RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

INTEGRIDADE DA DECISÃO ALGORÍTMICA EM SAÚDE: EDUCAÇÃO, AUTONOMIA E GOVERNANÇA

INTEGRITY OF ALGORITHMIC DECISION-MAKING IN HEALTHCARE: EDUCATION, AUTONOMY, AND GOVERNANCE

INTEGRIDAD DE LA DECISIÓN ALGORÍTMICA EN SALUD: EDUCACIÓN, AUTONOMÍA Y GOBERNANZA

Ademar Alves dos Santos
Doutor em Educação. Universidade Federal de Uberlândia (UFU). Minas Gerais. Brasil.
E-mail: ademar.santos@ufu.br

Resumo
A inteligência artificial amplia capacidades diagnósticas, analíticas e organizacionais em saúde, mas sua utilidade técnica não assegura, por si só, a integridade de decisões de alto impacto. Este artigo teórico-conceitual analisa como a mediação algorítmica reconfigura autonomia, responsabilidade e governança, tomando a saúde como objeto central e a educação e a democracia como interfaces auxiliares. A análise articula literatura científica recente e documentos normativos e institucionais, com ênfase no período de 2020 a setembro de 2026. A contribuição teórica consiste em passar da ética formulada como conjunto de princípios para a integridade da decisão concreta. Propõe-se uma matriz composta por seis dimensões articuladas, validade clínica, segurança e proporcionalidade; autonomia, competência e supervisão substantiva; transparência, explicabilidade e justificabilidade; responsabilidade, contestabilidade e revisão; justiça, equidade e integridade dos dados; auditabilidade e governança do ciclo de vida. A educação qualifica a formação das capacidades necessárias à supervisão crítica, enquanto a democracia acrescenta critérios de justificabilidade, recurso e responsabilidade pública. Conclui-se que a integridade decisória depende de mecanismos institucionais verificáveis e de responsáveis identificáveis, não apenas do desempenho do sistema de IA.

Palavras-chave: Inteligência Artificial; Bioética; Tomada de Decisões; Política de Saúde; Governança.

Abstract
Artificial intelligence expands diagnostic, analytical, and organizational capacities in healthcare, but technical usefulness alone does not ensure the integrity of high-impact decisions. This theoretical-conceptual article examines how algorithmic mediation reshapes autonomy, responsibility, and governance, taking healthcare as the central object and education and democracy as auxiliary analytical interfaces. The analysis combines recent scientific literature with normative and institutional documents, emphasizing the period from 2020 to September 2026. Its theoretical contribution is to move from ethics framed as a set of principles toward the integrity of concrete decisions. A six-dimension matrix is proposed, encompassing clinical validity, safety, and proportionality; autonomy, competence, and substantive human oversight; transparency, explainability, and justifiability; responsibility, contestability, and review; justice, equity, and data integrity; and auditability with lifecycle governance. Education clarifies the capacities required for critical oversight, while democracy adds criteria of justifiability, appeal, and public accountability. The article concludes that decision integrity depends on verifiable institutional mechanisms and identifiable responsible actors, not only on algorithmic performance.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

Keywords: Artificial Intelligence; Bioethics; Decision Making; Health Policy; Governance.
Resumen
La inteligencia artificial amplía las capacidades diagnósticas, analíticas y organizativas en salud, pero su utilidad técnica no garantiza, por sí sola, la integridad de decisiones de alto impacto. Este artículo teórico-conceptual analiza cómo la mediación algorítmica reconfigura la autonomía, la responsabilidad y la gobernanza, tomando la salud como objeto central y la educación y la democracia como interfaces auxiliares. El análisis articula literatura científica reciente y documentos normativos e institucionales, con énfasis en el período de 2020 a septiembre de 2026. La contribución teórica consiste en pasar de una ética formulada como conjunto de principios a la integridad de la decisión concreta. Se propone una matriz compuesta por seis dimensiones articuladas: validez clínica, seguridad y proporcionalidad; autonomía, competencia y supervisión sustantiva; transparencia, explicabilidad y justificabilidad; responsabilidad, contestabilidad y revisión; justicia, equidad e integridad de los datos; auditabilidad y gobernanza del ciclo de vida. La educación cualifica la formación de las capacidades necesarias para la supervisión crítica, mientras que la democracia añade criterios de justificabilidad, recurso y responsabilidad pública. Se concluye que la integridad decisoria depende de mecanismos institucionales verificables y de responsables identificables, y no únicamente del desempeño del sistema de IA.

Palabras clave: Inteligencia Artificial; Bioética; Toma de Decisiones; Política de Salud; Gobernanza.

1. Introdução

A inteligência artificial (IA) deixou o campo estritamente experimental e passou a integrar processos decisórios em serviços de saúde, sistemas educacionais, administração pública e ambientes informacionais. Na saúde, modelos preditivos, sistemas de apoio diagnóstico, ferramentas de triagem, processamento de linguagem natural e modelos generativos já são empregados ou avaliados em atividades clínicas, administrativas e de saúde pública. Esse avanço amplia possibilidades de análise e organização do cuidado, mas também introduz riscos relacionados a erro, viés, opacidade, segurança e uso acrítico de recomendações automatizadas. A OMS mantém orientação específica para IA em saúde e, mais recentemente, para grandes modelos multimodais, reforçando segurança, autonomia, transparência e responsabilidade (WORLD HEALTH ORGANIZATION, 2021, 2024; NING et al., 2024).

O problema ético não se resume à capacidade técnica de um sistema produzir respostas corretas. Importa compreender a posição que ele ocupa na cadeia decisória e como sua presença reorganiza autoridade, deveres profissionais, direitos dos usuários e responsabilidade institucional. Em uma

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

decisão clínica, a recomendação algorítmica pode ampliar a informação disponível ao profissional e ao paciente. Também pode induzir dependência cognitiva, dificultar a explicação do raciocínio adotado ou favorecer confiança excessiva em resultados automatizados. A responsabilidade continua vinculada a pessoas e organizações capazes de justificar escolhas e responder por suas consequências (CHEN et al., 2021; GROSSER et al., 2025).

Nos sistemas públicos de saúde, a questão ganha outra escala. No Brasil, a incorporação da IA ocorre em uma infraestrutura digital marcada por avanços em interoperabilidade, prontuários, telessaúde e uso de dados, mas também por desigualdades territoriais e institucionais. Estudo recente sobre IA na saúde pública brasileira identifica oportunidades de aplicação no Sistema Único de Saúde e, simultaneamente, problemas associados à fragmentação de sistemas, formação profissional, equidade, explicabilidade e soberania tecnológica (SOARES; CHIAVEGATTO FILHO, 2026). Surge, portanto, uma questão substantiva. Como avaliar a integridade de uma decisão quando a IA participa de práticas que afetam cuidado, priorização de recursos, acesso a serviços e proteção de direitos?

A literatura ética sobre IA formulou princípios recorrentes, entre eles transparência, justiça, privacidade, não discriminação, supervisão humana e responsabilidade. O problema não é a ausência desses princípios, mas sua tradução ainda incompleta para a decisão concreta. Mittelstadt (2019) demonstra que princípios gerais são insuficientes para assegurar sistemas éticos, enquanto Morley et al. (2020) identificam dificuldades na passagem de valores abstratos para instrumentos e procedimentos. Permanece, portanto, uma lacuna de integração entre dimensão ética, mecanismo institucional, evidência verificável e responsável identificável.

A saúde constitui o objeto central deste artigo. Educação e democracia não são tratadas como campos paralelos de comparação, mas como interfaces auxiliares que refinam problemas próprios da decisão em saúde. A educação

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

permite compreender como se formam competências para leitura crítica, preservação do julgamento profissional e supervisão substantiva. A democracia acrescenta justificabilidade, possibilidade de recurso e responsabilidade pública quando decisões tecnológicas afetam direitos e acesso a serviços.

A questão que orienta o estudo pergunta quais condições precisam coexistir para que uma decisão em saúde influenciada por IA permaneça clinicamente válida, profissionalmente autônoma, institucionalmente responsável, distributivamente justa e passível de contestação. O objetivo é construir uma interpretação teórica que relacione princípios éticos a condições humanas e institucionais identificáveis em processos decisórios apoiados por IA.

Sustenta-se que a integridade dessas decisões depende menos da simples presença humana e mais da preservação de autoridade decisória substantiva, da definição prévia de responsabilidades e da existência de mecanismos verificáveis de supervisão, documentação, auditoria, contestação e monitoramento. O salto teórico proposto consiste em retirar a análise do plano da ética abstrata e situá-la na integridade da decisão concreta. A matriz desenvolvida adiante organiza esse movimento sem se apresentar como escala validada, protocolo clínico ou instrumento regulatório.

2 Percurso teórico-metodológico

O estudo possui natureza teórico-conceitual, sustentada por pesquisa bibliográfica e documental. A pesquisa bibliográfica foi empregada para construir o problema, confrontar categorias e integrar contribuições produzidas em Saúde, Educação, Bioética, Direito e governança da inteligência artificial. Gil (2008) destaca que a pesquisa bibliográfica permite trabalhar criticamente com conhecimento já produzido, ultrapassando a reunião descritiva de referências. Grazziotin, Klaus e Pereira (2022) reforçam que pesquisa bibliográfica e pesquisa documental possuem objetos e percursos próprios, embora possam ser articuladas em estudos teóricos. Essa distinção orientou o desenho do artigo e evita tratar o percurso realizado como revisão sistemática destinada a estimar

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

efeitos, frequências ou prevalência de temas.

A busca bibliográfica foi orientada por publicações recentes, com ênfase no período de 2020 a 29 de setembro de 2026, sem excluir obras anteriores necessárias à fundamentação metodológica e conceitual. Foram priorizados artigos indexados e documentos institucionais sobre decisão clínica apoiada por IA, autonomia profissional, supervisão humana, viés de automação, justiça algorítmica, explicabilidade, auditabilidade, formação e governança. A seleção foi guiada pela pertinência ao problema e pela capacidade de esclarecer mecanismos decisórios, e não por contagem de ocorrência de termos. A verificação normativa foi encerrada em 29 de setembro de 2026. Assim, a situação jurídica de leis, resoluções, referenciais e proposições legislativas é descrita conforme o estado identificado nessa data de corte.

No campo documental, foram considerados materiais da Organização Mundial da Saúde, UNESCO, União Europeia, Conselho da Europa e documentos brasileiros de proteção de dados, saúde digital, regulação profissional e educação. Cellard (2008) adverte que documentos precisam ser interpretados em seu contexto de produção, autoria, finalidade e condições institucionais, e não tomados como registros neutros. Em direção semelhante, Fávero e Centenaro (2019) mostram que a pesquisa documental aplicada às políticas educacionais exige atenção aos limites da fonte e às relações entre texto normativo, contexto e implementação. Por isso, leis, resoluções, pareceres e referenciais foram utilizados como fontes primárias para reconstruir o ambiente institucional da IA, sem serem convertidos em evidência de validade empírica da matriz proposta.

No plano analítico, as fontes foram lidas a partir de quatro perguntas articuladas. A primeira examinou o que permanece sob decisão humana quando a IA produz recomendações, classificações ou conteúdos. A segunda identificou riscos capazes de comprometer validade, autonomia, equidade e responsabilidade. A terceira buscou mecanismos institucionais que tornassem

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

esses riscos observáveis e passíveis de resposta. A quarta verificou quais capacidades formativas, jurídicas e organizacionais permitem transformar supervisão formal em supervisão substantiva. O procedimento foi interpretativo e comparativo em nível conceitual, não uma comparação empírica entre setores.

A passagem das fontes às seis dimensões ocorreu por refinamento conceitual sucessivo. Primeiro, foram identificados conceitos recorrentes ou convergentes associados às quatro perguntas orientadoras, como validade, segurança, autonomia, competência, supervisão, transparência, explicabilidade, responsabilidade, contestabilidade, equidade, integridade dos dados, auditabilidade e monitoramento. Em seguida, conceitos com função analítica próxima foram agrupados quando respondiam ao mesmo problema decisório. Validade e segurança foram associadas à proporcionalidade porque a intensidade do controle depende do risco e da reversibilidade. Autonomia e competência foram articuladas à supervisão substantiva porque presença humana sem conhecimento, tempo e autoridade não garante controle efetivo. Transparência e explicabilidade foram diferenciadas de justificabilidade, mas reunidas na mesma dimensão por constituírem etapas relacionadas da produção de razões compreensíveis e defensáveis. Responsabilidade foi articulada à contestabilidade e à revisão por depender da existência de atores identificáveis e canais de correção. Justiça foi distinguida de métricas isoladas de fairness e associada à equidade e à integridade dos dados. Auditabilidade foi ligada à governança do ciclo de vida por exigir registros, monitoramento, reavaliação e critérios de atualização ou suspensão.

Esse refinamento também envolveu exclusão de sobreposições. Categorias que funcionavam como mecanismos ou evidências, e não como dimensões de integridade, foram mantidas em nível operacional e aparecem no Quadro 3. Foram preservadas seis dimensões porque esse arranjo apresentou maior capacidade de abranger os problemas identificados sem duplicação conceitual e sem transformar a matriz em catálogo excessivamente fragmentado.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

A literatura científica forneceu principalmente problemas, conceitos e evidências sobre desempenho, interação humano-IA, autonomia, equidade e governança. Os documentos normativos contribuíram de modo distinto, ao explicitar deveres, direitos, competências institucionais, mecanismos de supervisão e formas de responsabilização. A configuração final é, portanto, uma síntese teórico-analítica dos autores construída a partir desses dois conjuntos de fontes, e não uma taxonomia previamente existente na literatura.

3 Da ferramenta à mediação decisória

Uma parte do debate público sobre IA ainda a descreve como ferramenta neutra cuja adequação dependeria apenas do usuário. Essa formulação é insuficiente para aplicações que classificam riscos, priorizam demandas, geram recomendações clínicas, selecionam informações relevantes ou produzem respostas que influenciam condutas. Nesses casos, a tecnologia participa da arquitetura da decisão. Mesmo quando a palavra final permanece formalmente com um profissional, o modo como informações são ordenadas, apresentadas e reputadas como confiáveis pode alterar o espaço de julgamento.

A noção de decisão mediada por IA empregada neste artigo designa situações em que um sistema algorítmico influencia de forma material a formação de uma decisão humana ou institucional. Não exige que o sistema decida de maneira autônoma. Basta que sua recomendação, classificação, previsão, síntese ou geração de conteúdo integre o processo que conduz à ação. Essa definição permite analisar sistemas de apoio clínico, algoritmos de triagem, ferramentas de priorização, modelos generativos utilizados por profissionais, sistemas de avaliação educacional e aplicações públicas baseadas em classificação ou recomendação.

A mediação algorítmica produz três reconfigurações articuladas. No plano epistêmico, o sistema modifica quais evidências se tornam visíveis, como são hierarquizadas e com que grau de autoridade chegam ao profissional. No plano organizacional, sua implantação redistribui tarefas entre profissionais, gestores,

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

fornecedores e equipes técnicas, alterando fluxos de trabalho e pontos de controle. No plano normativo, a influência de uma saída algorítmica sobre decisões relevantes exige identificar quem deve justificá-la, revisá-la e responder por suas consequências.

Na saúde, essas reconfigurações são especialmente sensíveis porque a decisão pode repercutir sobre diagnóstico, tratamento, acesso, tempo de espera, priorização e proteção de informações pessoais. A OMS recomenda que sistemas de IA preservem autonomia humana, promovam bem-estar e segurança, assegurem transparência e responsabilidade, protejam equidade e sejam responsivos às necessidades dos contextos de uso (WORLD HEALTH ORGANIZATION, 2021). A literatura mais recente acrescenta que sistemas de apoio clínico precisam ser acompanhados por monitoramento, validação contextual e supervisão humana capaz de reconhecer limites e incertezas.

Esses elementos indicam que a qualidade ética não reside apenas no algoritmo. Ela depende da relação entre desempenho técnico, contexto de uso, integridade dos dados, competências profissionais, regras institucionais e direitos das pessoas afetadas. Por essa razão, a unidade de análise adotada não é o modelo isolado, mas a decisão sociotécnica da qual ele participa. Essa escolha permite distinguir integridade do sistema e integridade da decisão. Um sistema pode apresentar bom desempenho e documentação adequada e ainda participar de uma decisão pouco justificável, excessivamente dependente ou institucionalmente frágil.

4 Saúde e decisões de alto impacto

A saúde oferece um contexto privilegiado para examinar a integridade de decisões mediadas por IA. Diferentes aplicações podem ampliar capacidade diagnóstica, reduzir tarefas administrativas, apoiar vigilância epidemiológica e organizar fluxos assistenciais. Entretanto, a utilidade de um sistema não equivale à integridade da decisão da qual participa. Um modelo pode apresentar bom desempenho agregado e ainda produzir consequências problemáticas quando

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

aplicado em população distinta, inserido em fluxo inadequado ou utilizado por profissionais sem condições de verificar suas recomendações.

A experiência pós-implantação confirma essa necessidade de cautela. Akhlaghi et al. (2024), ao avaliar em tempo real um sistema de predição de internação em emergência, identificaram desempenho inferior ao observado originalmente e diferenças relevantes entre especialidades. O estudo reforça a necessidade de avaliação continuada após a entrada do sistema no serviço. A validade deve ser compreendida de maneira clínica, contextual, distributiva e temporal. Não basta demonstrar que o modelo funciona em algum lugar. É necessário examinar se permanece adequado à população, ao serviço, ao momento e à finalidade em que participa da decisão.

A autonomia profissional constitui uma condição central. Ela não se resume ao fato de um médico permanecer formalmente responsável pelo ato. Supervisão substantiva exige informação adequada, competência para interpretar incerteza, tempo para verificar a recomendação e autoridade real para discordar. Grosser et al. (2025) mostram que a IA pode ampliar informação e apoiar decisões, mas também pode favorecer dependência e padronização de condutas. A Resolução CFM nº 2.454/2026 reforça esse entendimento ao assegurar que o médico possa recusar sistemas inadequados, preservar julgamento profissional e rejeitar recomendações algorítmicas quando necessário (CONSELHO FEDERAL DE MEDICINA, 2026).

A confiança também precisa ser calibrada. O objetivo institucional não deve ser maximizar confiança na IA, mas aproximá-la da confiabilidade demonstrada pelo sistema. Sobreconfiança pode favorecer aceitação automática. Subconfiança pode levar ao abandono de recomendações úteis. Entre esses extremos, confiança calibrada exige conhecimento sobre limites, incerteza, contexto e consequências possíveis. A supervisão humana deixa de ser variável binária e passa a depender da qualidade da interação entre profissional, sistema e organização.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

A responsabilidade torna-se mais complexa quando desenvolvimento, contratação, implantação e uso clínico são distribuídos entre diferentes agentes. A resposta ética não consiste em atribuir agência moral plena ao sistema. É necessário reconhecer responsabilidades profissionais, institucionais, técnicas e regulatórias. A Resolução CFM nº 2.454/2026 mantém o médico responsável pelos atos praticados com auxílio de IA, mas também impõe deveres de governança às instituições e prevê auditoria, monitoramento e comunicação de falhas (CONSELHO FEDERAL DE MEDICINA, 2026). Essa arquitetura reduz o risco de responsabilização concentrada apenas no profissional que recebe a recomendação final.

A transparência adquire valor quando produz compreensão suficiente para ação. Uma explicação técnica pode ser adequada a desenvolvedores e insuficiente para médicos, pacientes ou auditores. Por isso, a explicabilidade precisa ser relacionada ao destinatário e à finalidade. Economou-Zavlanos et al. (2024) propõem traduzir princípios de segurança, justiça, transparência e accountability em critérios avaliáveis para implementação de IA em saúde. Para o profissional, a informação deve apoiar avaliação clínica. Para o paciente, precisa permitir compreender a participação relevante da IA no cuidado. Para a instituição, deve contribuir para reconstrução do processo e revisão de responsabilidades. A justificabilidade representa um passo adicional, pois exige que a decisão possa ser defendida com razões clínicas e institucionais, e não apenas descrita em termos computacionais.

A justiça amplia a análise do desempenho médio para a distribuição de erros, benefícios e oportunidades. Métricas de fairness podem identificar determinadas disparidades, mas não definem sozinhas o que deve ser considerado justo. A escolha da métrica depende do tipo de decisão, do dano, do grupo protegido, do resultado analisado e das consequências de falsos positivos e falsos negativos. A avaliação ética precisa articular fairness algorítmica, equidade clínica e justiça institucional. No SUS, essa preocupação envolve

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

também desigualdades territoriais, infraestrutura, conectividade, interoperabilidade e capacidade de formação.

A proporcionalidade do controle completa esse quadro. Aplicações administrativas de baixo impacto não exigem o mesmo nível de governança de sistemas que influenciam diagnóstico, prognóstico, triagem, priorização ou tratamento. A reversibilidade também importa. Quanto menor a possibilidade de corrigir uma decisão antes que o dano se consolide, maior deve ser a intensidade de validação, supervisão e monitoramento. A Resolução CFM nº 2.454/2026 adota classificação de risco e determina controles mais rigorosos para aplicações de maior impacto, aproximando regulação profissional e governança baseada em risco (CONSELHO FEDERAL DE MEDICINA, 2026).

5 Educação como interface formativa da integridade decisória

A educação entra como interface formativa porque o uso responsável da IA depende de pessoas capazes de compreender o que a tecnologia faz, o que ela não faz e quando sua recomendação deve ser questionada. Essa questão ultrapassa a formação de profissionais da saúde. Professores, estudantes e famílias convivem com sistemas que produzem textos, classificações, recomendações, avaliações e trajetórias personalizadas. A UNESCO (2023) recomenda que a incorporação de IA generativa em educação preserve agência humana, proteção de dados, adequação etária e capacidade crítica. O problema educacional, portanto, não é apenas aprender a operar ferramentas. É desenvolver condições para interpretar resultados, reconhecer incertezas, preservar autoria, identificar vieses e manter autonomia intelectual diante de respostas produzidas por sistemas de IA.

Na formação em saúde, o uso de IA pode apoiar planejamento, simulação, feedback e aprendizagem, mas também favorecer dependência excessiva e enfraquecimento do raciocínio independente. Ceccim (2005) permite compreender a educação permanente como produção de capacidade no próprio trabalho. Allen e Kendeou (2024) tratam alfabetização em IA como competência

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

interdisciplinar necessária à compreensão crítica de sistemas e resultados, enquanto Malerbi et al. (2023) destacam a educação digital como condição para implantação de IA em saúde. Woodhouse et al. (2010) e Fixsen, Ward e Blase (2025) reforçam a articulação entre competência, prática e capacidade de implementação. Esse argumento é especialmente importante para formação inicial e continuada e ganha relevância no Brasil diante da heterogeneidade de uso de IA entre profissionais de saúde descrita por Portilho e Marin (2026).

No Brasil, essa interface encontra apoio em dois movimentos recentes. A Resolução CFM nº 2.454/2026 exige atualização profissional sobre capacidades, limitações, riscos e vieses dos sistemas utilizados na medicina. No campo educacional, o Referencial para Desenvolvimento e Uso Responsáveis de Inteligência Artificial na Educação, do MEC, e as diretrizes discutidas no CNE em 2026 enfatizam centralidade humana, transparência, proteção de dados, equidade, formação e supervisão. Embora pertençam a campos normativos distintos, esses instrumentos convergem na ideia de que a tecnologia deve apoiar e não substituir julgamento, autoria, mediação pedagógica e responsabilidade humana.

Para professores e famílias, essa convergência possui consequência prática. Sistemas de IA podem apoiar aprendizagem e ampliar acessibilidade, mas seu uso precisa ser proporcional à idade, à finalidade pedagógica e ao risco envolvido. Crianças e adolescentes não devem ser tratados como usuários abstratos de tecnologia. Dados pessoais, perfilização, dependência, produção de conteúdo sintético e decisões automatizadas sobre desempenho podem repercutir sobre direitos e trajetórias escolares. A mediação adulta, por isso, precisa combinar orientação pedagógica, proteção de dados, diálogo com as famílias e critérios institucionais claros para adoção, avaliação e revisão das ferramentas utilizadas.

A contribuição da Educação ao argumento pode ser sintetizada em uma proposição. Competência algorítmica medeia a passagem da supervisão formal

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

para a supervisão substantiva. A presença de um profissional, professor ou responsável no circuito decisório só preserva autonomia quando existem conhecimento, tempo, informação e autoridade reais para verificar, contrariar e justificar a relação com a recomendação algorítmica. Assim, educação não constitui tema lateral do artigo. Ela fornece a infraestrutura formativa sem a qual autonomia, contestabilidade e responsabilidade permanecem apenas formais.

6 Democracia como interface de legitimidade decisória

A democracia entra como interface de legitimidade, não como terceiro setor equivalente à saúde. No plano jurídico brasileiro, essa leitura encontra base na Constituição de 1988. Os direitos à vida, à liberdade, à igualdade e à segurança previstos no art. 5º limitam decisões tecnológicas que possam produzir discriminação ou reduzir autonomia. O art. 37 submete a administração pública aos princípios de legalidade, impessoalidade, moralidade, publicidade e eficiência. Na saúde, os arts. 196 a 198 reconhecem o direito universal e igualitário, qualificam ações e serviços como de relevância pública e incluem a participação da comunidade entre as diretrizes do SUS. Esses fundamentos impedem que eficiência algorítmica seja utilizada como justificativa autossuficiente para decisões públicas de alto impacto (BRASIL, 1988).

A Recomendação da UNESCO sobre Ética da Inteligência Artificial também situa proteção de direitos humanos, diversidade, inclusão, transparência e responsabilização como condições para governança socialmente legítima da IA (UNESCO, 2021). Esses princípios funcionam como referência internacional e não substituem as obrigações constitucionais e legais brasileiras discutidas a seguir.

A participação comunitária prevista constitucionalmente foi institucionalizada pela Lei nº 8.142/1990, que organiza Conselhos e Conferências de Saúde como instâncias permanentes de participação na gestão do SUS. Em uma agenda de IA, essa tradição jurídica amplia o problema da legitimidade. Critérios utilizados em triagem, classificação de risco, filas, prioridades, alocação

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

de recursos ou acesso a serviços não dizem respeito apenas à relação entre usuário e software. Eles envolvem opções institucionais que podem afetar coletividades e, por isso, precisam permanecer sujeitas a controle, publicidade compatível com a proteção de dados e formas institucionais de participação e revisão. (BRASIL, 1990).

A contestabilidade possui fundamento jurídico específico na proteção de dados, mas a categoria utilizada na matriz é mais ampla do que o direito previsto no art. 20 da LGPD. Esse dispositivo assegura ao titular o direito de solicitar revisão de decisões tomadas unicamente com base em tratamento automatizado que afetem seus interesses e determina o fornecimento de informações claras e adequadas sobre critérios e procedimentos utilizados, ressalvados os segredos comercial e industrial. Em decisões clínicas apoiadas por IA, a decisão final pode permanecer humana e, por isso, o art. 20 não abrange automaticamente todas as situações analisadas. A contestabilidade aqui adotada possui sentido ético-institucional mais abrangente. Refere-se à existência de condições para questionar a influência algorítmica, obter reavaliação por autoridade competente e receber resposta institucional, mesmo quando não se configura uma decisão unicamente automatizada. Esse fundamento jurídico decorre da Lei nº 13.709/2018 (BRASIL, 2018). A Resolução CFM nº 2.454/2026 aproxima essa exigência da prática médica ao preservar autonomia profissional, informação ao paciente, supervisão, monitoramento e possibilidade de questionamento dos resultados produzidos por IA.

A interface democrática acrescenta responsabilidade pública e devido processo institucional. Transparência só possui valor quando produz inteligibilidade suficiente para profissionais, pacientes, famílias, gestores e órgãos de controle. Justificabilidade exige razões que possam ser apresentadas e examinadas. Contestabilidade exige canal real de questionamento. Revisão exige autoridade capaz de corrigir ou interromper uma decisão. Responsabilização exige que os papéis dos diferentes agentes sejam identificáveis. Essa cadeia

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

aproxima governança algorítmica de princípios jurídicos de proteção de direitos, controle da administração e acesso a mecanismos de tutela individual e coletiva.

Em termos de desenho institucional, a contestabilidade também pode ser incorporada desde a concepção dos processos. Alfrink et al. (2023) argumentam que sistemas contestáveis precisam prever pontos de intervenção e condições para questionamento, e não apenas remédios posteriores ao dano. Na saúde, essa orientação reforça a necessidade de definir previamente quem recebe a contestação, quais registros sustentam a reavaliação e que autoridade pode alterar ou interromper a decisão.

Saúde, Educação e Democracia convergem, assim, sem formar três objetos equivalentes. A Saúde oferece o contexto de alto impacto em que decisões podem afetar vida, cuidado e acesso. A Educação explica a formação das capacidades necessárias para compreender e supervisionar a tecnologia. A Democracia, apoiada em fundamentos constitucionais, sanitários e de proteção de dados, qualifica legitimidade, participação, contestação e responsabilidade pública. Essa articulação permite avaliar a integridade decisória sem reduzir o problema à precisão do modelo ou à mera presença formal de um ser humano.

No plano internacional, a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito reforça a vinculação entre governança de IA, direitos fundamentais, transparência, accountability e mecanismos de recurso (COUNCIL OF EUROPE, 2024). Esse referencial não substitui o direito brasileiro, mas confirma que contestabilidade e supervisão possuem dimensão institucional e democrática que ultrapassa a explicação técnica do sistema.

7 Matriz de integridade da decisão mediada por inteligência artificial

O diálogo entre saúde, educação e democracia permite formular uma matriz teórico-analítica para examinar a integridade de decisões mediadas por IA em saúde. Neste artigo, integridade decisória designa a propriedade do processo sociotécnico e institucional pela qual uma decisão influenciada por IA permanece

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

clinicamente válida, proporcional ao risco, submetida a autoridade humana substantiva, justificável, contestável, equitativa, rastreável e vinculada a responsabilidades identificáveis. Não é atributo intrínseco do algoritmo. Resulta da relação entre sistema, profissionais, instituição, regras de governança e pessoas afetadas. A definição possui pretensão conceitual geral para decisões de alto impacto, enquanto o contexto brasileiro funciona como aplicação normativa e institucional específica.

Quadro 1. Evidências e tensões da decisão clínica mediada por inteligência artificial

Eixo	Síntese da literatura	Risco para a decisão	Implicação teórica
Apoio clínico	A IA pode ampliar precisão, padronização e capacidade analítica em tarefas específicas.	Aceitação da saída como substituta do julgamento clínico.	Utilidade clínica não equivale a integridade decisória.
Pós-implantação	O desempenho pode variar após implantação e entre contextos ou especialidades.	Perda de validade contextual e falsa segurança.	Validação deve ser contínua e sensível ao contexto.
Interação humano IA	Pressão de tempo, confiança excessiva e desenho da interface influenciam concordância com a IA.	Automação acrítica, sobredependência e enfraquecimento do raciocínio independente.	Supervisão precisa ser substantiva e apoiada por competência algorítmica.
Justiça e dados	Erros e desempenho podem variar entre grupos, territórios e populações.	Reprodução de desigualdades e exclusões.	Equidade exige avaliação distributiva e integridade dos dados.
Governança	Transparência isolada não assegura responsabilidade.	Opacidade organizacional e indefinição de deveres.	Auditoria, contestação e monitoramento devem acompanhar o ciclo de vida.

Fonte. Elaboração dos autores com base na literatura analisada.

A leitura conjunta do Quadro 1 permite identificar um padrão. Quanto mais a IA participa da formação da decisão, menos adequado é avaliá-la apenas por precisão média. Validade pós-implantação, capacidade de recusa, distribuição de erros e definição de responsabilidades passam a integrar a própria qualidade da decisão. Essa passagem justifica a matriz apresentada no Quadro 2, que reorganiza as tensões da literatura em perguntas de integridade.

O Quadro 2 organiza o núcleo conceitual da proposta em seis dimensões.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

Seus componentes possuem antecedentes na literatura e nos documentos analisados. Validade, segurança, autonomia profissional, supervisão humana, transparência, explicabilidade, responsabilidade, equidade, auditabilidade e governança do ciclo de vida não são categorias criadas por este artigo. A operação teórica original consiste em reuni-las, diferenciá-las e relacioná-las sob a categoria de integridade decisória, tomando a decisão sociotécnica, e não o sistema de IA isolado, como unidade de análise. Cada dimensão corresponde a uma pergunta que precisa ser respondida quando a IA influencia uma decisão de alto impacto.

As seis dimensões são interdependentes e não possuem lógica aditiva ou compensatória. A matriz não autoriza somar itens nem produzir pontuação global. Em decisões de maior risco, uma deficiência grave em dimensão essencial pode comprometer a integridade do processo mesmo quando outras dimensões apresentam desempenho satisfatório. Elevada acurácia não compensa ausência de validade no contexto de uso, assim como boa explicabilidade não neutraliza falta de possibilidade efetiva de revisão. Validade mínima para a finalidade, supervisão humana substantiva, possibilidade de revisão em decisões de alto impacto e rastreabilidade suficiente para responsabilização funcionam como condições que não devem ser presumidas compensáveis. A definição de limiares, pesos ou critérios formais de suficiência exigiria validação empírica posterior e não integra a presente proposição.

Quadro 2. Matriz teórico-analítica de integridade da decisão algorítmica em saúde

Dimensão	Pergunta de integridade	Foco de avaliação
Validade clínica, segurança e proporcionalidade	O desempenho sustenta o uso pretendido e o nível de controle corresponde ao risco e à reversibilidade?	Validação local, segurança, criticidade, contexto e benefício clínico.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

Autonomia, competência e supervisão substantiva	O profissional consegue compreender, verificar, rejeitar ou interromper a recomendação?	Autoridade humana real, formação, tempo, confiança calibrada e registro de divergência.
Transparência, explicabilidade e justificabilidade	A participação da IA pode ser compreendida e a decisão pode ser defendida com razões relevantes?	Informação adequada ao destinatário, limitações, critérios e fundamento clínico.
Responsabilidade, contestabilidade e revisão	Estão definidos responsáveis e meios efetivos de questionamento, recurso e correção?	Papéis profissionais e institucionais, segunda avaliação, canais de revisão e resposta.
Justiça, equidade e integridade dos dados	Os erros e efeitos são avaliados entre grupos, territórios e contextos?	Representatividade, qualidade dos dados, desempenho por subgrupos e efeitos distributivos.
Auditabilidade e governança do ciclo de vida	É possível reconstruir a decisão e acompanhar o sistema antes, durante e depois do uso?	Registros, versão do modelo, monitoramento, reavaliação, atualização e descontinuação.

Fonte. Elaboração dos autores com base na literatura e nos documentos normativos analisados.

O Quadro 2 concentra o salto teórico do artigo. A unidade de análise deixa de ser o sistema de IA isolado e passa a ser a decisão mediada por IA. Isso permite observar simultaneamente o sistema, o profissional, a instituição e a pessoa afetada. A matriz também explicita por que Educação e Democracia são necessárias ao argumento. Competência e autonomia dependem de formação, enquanto justificabilidade, contestabilidade e responsabilidade dependem de condições jurídicas e institucionais de participação e revisão.

O Quadro 3 responde à principal dificuldade identificada na literatura sobre ética da IA, a distância entre princípios e prática institucional. Para cada dimensão, o quadro relaciona um risco predominante, um mecanismo de resposta, um indício documental passível de verificação e atores com responsabilidade principal. Esses registros não equivalem a provas suficientes de integridade ética ou efetividade da governança. Um comprovante de treinamento demonstra que houve atividade formativa, mas não comprova competência efetiva. Um log demonstra rastreabilidade, mas não assegura supervisão substantiva. Uma métrica estratificada identifica determinada distribuição de

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

desempenho, mas não resolve, isoladamente, a questão normativa da equidade. A função dessas evidências é documentar mecanismos e permitir auditoria crítica.

Quadro 3. Da dimensão ética ao mecanismo institucional verificável

Dimensão	Risco principal	Mecanismo institucional	Indício documental verificável	Atores com responsabilidade principal
Validade e segurança	Desempenho inadequado ou perda de validade contextual.	Validação local e reavaliação periódica.	Relatórios de desempenho, calibração e eventos adversos.	Instituição, equipe técnica e responsáveis clínicos.
Autonomia e competência	Aceitação automática e dependência excessiva.	Formação continuada e autoridade para recusa.	Registro de treinamento, divergências e override.	Instituição e profissional.
Justificabilidade	Opacidade e explicação inútil ao destinatário.	Informação adequada ao contexto e documentação decisória.	Relatórios, prontuário e registro da participação da IA.	Profissional e instituição.
Responsabilidade e revisão	Lacunas de responsabilização e ausência de recurso.	Distribuição prévia de papéis e canais de contestação.	Protocolos, registros de revisão e respostas a reclamações.	Instituição, profissional, fornecedor e órgãos competentes.
Justiça e dados	Erros desiguais e reprodução de desigualdades.	Auditoria por grupos e governança de dados.	Métricas estratificadas, qualidade e representatividade dos dados.	Instituição, equipe técnica e governança de dados.
Ciclo de vida	Degradação, drift, atualização não controlada.	Monitoramento contínuo, revalidação e critérios de suspensão.	Logs, histórico de versões, auditorias e decisões de atualização.	Instituição, desenvolvedor e instâncias de governança.

Fonte. Elaboração dos autores.

A interpretação do Quadro 3 exige atenção à distribuição de responsabilidades. A coluna relativa aos atores não estabelece responsabilidade jurídica exclusiva nem esgota a cadeia de accountability. Desenvolvimento, contratação, implantação, uso, monitoramento e regulação podem envolver agentes distintos, cujas obrigações dependem do contexto e do regime jurídico aplicável. Nenhuma coluna pode ser lida isoladamente. Formação continuada

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

sem autoridade para recusa não assegura autonomia. Auditoria sem dados estratificados não sustenta equidade. Registro sem canal de revisão não garante contestabilidade. Monitoramento sem critérios de suspensão não configura governança do ciclo de vida. Os indícios documentais tornam mecanismos auditáveis, mas precisam ser confrontados com sua efetividade prática. O quadro reforça, portanto, que integridade decisória é propriedade institucionalmente produzida e depende da articulação entre pessoas, procedimentos, registros e instâncias de controle.

8 Interpretação da matriz e implicações para a saúde

A matriz reorganiza a avaliação ética em torno da decisão concreta. Em vez de perguntar de forma genérica se um sistema é ético, examina se o processo decisório reúne condições suficientes de validade, supervisão, justificabilidade, contestação, equidade e auditabilidade. A utilidade da proposta está em tornar visíveis fragilidades que métricas de desempenho, isoladamente, não capturam.

A distinção entre integridade do sistema e integridade da decisão é fundamental. Integridade do sistema se refere a dados, segurança, desempenho, documentação e comportamento do modelo. Integridade da decisão considera como esse sistema participa de uma escolha concreta, com determinada população, profissional, finalidade, consequência e possibilidade de revisão. A unidade principal da proposta é a segunda, sem ignorar que decisões íntegras dependem de sistemas tecnicamente adequados.

A supervisão humana constitui condição operacional, mas sua simples presença formal é insuficiente. Para que o controle seja substantivo, precisam existir competência, informação relevante, tempo institucional, autoridade para discordar e possibilidade de interromper ou modificar a recomendação. A Resolução CFM nº 2.454/2026 afirma expressamente que soluções de IA não são soberanas, exige supervisão humana e preserva a autoridade final do médico (CONSELHO FEDERAL DE MEDICINA, 2026).

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

A discordância entre profissional e sistema de IA também pode produzir aprendizagem institucional. Registros recorrentes de rejeição, correção ou revisão podem revelar problemas de calibração, mudanças epidemiológicas, inadequação contextual, falhas de interface ou situações clínicas pouco representadas nos dados originais. O override deixa, assim, de ser apenas mecanismo de segurança e passa a oferecer informação para governança e melhoria contínua.

A rastreabilidade precisa permitir reconstruir a proveniência da decisão. Essa reconstrução envolve dados utilizados, versão do modelo, recomendação produzida, informação apresentada ao profissional, decisão humana adotada, eventual revisão e resultado posterior. Raji et al. (2020) mostram que auditoria algorítmica efetiva exige estruturas de accountability ao longo do ciclo de desenvolvimento e uso, e não apenas inspeção pontual do modelo. Sem uma cadeia reconstruível, auditoria e responsabilização permanecem limitadas. Essa exigência aproxima a governança de uma lógica centrada em evidências documentais, e não apenas em declarações de princípios.

A equidade exige distinguir fairness algorítmica, equidade clínica e justiça institucional. Métricas estatísticas ajudam a identificar disparidades, mas sua escolha possui pressupostos normativos. Uma mesma aplicação pode apresentar boa acurácia média e desempenho inadequado para um subgrupo. Também pode ser tecnicamente uniforme e ainda ampliar desigualdades quando depende de conectividade, infraestrutura ou competências distribuídas de maneira desigual. A avaliação precisa considerar dados, resultados e condições institucionais de uso.

A governança é temporal. Antes da implantação, requer validação, análise de risco e definição de responsabilidades. Durante a decisão, exige informação, supervisão e possibilidade de recusa. Depois da decisão, precisa garantir registro, explicação, contestação e revisão. Ao longo do uso institucional, requer monitoramento de desempenho, viés, mudanças no contexto, incidentes e

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

necessidade de atualização ou descontinuação. Integridade decisória é, portanto, processual e não episódica.

A criticidade temporal e a carga cognitiva também importam. Transparência adequada não significa oferecer o maior volume possível de informação. Significa disponibilizar informação relevante, inteligível e utilizável no momento em que o profissional precisa decidir. Interfaces excessivamente complexas, alertas reiterados ou explicações pouco conectadas ao problema clínico podem dificultar revisão crítica e favorecer aceitação automática.

Com essa organização, Educação e Democracia refinam a análise sem disputar centralidade com a Saúde. A primeira explica como capacidades formativas sustentam autonomia e supervisão. A segunda qualifica legitimidade, contestação e responsabilidade pública. O objeto permanece a decisão em saúde e a pergunta permanece institucional, quais condições tornam verificável sua integridade quando a IA influencia o processo.

9 Implicações para políticas públicas brasileiras

O contexto brasileiro combina proteção de dados, saúde digital, telessaúde, formação profissional e regulação específica da prática médica. Não existe um único regime jurídico que concentre todas essas dimensões. O desenho é incremental e setorial. A LGPD fornece base para o tratamento de dados pessoais sensíveis (BRASIL, 2018). A Lei nº 14.510/2022 consolidou a telessaúde de forma permanente em todo o território nacional (BRASIL, 2022). O Programa SUS Digital organiza uma agenda de transformação digital do sistema público. O Projeto de Lei nº 2.338/2023 permanece em tramitação na Câmara dos Deputados e deve ser tratado como proposta regulatória, não como norma vigente (BRASIL, 2023). A versão aprovada pelo Senado adota abordagem baseada em risco e articula direitos de informação, explicação, contestação e revisão humana com mecanismos de avaliação de impacto e governança, em complementaridade com a proteção de dados pessoais (PEREIRA et al., 2026).

A matriz não foi concebida como instrumento exclusivamente brasileiro.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

Sua arquitetura deriva de literatura e referenciais internacionais, enquanto esta seção examina como seus componentes encontram correspondência, tensões e limites no ambiente institucional brasileiro. O Brasil funciona, portanto, como contexto de aplicação normativa da proposição, e não como limite de sua pretensão conceitual.

Em 2026, a Resolução CFM nº 2.454 tornou-se referência central para o uso de IA na medicina. A norma estabelece regras sobre pesquisa, desenvolvimento, governança, auditoria, monitoramento, capacitação e uso responsável. Reconhece a IA como apoio à decisão, preserva a autoridade final do médico, exige informação ao paciente quando a tecnologia participa de forma relevante do cuidado e prevê classificação de risco, auditabilidade, contestabilidade e gestão do ciclo de vida (CONSELHO FEDERAL DE MEDICINA, 2026).

Essa resolução possui forte correspondência com a matriz teórica proposta. Validação científica e segurança dialogam com a primeira dimensão. Autonomia e atualização profissional aproximam regulação e competência algorítmica. Explicabilidade, informação ao paciente e possibilidade de revisão reforçam justificabilidade e contestabilidade. A previsão de auditoria, monitoramento contínuo, avaliação de risco e mitigação de vieses aproxima a norma das dimensões de justiça e governança do ciclo de vida.

No arranjo federal, a coordenação da saúde digital permanece distribuída. O Ministério da Saúde conduz a política nacional por meio da Secretaria de Informação e Saúde Digital, responsável pelo Programa SUS Digital. O DataSUS integra essa estrutura como departamento de informação e informática do SUS e sustenta parte relevante da infraestrutura informacional utilizada para gestão, planejamento e serviços digitais. Quando a aplicação de IA se enquadra como software destinado a finalidade médica, a Anvisa exerce a regulação sanitária aplicável aos softwares como dispositivos médicos. Esse desenho confirma que a governança da IA no SUS opera em rede e combina direção política,

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

infraestrutura de dados, regulação sanitária e capacidades institucionais, sem uma autoridade única exclusiva para todas as aplicações de IA em saúde (HADDAD et al., 2025; GUARITA JAMBOR, 2025).

Essa distribuição de competências também ajuda a delimitar responsabilidades. A coordenação ministerial não substitui a avaliação técnica local, a regulação sanitária não esgota a governança institucional e a disponibilidade de dados não assegura, por si só, validade clínica ou equidade. A integridade da decisão depende da articulação entre essas camadas e da capacidade de cada instituição documentar critérios, validar o uso em seu contexto, monitorar efeitos e interromper aplicações quando riscos superarem benefícios.

A educação permanente constitui parte dessa capacidade pública. UNA-SUS, AVASUS e outras estruturas de formação não são programas de IA, mas formam uma base que pode apoiar atualização profissional em saúde digital. A literatura brasileira também registra adoção desigual de ferramentas de IA entre profissionais e instituições, o que reforça a necessidade de capacitação, governança ética e infraestrutura adequada. Formação precisa ser tratada como condição de implementação e não como atividade acessória.

No campo educacional, o cenário brasileiro passou por uma inflexão em 2026. O Referencial para Desenvolvimento e Uso Responsáveis de Inteligência Artificial na Educação, do Ministério da Educação, consolidou orientações nacionais sobre centralidade dos educadores, transparência, explicabilidade, proteção de dados, equidade, formação docente, avaliação de impacto e governança institucional (BRASIL, 2026). Em setembro do mesmo ano, o Conselho Nacional de Educação apresentou diretrizes orientadoras para a utilização da IA em todos os níveis, etapas e modalidades da educação brasileira. A documentação da Comissão Bicameral explicita centralidade da regência humana, proporcionalidade regulatória, supervisão humana significativa, documentação, rastreabilidade, contestação e revisão. Na data de corte de 29 de

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

setembro de 2026, a correspondente resolução normativa não foi identificada como ato publicado, razão pela qual o documento é tratado como etapa de institucionalização e não como norma regulamentar plenamente vigente (CONSELHO NACIONAL DE EDUCAÇÃO, 2026).

A articulação entre Saúde e Educação deixa, assim, de aparecer apenas como aproximação conceitual. Em 2026, a Resolução CFM nº 2.454 e os referenciais educacionais federais passaram a expressar princípios convergentes para decisões mediadas por IA. Em ambos os campos aparecem centralidade humana, proporcionalidade ao risco, formação, transparência, proteção de dados, documentação, monitoramento e possibilidade de revisão. Essa convergência não elimina diferenças de finalidade e força normativa, mas permite compreender capacidade institucional para IA como combinação entre capacidade técnica, capacidade formativa e capacidade de governança.

A principal implicação para políticas públicas é que adoção responsável não depende apenas de aquisição tecnológica. Serviços precisam avaliar se possuem dados adequados, infraestrutura, profissionais preparados, mecanismos de auditoria, canais de revisão e responsabilidades definidas. Em sistemas públicos marcados por desigualdades territoriais, uma tecnologia pode ser segura em uma instituição e inadequada em outra. A decisão sobre implantação precisa considerar essas diferenças.

A literatura de implementação ajuda a qualificar esse argumento. Asada et al. (2023) mostram que políticas orientadas à equidade precisam examinar simultaneamente desenho, implementação e efeitos distributivos. Aplicada à governança da IA, essa perspectiva indica que direitos, protocolos e salvaguardas só se tornam efetivos quando existem capacidades institucionais para sustentá-los no cotidiano. A contribuição dessas abordagens ao presente artigo é teórica. Elas ajudam a explicar por que a distância entre norma e prática deve ser tratada como dimensão da governança, sem converter o estudo em avaliação empírica de implementação.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

A participação social completa esse desenho. Conselhos, usuários, entidades profissionais e instâncias de controle podem discutir critérios de transparência, usos de maior impacto e condições de aceitabilidade de sistemas que afetam acesso, cuidado e distribuição de recursos. A participação não substitui avaliação técnica, mas amplia legitimidade e responsabilidade pública. No Brasil, esse aspecto é especialmente relevante porque a institucionalização da IA ocorre em ritmos diferentes entre setores e territórios.

10 Conclusão

A expansão da inteligência artificial em saúde exige que o debate ético ultrapasse listas de princípios e métricas isoladas. A questão central é em que condições a influência de um sistema de IA pode integrar uma decisão legítima e segura. A análise desenvolvida indica que essa resposta depende da articulação entre validade clínica, proporcionalidade ao risco, autoridade profissional efetiva, justificabilidade, possibilidade de revisão, equidade, rastreabilidade e governança ao longo do ciclo de vida.

A análise respondeu a esse problema por meio de seis dimensões articuladas. Validade clínica, segurança e proporcionalidade situam o desempenho no contexto real de uso. Autonomia, competência e supervisão substantiva verificam se o profissional dispõe de conhecimento, tempo e autoridade para aceitar, rejeitar ou interromper uma recomendação. Transparência, explicabilidade e justificabilidade qualificam as razões que sustentam a decisão. Responsabilidade, contestabilidade e revisão estruturam caminhos de correção. Justiça, equidade e integridade dos dados examinam a distribuição de riscos e benefícios. Auditabilidade e governança do ciclo de vida asseguram que a influência algorítmica possa ser reconstruída, monitorada e, quando necessário, suspensa.

A Educação mostrou que essas condições não surgem automaticamente com a presença de um profissional ou de um responsável. Elas precisam ser formadas. Letramento em IA, pensamento crítico, domínio disciplinar, autonomia

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

intelectual e compreensão dos limites dos sistemas são componentes da própria governança. Para profissionais da saúde, professores, estudantes e famílias, isso significa substituir uma cultura de uso instrumental por uma cultura de mediação crítica. A tecnologia pode apoiar cuidado e aprendizagem, mas não deve obscurecer autoria, julgamento, responsabilidade ou direitos das pessoas afetadas.

A Democracia acrescentou a dimensão de legitimidade. No SUS, participação social, igualdade de acesso, publicidade administrativa e proteção de direitos possuem base constitucional e legal. A LGPD introduz direitos de informação e revisão em decisões exclusivamente automatizadas, enquanto a Lei nº 8.142/1990 institucionaliza participação comunitária na gestão da saúde. Esses fundamentos jurídicos mostram que governança da IA não pode ser reduzida a boas práticas voluntárias. Em contextos públicos, decisões que afetam acesso, prioridade ou direitos precisam estar sujeitas a justificativa, controle, contestação e responsabilização.

A contribuição teórica do artigo reside, portanto, na matriz de integridade da decisão algorítmica em saúde. Seu diferencial é relacionar dimensões éticas a mecanismos institucionais, evidências verificáveis e responsáveis identificáveis. Os três quadros cumprem funções complementares nesse percurso. O Quadro 1 organiza tensões que tornam insuficiente avaliar apenas o desempenho técnico. O Quadro 2 transforma essas tensões em perguntas de integridade. O Quadro 3 mostra como princípios podem ser relacionados a procedimentos, registros e responsabilidades. A sequência evidencia a passagem da ética abstrata para a decisão concreta sem converter a proposta em escala, protocolo clínico ou instrumento regulatório validado.

Essa contribuição deve ser lida à luz dos limites do próprio desenho. A pesquisa bibliográfica foi orientada e não exaustiva, o que impede inferências sobre frequência ou prevalência de temas. Educação e Democracia foram mobilizadas como interfaces analíticas e não como unidades empíricas

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

comparadas. A matriz permanece uma proposição teórico-analítica ainda não validada empiricamente. Seu desenvolvimento futuro pode incluir avaliação de conteúdo por painel multidisciplinar, técnica Delphi, aplicação piloto em decisões clínicas com diferentes níveis de risco, análise de concordância entre avaliadores e estudos de implementação em instituições e territórios distintos. Essas etapas poderão testar clareza, suficiência, aplicabilidade e sensibilidade das dimensões sem pressupor, nesta fase, uma lógica de escore ou certificação. O ambiente regulatório e tecnológico também muda rapidamente, razão pela qual documentos e proposições legislativas devem ser revistos segundo a data de corte informada no percurso metodológico.

A conclusão central é que decidir com algoritmos não significa transferir a decisão à tecnologia. Quanto maior a capacidade da IA de influenciar diagnóstico, prioridade, acesso ou conduta, maior deve ser a capacidade humana e institucional de compreender, verificar, contrariar, justificar, revisar e responder por essa influência. Em saúde, essa exigência protege segurança e autonomia. Na educação, sustenta formação crítica e mediação responsável. Na democracia, preserva legitimidade, participação e controle. A integridade decisória emerge exatamente da articulação entre essas dimensões.

Declarações editoriais

Conflitos de interesse. Os autores declaram inexistência de conflitos de interesse relacionados ao manuscrito.

Aspectos éticos. Não se aplica apreciação por Comitê de Ética em Pesquisa, pois o estudo é teórico-conceitual e não utiliza participantes humanos, dados pessoais identificáveis, prontuários ou material biológico.

Disponibilidade de dados. Não se aplica, pois o estudo não produziu conjunto de dados empíricos.

Uso de inteligência artificial. Em consonância com a Política de Integridade na Atividade Científica do CNPq, instituída pela Portaria CNPq nº 2.664/2026, declara-se que, na elaboração deste artigo, foi utilizada a ferramenta de

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

Inteligência Artificial Generativa ChatGPT nas fases de revisão e adequação normativa (ABNT). A ferramenta não foi empregada para geração autônoma de resultados ou substituição da responsabilidade autoral. Todo o conteúdo foi integralmente revisado, verificado e validado pelo autor, que assume responsabilidade pela integridade, originalidade e exatidão do texto final (CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO, 2026).

Referências

AKHLAGHI, H.; FREEMAN, S.; VARI, C.; MCKENNA, B.; BRAITBERG, G.; KARRO, J.; TAHAYORI, B. Machine learning in clinical practice: evaluation of an artificial intelligence tool after implementation. **Emergency Medicine Australasia**, v. 36, n. 1, p. 118-124, 2024. DOI: https://doi.org/10.1111/1742-6723.14325.

ALFRINK, K.; KELLER, I.; KORTUEM, G.; DOORN, N. Contestable AI by design: towards a framework. **Minds and Machines**, v. 33, p. 613-639, 2023. DOI: https://doi.org/10.1007/s11023-022-09611-z.

ALLEN, L. K.; KENDEOU, P. ED-AI Lit: an interdisciplinary framework for AI literacy in education. **Policy Insights from the Behavioral and Brain Sciences**, v. 11, n. 1, p. 3-10, 2024. DOI: https://doi.org/10.1177/23727322231220339.

ASADA, Yuka; KROLL-DESROSIERS, Aimee; CHRIQUI, Jamie F.; CURRAN, Geoffrey M.; EMMONS, Karen M.; HAIRE-JOSHU, Debra; BROWNSON, Ross C. Applying hybrid effectiveness-implementation studies in equity-centered policy implementation science. **Frontiers in Health Services**, v. 3, 1220629, 2023. DOI: https://doi.org/10.3389/frhs.2023.1220629.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 29 set. 2026.

BRASIL. **Lei nº 8.142, de 28 de dezembro de 1990**. Dispõe sobre a participação da comunidade na gestão do Sistema Único de Saúde (SUS) e sobre as transferências intergovernamentais de recursos financeiros na área da saúde e dá outras providências. Brasília, DF: Presidência da República, 1990. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8142.htm. Acesso em: 29 set. 2026.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em:

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 29 set. 2026.

BRASIL. **Lei nº 14.510, de 27 de dezembro de 2022**. Altera a Lei nº 8.080, de 19 de setembro de 1990, para autorizar e disciplinar a prática da telessaúde em todo o território nacional, e dá outras providências. Brasília, DF: Presidência da República, 2022. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2019-2022/2022/lei/l14510.htm. Acesso em: 29 set. 2026.

BRASIL. Senado Federal. **Projeto de Lei nº 2.338, de 2023**. Dispõe sobre o uso da Inteligência Artificial. Brasília, DF: Senado Federal, 2023. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 29 set. 2026.

BRASIL. Ministério da Educação. **Referencial para Desenvolvimento e Uso Responsáveis de Inteligência Artificial na Educação**. 1. ed. Brasília, DF: MEC, 2026.

CECCIM, Ricardo Burg. Educação Permanente em Saúde: descentralização e disseminação de capacidade pedagógica na saúde. **Ciência & Saúde Coletiva**, v. 10, n. 4, p. 975-986, 2005. DOI: https://doi.org/10.1590/S1413-81232005000400020.

CELLARD, André. A análise documental. In: POUPART, Jean et al. **A pesquisa qualitativa: enfoques epistemológicos e metodológicos**. Petrópolis: Vozes, 2008. p. 295-316.

CHEN, Irene Y.; PIERSON, Emma; ROSE, Sherri; JOSHI, Shalmali; FERRYMAN, Kadija; GHASSEMI, Marzyeh. Ethical machine learning in healthcare. **Annual Review of Biomedical Data Science**, v. 4, p. 123-144, 2021. DOI: https://doi.org/10.1146/annurev-biodatasci-092820-114757.

CONSELHO FEDERAL DE MEDICINA. **Resolução CFM nº 2.454, de 11 de fevereiro de 2026**. Normatiza o uso da inteligência artificial na medicina. Diário Oficial da União: seção 1, Brasília, DF, 27 fev. 2026, p. 158. Em vigor desde 26 ago. 2026.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Portaria CNPq nº 2.664, de 6 de março de 2026**. Institui a Política de Integridade na Atividade Científica do CNPq. Brasília, DF: CNPq, 2026. Disponível em: https://www.gov.br/cnpq/pt-br/composicao/comissao-de-integridade/documentos. Acesso em: 29 set. 2026.

CONSELHO NACIONAL DE EDUCAÇÃO. **Diretrizes para a Utilização da Inteligência Artificial na Educação Brasileira**. Apresentação da Comissão

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

Bicameral. Brasília, DF: CNE, 2026. Documento institucional utilizado como referência de processo de institucionalização; verificação normativa encerrada em 29 set. 2026.

COUNCIL OF EUROPE. **Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law**. CETS No. 225. Vilnius, 5 Sept. 2024. Disponível em: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence. Acesso em: 29 set. 2026.

ECONOMOU-ZAVLANOS, N. J.; BESSIAS, S.; CARY, M. P. et al. Translating ethical and quality principles for the effective, safe and fair development, deployment and use of artificial intelligence technologies in healthcare. **Journal of the American Medical Informatics Association**, v. 31, n. 3, p. 705-713, 2024. DOI: https://doi.org/10.1093/jamia/ocad221.

FÁVERO, Altair Alberto; CENTENARO, Junior Bufon. A pesquisa documental nas investigações de políticas educacionais: potencialidades e limites. **Contrapontos**, v. 19, n. 1, p. 170-184, 2019. DOI: https://doi.org/10.14210/contrapontos.v19n1.p170-184.

FIXSEN, Dean L.; WARD, Caryn S.; BLASE, Karen A. Developing Implementation & Scaling Capacity in Education. **Global Implementation Research and Applications**, v. 5, p. 538-555, 2025. DOI: https://doi.org/10.1007/s43477-025-00186-2.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GRAZZIOTIN, Luciane Sgarbi; KLAUS, Viviane; PEREIRA, Ana Paula Marques. Pesquisa documental histórica e pesquisa bibliográfica: focos de estudo e percursos metodológicos. **Pro-Posições**, v. 33, e20200141, 2022. DOI: https://doi.org/10.1590/1980-6248-2020-0141.

GROSSER, John; DÜVEL, Juliane; HASEMANN, Lena; SCHNEIDER, Emilia; GREINER, Wolfgang. Studying the potential effects of artificial intelligence on physician autonomy: scoping review. **JMIR AI**, v. 4, e59295, 2025. DOI: https://doi.org/10.2196/59295.

GUARITA JAMBOR, Daniela. Regulatory assessment of software as medical device with a focus on inclusion and equity in artificial intelligence applications in healthcare. **Revista de Direito Sanitário**, v. 25, n. 1, e0017, 2025. DOI: https://doi.org/10.11606/issn.2316-9044.rdisan.2025.231708.

HADDAD, A. E.; GARRIDO, D.; SHIBATA, L.; SANTOS, A. F.; D'AGOSTINO, M. Evolution of Brazilian Telehealth Regulation: from the National Telehealth Program

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

to SUS Digital. **Revista de Direito Sanitário**, v. 25, n. 1, e0008, 2025. DOI: https://doi.org/10.11606/issn.2316-9044.rdisan.2025.238280.

MALERBI, F. K.; NAKAYAMA, L. F.; DYCHIAO, R. G.; RIBEIRO, L. Z.; VILLANUEVA, C.; CELI, L. A.; REGATIERI, C. V. Digital education for the deployment of artificial intelligence in health care. **Journal of Medical Internet Research**, v. 25, e43333, 2023. DOI: https://doi.org/10.2196/43333.

MITTELSTADT, Brent. Principles alone cannot guarantee ethical AI. **Nature Machine Intelligence**, v. 1, p. 501-507, 2019. DOI: https://doi.org/10.1038/s42256-019-0114-4.

MORLEY, Jessica; FLORIDI, Luciano; KINSEY, Libby; ELHALAL, Anat. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. **Science and Engineering Ethics**, v. 26, p. 2141-2168, 2020. DOI: https://doi.org/10.1007/s11948-019-00165-5.

NING, Yilin et al. Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist. **The Lancet Digital Health**, v. 6, n. 11, p. e848-e856, 2024. DOI: https://doi.org/10.1016/S2589-7500(24)00143-2.

PEREIRA, L. H. F. E.; SPÓSITO, S. L.; DE MENDONCA, F. L. L.; CANEDO, E. D. Privacy and Artificial Intelligence in Brazil: An Analysis of Bill 2,338/2023 and Its Complementarity with the LGPD. In: INTERNATIONAL CONFERENCE ON ENTERPRISE INFORMATION SYSTEMS, 2026. **Proceedings [...]**. DOI: https://doi.org/10.5220/0014730300004018.

PORTILHO, L.; MARIN, H. F. Artificial intelligence: the use and application by Brazilian healthcare professionals. **Studies in Health Technology and Informatics**, v. 336, p. 705-709, 2026. DOI: https://doi.org/10.3233/SHTI260262.

RAJI, Inioluwa Deborah et al. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY, 2020. **Proceedings [...]**. New York: ACM, 2020. p. 33-44. DOI: https://doi.org/10.1145/3351095.3372873.

SOARES, Marcela Quaresma; CHIAVEGATTO FILHO, Alexandre Dias Porto. Inteligência artificial na saúde pública brasileira: potencialidades, desafios e implicações éticas para o Sistema Único de Saúde. **Revista Brasileira de Epidemiologia**, v. 29, e260006, 2026. DOI: https://doi.org/10.1590/1980-549720260006.

UNESCO. **Recommendation on the Ethics of Artificial Intelligence**. Paris: UNESCO, 2021. Disponível em: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics. Acesso em: 29 set. 2026.

CC BY

RSV
ISSN 2674-8584

Received: 11/09/2026 - **Accepted:** 30/09/2026
Vol: 08.09
DOI: 10.66104/my5qpa94
Pages: 1-33

UNESCO. **Guidance for generative AI in education and research**. Paris: UNESCO, 2023. Disponível em: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research. Acesso em: 29 set. 2026.

WOODHOUSE, Lynn D.; AULD, M. Elaine; MINER, Kathleen; ALLEY, Kelly Bishop; LYSOBY, Linda; LIVINGOOD, William C. Crosswalking public health and health education competencies: implications for professional preparation and practice. **Journal of Public Health Management and Practice**, v. 16, n. 3, p. E20-E28, 2010. DOI: https://doi.org/10.1097/PHH.0b013e3181b3a3d2.

WORLD HEALTH ORGANIZATION. **Ethics and governance of artificial intelligence for health: WHO guidance**. Geneva: WHO, 2021. Disponível em: https://www.who.int/publications/i/item/9789240029200. Acesso em: 29 set. 2026.

WORLD HEALTH ORGANIZATION. **Ethics and governance of artificial intelligence for health: guidance on large multi-modal models**. Geneva: WHO, 2024. Disponível em: https://www.who.int/publications/i/item/9789240084759. Acesso em: 29 set. 2026

CC BY