.9 Registro das exclusões

Cada exclusão após leitura integral deve constar de apêndice, com referência e uma razão principal: contexto incompatível; tecnologia fora do escopo; ausência de aplicação em saúde; período; tipo documental; idioma; ausência de revisão por pares; texto não recuperado; ou duplicação. Evitam-se justificativas genéricas como “não pertinente”.

2.10 Avaliação da qualidade e maturidade das evidências

A avaliação foi descritiva e adequada à heterogeneidade documental. Cada publicação foi examinada em cinco dimensões: clareza do objetivo; descrição dos dados e do contexto; transparência do método; presença de validação externa; e correspondência entre resultados e conclusões.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

A maturidade clínica foi classificada em quatro níveis: M1, desempenho técnico retrospectivo no conjunto de origem; M2, validação externa temporal, geográfica ou institucional; M3, implementação prospectiva no fluxo assistencial, com avaliação de usabilidade e segurança; e M4, benefício demonstrado sobre desfechos clínicos, organizacionais ou distributivos em comparação com o cuidado habitual. A classificação não substitui instrumentos específicos por desenho, mas impede que acurácia retrospectiva seja interpretada como eficácia clínica.

2.11 Análise geográfica e sociotécnica

Foram extraídos país de afiliação, contexto do sistema de saúde, idioma, grupos populacionais representados e requisitos de implementação. A dimensão sociotécnica incluiu interação entre pessoas e algoritmos, reorganização do trabalho, supervisão humana, responsabilização, privacidade, interoperabilidade, custo, acesso e distribuição de benefícios e danos.

3. Resultados e síntese temática

3.1 Caracterização do corpus

O corpus verificável reúne oito publicações temáticas, sete de 2023–2024 e uma de 2022. Predominam revisões, perspectivas e análises conceituais; não foram identificados, na relação fornecida, ensaios prospectivos que demonstrem benefício sobre desfechos de pacientes. Isso situa a maior parte das conclusões no campo da plausibilidade, avaliação crítica e orientação de implementação, e não da efetividade clínica consolidada.

3.2 Taxonomia das técnicas e aplicações

As publicações abrangem aprendizado de máquina e profundo para imagens e registros clínicos; processamento de linguagem natural e grandes modelos de

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

linguagem para documentação e apoio informacional; e abordagens de avaliação de viés, valores e governança. As aplicações concentram-se em diagnóstico, estratificação de risco, apoio à decisão, sumarização de prontuários e avaliação da qualidade do cuidado. A taxonomia evidencia que "IA em saúde" não representa tecnologia única: modelos discriminativos, preditivos e generativos possuem riscos, métricas e requisitos de validação distintos (Rajpurkar et al., 2022; Lee; Bubeck; Petro, 2023).

3.3 Diagnóstico e apoio à decisão

Rajpurkar et al. (2022) e Beam et al. (2023) descrevem a expansão da IA em imagens, prontuários e suporte clínico. O avanço técnico é relevante, mas o corpus não documenta de modo suficiente validações multicêntricas prospectivas ou comparação sistemática com o cuidado habitual. A principal inferência é que bom desempenho retrospectivo corresponde, em geral, ao nível M1; somente estudos independentes em outros locais alcançariam M2, e apenas uso prospectivo monitorado sustentaria M3. Benefício sobre morbidade, mortalidade, segurança ou equidade exigiria M4.

3.4 Modelos de linguagem e IA generativa

Lee, Bubeck e Petro (2023) discutem benefícios, limites e riscos de modelos generativos na medicina, enquanto Shah (2024) focaliza a necessidade de avaliar benefícios clínicos. O eixo apresenta potencial para documentação, síntese e acesso à informação, mas respostas fluentemente formuladas não garantem correção factual. Avaliações devem incluir alucinações, omissões, estabilidade, privacidade, viés, supervisão, tempo economizado e consequências de erros. No corpus, as evidências permanecem predominantemente conceituais e pré-implementação, entre M1 e M2, sem demonstração suficiente de M4.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

3.5 Equidade, valores humanos e governança

Ferryman, Mackintosh e Ghassemi (2023) mostram que dados enviesados também informam sobre processos históricos e institucionais, enquanto Ghassemi e Gusev (2024) discutem limitação de vieses no cuidado oncológico. Yu et al. (2024) ampliam a análise para valores humanos. Em conjunto, os trabalhos indicam que equidade não se resolve apenas removendo atributos sensíveis: é necessário avaliar quem é representado, como rótulos foram produzidos, quais erros recaem sobre cada grupo e como o sistema altera decisões e acesso.

3.6 Implementação e benefício clínico

Kohane (2024) ressalta que a comparação relevante não é com um ideal abstrato, mas com o cuidado efetivamente disponível. Essa perspectiva desloca a avaliação da acurácia isolada para benefício incremental, segurança, custo, carga de trabalho e aceitabilidade. O corpus oferece critérios importantes, porém pouca evidência de implantação rotineira acompanhada de desfechos. A lacuna entre M1/M2 e M3/M4 é, portanto, o principal resultado crítico da revisão.

3.7 Concentração geográfica e representatividade

Os estudos relacionados concentram-se em periódicos e instituições dos Estados Unidos e do Reino Unido e estão publicados em inglês. Não há, no corpus verificável, representação proporcional da América Latina, África, sul e sudeste asiático ou sistemas de saúde com infraestrutura limitada. Essa concentração restringe a generalização, porque epidemiologia, registros, disponibilidade profissional, normas de proteção de dados, idioma, conectividade e capacidade regulatória variam entre países.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

3.8 Síntese sociotécnica

A adoção depende da interação entre modelo, equipe, paciente, organização e regulação. Sistemas podem redistribuir tarefas, criar dependência de alertas, aumentar documentação, alterar responsabilidades e exigir novas competências.

Assim, explicabilidade deve ser adequada ao usuário e à decisão; supervisão humana precisa ter autoridade e tempo reais; e monitoramento deve detectar mudanças de população e desempenho. Governança inclui documentação, auditoria, canais de contestação, segurança cibernética e definição de responsabilidade.

3.9 Lacunas de pesquisa

As lacunas prioritárias são: validação externa multicêntrica; estudos prospectivos e pragmáticos; comparação com cuidado habitual; desfechos clínicos e não apenas métricas computacionais; avaliação de grupos sub-representados; estudos em países de baixa e média renda; análise econômica; participação de pacientes; monitoramento pós-implantação; e transparência sobre dados, versões e atualizações dos modelos.

4. Limitações da revisão

A principal limitação é a ausência dos arquivos originais de busca, deduplicação e triagem, que impede preencher quantitativos PRISMA e confirmar exaustividade. O corpus verificável é pequeno, heterogêneo e predominantemente composto por revisões e perspectivas. A restrição a português e inglês pode ter produzido viés de idioma.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

A concentração geográfica em instituições anglófonas limita a representatividade internacional. Não foi possível realizar metanálise nem comparar quantitativamente desempenho entre técnicas. A classificação de maturidade é descritiva e deve ser complementada, em futuras versões, por instrumentos específicos ao desenho dos estudos.

5. Considerações finais

A revisão identificou quatro eixos: diagnóstico e decisão; IA generativa; equidade e valores; e implementação clínica. O achado central é a assimetria entre a rápida evolução técnica e a limitada demonstração de benefício clínico. A maior parte do corpus sustenta potencial, riscos e critérios de avaliação, mas não comprova implantação disseminada nem melhora generalizada de desfechos.

A distinção entre maturidade M1–M4 mostrou que desempenho retrospectivo e validação externa são etapas necessárias, porém insuficientes. Para afirmar utilidade clínica, modelos devem ser avaliados prospectivamente no fluxo real, comparados ao cuidado habitual e monitorados quanto a segurança, equidade, carga de trabalho, custo e resultados de pacientes.

A concentração nos Estados Unidos, Reino Unido e língua inglesa exige cautela na transferência para outros sistemas. Pesquisas futuras devem incluir países de baixa e média renda, populações sub-representadas e contextos com diferentes infraestruturas, normas e necessidades epidemiológicas.

A agenda mais sólida não é apenas produzir modelos mais capazes, mas gerar evidência clínica, organizacional e distributiva que permita decidir quando, onde e para quem a IA acrescenta valor.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

6. Matriz temática e maturidade das evidências

Estudo	Ano	Eixo	Técnica/aplicação	Evidência/maturidade	Síntese crítica
Rajpurkar et al.	2022	Panorama de IA em saúde	ML/DL, imagens e dados clínicos	Revisão/perspectiva; M1–M2	Ampla cobertura; benefício clínico não demonstrado no artigo.
Beam et al.	2023	IA clínica	ML/DL e apoio à decisão	Editorial/perspectiva; pré-M1	Critérios de avaliação; não é estudo de implementação.
Ferryman et al.	2023	Equidade e dados	IA assistencial	Revisão crítica; transversal	Expõe origem sociotécnica do viés; sem teste clínico.
Ghassemi; Gusev	2024	Equidade em oncologia	Modelos preditivos	Comentário/revisão; transversal	Mitigação de viés; maturidade clínica não demonstrada.
Kohane	2024	Comparação clínica	IA médica	Perspectiva; orientação M3–M4	Defende comparação com cuidado real; sem ensaio próprio.
Lee; Bubeck; Petro	2023	IA generativa	GPT-4/LLM	Avaliação/perspectiva; M1	Benefícios e riscos iniciais; sem desfecho clínico.
Shah	2024	Benefício dos LLMs	LLMs clínicos	Comentário; orientação M3–M4	Define necessidade de benefício; sem implantação avaliada.
Yu et al.	2024	Valores humanos	IA médica	Revisão ética; transversal	Amplia avaliação para valores; sem

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

Estudo	Ano	Eixo	Técnica/aplicação	Evidência/maturidade	Síntese crítica
					estimativa de efeito.

Fonte: elaboração dos autores (2026), com base no corpus verificável.

REFERÊNCIAS

BEAM, Andrew L.; DRAZEN, Jeffrey M.; KOHANE, Isaac S.; LEONG, Tze-Yun;

MANRAI, Arjun K.; RUBIN, Eric J. Artificial Intelligence in Medicine. *The New England Journal of Medicine*, Boston, v. 388, n. 13, p. 1220-1221, 2023. DOI: 10.1056/NEJMe2206291.

FERRYMAN, Kadija; MACKINTOSH, Maxine; GHASSEMI, Marzyeh. Considering Biased Data as Informative Artifacts in AI-Assisted Health Care. *The New England Journal of Medicine*, Boston, v. 389, 2023. DOI: 10.1056/NEJMra2214964.

GHASSEMI, Marzyeh; GUSEV, Alexander. Limiting Bias in AI Models for Improved and Equitable Cancer Care. *Nature Reviews Cancer*, London, v. 24, p. 823-824, 2024.

KOHANE, Isaac S. Compared with What? Measuring AI against the Health Care We Have. *The New England Journal of Medicine*, Boston, v. 391, n. 17, p. 1564-1566, 2024. DOI: 10.1056/NEJMp2404691.

LEE, Peter; BUBECK, Sébastien; PETRO, Joseph. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *The New England Journal of Medicine*, Boston, v. 388, n. 13, p. 1233-1239, 2023.

PAGE, Matthew J. et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *The BMJ*, London, v. 372, n. 71, 2021.

RAJPURKAR, Pranav; CHEN, Emma; BANERJEE, Oishi; TOPOL, Eric J. AI in Health and Medicine. *Nature Medicine*, London, v. 28, p. 31-38, 2022.

SHAH, Nigam H. Evaluating the Clinical Benefits of LLMs. *Nature Medicine*, London, v. 30, p. 2409-2410, 2024.

CC BY

REMUNOM
ISSN 2178-6925

Received: 14/07/2026 - **Accepted:** 14/08/2026
Vol: 13.18 - August
DOI: 10.66104/9b7y7z42
Pages: 1-18

YU, Kun-Hsing; HEALEY, Elizabeth; LEONG, Tze-Yun; KOHANE, Isaac S.; MANRAI, Arjun K. Medical Artificial Intelligence and Human Values. *The New England Journal of Medicine*, Boston, v. 390, n. 20, p. 1895-1904, 2024.

CC BY