

**INTELIGÊNCIA ARTIFICIAL NA ATENÇÃO PRIMÁRIA À SAÚDE BRASILEIRA:
POTENCIALIDADES, RISCOS ÉTICOS E IMPLICAÇÕES PARA A EQUIDADE NO SUS**

**ARTIFICIAL INTELLIGENCE IN BRAZILIAN PRIMARY HEALTH CARE: POTENTIALS,
ETHICAL RISKS AND IMPLICATIONS FOR EQUITY IN THE SUS**

**INTELIGENCIA ARTIFICIAL EN LA ATENCIÓN PRIMARIA DE SALUD BRASILEÑA:
POTENCIALIDADES, RIESGOS ÉTICOS E IMPLICACIONES PARA LA EQUIDAD EN
EL SUS**

Vinicius de Lima Lovadini

Doutor em Ciências pela Universidade de São Paulo – USP

viniciuslovadini@hotmail.com

ORCID: <https://orcid.org/0000-0001-9066-2160>

Ana Carolina Buchegger Marcelino Moura

Graduanda em Odontologia pela Universidade Federal da Paraíba

ana.marcelino@academico.ufpb.br

ORCID: <https://orcid.org/0009-0001-9829-6545>

Isabelly Almeida Costa

Acadêmica de Medicina pela Universidade Brasil

isaalmeidacosta@hotmail.com

ORCID: <https://orcid.org/0009-0000-6496-7972>

André Wilian Lozano

Doutor em Enfermagem pela Universidade Federal de São Carlos- UFSCAR

lozanoenf@gmail.com

ORCID: <https://orcid.org/0000-0002-5721-7054>

Valéria Albuquerque Vaz Rodrigues

Especialista em Educação Permanente em Saúde pela FIOCRUZ

v_vaz2005@yahoo.com.br

ORCID: <https://orcid.org/0009-0009-5500-8636>

Ana Paula de Lima

Mestranda em Engenharia Biomédica pela Universidade Brasil – UB

dranapaulalima@gmail.com

ORCID: <https://orcid.org/0009-0007-5146-8657>

Wagner Rafael da Silva

Doutor em Engenharia Biomédica pela Universidade Brasil – UB

wagner.silva@ub.edu.br

ORCID: <https://orcid.org/0000-0002-0952-4877>

Vanessa Dias de Oliveira Justi

Doutoranda em Engenharia Biomédica pela Universidade Brasil – UB

vanessad.o.justi@gmail.com

ORCID: <https://orcid.org/0000-0003-4853-8136>

Alessandra Cristiane Alves do Nascimento

Enfermeira pelo UNISALESIANO

profalessandran@outlook.com

ORCID: <https://orcid.org/0009-0003-6315-4832>

Patrícia Michelassi Carrinho Aureliano

Doutora em Engenharia Biomédica pela Universidade Brasil

patricia.aureliano@ub.edu.br

ORCID: <https://orcid.org/0000-0002-8642-479X>

José Martins Pinto Neto

Doutor em Ciências pela Universidade de São Paulo – USP

jmpintoneto@gmail.com

ORCID: <https://orcid.org/0000-0002-6794-8379>

Nicezia Vilela Junqueira Franqueiro

Doutora em Engenharia Biomédica pela Universidade Brasil

junqueira.franqueiro@gmail.com

ORCID: <https://orcid.org/0000-0001-5721-3320>

Bianca Ortunho Boato

Especialista em Gestão e Acreditação com Ênfase na Metodologia ONA pelo Instituto

Albert Einstein

biancaortunho@hotmail.com

ORCID: <https://orcid.org/0009-0008-5513-3800>

Valter Mariano dos Santos Junior

Doutor em Ciências pela Universidade Federal de São Carlos – UFSCar

profvaltermariano@gmail.com

ORCID: <https://orcid.org/0000-0002-2650-9441>

Resumo

A incorporação da inteligência artificial (IA) aos sistemas de saúde tem ampliado as possibilidades de apoio à decisão clínica, estratificação de riscos, automação de tarefas, vigilância epidemiológica e organização do cuidado. Entretanto, na Atenção Primária à Saúde (APS), tais aplicações incidem sobre contextos marcados por desigualdades territoriais, raciais, socioeconômicas e digitais, o que pode converter ganhos de eficiência em novas formas de exclusão. Este estudo teve como objetivo analisar criticamente as potencialidades, os riscos éticos e as implicações para a equidade decorrentes do uso da IA na APS brasileira e no Sistema Único de Saúde (SUS). Trata-se de revisão integrativa, com recorte temporal de 2016 a 2026, baseada em literatura científica nacional e internacional e em documentos normativos pertinentes à saúde digital, proteção de dados e governança algorítmica. A síntese foi organizada em quatro eixos: aplicações e benefícios potenciais; fontes de viés e desigualdade; proteção de dados, transparência e responsabilização; e requisitos de governança para adoção no SUS. Os achados indicam que a IA pode apoiar a identificação precoce de riscos, reduzir carga administrativa, ampliar a capacidade analítica das equipes e favorecer a coordenação do cuidado. Contudo, dados incompletos, sub-representação de grupos populacionais, escolhas

inadequadas de variáveis-alvo, baixa explicabilidade, dependência tecnológica e ausência de monitoramento por subgrupos podem reproduzir ou ampliar iniquidades. Conclui-se que a adoção eticamente aceitável da IA na APS depende de validação local, supervisão humana, avaliação contínua de desempenho e equidade, participação social, proteção de dados e responsabilização institucional. No SUS, a inovação tecnológica deve permanecer subordinada aos princípios de universalidade, integralidade e equidade, não podendo substituir investimentos estruturais, relações de cuidado e capacidade pública de governança.

Palavras-chave: Inteligência artificial; Atenção Primária à Saúde; Equidade em saúde; Sistema Único de Saúde; Ética.

Abstract

The incorporation of artificial intelligence (AI) into health systems has expanded opportunities for clinical decision support, risk stratification, task automation, epidemiological surveillance and care organization. However, in Primary Health Care (PHC), these applications operate in settings marked by territorial, racial, socioeconomic and digital inequalities, which may transform efficiency gains into new forms of exclusion. This study critically analyzed the potential benefits, ethical risks and equity implications arising from the use of AI in Brazilian PHC and the Unified Health System (SUS). An integrative review covering 2016–2026 was conducted using national and international scientific literature and regulatory documents on digital health, data protection and algorithmic governance. The synthesis addressed potential applications and benefits; sources of bias and inequality; data protection, transparency and accountability; and governance requirements for adoption in the SUS. AI may support early risk identification, reduce administrative workload, enhance teams' analytical capacity and improve care coordination. Nevertheless, incomplete data, underrepresentation of population groups, inappropriate target variables, limited explainability, technological dependence and lack of subgroup monitoring may reproduce or intensify inequities. Ethical adoption in PHC therefore requires local validation, human oversight, continuous assessment of performance and equity, social participation, data protection and institutional accountability. In the SUS, technological innovation must remain subordinate to universality, comprehensiveness and equity and cannot replace structural investment, care relationships or public governance capacity.

Keywords: Artificial intelligence; Primary Health Care; Health equity; Unified Health System; Ethics.

Resumen

La incorporación de la inteligencia artificial (IA) a los sistemas de salud ha ampliado las posibilidades de apoyo a la decisión clínica, estratificación de riesgos, automatización de tareas, vigilancia epidemiológica y organización del cuidado. Sin embargo, en la Atención Primaria de Salud (APS), estas aplicaciones operan en

contextos marcados por desigualdades territoriales, raciales, socioeconómicas y digitales, lo que puede convertir las ganancias de eficiencia en nuevas formas de exclusión. Este estudio analizó críticamente las potencialidades, los riesgos éticos y las implicaciones para la equidad derivadas del uso de la IA en la APS brasileña y en el Sistema Único de Salud (SUS). Se realizó una revisión integradora, con recorte temporal de 2016 a 2026, basada en literatura científica nacional e internacional y documentos normativos sobre salud digital, protección de datos y gobernanza algorítmica. La síntesis abordó aplicaciones y beneficios potenciales; fuentes de sesgo y desigualdad; protección de datos, transparencia y rendición de cuentas; y requisitos de gobernanza para su adopción en el SUS. La IA puede favorecer la identificación temprana de riesgos, reducir la carga administrativa, ampliar la capacidad analítica de los equipos y mejorar la coordinación del cuidado. No obstante, los datos incompletos, la subrepresentación de grupos poblacionales, las variables objetivo inadecuadas, la baja explicabilidad, la dependencia tecnológica y la ausencia de monitoreo por subgrupos pueden reproducir o ampliar inequidades. Su adopción ética exige validación local, supervisión humana, evaluación continua del desempeño y de la equidad, participación social, protección de datos y responsabilidad institucional. En el SUS, la innovación tecnológica debe permanecer subordinada a los principios de universalidad, integralidad y equidad.

Palabras clave: Inteligencia artificial; Atención Primaria de Salud; Equidad en salud; Sistema Único de Salud; Ética.

1 INTRODUÇÃO

A inteligência artificial compreende um conjunto de métodos computacionais capazes de executar tarefas que envolvem reconhecimento de padrões, classificação, predição, geração de conteúdo e apoio à tomada de decisão. Na saúde, essas tecnologias vêm sendo aplicadas à interpretação de imagens, à estratificação de risco, ao suporte clínico, à vigilância epidemiológica, à gestão de filas, à automação documental e à comunicação com usuários. Seu potencial, porém, não decorre apenas da capacidade técnica dos modelos, mas da forma como são desenhados, validados, integrados aos processos de trabalho e submetidos à governança pública (WORLD HEALTH ORGANIZATION, 2021; LIN, 2022).

Na Atenção Primária à Saúde (APS), a relevância da IA é particularmente expressiva. Como primeiro nível de atenção e coordenadora do cuidado, a APS lida com grande volume de informações, elevada diversidade clínica, acompanhamento longitudinal, demandas administrativas e necessidade permanente de priorização. Ferramentas algorítmicas podem apoiar o rastreamento de condições crônicas, a

identificação de pessoas com maior risco de hospitalização, a qualificação de registros, a busca ativa, o planejamento territorial e a tomada de decisão compartilhada. Entretanto, a mesma capilaridade que amplia o potencial de uso também aumenta a possibilidade de danos em larga escala quando modelos inadequados são incorporados à rotina (WANG et al., 2021; SASSEVILLE et al., 2023).

A literatura contemporânea demonstra que sistemas de IA não são neutros. Eles aprendem padrões presentes nos dados, nas classificações e nas decisões históricas que os alimentam. Quando esses registros refletem acesso desigual, subdiagnóstico, racismo institucional, diferenças territoriais ou menor utilização de serviços por determinados grupos, o algoritmo pode transformar desigualdades sociais em regularidades matemáticas e reproduzi-las como se fossem estimativas objetivas. O problema não se limita à inclusão explícita de raça, sexo ou renda como variáveis, pois proxies geográficos, econômicos e assistenciais podem reconstituir essas desigualdades de modo indireto (RAJKOMAR et al., 2018; OBERMEYER et al., 2019; CHEN et al., 2023).

No Brasil, essa discussão deve ser situada nas características do Sistema Único de Saúde (SUS), que combina universalidade formal com persistentes desigualdades de acesso, infraestrutura, conectividade, disponibilidade de profissionais e qualidade da informação. A Estratégia de Saúde Digital para o Brasil 2020–2028 reconhece a necessidade de interoperabilidade, governança, segurança e uso qualificado de dados, enquanto a Política Nacional de Informação e Informática em Saúde estabelece princípios para a transformação digital centrada nas necessidades das pessoas e do sistema público (BRASIL, 2020; BRASIL, 2021a). A aplicação de IA na APS, portanto, não pode ser tratada como simples aquisição de software, mas como intervenção sociotécnica com efeitos clínicos, distributivos, organizacionais e políticos.

A proteção de dados constitui outro eixo crítico. Informações de saúde são dados pessoais sensíveis e seu tratamento está submetido à Lei Geral de Proteção de Dados Pessoais (LGPD), que estabelece princípios como finalidade, adequação, necessidade, transparência, segurança, prevenção, não discriminação e

responsabilização. Em sistemas algorítmicos, esses princípios exigem atenção ao ciclo completo dos dados: coleta, integração, treinamento, compartilhamento, armazenamento, inferência, auditoria e descarte. O uso secundário de bases assistenciais sem controles adequados pode expor usuários a riscos de reidentificação, discriminação e decisões automatizadas opacas (BRASIL, 2018).

Além disso, a expansão recente de modelos generativos e multimodais acrescentou desafios relacionados à produção de informações incorretas, rastreabilidade das respostas, confidencialidade, propriedade intelectual, dependência de fornecedores e uso não supervisionado por profissionais e usuários. A Organização Mundial da Saúde recomenda cautela, avaliação proporcional ao risco, transparência, supervisão humana e mecanismos de responsabilização, especialmente quando ferramentas de uso geral são incorporadas a contextos clínicos ou de saúde pública (WORLD HEALTH ORGANIZATION, 2025).

Diante desse cenário, torna-se necessário ultrapassar abordagens tecnocêntricas e examinar a IA a partir dos princípios do SUS. A equidade não deve ser tratada como atributo opcional ou etapa posterior de correção, mas como objetivo explícito desde a formulação do problema até o monitoramento pós-implantação. Assim, esta revisão teve como objetivo analisar criticamente as potencialidades, os riscos éticos e as implicações para a equidade relacionados ao uso da inteligência artificial na APS brasileira, identificando condições necessárias para sua adoção responsável no SUS.

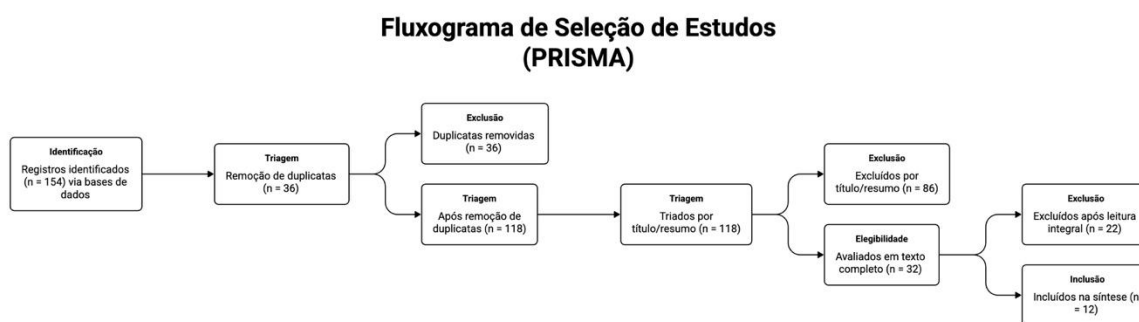
2 MÉTODO

Trata-se de revisão integrativa da literatura, método que permite reunir evidências empíricas, estudos teóricos e documentos normativos para compreender fenômenos complexos e produzir síntese aplicável à prática e à formulação de políticas. A revisão foi estruturada pelas etapas de definição do problema, estabelecimento dos critérios de elegibilidade, busca e seleção das fontes, extração padronizada, avaliação crítica e síntese temática.

A questão orientadora foi: quais são as potencialidades, os riscos éticos e as implicações para a equidade associados ao desenvolvimento e ao uso de sistemas de

inteligência artificial na Atenção Primária à Saúde, especialmente no contexto do SUS? A seleção dos estudos foi conduzida de acordo com as recomendações do Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020), contemplando as etapas de identificação, triagem, elegibilidade e inclusão. Inicialmente, os registros recuperados nas bases de dados foram organizados e submetidos à remoção de duplicatas. Em seguida, realizou-se a leitura dos títulos e resumos para verificar a aderência à questão norteadora e aos critérios de elegibilidade previamente estabelecidos. Os estudos potencialmente relevantes foram avaliados na íntegra, sendo excluídos aqueles que não apresentavam relação direta com a inteligência artificial na Atenção Primária à Saúde, com a equidade, com os aspectos éticos ou com a governança no SUS. Ao final, foram incluídas as publicações que apresentaram consistência temática e contribuição científica para a síntese proposta, com o percurso de seleção representado por meio do fluxograma PRISMA, conforme demonstrado na Figura 1.

Figura 1- Fluxograma de seleção dos estudos de acordo com modelo PRISMA, 2026.



Fonte: Elaborado pelos autores.

O recorte temporal principal compreendeu publicações de 2016 a junho de 2026. Foram considerados artigos científicos indexados, revisões, estudos metodológicos, documentos de consenso e diretrizes de organismos nacionais e internacionais. Documentos anteriores a 2016 foram admitidos exclusivamente

quando indispensáveis à contextualização histórica e conceitual de desigualdades estruturais no SUS. A busca complementar utilizou combinações dos descritores e termos livres “artificial intelligence”, “machine learning”, “primary health care”, “primary care”, “health equity”, “algorithmic bias”, “ethics”, “digital health”, “Sistema Único de Saúde” e “Atenção Primária à Saúde”, em português e inglês.

Foram incluídas fontes que: a) abordassem aplicações de IA potencialmente transferíveis ou diretamente relacionadas à APS; b) analisassem vieses, desigualdades ou desempenho diferencial entre grupos; c) discutissem privacidade, transparência, explicabilidade, responsabilização ou governança; ou d) apresentassem diretrizes normativas relevantes ao SUS e à saúde digital. Foram excluídos textos sem relação com saúde, publicações meramente promocionais, estudos estritamente técnicos sem discussão de impacto assistencial e materiais sem autoria ou procedência verificável.

O corpus científico central foi composto por estudos sobre IA e equidade em saúde, IA na atenção primária, justiça algorítmica, monitoramento pós-implantação e mitigação de vieses. A esses foram agregados documentos normativos e institucionais para contextualização brasileira, entre eles a LGPD, a Estratégia de Saúde Digital para o Brasil 2020–2028, a Política Nacional de Informação e Informática em Saúde e orientações da Organização Mundial da Saúde. Reportagens e textos de divulgação presentes na versão inicial foram substituídos, sempre que possível, pelos estudos científicos originais correspondentes.

A extração considerou autoria, ano, desenho, cenário, finalidade da tecnologia, população analisada, benefícios potenciais, fontes de risco, medidas de equidade e recomendações de governança. A análise crítica não atribuiu equivalência entre diferentes tipos de evidência: estudos empíricos foram utilizados para sustentar efeitos observados; revisões e artigos conceituais, para organizar mecanismos e categorias; e documentos normativos, para fundamentar deveres éticos e institucionais.

A síntese foi conduzida por análise temática e organizada em quatro eixos: potencialidades para o cuidado e a gestão; mecanismos de viés e produção de iniquidades; privacidade, transparência e responsabilização; e requisitos de

governança para adoção no SUS. Por não se tratar de metanálise, não foram estimadas medidas combinadas de efeito.

3 RESULTADOS

A literatura analisada revela um campo em rápida expansão, porém marcado por assimetria entre o desenvolvimento tecnológico e a produção de evidências de implementação em cenários reais da APS. Predominam revisões, estudos metodológicos e análises conceituais; são menos frequentes avaliações prospectivas, multicêntricas e orientadas por desfechos de equidade. A maior parte das fontes converge ao reconhecer que o impacto de uma ferramenta depende não apenas de sua acurácia, mas também da qualidade dos dados, do contexto de implantação, da interação com profissionais e usuários e da capacidade institucional de monitoramento.

3.1 Caracterização das principais fontes científicas

Autor/ano	Desenho	Foco	Contribuição principal
Rajkomar et al. (2018)	Artigo analítico	Equidade em aprendizado de máquina	Dados históricos, rótulos enviesados e diferenças de representação podem produzir desempenho desigual; recomenda avaliação por subgrupos e definição explícita de justiça.
Obermeyer et al. (2019)	Estudo empírico	Algoritmo de gestão populacional	Demonstrou viés racial decorrente do uso de gasto em saúde como proxy de necessidade clínica, evidenciando que variáveis-alvo podem incorporar desigualdades de acesso.
Morley et al. (2020)	Revisão sistemática	Ética da IA em saúde	Identificou desafios recorrentes de privacidade, confiança, responsabilidade, transparência, vieses e impacto na relação profissional-paciente.
Wang et al. (2021)	Revisão de escopo/protocolo aplicado à APS	Equidade em IA na atenção primária	Evidenciou a necessidade de examinar populações vulnerabilizadas, representatividade e estratégias de mitigação desde o desenvolvimento.
Chen et al. (2021)	Artigo de perspectiva	Aprendizado de máquina ético	Propôs abordar justiça social em todo o ciclo de vida: seleção do problema, dados, modelagem, implantação e monitoramento.
Lin (2022)	Revisão clínica	IA para profissionais da APS	Apresentou aplicações, limitações, riscos de viés e necessidade de alfabetização dos profissionais para uso crítico.
Sasseville et al. (2023)	Revisão de escopo	Grupos vulneráveis na APS comunitária	Mapeou métodos de identificação e mitigação de vieses, apontando baixa padronização e sub-representação de grupos diversos.
Chen et al. (2023)	Revisão metodológica	Justiça algorítmica em medicina	Mostrou que métricas agregadas podem ocultar desigualdades e que fairness exige escolhas normativas, validação contextual e monitoramento contínuo.

Chin et al. (2023)	Princípios orientadores	Disparidades raciais e étnicas	Defendeu equidade como objetivo fundamental, participação comunitária, transparência, prestação de contas e reparação de danos.
Sendak, Balu e Hernandez (2023)	Comentário científico	Monitoramento proativo	Sustentou que desempenho real e efeitos distributivos devem ser acompanhados após a implantação, com capacidade técnica institucional.
Berdahl et al. (2024)	Revisão	Estratégias para reduzir iniquidades	Organizou intervenções relativas a dados, desenvolvimento, validação, implantação, governança e engajamento de comunidades.
Coots et al. (2024)	Revisão crítica	Viés racial em algoritmos clínicos e populacionais	Evidenciou debates sobre proxies, causalidade, categorias raciais, contexto institucional e limites das métricas formais de fairness.
Martínez-Martínez et al. (2025)	Estudo qualitativo	Percepções de pacientes e profissionais da APS	Identificou expectativas de redução de carga burocrática e preocupações com privacidade, segurança, vieses e perda de vínculo humano.
Gao et al. (2026)	Revisão/agenda de saúde pública	IA em saúde pública	Enfaticizou a natureza sociotécnica da transformação, a necessidade de governança participativa e as desigualdades globais de produção de evidências.

Fonte: Dados da pesquisa.

3.2 Potencialidades da IA na APS

A primeira potencialidade refere-se ao apoio à identificação precoce de riscos. Modelos preditivos podem combinar dados clínicos, demográficos e de utilização para sinalizar pessoas com maior probabilidade de descompensação, internação evitável, baixa adesão ou necessidade de acompanhamento intensivo. Na APS, esse recurso pode qualificar a busca ativa e a gestão de condições crônicas, desde que os alertas sejam interpretados por profissionais e não convertidos em decisões automáticas sem contexto (LIN, 2022; WANG et al., 2021).

A segunda potencialidade envolve a redução de carga administrativa. Sistemas de processamento de linguagem natural e modelos generativos podem apoiar sumarização de prontuários, elaboração preliminar de documentos, codificação e organização de informações. Em serviços com elevada demanda, isso pode liberar tempo para atividades clínicas e relacionais. Contudo, a economia de tempo só se concretiza quando a ferramenta é integrada ao fluxo de trabalho, apresenta baixa taxa de erro e não transfere aos profissionais uma nova carga de revisão e correção (MARTÍNEZ-MARTÍNEZ et al., 2025; WORLD HEALTH ORGANIZATION, 2025).

A IA também pode ampliar a capacidade de análise territorial e populacional. A identificação de padrões de adoecimento, lacunas vacinais, absenteísmo, descontinuidade do cuidado e concentração espacial de riscos pode apoiar planejamento local, vigilância e alocação de recursos. Esse uso é coerente com a função da APS de conhecer a população adscrita e organizar ações segundo necessidades do território. Entretanto, áreas com registros incompletos ou menor conectividade podem aparecer artificialmente como territórios de menor necessidade, produzindo um paradoxo no qual populações menos documentadas recebem menor prioridade.

Outra aplicação potencial é o suporte à decisão clínica, incluindo recomendações, triagem e interpretação de dados. Quando adequadamente validados, esses sistemas podem reduzir variabilidade, oferecer lembretes e auxiliar profissionais diante de grande volume de informações. Não devem, porém, substituir julgamento clínico, escuta qualificada ou decisão compartilhada. O desempenho médio de um modelo não garante benefício individual, e alertas sem explicação podem favorecer confiança excessiva ou rejeição indiscriminada (RAJKOMAR et al., 2018; CHEN et al., 2023).

Por fim, a IA pode contribuir para detectar desigualdades previamente invisibilizadas, ao permitir auditorias de desempenho e comparação de resultados entre grupos. Quando a equidade é incorporada como objetivo explícito, ferramentas analíticas podem identificar subdiagnóstico, variações injustificadas de conduta e padrões de exclusão. Assim, a tecnologia pode ser utilizada não apenas para prever riscos individuais, mas para avaliar criticamente o próprio sistema de saúde (BERDAHL et al., 2024).

3.3 Mecanismos de viés e produção de iniquidades

O viés pode surgir antes mesmo da modelagem. A definição do problema determina quem será beneficiado, qual necessidade será considerada e qual desfecho será otimizado. Um sistema concebido para reduzir custos pode produzir prioridades distintas de outro orientado à redução de sofrimento ou internações evitáveis. O estudo de Obermeyer et al. (2019) mostrou que usar gasto em saúde

como proxy de necessidade clínica subestimou a gravidade de pacientes negros, porque menor gasto refletia barreiras de acesso e não menor adoecimento.

Os dados constituem outra fonte central de risco. Registros eletrônicos representam encontros com o sistema, e não a totalidade das necessidades de saúde. Pessoas com menor acesso, vínculos frágeis, baixa conectividade ou histórico de discriminação podem estar sub-representadas. Além disso, variáveis ausentes ou registradas de modo inconsistente reduzem a validade externa dos modelos. Na APS brasileira, diferenças entre municípios e equipes quanto à qualidade do cadastro e do prontuário tornam indispensável a validação local.

O chamado viés de rótulo ocorre quando o desfecho utilizado para treinar o modelo é imperfeito ou socialmente condicionado. Diagnóstico registrado, adesão, retorno ao serviço e utilização podem refletir barreiras, decisões profissionais e desigualdades institucionais. Um modelo pode aprender que determinado grupo apresenta menor probabilidade de acompanhamento simplesmente porque esse grupo enfrentou maior dificuldade de acesso. A exclusão da variável raça não elimina o problema, pois outros atributos podem operar como proxies (RAJKOMAR et al., 2018; COOTS et al., 2024).

As métricas de desempenho também exigem cautela. Acurácia, área sob a curva e sensibilidade globais podem ocultar diferenças importantes entre subgrupos. Um modelo considerado satisfatório na população total pode apresentar maior taxa de falsos negativos justamente entre pessoas que já enfrentam maior risco social. Por isso, a avaliação deve ser estratificada por sexo, raça/cor, idade, território, deficiência, condição socioeconômica e outros marcadores pertinentes, sem naturalizar categorias sociais como explicações biológicas (CHEN et al., 2023).

Na implantação, surgem riscos de viés de automação, quando profissionais atribuem autoridade excessiva ao sistema, e de viés de rejeição, quando alertas são sistematicamente ignorados. Ambos dependem do desenho da interface, da carga de trabalho, da formação e da cultura organizacional. A responsabilização não pode ser transferida individualmente ao profissional quando a instituição adota uma tecnologia sem validação, treinamento ou mecanismos de contestação.

A exclusão digital é igualmente relevante. Soluções baseadas em aplicativos, portais ou comunicação automatizada podem ampliar acesso para alguns grupos e criar barreiras para pessoas sem conectividade, dispositivos, alfabetização digital ou acessibilidade. Na APS, canais digitais devem complementar, e não substituir, modalidades presenciais e comunitárias. A transformação digital que condiciona acesso a competências ou recursos privados contradiz o princípio da universalidade.

3.4 Privacidade, transparência e responsabilização

A utilização de dados de saúde exige base legal, finalidade definida, minimização, segurança e governança. A LGPD classifica dados referentes à saúde como sensíveis e estabelece o princípio da não discriminação, particularmente relevante para sistemas que produzem inferências sobre risco, comportamento ou utilização de serviços. Mesmo quando o tratamento dispensa consentimento em hipóteses legais específicas, permanecem os deveres de transparência, segurança e prestação de contas (BRASIL, 2018).

Explicabilidade não significa necessariamente revelar todo o código-fonte, mas permitir compreensão proporcional ao risco: quais dados são utilizados, qual finalidade, que população sustentou o treinamento, quais limitações foram identificadas, como o resultado influencia a conduta e como uma decisão pode ser contestada. Em sistemas de alto impacto, a opacidade comercial não deve impedir auditoria independente ou supervisão pública.

A responsabilização deve abranger desenvolvedores, fornecedores, gestores e instituições usuárias. É insuficiente atribuir ao profissional a decisão final quando ele não dispõe de informação, tempo ou autonomia para contestar o sistema. Contratos de aquisição precisam definir acesso a documentação, requisitos de segurança, atualização, monitoramento, comunicação de incidentes, portabilidade e condições de descontinuação.

Modelos generativos acrescentam riscos de fabricação de informações, referências inexistentes, vazamento de dados e respostas não reproduzíveis. Seu uso em documentos clínicos exige revisão humana integral, proibição de inserção indevida de dados identificáveis em plataformas não autorizadas e registro de quando o conteúdo foi produzido ou auxiliado por IA. A supervisão deve ser mais

rigorosa quanto maior o potencial de dano (WORLD HEALTH ORGANIZATION, 2025).

4 DISCUSSÃO

Os achados demonstram que a pergunta central não é se a IA é intrinsecamente benéfica ou prejudicial, mas quais finalidades orientam seu uso, quais grupos foram representados no desenvolvimento, quais riscos são aceitáveis e quem responde por danos. A IA é uma tecnologia sociotécnica: seus efeitos emergem da interação entre dados, modelos, infraestrutura, profissionais, normas, interesses econômicos e condições locais. Essa compreensão é particularmente importante na APS, onde decisões aparentemente simples podem alterar acesso, seguimento e priorização de grandes contingentes populacionais (CHIN et al., 2023).

A adoção de IA no SUS deve ser orientada por necessidade pública demonstrável. Tecnologias não devem ser incorporadas apenas por novidade ou promessa de eficiência. O problema assistencial precisa ser claramente definido, e alternativas menos complexas devem ser consideradas. Em muitos cenários, melhorias de cadastro, interoperabilidade, dimensionamento de equipes e organização de processos podem produzir resultados superiores aos de um modelo sofisticado aplicado sobre infraestrutura precária (CHIN et al., 2023).

A equidade precisa ser operacionalizada. Isso implica definir previamente quais grupos podem ser afetados, quais disparidades serão medidas, quais limites de desempenho são aceitáveis e quais ações serão tomadas quando houver diferença injusta. O monitoramento deve ocorrer antes e após a implantação, pois modelos podem perder desempenho com mudanças epidemiológicas, alterações no registro ou transformação do perfil populacional. Auditoria não é etapa única, mas processo permanente (CHIN et al., 2023; SENDAK; BALU; HERNANDEZ, 2023).

No contexto brasileiro, raça/cor deve ser tratada com rigor. Sua ausência nos bancos impede avaliar desigualdades; seu uso acrítico pode reforçar essencialismos. A categoria deve ser compreendida como marcador social associado a exposições, oportunidades e discriminação, e não como variável biológica autossuficiente. A qualidade do preenchimento, a participação de

populações afetadas e a interpretação interdisciplinar são condições para uma análise responsável (SENDAK; BALU; HERNANDEZ, 2023).

A participação social é requisito substantivo, não meramente consultivo. Usuários, trabalhadores, gestores e comunidades devem participar da definição de prioridades, avaliação de aceitabilidade e construção de mecanismos de contestação. Na estrutura do SUS, conselhos de saúde, comissões intergestores e instâncias de ética e proteção de dados podem contribuir para a governança, desde que recebam informação compreensível e capacidade técnica (BRASIL, 2018).

A soberania e a capacidade pública também merecem atenção. Dependência de soluções proprietárias pode limitar auditoria, interoperabilidade e continuidade. O SUS necessita de critérios de contratação que protejam o interesse público, evitem aprisionamento tecnológico e garantam acesso a documentação, métricas e registros de alteração. A infraestrutura nacional de saúde digital deve favorecer padrões abertos, segurança, rastreabilidade e reutilização responsável dos dados (BRASIL, 2018).

A alfabetização em IA dos profissionais é indispensável. Não se espera que todos dominem programação, mas que compreendam limites, incerteza, vieses, privacidade e responsabilidade. A formação deve incluir leitura crítica de métricas, reconhecimento de resultados implausíveis e comunicação transparente com usuários. Sistemas que não permitem contestação tendem a reduzir autonomia profissional e enfraquecer a relação de cuidado (COOTS et al., 2024).

As aplicações generativas exigem política institucional específica. Seu uso informal já ocorre em ambientes educacionais e assistenciais, e a ausência de regras não elimina o risco. Protocolos devem indicar ferramentas autorizadas, tipos de informação permitidos, requisitos de revisão, registro de uso e resposta a incidentes. A geração de texto pode apoiar atividades administrativas, mas não deve produzir diagnósticos ou condutas sem supervisão qualificada (COOTS et al., 2024).

Em síntese, a IA pode contribuir para uma APS mais responsiva, mas não corrige automaticamente desigualdades estruturais. Sem governança, pode acelerar decisões injustas e conferir aparência técnica a escolhas políticas. A inovação compatível com o SUS deve ampliar capacidade de cuidado, reduzir barreiras e

fortalecer o trabalho das equipes, mantendo a pessoa e o território como referências centrais.

4.1 Proposta de requisitos mínimos para adoção responsável no SUS

Requisito	Aplicação prática
1. Justificativa sanitária	Definir o problema, a população beneficiária e o valor público esperado; comparar a IA com alternativas organizacionais e tecnológicas mais simples.
2. Governança de dados	Assegurar a qualidade, bem como a interoperabilidade, minimização, segurança, rastreabilidade e conformidade com a LGPD.
3. Representatividade	Verificar composição das bases e documentar ausências, desequilíbrios e limitações relativas a grupos e territórios.
4. Validação local	Avaliar desempenho no cenário real de uso e em populações comparáveis às da implantação.
5. Avaliação de equidade	Relatar sensibilidade, especificidade, falsos positivos e falsos negativos por subgrupos relevantes.
6. Supervisão humana	Definir responsabilidades, limites de automação e possibilidade efetiva de revisão e contestação.
7. Transparência	Disponibilizar documentação sobre finalidade, dados, limitações, atualizações e influência do sistema nas decisões.
8. Participação social	Incluir trabalhadores, usuários e comunidades na definição, avaliação e monitoramento.
9. Monitoramento contínuo	Acompanhar desempenho, incidentes, mudanças de distribuição e efeitos distributivos após a implantação.
10. Plano de resposta e descontinuação	Estabelecer correção, comunicação, reparação e retirada quando riscos superarem benefícios.

Fonte: Dados da pesquisa.

5 LIMITAÇÕES

Esta revisão apresenta limitações. A heterogeneidade de desenhos e contextos impede comparação direta entre aplicações. Parte relevante da literatura discute princípios e riscos potenciais, enquanto ainda são escassas avaliações prospectivas de impacto na APS brasileira. A rápida evolução tecnológica também pode tornar ferramentas e práticas obsoletas em curto período. Além disso, a síntese privilegiou equidade, ética e governança, não avaliando de forma exaustiva desempenho técnico de aplicações específicas. Tais limitações reforçam a

necessidade de pesquisas brasileiras multicêntricas, transparentes e orientadas por desfechos clínicos, organizacionais e distributivos.

6 CONSIDERAÇÕES FINAIS

A inteligência artificial oferece possibilidades relevantes para a APS, especialmente no apoio à estratificação de riscos, organização de informações, planejamento territorial, coordenação do cuidado e redução de tarefas administrativas. Todavia, esses benefícios são condicionais e não podem ser presumidos a partir de desempenho técnico isolado.

Os principais riscos decorrem de dados incompletos ou não representativos, variáveis-alvo que incorporam desigualdades, métricas agregadas que ocultam desempenho diferencial, opacidade, fragilidade de proteção de dados, exclusão digital e ausência de responsabilização. Em um sistema universal, erros distribuídos de maneira desigual têm significado ético e sanitário particular.

Para o SUS, a adoção responsável exige validação local, avaliação contínua por subgrupos, supervisão humana, participação social, transparência, segurança e capacidade pública de governança. A IA deve fortalecer o trabalho das equipes e ampliar direitos, sem substituir vínculos, julgamento clínico ou investimentos estruturais. O critério decisivo não deve ser a modernidade da ferramenta, mas sua contribuição comprovada para um cuidado mais seguro, integral e equitativo.

REFERÊNCIAS

BERDAHL, C. T. et al. Strategies to improve the impact of artificial intelligence on health equity: scoping review. *JMIR AI*, Toronto, v. 3, e52936, 2024. DOI: 10.2196/52936.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). *Diário Oficial da União*: seção 1, Brasília, DF, 15 ago. 2018.

BRASIL. Ministério da Saúde. *Estratégia de Saúde Digital para o Brasil 2020–2028*. Brasília, DF: Ministério da Saúde, 2020.

BRASIL. Ministério da Saúde. Portaria GM/MS nº 1.768, de 30 de julho de 2021. Altera o Anexo XLII da Portaria de Consolidação GM/MS nº 2, de 28 de setembro de 2017, para dispor sobre a Política Nacional de Informação e Informática em Saúde. *Diário Oficial da União*: seção 1, Brasília, DF, 2 ago. 2021a.

BRASIL. Conselho Nacional de Saúde. Resolução nº 659, de 26 de julho de 2021. Dispõe sobre a Política Nacional de Informação e Informática em Saúde. Brasília, DF: Conselho Nacional de Saúde, 2021b.

CHEN, I. Y.; PIERSON, E.; ROSE, S.; JOSHI, S.; FERRYMAN, K.; GHASSEMI, M. Ethical machine learning in healthcare. *Annual Review of Biomedical Data Science*, Palo Alto, v. 4, p. 123–144, 2021. DOI: 10.1146/annurev-biodatasci-092820-114757.

CHEN, R. J. et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering*, London, v. 7, p. 719–742, 2023. DOI: 10.1038/s41551-023-01056-8.

CHIN, M. H. et al. Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. *JAMA Network Open*, Chicago, v. 6, n. 12, e2345050, 2023. DOI: 10.1001/jamanetworkopen.2023.45050.

COOTS, M.; LINN, K. A.; GOEL, S.; et al. Racial bias in clinical and population health algorithms: a critical review of current debates. *Annual Review of Public Health*, Palo Alto, v. 45, p. 173–192, 2024. DOI: 10.1146/annurev-publhealth-071823-112058.

GAO, Q. et al. Opportunities and challenges of artificial intelligence in public health: a systematic review and research agenda. *Frontiers in Public Health*, Lausanne, v. 14, 2026.

LIN, S. Y. A clinician's guide to artificial intelligence: why and how primary care should lead the health care AI revolution. *Journal of the American Board of Family Medicine*, Lexington, v. 35, n. 1, p. 175–184, 2022. DOI: 10.3122/jabfm.2022.01.210226.

MARTÍNEZ-MARTÍNEZ, H. et al. Perceptions of, barriers to, and facilitators of the use of artificial intelligence in primary care: qualitative study with professionals and patients. *Journal of Medical Internet Research*, Toronto, v. 27, 2025.

MORLEY, J. et al. The ethics of AI in health care: a mapping review. *Social Science & Medicine*, Oxford, v. 260, 113172, 2020. DOI: 10.1016/j.socscimed.2020.113172.

OBERMEYER, Z.; POWERS, B.; VOGELI, C.; MULLAINATHAN, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, Washington, DC, v. 366, n. 6464, p. 447–453, 2019. DOI: 10.1126/science.aax2342.

RAJKOMAR, A.; HARDT, M.; HOWELL, M. D.; CORRADO, G.; CHIN, M. H. Ensuring fairness in machine learning to advance health equity. *Annals of Internal Medicine*, Philadelphia, v. 169, n. 12, p. 866–872, 2018. DOI: 10.7326/M18-1990.

SASSEVILLE, M. et al. Risk of bias mitigation for vulnerable and diverse groups in community-based primary health care artificial intelligence models: a scoping review. *Journal of Medical Internet Research*, Toronto, v. 25, e46679, 2023. DOI: 10.2196/46679.

SENDAK, M.; BALU, S.; HERNANDEZ, A. F. Proactive algorithm monitoring to ensure health equity. *JAMA Network Open*, Chicago, v. 6, n. 12, e2345022, 2023. DOI: 10.1001/jamanetworkopen.2023.45022.

WANG, J. X. et al. Health equity in artificial intelligence and primary care research: scoping review. *Journal of Medical Internet Research*, Toronto, v. 23, n. 9, e27799, 2021. DOI: 10.2196/27799.

WORLD HEALTH ORGANIZATION. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: WHO, 2021.

WORLD HEALTH ORGANIZATION. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: WHO, 2025.