LUMEN ET VIRTUS
ISSN: 2177-2789

GOVERNANÇA DA INTELIGÊNCIA ARTIFICIAL NA SAÚDE: DESAFIOS PARA O MONITORAMENTO DE VIESES, ERROS ALGORÍTMICOS E SEGURANÇA DO PACIENTE

ARTIFICIAL INTELLIGENCE GOVERNANCE IN HEALTHCARE: CHALLENGES FOR MONITORING BIAS, ALGORITHMIC ERRORS, AND PATIENT SAFETY

GOBERNANZA DE LA INTELIGENCIA ARTIFICIAL EN LA ATENCIÓN MÉDICA: DESAFÍOS PARA EL MONITOREO DE SESGOS, ERRORES ALGORÍTMICOS Y SEGURIDAD DEL PACIENTE

https://doi.org/10.56238/levv17n65-019

Data de submissão: 08/09/2026 **Data de publicação:** 08/10/2026

Artur Lima da Silva
Especialista em Segurança de Informação
Instituição: Centro Universitário Venda Nova do Imigrante (UNIFAVENI)
E-mail: arturacademix@gmail.com

Veronica do Carmo Santos Gomes
Graduada em Enfermagem
Instituição: Universidade Estácio de Sá
E-mail: veronica.hud@hotmail.com

Thaís Fernandes do Ramo
Graduada em Fisioterapia
Instituição: Centro Universitário Unifacisa
E-mail: thaisfernandesdoramo@gmail.com

Evelyn Fernandes Santos
Graduada em Medicina
Instituição: Universidad de Aquino de Bolívia (UDABOL)
E-mail: efernandes-es@udabol.edu.bo

Dilce Maria Mendes Magioli
Graduada em Fisioterapia
Instituição: Faculdade Estácio de Sá
E-mail: dilcemmm@gmail.com

Marilene Alves Tavares da Silva
Graduada em Tecnologia em Gestão Hospitalar
Instituição: Unopar
E-mail: tavaresrodrigues@hotmail.com

Maria Aparecida Vital de Lima
Graduada em Farmácia
Instituição: Universidade de Nova Iguaçu (UNIG)
E-mail: Vitallfarmacia@hotmail.com

Michele Santos de Araújo Ferreira
Pós-Graduada em Enfermagem do Trabalho
Instituição: AVM Faculdade Integrada
E-mail: favodemisara@yahoo.com.br

Regianne Antunes Dourado
Especialista em Enfermagem Neonatal
Instituição: Universidade Federal do Ceará (UFC)
E-mail: regiannedourado@gmail.com

Andréia da Costa Soares
Mestranda em Ensino na Sáude
Instituição: Universidade Federal Fluminense (UFF)
E-mail: andreiadacostasoares@gmail.com

RESUMO

OBJETIVO: Analisar os desafios da governança da IA na saúde, com ênfase no monitoramento de vieses, na identificação de erros algorítmicos e na segurança do paciente. MÉTODOS: Revisão narrativa da literatura, de abordagem qualitativa e caráter descritivo-analítico. A busca contemplou a Biblioteca Virtual em Saúde, incluindo LILACS e MEDLINE, além da SciELO e PubMed. Foram utilizados os descritores DeCS/MeSH Inteligência Artificial (Artificial Intelligence), Segurança do Paciente (Patient Safety), Algoritmos (Algorithms), Gestão de Riscos (Risk Management) e Sistemas de Apoio à Decisão Clínica (Decision Support Systems, Clinical), combinados pelos operadores booleanos AND e OR. Dos oito documentos avaliados, sete artigos publicados entre 2025 e 2026 foram incluídos, submetidos à leitura integral e interpretação crítica. RESULTADOS: A baixa representatividade das bases de treinamento e as diferenças de desempenho entre grupos populacionais constituíram fatores associados aos vieses algorítmicos. Foram identificados riscos de falsos negativos, alertas clínicos inadequados e deterioração do desempenho dos modelos após sua implantação. A indefinição das responsabilidades institucionais e a insuficiência de mecanismos de monitoramento dificultam a identificação e correção dessas falhas. No SUS, a fragmentação dos registros clínicos, as desigualdades tecnológicas e as limitações da formação profissional representam obstáculos adicionais à supervisão dos sistemas inteligentes. CONCLUSÃO: A governança da IA requer validação local, auditorias periódicas, indicadores de equidade e supervisão humana contínua. A pesquisa identifica requisitos técnicos, éticos e institucionais para o monitoramento dos algoritmos, contribuindo para a organização de mecanismos de prevenção de riscos assistenciais e responsabilização pelas decisões clínicas.

Palavras-chave: Algoritmos. Gestão de Riscos. Inteligência Artificial. Segurança do Paciente. Sistemas de Apoio à Decisão Clínica.

ABSTRACT

OBJECTIVE: To analyze the challenges of AI governance in healthcare, emphasizing bias monitoring, the identification of algorithmic errors, and patient safety. METHODS: A narrative literature review with a qualitative approach and a descriptive-analytical nature. The search covered the Virtual Health Library (including LILACS and MEDLINE), SciELO, and PubMed. The DeCS/MeSH descriptors used were Artificial Intelligence, Patient Safety, Algorithms, Risk Management, and Decision Support Systems, Clinical, combined using the Boolean operators AND and OR. Of the eight documents

evaluated, seven articles published between 2025 and 2026 were included and subjected to full-text reading and critical interpretation. RESULTS: Low representativeness in training datasets and performance disparities across population groups were factors associated with algorithmic bias. Risks such as false negatives, inappropriate clinical alerts, and model performance degradation post-deployment were identified. Ambiguity regarding institutional responsibilities and insufficient monitoring mechanisms hinder the identification and correction of these failures. Within the SUS (Brazil's Unified Health System), fragmented clinical records, technological inequalities, and limitations in professional training pose additional obstacles to the oversight of intelligent systems. CONCLUSION: AI governance requires local validation, periodic audits, equity indicators, and continuous human oversight. The study identifies technical, ethical, and institutional requirements for algorithm monitoring, contributing to the establishment of mechanisms for preventing care-related risks and ensuring accountability for clinical decisions.

Keywords: Algorithms. Risk Management. Artificial Intelligence. Patient Safety. Clinical Decision Support Systems.

RESUMEN

OBJETIVO: Analizar los desafíos de la gobernanza de la IA en la atención médica, con énfasis en el monitoreo de sesgos, la identificación de errores algorítmicos y la seguridad del paciente. MÉTODOS: Revisión narrativa de la literatura, con un enfoque cualitativo y descriptivo-analítico. La búsqueda incluyó la Biblioteca Virtual en Salud, incluyendo LILACS y MEDLINE, así como SciELO y PubMed. Se utilizaron los descriptores DeCS/MeSH Inteligencia Artificial, Seguridad del Paciente, Algoritmos, Gestión de Riesgos y Sistemas de Apoyo a la Decisión Clínica, combinados con los operadores booleanos AND y OR. De los ocho documentos evaluados, se incluyeron siete artículos publicados entre 2025 y 2026, los cuales fueron sometidos a lectura completa e interpretación crítica. RESULTADOS: La baja representatividad de los conjuntos de datos de entrenamiento y las diferencias de rendimiento entre grupos de población fueron factores asociados con sesgos algorítmicos. Se identificaron riesgos de falsos negativos, alertas clínicas inadecuadas y deterioro del rendimiento del modelo después de la implementación. La falta de responsabilidades institucionales claras y mecanismos de monitoreo insuficientes dificultan la identificación y corrección de estas fallas. En el Sistema Único de Salud (SUS) de Brasil, la fragmentación de los registros clínicos, las desigualdades tecnológicas y las limitaciones en la formación profesional representan obstáculos adicionales para la supervisión de los sistemas inteligentes. CONCLUSIÓN: La gobernanza de la IA requiere validación local, auditorías periódicas, indicadores de equidad y supervisión humana continua. La investigación identifica los requisitos técnicos, éticos e institucionales para el monitoreo de algoritmos, contribuyendo a la organización de mecanismos para la prevención de riesgos en la atención médica y la rendición de cuentas por las decisiones clínicas.

Palabras clave: Algoritmos. Gestión de Riesgos. Inteligencia Artificial. Seguridad del Paciente. Sistemas de Apoyo a la Toma de Decisiones Clínicas.

1 INTRODUÇÃO

A incorporação da inteligência artificial (IA) aos serviços de saúde tem ampliado as possibilidades de utilização de tecnologias digitais no diagnóstico, monitoramento clínico, planejamento terapêutico e organização da assistência. Por meio do processamento de grandes volumes de informações, esses sistemas auxiliam na identificação de padrões e no apoio à tomada de decisão, modificando a maneira como os dados são produzidos e utilizados pelos profissionais e introduzindo novas exigências para a gestão dos serviços e a qualidade do cuidado (Murakami; Costa, 2026).

A aplicação da IA contempla diferentes contextos assistenciais, incluindo interpretação de exames, identificação de riscos clínicos, classificação de pacientes e acompanhamento de condições de saúde. Sua utilização envolve mudanças nos processos de avaliação e decisão profissional, tornando necessária a organização de mecanismos de governança capazes de estabelecer responsabilidades, critérios de utilização e procedimentos de acompanhamento dos sistemas computacionais, especialmente quando suas recomendações podem interferir diretamente na definição de condutas clínicas (Pacheco *et al*., 2026).

A governança da IA na saúde compreende aspectos relacionados ao planejamento, desenvolvimento, validação, implantação e monitoramento dos sistemas, considerando a complexidade tecnológica e as particularidades dos ambientes assistenciais. Sua organização envolve a definição de responsabilidades entre gestores, profissionais de saúde e equipes técnicas, além da necessidade de contemplar requisitos éticos, institucionais e regulatórios durante o ciclo de vida das ferramentas, de modo a integrar o gerenciamento de riscos aos processos de utilização das tecnologias (Soriano-Vargas; Oliveira, 2026).

Entre os desafios associados à utilização de algoritmos na saúde, destaca-se a qualidade das informações empregadas no desenvolvimento dos modelos computacionais, considerando que os registros clínicos podem apresentar diferenças quanto à completude, padronização e representatividade populacional. A composição das bases de dados, a seleção das variáveis e os critérios de treinamento constituem elementos relevantes para compreender possíveis distorções nas recomendações automatizadas, sobretudo quando existem diferenças entre as populações utilizadas no desenvolvimento dos sistemas e aquelas atendidas nos serviços de saúde (Pinheiro *et al*., 2026).

Os vieses algorítmicos constituem uma preocupação relacionada à aplicação da IA em saúde, especialmente diante da possibilidade de diferenças no desempenho dos modelos conforme características sociodemográficas, clínicas e territoriais dos pacientes. A distribuição desigual de informações entre grupos populacionais e as limitações de representatividade dos dados introduzem questões sobre a equidade das decisões automatizadas, tornando necessário considerar os princípios de justiça, não discriminação e proteção da dignidade humana na organização da governança tecnológica (Amorim, 2026).

A ocorrência de erros algorítmicos representa outra dimensão relevante para a segurança do paciente, particularmente em atividades que envolvem classificação de gravidade, identificação de condições clínicas e definição de prioridades assistenciais. Na triagem, os sistemas inteligentes podem utilizar sinais vitais, sintomas e informações do histórico clínico para auxiliar a avaliação inicial, exigindo atenção às possibilidades de classificações inadequadas, falsos negativos e falsos positivos que podem interferir no direcionamento do atendimento e na continuidade da assistência (Pinheiro *et al*., 2026).

A dimensão institucional dessa problemática pode ser contextualizada por um levantamento realizado em 2023, com 2.425 hospitais dos Estados Unidos, no qual 65% das instituições utilizavam IA ou modelos preditivos, enquanto os percentuais informados para avaliação local da acurácia e dos vieses foram de 61% e 44%, respectivamente. Esses indicadores, referentes ao contexto hospitalar norte-americano, contribuem para contextualizar a importância de mecanismos institucionais de acompanhamento do desempenho e da confiabilidade das tecnologias utilizadas na assistência (Pinheiro *et al*., 2026).

O monitoramento da IA não se restringe à avaliação inicial de seu funcionamento, considerando que as características dos dados, os perfis populacionais e as condições assistenciais podem sofrer modificações ao longo do tempo. A governança dessas ferramentas envolve a definição de procedimentos para avaliação periódica, identificação de alterações no desempenho e acompanhamento dos riscos relacionados à utilização dos algoritmos, integrando aspectos técnicos, éticos e organizacionais aos processos institucionais de gerenciamento de riscos em saúde (Soriano-Vargas; Oliveira, 2026).

A transparência algorítmica também integra as questões relacionadas à utilização da IA em ambientes clínicos, uma vez que determinados sistemas apresentam processos decisórios de difícil interpretação. Nesse contexto, a IA explicável compreende métodos destinados a tornar as recomendações automatizadas mais compreensíveis para profissionais de saúde e pacientes, introduzindo a necessidade de considerar a explicabilidade, a rastreabilidade e a possibilidade de avaliação das decisões entre os componentes da governança tecnológica e da segurança assistencial (Cavalcante *et al*., 2026).

A utilização de algoritmos nos processos de decisão clínica envolve questões relacionadas à autonomia do paciente, ao consentimento informado e à delimitação das responsabilidades profissionais e institucionais. A incorporação de recomendações automatizadas à assistência exige atenção às condições em que essas informações são interpretadas e utilizadas, especialmente quando interferem na participação do indivíduo nas decisões sobre seu tratamento, tornando a supervisão humana e a governança ética dos dados elementos relevantes para a preservação dos direitos dos pacientes (Malta; Lamy, 2025).

A realização deste estudo justifica-se pela necessidade de compreender os desafios envolvidos na organização de mecanismos institucionais de acompanhamento da IA, considerando sua relação com os vieses, erros algorítmicos e a segurança do paciente. A problemática da pesquisa concentra-se na necessidade de compreender como a governança dessas tecnologias pode contemplar o monitoramento contínuo dos riscos, a identificação de disparidades entre grupos populacionais e a delimitação de responsabilidades sobre decisões apoiadas por algoritmos nos serviços de saúde (Soriano-Vargas; Oliveira, 2026).

Diante desse contexto, o presente estudo tem como objetivo analisar os desafios da governança da IA na saúde, com ênfase no monitoramento de vieses, na identificação de erros algorítmicos e na segurança do paciente. Especificamente, busca-se compreender os fatores relacionados à ocorrência de falhas nas decisões automatizadas, identificar os requisitos de transparência, validação e supervisão humana e examinar os mecanismos de governança e auditoria aplicáveis ao acompanhamento dos sistemas inteligentes, considerando sua relação com a qualidade assistencial e a equidade no cuidado.

2 METODOLOGIA

O presente estudo caracteriza-se como uma revisão narrativa da literatura, de abordagem qualitativa e caráter descritivo-analítico, destinada a examinar os desafios da governança da IA na saúde, com ênfase no monitoramento de vieses, na identificação de erros algorítmicos e na segurança do paciente. Esse delineamento permite reunir diferentes abordagens metodológicas, contemplando aspectos técnicos, éticos e institucionais relacionados à implementação e à supervisão de sistemas inteligentes na assistência à saúde.

Para a identificação das publicações científicas, foram estabelecidas como fontes de consulta a Biblioteca Virtual em Saúde (BVS), contemplando as bases Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS) e *Medical Literature Analysis and Retrieval System Online* (MEDLINE), além da *Scientific Electronic Library Online* (SciELO) e da PubMed. A estratégia de busca foi estruturada a partir de descritores controlados dos Descritores em Ciências da Saúde (DeCS) e do *Medical Subject Headings* (MeSH), considerando seus correspondentes nos idiomas português e inglês.

Foram selecionados os descritores Inteligência Artificial (*Artificial Intelligence*), Segurança do Paciente (*Patient Safety*), Algoritmos (*Algorithms*), Gestão de Riscos (*Risk Management*) e Sistemas de Apoio à Decisão Clínica (*Decision Support Systems*, Clinical). Os termos foram organizados mediante os operadores booleanos *AND* e *OR*, utilizando combinações como "*Artificial Intelligence*" *AND* "*Patient Safety*" e "*Artificial Intelligence*" *AND* "*Algorithms*" *AND* "*Risk Management*", com adequação das estratégias às características de cada fonte de consulta.

Os critérios de inclusão compreenderam artigos científicos completos, publicados em periódicos nacionais e internacionais, nos idiomas português e inglês, relacionados à governança da IA na saúde. Foram consideradas pesquisas qualitativas, revisões da literatura e análises teórico-conceituais que abordassem vieses algorítmicos, validação de sistemas inteligentes, monitoramento de desempenho e segurança assistencial. Foram excluídas publicações duplicadas, documentos sem acesso ao texto integral e trabalhos sem relação direta com o objetivo da pesquisa.

A seleção do material científico compreendeu a avaliação de oito documentos disponibilizados para a revisão, dos quais sete artigos foram incluídos e um foi excluído por não atender aos critérios de elegibilidade. As publicações selecionadas foram produzidas entre 2025 e 2026, abrangendo diferentes delineamentos metodológicos relacionados à governança institucional, à equidade algorítmica e à supervisão ética. A caracterização dos artigos foi organizada em uma tabela contendo autoria, ano de publicação, delineamento metodológico e principais contribuições relacionadas ao objeto investigado.

A análise dos dados foi realizada mediante leitura integral e interpretação crítica das publicações selecionadas, considerando as convergências e divergências entre as abordagens apresentadas. Foram examinados os fatores associados aos vieses e erros algorítmicos, os procedimentos de validação e auditoria, os mecanismos de monitoramento contínuo e a definição de responsabilidades institucionais. A interpretação do material bibliográfico considerou as relações entre governança tecnológica, segurança do paciente, equidade assistencial e particularidades da implementação da IA.

Por utilizar exclusivamente informações provenientes de publicações científicas, sem participação direta de seres humanos ou acesso a dados individuais identificáveis, a pesquisa não necessitou de submissão ao Comitê de Ética em Pesquisa. Entre as limitações metodológicas, destacam-se o caráter não sistemático da revisão narrativa, a heterogeneidade dos delineamentos incluídos e a ausência de avaliação padronizada da qualidade metodológica dos artigos, aspectos que restringem a comparação dos resultados e a generalização das conclusões. A rápida evolução das tecnologias também pode limitar a abrangência temporal das discussões apresentadas.

3 RESULTADOS E DISCUSSÕES

A análise dos oito documentos disponibilizados resultou na inclusão de sete artigos científicos, sendo um excluído por não atender aos critérios de elegibilidade estabelecidos para a revisão. As publicações selecionadas compreenderam uma pesquisa qualitativa, revisões da literatura e análises teórico-conceituais, contemplando aspectos relacionados à governança da IA na saúde, identificação de vieses, monitoramento de erros algorítmicos e segurança do paciente.

Tabela 1. Caracterização dos artigos incluídos na revisão sobre governança da inteligência artificial na saúde.

Autores/ano	Delineamento do estudo	Principais resultados e contribuições
Hasanzadeh *et al*. (2025)	Revisão crítica da literatura	Identificação de vieses relacionados à seleção de dados, representatividade populacional e desenvolvimento dos modelos, com estratégias de mitigação e monitoramento contínuo.
Lin *et al*. (2025)	Revisão da literatura	Descrição de procedimentos de governança institucional, avaliação multidisciplinar, validação local e acompanhamento do desempenho dos sistemas de apoio à decisão clínica.
Kumar *et al*. (2025)	Estudo teórico-propositivo	Proposição do modelo CAOS, fundamentado na classificação de riscos, proteção de dados, avaliação da equidade e supervisão de ferramentas utilizadas fora dos mecanismos formais de governança.
Owens *et al*. (2025)	Pesquisa qualitativa com 21 especialistas	Identificação de responsabilidades indefinidas na manutenção dos sistemas, insuficiência de estruturas institucionais de monitoramento e fragmentação das atividades de supervisão.
Abuchaim e Araujo (2026)	Análise crítica da literatura bioética e normativa	Proposição de diretrizes éticas e indicadores de equidade, transparência, privacidade, autonomia, responsabilização e sustentabilidade ética.
Palama *et al*. (2026)	Revisão estruturada e proposição de modelo normativo	Desenvolvimento de uma estrutura de auditoria que integra identificação de vieses, explicabilidade, monitoramento do desempenho, segurança e conformidade regulatória.
Soares e Chiavegatto Filho (2026)	Estudo teórico-analítico	Identificação de desafios relacionados à fragmentação dos sistemas de informação, desigualdades regionais, formação profissional e justiça algorítmica no Sistema Único de Saúde.

Fonte: Elaborado pelos autores, 2026

A identificação de vieses algorítmicos envolve limitações relacionadas à composição das bases de treinamento e à representação dos diferentes grupos populacionais. Hasanzadeh *et al*. (2025) descrevem distorções decorrentes da seleção dos participantes e da mensuração das variáveis, capazes de comprometer a confiabilidade das previsões clínicas. No Brasil, Soares e Chiavegatto Filho (2026) acrescentam que a fragmentação dos sistemas de informação e as desigualdades territoriais dificultam a obtenção de dados representativos da população atendida.

O monitoramento da equidade exige procedimentos capazes de identificar diferenças de desempenho ocultas nas avaliações globais de acurácia. Hasanzadeh *et al*. (2025) abordam métricas de justiça algorítmica, incluindo paridade demográfica e igualdade de oportunidades, considerando os resultados produzidos para diferentes grupos populacionais. Palama *et al*. (2026) complementam essa

perspectiva ao propor auditorias quantitativas e qualitativas, ressaltando que a escolha das métricas deve considerar a finalidade assistencial e os riscos clínicos.

A validação institucional constitui uma etapa distinta da aprovação inicial de uma ferramenta de IA, pois o desempenho obtido durante seu desenvolvimento pode não corresponder às condições dos serviços assistenciais. Lin *et al*. (2025) descrevem a necessidade de avaliação local, considerando as características da população e os processos de trabalho. Kumar *et al*. (2025), por sua vez, propõem classificar os sistemas conforme os riscos associados à sua utilização, vinculando a intensidade da supervisão ao impacto potencial sobre o cuidado.

Os erros algorítmicos apresentam repercussões particulares quando recomendações automatizadas interferem na identificação de condições graves e na definição das prioridades assistenciais. Kumar *et al*. (2025) recuperam avaliações anteriores de ferramentas preditivas de sepse, nas quais foram identificados falsos negativos e alertas inadequados. Essas limitações demonstram a necessidade de verificar sensibilidade, especificidade e capacidade de generalização dos modelos antes de incorporar suas recomendações às decisões clínicas.

A estabilidade do desempenho algorítmico depende do acompanhamento das modificações ocorridas após a implantação dos sistemas nos serviços de saúde. Palama *et al*. (2026) distinguem alterações na distribuição dos dados de entrada daquelas relacionadas às associações entre variáveis e resultados clínicos, propondo mecanismos de detecção de desvios e critérios para recalibração. O monitoramento deve contemplar indicadores de desempenho e limites previamente estabelecidos para identificar alterações que exijam ajustes ou reavaliação das ferramentas.

A delimitação das responsabilidades institucionais representa uma dificuldade para a continuidade da supervisão tecnológica. Owens *et al*. (2025) identificaram, nas entrevistas com 21 especialistas, uma lacuna de responsabilização caracterizada pela indefinição das atribuições relacionadas à manutenção e ao monitoramento dos sistemas. A insuficiência de investimentos nessas atividades contribui para que a identificação de falhas e a atualização dos modelos permaneçam fragmentadas entre diferentes profissionais e setores administrativos.

A organização da governança institucional requer procedimentos capazes de acompanhar os sistemas de apoio à decisão clínica desde sua aquisição até a eventual descontinuação de seu uso. Lin *et al*. (2025) descrevem uma estrutura multidisciplinar que contempla o registro das ferramentas, a avaliação de riscos e a definição de responsabilidades. A integração entre profissionais clínicos, especialistas em informática e instâncias administrativas permite estabelecer mecanismos de supervisão compatíveis com as particularidades dos serviços de saúde.

A preservação da autonomia do paciente e a proteção das informações clínicas constituem dimensões específicas da supervisão ética da IA. Abuchaim e Araujo (2026) propõem avaliar a concordância entre recomendações automatizadas e decisões médicas, o envolvimento dos pacientes

no processo decisório e os mecanismos de proteção dos dados. A documentação das decisões e a transparência sobre as limitações das ferramentas integram os critérios apresentados para o acompanhamento ético dos sistemas inteligentes.

A incorporação da IA ao Sistema Único de Saúde (SUS) exige considerar as diferenças existentes entre os serviços quanto à infraestrutura tecnológica e à disponibilidade de informações. Soares e Chiavegatto Filho (2026) relacionam a baixa interoperabilidade e a fragmentação dos registros clínicos às dificuldades de implementação dessas tecnologias. A insuficiência de formação profissional também compromete a capacidade institucional de interpretar recomendações automatizadas e identificar possíveis distorções nos resultados apresentados.

A explicabilidade dos sistemas inteligentes apresenta implicações para a compreensão das recomendações automatizadas e a identificação de decisões potencialmente inadequadas. Abuchaim e Araujo (2026) propõem indicadores referentes ao tempo necessário para produzir explicações compreensíveis, à completude das informações e à possibilidade de reconstruir a cadeia decisória. A transparência sobre as variáveis consideradas pelo algoritmo permite examinar a fundamentação das recomendações, sem eliminar a necessidade de avaliação profissional independente.

A continuidade da supervisão algorítmica requer que a manutenção dos sistemas seja incorporada às atividades regulares das instituições de saúde. Owens *et al*. (2025) identificaram que, embora os profissionais reconhecessem a possibilidade de deterioração do desempenho dos modelos, as atividades de identificação e correção dessas alterações permaneciam insuficientemente estruturadas. A definição de responsabilidades, a documentação de falhas e a investigação de resultados inesperados constituem medidas necessárias para acompanhar a segurança e a efetividade das ferramentas ao longo do tempo.

4 CONCLUSÃO

A governança da IA na saúde exige mecanismos institucionais que ultrapassem a validação inicial dos sistemas. A baixa representatividade das bases de treinamento, as alterações no desempenho dos algoritmos e a indefinição das responsabilidades institucionais dificultam a identificação de vieses e erros algorítmicos. A avaliação isolada da acurácia não permite identificar diferenças de desempenho entre grupos populacionais nem assegurar a confiabilidade das recomendações clínicas.

A implementação de auditorias periódicas, indicadores de equidade e critérios de recalibração constitui uma medida necessária à prevenção de riscos assistenciais. No SUS, a fragmentação dos registros clínicos, as desigualdades tecnológicas e as dificuldades de formação profissional comprometem esse processo. A contribuição da pesquisa consiste na identificação de requisitos técnicos, éticos e institucionais para o monitoramento dos sistemas inteligentes, considerando a segurança do paciente, a autonomia individual e a responsabilização pelas decisões clínicas.

As limitações da pesquisa compreendem o caráter narrativo da revisão, o número restrito de publicações incluídas, a heterogeneidade metodológica e a ausência de avaliação padronizada da qualidade dos artigos. A predominância de abordagens teóricas impossibilita determinar a efetividade dos mecanismos de governança na redução de eventos adversos. A rápida evolução das tecnologias também restringe a aplicabilidade temporal das conclusões apresentadas.

Recomenda-se a realização de pesquisas multicêntricas e longitudinais que avaliem a implementação de protocolos de auditoria algorítmica em diferentes serviços de saúde, especialmente no contexto brasileiro. Investigações futuras deverão examinar a frequência de erros clínicos, as diferenças de desempenho entre grupos populacionais e os efeitos da supervisão institucional sobre a segurança assistencial. O desenvolvimento de indicadores adaptados às condições operacionais do SUS poderá subsidiar práticas de monitoramento compatíveis com as necessidades dos serviços e da população.

REFERÊNCIAS

AMORIM, Ana Rosa Lima Loureiro de. Entre algoritmos e humanidade: desafios bioéticos na inteligência artificial e no Big Data em saúde. Revista JRG de Estudos Acadêmicos, v. 9, n. 20, 2026. DOI: https://doi.org/10.55892/jrg.v9i20.2912. Disponível em: https://www.revistajrg.com/index.php/jrg/article/view/2912.

ABUCHAIM, Ricardo Rheingantz; ARAUJO, Daniel Brito de. Supervisão ética em sistemas inteligentes de apoio à saúde. Revista Bioética, v. 34, 2026. DOI: https://doi.org/10.1590/1983-803420263995PT. Disponível em: https://scielo.br/j/bioet/a/sVCCn5LvSdDyVRFpjm5ZBGn/abstract/?ilang=pt_BR&lang=pt.

CAVALCANTE, Maryane Francisca Araujo de Freitas *et al.* Inteligência artificial explicável (XAI) em saúde digital: tensões entre transparência algorítmica, propriedade intelectual e ética do cuidado. Revista Ibero-Americana de Humanidades, Ciências e Educação, v. 12, n. 1, 2026. DOI: https://doi.org/10.51891/rease.v12i1.23937. Disponível em: https://periodicorease.pro.br/rease/article/view/23937.

HASANZADEH, Fereshteh *et al.* Bias recognition and mitigation strategies in artificial intelligence healthcare applications. npj Digital Medicine, v. 8, art. 154, 2025. DOI: https://doi.org/10.1038/s41746-025-01503-7. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC11897215/.

KUMAR, Rahul *et al.* Navigating healthcare AI governance: the comprehensive algorithmic oversight and stewardship framework for risk and equity. Health Care Analysis, 2025. DOI: https://doi.org/10.1007/s10728-025-00537-y. Disponível em: https://link.springer.com/article/10.1007/s10728-025-00537-y.

LIN, Anthony L. *et al.* Algorithm-based clinical decision support: evolving regulatory landscape and best practices for local oversight. Annual Review of Biomedical Data Science, v. 8, p. 491-507, 2025. DOI: https://doi.org/10.1146/annurev-biodatasci-103123-094601. Disponível em: https://www.annualreviews.org/content/journals/10.1146/annurev-biodatasci-103123-094601.

MALTA, Klauss Carvalho de; LAMY, Marcelo. Os desafios da autonomia do paciente frente ao uso da inteligência artificial na saúde. Cadernos Ibero-Americanos de Direito Sanitário, Brasília, v. 14, n. 4, 2025. DOI: https://doi.org/10.17566/ciads.v14i4.1348. Disponível em: https://www.cadernos.prodisa.fiocruz.br/index.php/cadernos/article/view/1348.

MURAKAMI, Rafael Guem; COSTA, Bruno Andrade. Inovações tecnológicas na área da saúde: inteligência artificial, saúde digital e desafios para os serviços de saúde. Revista Sociedade e Vida, v. 8, n. 4, 2026. DOI: https://doi.org/10.66104/xryz3127. Disponível em: https://rsv.ojsbr.com/rsv/pt_BR/article/view/6685.

OWENS, Kellie *et al.* Managing a “responsibility vacuum” in AI monitoring and governance in healthcare: a qualitative study. BMC Health Services Research, v. 25, art. 1217, 2025. DOI: https://doi.org/10.1186/s12913-025-13388-z. Disponível em: https://link.springer.com/article/10.1186/s12913-025-13388-z.

PACHECO, Milton Soares *et al.* Tecnologias, inovação, inteligência artificial e os impactos na tomada de decisão clínica em saúde: uma revisão de escopo. Veredas do Direito, v. 23, n. 14, 2026. DOI: https://doi.org/10.18623/rvd.v23.8081. Disponível em: https://revista.domhelder.edu.br/index.php/veredas/article/view/8081.

PINHEIRO, Maria Tarcila Rabelo *et al.* Gestão e auditoria de IA na saúde: modelos de governança para monitorar vieses e erros na triagem de pacientes. Revista DCS, v. 23, n. 96, 2026. DOI: https://doi.org/10.54899/dcs.v23i96.7009. Disponível em: https://ojs.revistadcs.com/index.php/revista/article/view/7009.

PALAMA, Valentina *et al.* Auditing and monitoring artificial intelligence systems in healthcare: a multilayer framework for bias detection, explainability, and regulatory compliance. Cureus, v. 18, n. 3, e104547, 2026. DOI: https://doi.org/10.7759/cureus.104547. Disponível em: https://pubmed.ncbi.nlm.nih.gov/41930073/.

SORIANO-VARGAS, Aurea; OLIVEIRA, Ana Beatriz de Figueiredo. Um framework de governança do ciclo de vida para o gerenciamento de riscos em projetos multidisciplinares de inteligência artificial em saúde: evidências de um estudo de caso brasileiro. Revista de Gestão e Projetos, v. 17, n. 2, 2026. DOI: https://doi.org/10.5585/2026.31451. Disponível em: https://periodicos.uninove.br/gep/article/view/31451.

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, 2026. DOI: https://doi.org/10.1590/1980-549720260006.2. Disponível em: https://scielo.br/j/rbepid/a/wd44hqsJjHM89cyFgHdhQdr/?lang=pt.