Revista de Políticas Públicas da UFPE, Recife/PE, v. 11, n. 2, ano 2026. ISSN-e 2595-5535

GOVERNANÇA DA INTELIGÊNCIA ARTIFICIAL EM SAÚDE
ACCOUNTABILITY, TRANSPARÊNCIA E SEGURANÇA ASSISTENCIAL

Maryane Francisca Araujo de Freitas Cavalcante[1]
João Gabriel Freitas Cavalcante[2]
Miranisia Aparecida de Araujo Freitas Lopes[3]
Gerardo Pereira de Sousa[4]

Resumo
A incorporação da Inteligência Artificial (IA) em contextos clínicos tem ampliado debates relacionados à transparência algorítmica, *accountability,* segurança assistencial e governança regulatória. Nesse cenário, o estudo objetiva analisar os fundamentos da governança da IA em saúde, com ênfase nos eixos de responsabilidade institucional, transparência e mitigação de riscos clínicos. Metodologicamente, trata-se de investigação qualitativa, teórico-analítica, fundamentada em revisão narrativa crítica e análise conceitual interdisciplinar, envolvendo literatura científica, *frameworks* institucionais e documentos regulatórios internacionais publicados entre 2021 e 2026. Os resultados evidenciaram convergência na literatura quanto à necessidade de modelos sociotécnicos de governança baseados em supervisão contínua, gestão dinâmica de riscos, explicabilidade contextual e *accountability* distribuída. Além disso, identificaram-se limitações relevantes nos *frameworks* regulatórios atuais, especialmente diante da opacidade algorítmica, da deriva de desempenho e da insuficiência dos mecanismos tradicionais de supervisão pós-implantação. Conclui-se que a governança da IA em saúde constitui dimensão estruturante da segurança assistencial e da legitimidade institucional dos sistemas algorítmicos, exigindo arquiteturas regulatórias adaptativas capazes de integrar transparência, monitoramento contínuo e proteção dos direitos fundamentais no contexto da saúde digital contemporânea.

Palavras-chave: Inteligência Artificial em Saúde; Governança Algorítmica; *Accountability*; Transparência Algorítmica; Segurança do Paciente

[1] Mestranda em Propriedade Intelectual. Instituto Federal do Piauí - IFPI. E-mail: moren.afc@hotmail.com
[2] Discente Bacharelado em Ciências da Computação. Universidade Federal do Piauí - UFPI. E-mail: joaocavalcantejcavalcante@gmail.com
[3] Discente Bacharelado em Terapia Ocupacional. UNINASSAU Teresina-PI. E-mail: mira_nisia@hotmail.com
[4] Doutor em em Ciência da Propriedade Intelectual. Professor do Instituto Federal do Piauí - IFPI. E-mail: profgerardo.adm@ifpi.edu.br

GOVERNANCE OF ARTIFICIAL INTELLIGENCE IN HEALTHCARE:
ACCOUNTABILITY, TRANSPARENCY AND PATIENT SAFETY

Abstract
The incorporation of Artificial Intelligence (AI) into clinical contexts has intensified debates concerning algorithmic transparency, accountability, patient safety, and regulatory governance. In this context, the present study aims to analyze the foundations of AI governance in healthcare, with emphasis on institutional responsibility, transparency, and the mitigation of clinical risks. Methodologically, this is a qualitative and theoretical-analytical investigation grounded in critical narrative review and interdisciplinary conceptual analysis, involving scientific literature, institutional frameworks, and international regulatory documents published between 2020 and 2026. The findings revealed convergence in the literature regarding the need for sociotechnical governance models based on continuous oversight, dynamic risk management, contextual explainability, and distributed accountability. Furthermore, significant limitations were identified in current regulatory frameworks, particularly concerning algorithmic opacity, performance drift, and the insufficiency of traditional post-deployment oversight mechanisms. It is concluded that AI governance in healthcare constitutes a structuring dimension of patient safety and the institutional legitimacy of algorithmic systems, requiring adaptive regulatory architectures capable of integrating transparency, continuous monitoring, and the protection of fundamental rights within the contemporary digital health context.

Keywords: Artificial Intelligence in Healthcare; Algorithmic Governance; Accountability; Algorithmic Transparency; Patient Safety.

INTRODUÇÃO

A incorporação de sistemas de Inteligência Artificial (IA) no setor da saúde configura uma das transformações tecnológicas mais relevantes das últimas décadas. Algoritmos de aprendizado de máquina e modelos preditivos têm sido progressivamente integrados a processos clínicos, diagnósticos, administrativos e decisórios, frequentemente associados a promessas de maior eficiência, precisão e personalização do cuidado. Contudo, paralelamente às oportunidades tecnológicas, emergem preocupações substanciais relacionadas à segurança assistencial, à opacidade algorítmica, à responsabilidade decisória e aos desafios institucionais envolvidos na governança dessas tecnologias (Burrell, 2016).

Essas preocupações têm impulsionado o desenvolvimento do campo da governança algorítmica, que transcende abordagens regulatórias estritamente formais. De fato, a governança da IA envolve um conjunto de mecanismos institucionais, organizacionais e técnicos voltados à orientação do comportamento de sistemas automatizados, incorporando dimensões como supervisão humana significativa e gestão de riscos ao longo do ciclo de vida dos sistemas. Com isso, a interdependência entre desempenho algorítmico e estruturas institucionais reforça a compreensão de que a tecnologia não opera de maneira neutra, exigindo modelos de governança adaptativos e sensíveis aos contextos de aplicação (Wieringa, 2020).

Assim, existe um entendimento contemporâneo de que a IA não pode ser compreendida exclusivamente como artefato técnico, mas como tecnologia sociotécnica cujos efeitos emergem da interação dinâmica entre algoritmos, dados, instituições, profissionais e contextos normativos. Essa perspectiva amplia substancialmente o escopo analítico da governança, deslocando-a de uma abordagem centrada apenas na conformidade regulatória para uma arquitetura institucional mais abrangente. Nesse contexto, a IA deve ser compreendida como fenômeno capaz de produzir efeitos complexos e frequentemente imprevisíveis em ambientes de saúde.

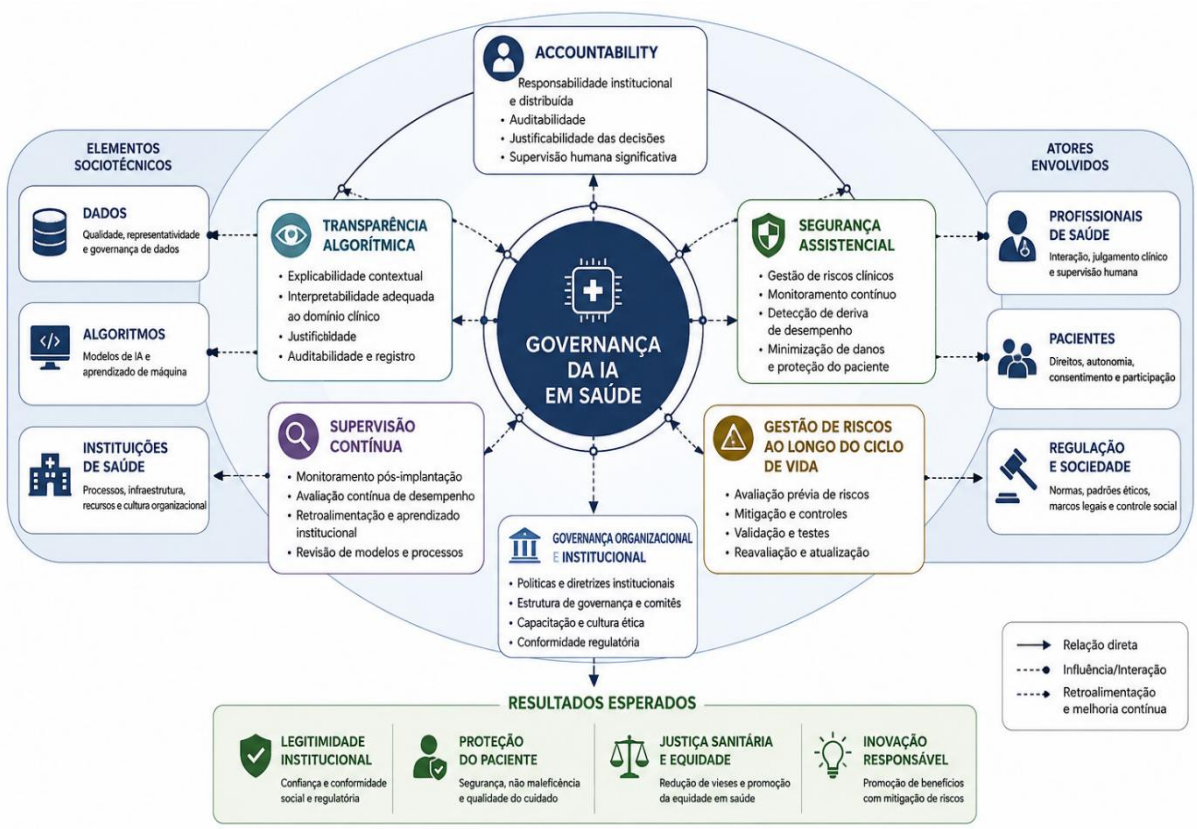
A literatura sobre governança algorítmica destaca a accountability como dimensão central da supervisão institucional da IA, envolvendo responsabilidade, auditabilidade e justificabilidade decisória (Wieringa, 2020). Entretanto, tais mecanismos enfrentam obstáculos relacionados à opacidade algorítmica e à complexidade dos modelos de aprendizado de máquina.

Nessa perspectiva, a opacidade algorítmica constitui um dos principais desafios contemporâneos à governança da IA em saúde. Esses modelos complexos frequentemente operam por meio de processos não intuitivos, dificultando a explicação causal das decisões produzidas e tensionando regimes tradicionais de responsabilidade. Conforme argumentado por Lipton, a interpretabilidade não constitui atributo monolítico, mas conceito multifacetado, frequentemente mobilizado de forma ambígua na literatura, o que evidencia a necessidade de abordagens analíticas mais refinadas sobre transparência e explicabilidade em sistemas algorítmicos (Lipton, 2017).

Além disso, é importante destacar que transparência deve ser concebida de maneira contextual, envolvendo níveis adequados de inteligibilidade, justificabilidade e auditabilidade conforme o domínio de aplicação. Em contextos clínicos, a transparência adquire relevância ampliada, uma vez que decisões assistenciais demandam possibilidade de contestação e validação por parte de profissionais e instituições.

Com isso, a Figura 1 apresenta uma síntese conceitual da arquitetura sociotécnica da governança da Inteligência Artificial em saúde, evidenciando a interdependência entre accountability, transparência algorítmica, segurança assistencial e gestão contínua de riscos. O modelo proposto busca representar a governança da IA como processo dinâmico e institucionalmente distribuído, articulando dimensões técnicas, organizacionais, regulatórias e éticas no contexto da saúde digital contemporânea.

Figura 1 – Arquitetura Sociotécnica da Governança da IA em Saúde



A representação evidencia que a confiabilidade dos sistemas algorítmicos não decorre exclusivamente do desempenho computacional, mas da integração entre supervisão humana, governança organizacional, proteção de dados, monitoramento contínuo e mecanismos regulatórios adaptativos. Dessa forma, a figura reforça a interpretação defendida neste estudo de que a governança da IA em saúde constitui condição estruturante para a promoção de inovação responsável, legitimidade institucional e proteção do paciente em ambientes clínicos mediados por tecnologias algorítmicas.

Nesse cenário, organismos internacionais e agências regulatórias reconhecem que os riscos associados à Inteligência Artificial em saúde ultrapassam dimensões estritamente técnicas, exigindo estruturas de governança capazes de assegurar transparência, segurança assistencial e proteção do paciente. Com isso, a literatura destaca que a principal tensão normativa reside no equilíbrio entre desempenho algorítmico, explicabilidade, accountability e mitigação de riscos clínicos. Diante dessas considerações, o presente estudo objetiva analisar os fundamentos da governança da IA em saúde, com ênfase nos eixos de responsabilidade, transparência e segurança assistencial.

1. METODOLOGIA

O presente estudo caracteriza-se como investigação teórico-analítica, de natureza qualitativa, orientada por revisão narrativa crítica e análise conceitual interdisciplinar. Esse delineamento revela-se adequado à análise de objetos marcados por elevada complexidade sociotécnica e normativa, como a governança da IA em saúde. Essa compreensão exige a integração de perspectivas provenientes de diferentes campos do conhecimento, incluindo saúde, regulação, ética, direito e tecnologia.

Nessa perspectiva, a opção por uma revisão narrativa justifica-se pela própria natureza do problema investigado, que demanda aprofundamento teórico, problematização conceitual e exame crítico de estruturas normativas emergentes. Com isso, a escolha deu-se em detrimento de abordagens centradas exclusivamente na quantificação de evidências.

Diferentemente de revisões sistemáticas, a revisão narrativa adotada neste estudo não se restringe a protocolos rígidos de exaustividade documental ou mensuração estatística, mas privilegia a densidade teórica, a relevância conceitual e a capacidade interpretativa das fontes examinadas. Esse tipo de abordagem é muito reconhecido na literatura metodológica como apropriado para investigações de caráter conceitual e analítico, especialmente em campos em consolidação, nos quais os debates ainda se encontram em processo de estruturação teórica e normativa. Nessa perspectiva, o objetivo não consiste em esgotar a produção existente, mas em mobilizar contribuições com elevada centralidade analítica e capacidade de sustentar uma reflexão crítica interdisciplinar.

O corpus analítico foi composto por literatura científica interdisciplinar relevante aos campos de governança algorítmica, ética em Inteligência Artificial, transparência, accountability, segurança do paciente e regulação tecnológica. Também foram considerados marcos regulatórios e documentos institucionais internacionais, dada sua centralidade na estruturação contemporânea dos regimes de governança da IA. Adicionalmente, mobilizaram-se referenciais teóricos consolidados para sustentar a análise conceitual, especialmente no que se refere às categorias de accountability, transparência algorítmica, opacidade, explicabilidade e gestão de riscos tecnológicos.

A seleção das fontes foi orientada por critérios de adequação temática, relevância teórica e impacto acadêmico, priorizando contribuições amplamente reconhecidas no debate internacional sobre IA e saúde. A estratégia de busca inicial contemplou a combinação de descritores em inglês e português, refletindo a natureza global e interdisciplinar da produção científica sobre o tema, incluindo termos como AI *governance, algorithmic accountability,*

transparency, explainability, patient safety e healthcare, com delimitação temporal compreendida entre 2021 e 2026. A priorização incidiu sobre artigos conceituais e revisões teóricas, bem como sobre diretrizes e frameworks institucionais, selecionados em função de sua elevada centralidade para os debates sobre governança, responsabilidade, transparência e risco tecnológico. Em consonância com o caráter narrativo da revisão, o foco metodológico concentrou-se na relevância analítica e na densidade teórica das fontes, e não na exaustividade documental.

A análise dos materiais foi conduzida por meio de uma abordagem interpretativa orientada à identificação de convergências conceituais entre distintas correntes teóricas e à problematização das tensões normativas decorrentes da incorporação de sistemas de Inteligência Artificial em contextos clínicos e assistenciais.

O exame analítico também buscou evidenciar lacunas teóricas e desafios epistemológicos presentes nos modelos contemporâneos de regulação e governança tecnológica. Portanto, esse procedimento permitiu compreender a IA não como artefato puramente técnico, mas como tecnologia sociotécnica de risco, cujos efeitos se vinculam a arranjos institucionais, práticas organizacionais e estruturas regulatórias.

Além disso, ressalta-se que o estudo não envolve coleta de dados empíricos, sendo caracterizado como investigação conceitual e crítica. Essa opção metodológica fundamenta-se na natureza do problema de pesquisa, que demanda refinamento teórico, clarificação analítica e integração interdisciplinar antes da formulação de proposições normativas ou avaliações empíricas. Logo, estudos teóricos dessa natureza desempenham papel relevante na consolidação de campos emergentes, especialmente em áreas marcadas por rápida transformação tecnológica, complexidade regulatória e incerteza institucional.

2. RESULTADOS E DISCUSSÃO

2.1 Governança Algorítmica, *Accountability* e Regulação da Inteligência Artificial em Saúde

A análise interpretativa do corpus evidenciou convergência sólida na literatura contemporânea ao situar a governança da Inteligência Artificial em saúde como dimensão estruturante, e não acessória, dos sistemas tecnológicos emergentes. Os trabalhos analisados rejeitam abordagens que tratam a governança como mera etapa regulatória ou burocrática, enfatizando seu papel constitutivo na segurança assistencial, confiabilidade sistêmica e legitimidade institucional das tecnologias algorítmicas. Essa inflexão teórica é consistente em diferentes tradições disciplinares, incluindo ética em IA, regulação tecnológica e segurança do paciente.

O principal modelo de governança defendido pelos estudos analisados fundamenta-se em uma perspectiva sociotécnica, multidimensional e centrada na confiança institucional, na qual transparência, *accountability*, segurança e equidade constituem elementos estruturantes do ciclo de vida da inteligência artificial em saúde. Hassan, Borycki e Kushniruk (2025) defendem um modelo de governança orientado pela confiança ("*trust-centred governance*"), capaz de integrar supervisão ética, mitigação de vieses, explicabilidade e segurança clínica como mecanismos permanentes de implementação e monitoramento da IA.

Os artigos convergem ao reconhecer que os *frameworks* atuais apresentam limitações críticas relacionadas à opacidade algorítmica, à fragmentação regulatória e à insuficiência dos mecanismos de supervisão pós-implantação. Kiseleva, Kotzinos e De Hert (2022) argumentam que a transparência não pode ser reduzida à mera explicabilidade técnica, devendo ser compreendida como um sistema multicamadas de *accountability* distribuída entre desenvolvedores, instituições de saúde e usuários clínicos.

Os achados presentes na literatura analisada indicam que validações técnicas prévias, embora indispensáveis, não asseguram a confiabilidade de longo prazo dos sistemas de IA, uma vez que o desempenho dos modelos tende a sofrer degradação em ambientes clínicos reais e dinâmicos. Essa lacuna, frequentemente descrita como *AI chasm*, decorre, em grande medida, do treinamento dos modelos em bases de dados não representativas, o que compromete sua validade externa e capacidade de generalização, conforme amplamente discutido na literatura sobre risco algorítmico e IA clínica (NIST, 2023).

De modo semelhante, o *framework* CAOS evidencia que os modelos regulatórios tradicionais permanecem excessivamente estáticos diante de sistemas adaptativos e autoatualizáveis, incapazes de monitorar adequadamente fenômenos como drift algorítmico, discriminação automatizada e degradação de desempenho clínico em tempo real. Além disso, essa formulação mostra-se coerente com marcos contemporâneos de governança e gestão de risco em IA, que enfatizam monitoramento contínuo, reavaliação contextual e supervisão institucional (Kiseleva; Kotzinos; De Hert, 2022; Kumar *et al.,* 2025).

O CAOS Framework oferece uma interpretação estruturada dessas vulnerabilidades ao enfatizar os riscos associados à chamada *shadow AI*, caracterizada pela adoção de sistemas não autorizados ou insuficientemente regulados em ecossistemas organizacionais complexos (Kumar *et al.,* 2025). Os achados descritos no *framework* demonstram que tais sistemas podem escapar a modelos regulatórios estáticos, expondo pacientes a riscos como deriva algorítmica, vieses não detectados e degradação silenciosa de desempenho, o que reforça a insuficiência de abordagens baseadas exclusivamente em validações prévias.

A análise do corpus evidenciou que a segurança assistencial, no contexto da Inteligência Artificial em saúde, não pode ser compreendida como atributo intrínseco do artefato tecnológico. Estudos recentes indicam que o desempenho algorítmico é fortemente condicionado por ecossistemas sociotécnicos, abrangendo fatores como qualidade dos dados, práticas institucionais e padrões de uso clínico (Bailo *et al.,* 2026). Essa perspectiva promove um deslocamento analítico relevante, ao redirecionar a atenção do algoritmo isolado para as condições estruturais que sustentam sua operação.

Sob esse enquadramento, o desempenho de ferramentas de IA em ambientes clínicos revela-se dependente da infraestrutura organizacional que lhes dá suporte, incluindo fluxos de trabalho, comportamento dos usuários e capacidade institucional de monitoramento contínuo. Esse entendimento reforça que a confiabilidade dos sistemas algorítmicos emerge de interações sistêmicas complexas, e não exclusivamente de propriedades técnicas dos modelos computacionais. Consequentemente, a governança organizacional e os mecanismos institucionais de supervisão assumem papel central na promoção da segurança assistencial e na mitigação de riscos tecnológicos.

Outro padrão conceitual recorrente identificado na análise refere-se à opacidade algorítmica como fator central de instabilidade normativa nos sistemas de Inteligência Artificial em saúde. A literatura converge ao indicar que a dificuldade de reconstruir trajetórias decisórias em modelos computacionais complexos compromete regimes tradicionais de responsabilidade, auditoria e justificabilidade decisória. Nessa linha, a interpretabilidade constitui conceito multifacetado e frequentemente mal compreendido, sendo inadequado assumir que determinados modelos são inerentemente transparentes ou explicáveis (Kiseleva; Kotzinos; De Hert, 2022; Nouis; Uren; Jariwala, 2025).

Ainda segundo Nouis, Uren e Jariwala (2025), a problemática da opacidade algorítmica é frequentemente associada ao chamado desafio da “caixa preta”, característico de arquiteturas de elevada complexidade, como redes neurais profundas. Nesses sistemas, a impossibilidade prática de interpretar os caminhos internos que conduzem às recomendações algorítmicas produz um abismo epistêmico e operacional, particularmente sensível em contextos clínicos de alta criticidade.

Além disso, os debates recentes ampliam a análise da opacidade algorítmica ao situá-la em transformações mais amplas das estruturas de poder informacional. A transição de regimes centrados na biopolítica tradicional para formas de governança mediadas por dados (biodata), evidencia que os riscos tecnológicos não se restringem à esfera técnica, mas envolvem novas configurações de vulnerabilidade, controle e assimetria informacional (Bailo *et al.,* 2026). Sob essa perspectiva, a opacidade algorítmica adquire relevância não apenas regulatória, mas também social, reconfigurando os debates sobre responsabilidade e segurança em saúde digital.

Nesse cenário, a governança algorítmica emerge como mecanismo compensatório fundamental. O CAOS *Framework* enfatiza a necessidade de documentação contínua, auditorias regulares e sistemas de supervisão humana significativa como estratégias de reconstrução da *accountability*. Esse entendimento é compatível com abordagens de governança distribuída discutidas na literatura contemporânea (Kumar *et al*., 2025).

A *accountability*, portanto, é reinterpretada como propriedade relacional e institucionalmente distribuída. Estudos recentes indicam que responsabilidade em sistemas de IA não pode ser atribuída exclusivamente ao desenvolvedor ou ao modelo computacional, mas envolve arranjos organizacionais complexos. Essa concepção amplia significativamente os modelos tradicionais de responsabilidade tecnológica (Bailo *et al.,* 2026).

No domínio da transparência, a análise enfatiza que a transparência deve ser compreendida como inteligibilidade funcional e contextual, exigindo níveis diferenciados de explicação conforme o domínio de aplicação e o perfil de risco envolvido (WHO, 2021; Kiseleva; Kotzinos; De Hert, 2022). Nessa perspectiva, marcos regulatórios e referenciais recentes confirmam que a transparência em sistemas de IA constitui um arranjo multicamadas de prestação de contas, incorporando requisitos como auditabilidade, rastreabilidade e governança de dados (NIST, 2023).

Para Papagiannidis, Mikalef e Conboy (2025) esse entendimento conecta-se diretamente à necessidade de transformar a opacidade algorítmica em risco gerenciável, por meio de documentação rigorosa, monitoramento contínuo e mecanismos institucionais de supervisão. Entretanto, a literatura identifica de forma recorrente a tensão entre explicabilidade e desempenho, observando que modelos de elevada complexidade frequentemente oferecem ganhos preditivos à custa de reduzida interpretabilidade . Esse conflito é interpretado não como problema exclusivamente técnico, mas como dilema normativo e regulatório, cuja resolução depende de critérios contextuais, avaliação de risco e estruturas de governança adequadas (WHO, 2021; NIST, 2023).

No entanto, marcos regulatórios e referenciais institucionais contemporâneos reforçam que sistemas de IA em saúde devem incorporar mecanismos de explicabilidade proporcionais ao grau de risco, à criticidade decisória e ao impacto assistencial envolvidos. Documentos da WHO (2021) e do NIST (2023) indicam que a explicabilidade não constitui requisito uniforme, mas dimensão contextual e graduada, dependente das características específicas de aplicação dos sistemas algorítmicos.

Outro eixo de convergência identificado na análise refere-se à adoção de modelos regulatórios baseados em risco como princípio estruturante da governança da Inteligência Artificial em saúde. O CAOS Framework propõe mecanismos de estratificação algorítmica compatíveis com abordagens contemporâneas de regulação adaptativa, enfatizando que diferentes aplicações tecnológicas demandam níveis proporcionais de controle e supervisão (Kumar *et al*., 2025). Essa lógica encontra respaldo consistente em marcos regulatórios internacionais e em frameworks institucionais voltados à gestão de riscos em sistemas de IA (WHO, 2021; NIST, 2023).

A regulação baseada em risco representa deslocamento conceitual relevante ao substituir concepções homogêneas de tecnologia pelo reconhecimento da heterogeneidade dos

impactos clínicos e potenciais de dano associados aos sistemas algorítmicos. Tal abordagem é central em marcos regulatórios recentes, notadamente no AI Act da União Europeia, que estabelece gradações normativas proporcionais aos níveis de risco, variando de categorias consideradas inaceitáveis a aplicações de risco mínimo (European Union, 2024). Essa mudança paradigmática evidencia que a governança da IA em saúde demanda estruturas normativas dinâmicas e contextualmente sensíveis, capazes de alinhar inovação tecnológica, segurança assistencial e proteção do paciente.

Evidências recentes ilustram de maneira concreta tais vulnerabilidades, como observado em sistemas clínicos cujo desempenho foi significativamente afetado por alterações abruptas nos padrões assistenciais durante a pandemia de COVID-19. Nessas circunstâncias, a degradação de desempenho não decorre necessariamente de falhas computacionais, mas de mudanças estruturais na distribuição dos dados e nos contextos de decisão clínica (NIST, 2023). Assim, a confiabilidade dos modelos algorítmicos revela-se dependente de mecanismos institucionais capazes de sustentar revalidações contextuais e governança dinâmica de risco.

No domínio da privacidade, os achados indicam que a incorporação de IA amplia e complexifica os riscos tradicionais de vulnerabilidade informacional, particularmente em ambientes que lidam com dados sensíveis. Sistemas algorítmicos podem expandir superfícies de exposição de Informações de Saúde Protegidas (PHI), tanto por falhas técnicas quanto por práticas institucionais inadequadas, especialmente em cenários associados à chamada *shadow AI* (WHO, 2021; European Union, 2024). Documentos regulatórios internacionais enfatizam, portanto, a necessidade de estruturas robustas de governança de dados, capazes de assegurar proteção efetiva contra vazamentos, reidentificação e usos indevidos.

Estratégias tecnológicas como *federated learning* e *differential privacy* emergem na literatura como abordagens promissoras para mitigação de riscos de privacidade, ao permitirem o treinamento de modelos sem a transferência direta de dados brutos entre instituições (Kumar *et al*., 2025). Contudo, os estudos ressaltam que tais técnicas não eliminam integralmente os riscos, podendo introduzir compromissos entre privacidade e utilidade do modelo, especialmente em contextos clínicos que demandam elevada precisão preditiva (OECD, 2023). Consequentemente, a mitigação eficaz dos riscos associados à IA em saúde exige não apenas soluções técnicas, mas estruturas institucionais robustas de governança, validação contínua e responsabilidade organizacional (Bailo *et al*., 2026, NIST, 2023).

2.2 Segurança Assistencial, Governança Organizacional e Perspectivas Sociotécnicas da IA em Saúde

A análise do corpus evidenciou que a governança organizacional constitui elemento decisivo para a incorporação segura e responsável de sistemas de Inteligência Artificial em ambientes clínicos. O estudo empírico de Kim *et al.* (2025), conduzido em um grande sistema hospitalar canadense, demonstra que a governança efetiva depende da existência de estruturas institucionais formais, processos organizacionais claramente definidos e alinhamento entre múltiplos atores institucionais. A ausência desses mecanismos favorece práticas *ad hoc* e inconsistentes, comprometendo diretamente a segurança assistencial e os regimes de *accountability* institucional.

Sob essa perspectiva, a governança organizacional não deve ser interpretada como obstáculo à inovação tecnológica, mas como infraestrutura necessária à confiabilidade, transparência e equidade dos sistemas algorítmicos. Os achados indicam que modelos estruturados, como o PPTO (People, Process, Technology and Oversight), oferecem instrumentos práticos para traduzir princípios éticos e normativos abstratos em políticas

institucionais formalmente adotadas. Tal abordagem reforça que a sustentabilidade e a segurança da IA em saúde dependem de arranjos institucionais, capazes de articular governança, responsabilidade e mitigação de riscos tecnológicos (Kim *et al*., 2025).

A análise das fontes confirma que a Inteligência Artificial em saúde deve ser compreendida como tecnologia inerentemente sociotécnica, cujos riscos e benefícios não decorrem exclusivamente de propriedades computacionais, mas emergem da interação entre componentes técnicos e dinâmicas humanas, organizacionais e institucionais (Bailo *et al*., 2026; WHO, 2021). Nesse enquadramento, falhas sistêmicas frequentemente resultam de interações complexas envolvendo qualidade dos dados, arquitetura dos modelos, design das interfaces e padrões de uso pelos profissionais. Essa perspectiva amplia substancialmente o escopo da análise de risco tecnológico, exigindo que a avaliação considere o sistema completo, incluindo fluxos de trabalho, caminhos de escalonamento e capacidade organizacional de resposta a degradações de desempenho (NIST, 2023).

Os achados reforçam que a governança da IA deve ser compreendida como processo dinâmico e contínuo ao longo de todo o ciclo de vida tecnológico, e não como mecanismo exclusivamente normativo ou estático (OECD, 2023; NIST, 2023). Modelos contemporâneos de regulação adaptativa, como o mecanismo DRAM do CAOS *Framework*, enfatizam monitoramento contínuo, ajustes contextuais e detecção de deriva algorítmica como estratégias centrais de mitigação de riscos emergentes (Kumar *et al.,* 2025).

Sob essa perspectiva, a legitimidade da IA em saúde revela-se menos dependente de métricas isoladas de desempenho e mais condicionada à existência de arquiteturas institucionais robustas que sustentem sua operação segura e auditável (WHO, 2021; OECD, 2023). Transparência, *accountability* e segurança assistencial emergem como dimensões indissociáveis, uma vez que a opacidade algorítmica pode comprometer simultaneamente a atribuição de responsabilidade e a proteção do paciente (Papagiannidis; Mikalef; Conboy, 2025). Dessa forma, a governança efetiva requer não apenas soluções técnicas, mas arranjos institucionais capazes de integrar controle, monitoramento e responsabilidade organizacional.

A principal contribuição do presente estudo para o contexto brasileiro reside na articulação entre governança algorítmica, proteção de dados em saúde e fortalecimento institucional do Sistema Único de Saúde diante da expansão das tecnologias de IA. Diferentemente de abordagens predominantemente tecnicistas, o trabalho propõe interpretar a governança da IA em saúde como problema simultaneamente ético, jurídico, organizacional e político, considerando as desigualdades estruturais que caracterizam os sistemas públicos latino-americanos. Essa perspectiva aproxima-se da crítica desenvolvida por Capeller, Pedroso e Santos (2025), ao demonstrarem que a sociedade algorítmica contemporânea é atravessada por tensões entre direitos de cidadania, interesses corporativos e mercantilização dos biodados em saúde global (Capeller; Pedroso; Santos, 2025).

Além disso, os estudos analisados indicam que a governança eficaz da IA depende da institucionalização de mecanismos permanentes de monitoramento organizacional, gestão de riscos e supervisão interdisciplinar. O modelo PPTO, desenvolvido em contexto hospitalar canadense, demonstra que a governança algorítmica exige integração entre pessoas, processos, infraestrutura tecnológica e operações institucionais, superando abordagens centradas exclusivamente na validação técnica dos algoritmos. Nessa direção, o AI Risk Management Framework (AI RMF 1.0) do NIST reforça que sistemas confiáveis de IA devem incorporar simultaneamente segurança, transparência, explicabilidade, justiça e gerenciamento contínuo de riscos ao longo de todo o ciclo de vida tecnológico (Kim *et al.,* 2025; NIST, 2023).

A lacuna teórica que o trabalho pretende preencher consiste justamente na ausência de abordagens interdisciplinares capazes de integrar governança algorítmica, accountability institucional e justiça sanitária em uma perspectiva aplicada ao campo da saúde pública brasileira. Embora a literatura internacional avance na formulação de princípios éticos e

modelos regulatórios, ainda predominam análises fragmentadas entre dimensões técnicas, jurídicas e bioéticas, dificultando a compreensão sistêmica dos impactos sociopolíticos da IA na saúde. (Bailo *et al.,* 2026).

Assim, o estudo busca contribuir para a consolidação de uma abordagem crítico-regulatória da governança da inteligência artificial em saúde, compreendendo a transparência não apenas como requisito técnico de explicabilidade, mas como condição democrática de controle social, legitimidade institucional e proteção dos direitos fundamentais no ambiente algorítmico contemporâneo. Os resultados da investigação indicam que a governança da IA em saúde constitui elemento estruturante para o desenvolvimento de processos de inovação tecnológica ética, responsável e socialmente sustentável. Nesse sentido, a literatura contemporânea converge ao reconhecer que os desafios relacionados aos sistemas algorítmicos ultrapassam dimensões estritamente técnicas, exigindo mecanismos institucionais permanentes de supervisão, transparência e gestão de riscos.

Além disso, os estudos analisados evidenciam que a expansão da inteligência artificial em contextos clínicos demanda modelos regulatórios capazes de acompanhar a natureza dinâmica e adaptativa dessas tecnologias. A opacidade algorítmica, associada à elevada complexidade decisória dos sistemas de IA, amplia os desafios relacionados à *accountability*, à segurança clínica e à proteção de direitos fundamentais no campo da saúde digital. Portanto, o desafio central consiste na construção de arquiteturas institucionais aptas a governar sistemas automatizados de alto impacto clínico, assegurando equilíbrio entre inovação tecnológica, responsabilidade ética e justiça sanitária.

CONCLUSÃO

A análise interpretativa do corpus evidenciou convergência robusta na literatura contemporânea ao situar a governança da Inteligência Artificial em saúde como dimensão estruturante da segurança assistencial, da confiabilidade sistêmica e da legitimidade institucional dos sistemas algorítmicos. Os estudos examinados rejeitam abordagens restritas à conformidade regulatória formal, enfatizando modelos sociotécnicos de governança baseados em transparência, *accountability,* supervisão contínua e gestão dinâmica de riscos. Nesse contexto, *frameworks* contemporâneos, como CAOS e PPTO, reforçam a necessidade de arquiteturas institucionais adaptativas capazes de acompanhar sistemas algorítmicos autoatualizáveis e de elevado impacto clínico.

Os resultados também demonstraram que validações técnicas prévias, embora indispensáveis, não asseguram confiabilidade duradoura em ambientes clínicos dinâmicos. A literatura evidencia que fenômenos como deriva algorítmica, degradação de desempenho e AI chasm comprometem a validade externa dos modelos quando expostos à variabilidade dos contextos assistenciais reais. Além disso, os estudos analisados indicam que a opacidade algorítmica constitui importante fator de instabilidade normativa, dificultando processos de auditoria, justificabilidade e atribuição de responsabilidade institucional em sistemas de IA de elevada complexidade.

Outro eixo de convergência identificado refere-se à compreensão contextual da transparência e da explicabilidade algorítmica. A literatura contemporânea sustenta que transparência não corresponde à abertura irrestrita de códigos ou modelos computacionais, mas à existência de mecanismos adequados de inteligibilidade, rastreabilidade e prestação de contas compatíveis com o perfil de risco e o contexto de aplicação clínica. Nesse sentido, marcos regulatórios internacionais e referenciais institucionais enfatizam que a governança eficaz da IA depende de monitoramento contínuo, supervisão interdisciplinar e integração entre mecanismos técnicos, organizacionais e normativos de mitigação de riscos.

A análise também evidenciou que os riscos associados à Inteligência Artificial em saúde devem ser compreendidos sob perspectiva sociotécnica, considerando a interação entre dados, modelos computacionais, práticas organizacionais e estruturas institucionais.

A principal contribuição do presente estudo consiste em propor uma interpretação crítico-regulatória da governança algorítmica aplicada ao contexto da saúde pública, especialmente no que se refere à articulação entre *accountability* institucional, justiça sanitária e fortalecimento das capacidades organizacionais do Sistema Único de Saúde. Dessa forma, o trabalho busca preencher lacuna teórica relacionada à ausência de abordagens interdisciplinares capazes de integrar simultaneamente dimensões técnicas, jurídicas, bioéticas e organizacionais da governança da IA em saúde.

Por fim, é importante reconhecer algumas limitações metodológicas e teóricas do presente estudo. Por tratar-se de investigação teórico-analítica fundamentada em revisão narrativa crítica, o trabalho não pretendeu alcançar exaustividade documental, nem produzir generalizações empíricas, concentrando-se prioritariamente na relevância conceitual e interpretativa das fontes selecionadas. Além disso, a predominância de literatura internacional e a ausência de investigação empírica aplicada ao contexto brasileiro indicam a necessidade de estudos futuros voltados à validação prática de modelos de governança algorítmica em instituições de saúde, especialmente no âmbito das políticas públicas e da regulação nacional da Inteligência Artificial.

REFERENCIAS:

BAILO, Paolo; NITTARI, Giulio; PESEL, Giuliano; BASELLO, Emerenziana; SPASARI, Tommaso; RICCI, Giovanna. Governing healthcare AI in the real world: how fairness, transparency, and human oversight can coexist: a narrative review. **Sci,** Basel, v. 8, n. 2, p. 36, 2026. DOI: https://doi.org/10.3390/sci8020036.

BURRELL, Jenna. How the machine "thinks": understanding opacity in machine learning algorithms. **Big Data & Society**, London, v. 3, n. 1, p. 1-12, 2016. DOI: https://doi.org/10.1177/2053951715622512.

CAPELLER, Wanda Maria de Lemos; PEDROSO, João António Fernandes; SANTOS, Andreia Filipa Gonçalves dos. AI, global health and algorithmic society: biodata's democratic use and regulation. **Saúde e Sociedade,** São Paulo, v. 34, n. 3, e250066in, 2025. DOI: https://doi.org/10.1590/S0104-12902025250066in.

EUROPEAN UNION. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).* **Official Journal of the European Union, Brussels,** 2024. Disponível em: http://data.europa.eu/eli/reg/2024/1689/oj . Acesso: em 24 de maio de 2026.

HASSAN, Masooma; BORYCKI, Elizabeth M.; KUSHNIRUK, Andre W. Artificial intelligence governance framework for healthcare. *Healthcare Management Forum*, Thousand Oaks, v. 38, n. 2, p. 125-130, 2025. DOI: https://doi.org/10.1177/08404704241291226.

KIM, Jee Young; HASAN, Alifia; KUEPER, Jacqueline; TANG, Terence; HAYES, Chris; FINE, Benjamin; BALU, Suresh; SENDAK, Mark. Establishing organizational AI governance in healthcare: a case study in Canada. **NPJ Digital Medicine**, London, v. 8, p. 522, 2025. DOI: https://doi.org/10.1038/s41746-025-01909-3.

KISELEVA, Anastasiya; KOTZINOS, Dimitris; DE HERT, Paul. Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. **Frontiers in Artificial Intelligence**, Lausanne, v. 5, p. 879603, 2022. DOI: https://doi.org/10.3389/frai.2022.879603.

KUMAR, Rahul; SPORN, Kyle; WAISBERG, Ethan; ONG, Joshua; PALADUGU, Phani; VADHERA, Amar S.; AMIRI, Dylan; NGO, Alex; JAGADEESAN, Ram; TAVAKKOLI, Alireza; LOFTUS, Timothy; LEE, Andrew G. Navigating healthcare AI governance: the Comprehensive Algorithmic Oversight and Stewardship Framework for Risk and Equity. **Health Care Analysis**, Dordrecht, 2025. DOI: https://doi.org/10.1007/s10728-025-00537-y.

LIPTON, Zachary C. **The mythos of model interpretability.** Ithaca: Cornell University, 2017. Disponível em: https://arxiv.org/abs/1606.03490 .

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **Artificial Intelligence Risk Management Framework (AI RMF 1.0).** Gaithersburg: U.S. Department of Commerce, 2023. DOI: https://doi.org/10.6028/NIST.AI.100-1.

NOUIS, Saoudi CE; UREN, Victoria; JARIWALA, Srushti. Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: a qualitative study of healthcare professionals' perspectives in the UK. **BMC Medical Ethics**, London, v. 26, p. 89, 2025. DOI: https://doi.org/10.1186/s12910-025-01243-z.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). **Advancing accountability in AI: governing and managing risks throughout the lifecycle for trustworthy AI.** Paris: OECD Publishing, 2023. Disponível em: https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/advancing-accountability-in-ai_753bf8c8/2448f04b-en.pdf .

World Health Organization (WHO). **Ethics and governance of artificial intelligence for health: WHO guidance**. Geneva: World Health Organization, 2021. Disponível em: https://www.who.int/publications/i/item/9789240084759 .

WIERINGA, Maranke. **What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability**. In: CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY (FAT*), 2020, Barcelona. *Proceedings [...]*. New York: Association for Computing Machinery, 2020. p. 1-18. DOI: https://doi.org/10.1145/3351095.3372833 .