# GOVERNANÇA RESPONSÁVEL DA INTELIGÊNCIA ARTIFICIAL EM SAÚDE NO BRASIL: REVISÃO SISTEMATIZADA E PROPOSIÇÃO DE UM CICLO INSTITUCIONAL

*RESPONSIBLE GOVERNANCE OF ARTIFICIAL INTELLIGENCE IN HEALTHCARE IN BRAZIL: A SYSTEMATIZED REVIEW AND PROPOSAL FOR AN INSTITUTIONAL CYCLE*

**AGNES ALTMANN**
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL (PUCRS)

**EDIMARA MEZZOMO LUCIANO**
PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL (PUCRS)

**Comunicação:**
O XIV SINGEP foi realizado em conjunto com a 14th Conferência Internacional do CIK (CYRUS Institute of Knowledge) e com a SPARK (Society for Promoting Academic Research and Knowledge), em formato híbrido, com sede presencial na UNINOVE - Universidade Nove de Julho, no Brasil.

**Agradecimento à orgão de fomento:**
Os autores agradecem à Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS) pelo apoio institucional e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio ao desenvolvimento desta pesquisa.

# GOVERNANÇA RESPONSÁVEL DA INTELIGÊNCIA ARTIFICIAL EM SAÚDE NO BRASIL: REVISÃO SISTEMATIZADA E PROPOSIÇÃO DE UM CICLO INSTITUCIONAL

**Objetivo do estudo**
Analisar criticamente a produção acadêmica recente sobre governança da inteligência artificial em saúde no Brasil, caracterizando o corpus, identificando convergências, divergências e lacunas, e propondo um ciclo institucional orientado à priorização pública, diligência prévia, validação local, monitoramento contínuo e prestação de contas.

**Relevância/originalidade**
A relevância do estudo reside em integrar dimensões éticas, regulatórias, técnicas e organizacionais da governança da inteligência artificial em saúde. Sua originalidade está na proposição de um ciclo institucional de cinco movimentos, fundamentado na literatura brasileira e articulado a referenciais internacionais consolidados.

**Metodologia/abordagem**
O estudo adota abordagem qualitativa e revisão sistematizada da literatura, com buscas realizadas na BDTD, no Portal de Periódicos CAPES e na SciELO. Os registros foram submetidos à triagem, caracterização, leitura integral e análise temática de convergências, tensões, lacunas e implicações institucionais.

**Principais resultados**
Os resultados evidenciam convergência quanto à necessidade de proteção de dados, supervisão humana, explicabilidade, responsabilização e validação contextual. Também revelam tensões entre inovação e precaução, transparência e privacidade, padronização e adaptação local, fundamentando um ciclo institucional contínuo de governança responsável em saúde.

**Contribuições teóricas/metodológicas**
Teoricamente, o estudo articula governança, ética, regulação, capacidade institucional e equidade, superando abordagens fragmentadas. Metodologicamente, distingue o corpus descritivo do núcleo analítico, explicita critérios de seleção e organiza convergências, tensões e lacunas, oferecendo estrutura replicável para futuras revisões e pesquisas empíricas nacionais.

**Contribuições sociais/para a gestão**
O estudo oferece aos gestores de saúde um ciclo prático para orientar decisões sobre adoção, validação e monitoramento de sistemas de inteligência artificial. Socialmente, favorece transparência, responsabilização, segurança, participação e equidade, contribuindo para prevenir danos, reduzir desigualdades e fortalecer a confiança pública.

**Palavras-chave:** Governança da Inteligência Artificial, Equidade algorítmica., Saúde Digital, revisão sistematizada, Saúde

# *RESPONSIBLE GOVERNANCE OF ARTIFICIAL INTELLIGENCE IN HEALTHCARE IN BRAZIL: A SYSTEMATIZED REVIEW AND PROPOSAL FOR AN INSTITUTIONAL CYCLE*

**Study purpose**
Critically analyze recent academic production on artificial intelligence governance in Brazilian healthcare, characterize the corpus, identify convergences, divergences, and gaps, and propose an institutional cycle structured around public prioritization, due diligence, local validation, continuous monitoring, and accountability throughout the technology life cycle.

**Relevance / originality**
The study's relevance lies in integrating ethical, regulatory, technical, and organizational dimensions of artificial intelligence governance in healthcare. Its originality consists of proposing a five-stage institutional cycle grounded in Brazilian literature and connected with established international frameworks for responsible organizational decision making.

**Methodology / approach**
The study adopts a qualitative approach and a systematized literature review, with searches conducted in BDTD, the CAPES Journal Portal, and SciELO. Records underwent screening, characterization, full-text reading, and thematic analysis of convergences, tensions, gaps, and institutional implications for healthcare AI governance.

**Main results**
The results reveal convergence regarding the need for data protection, human oversight, explainability, accountability, and contextual validation. They also expose tensions between innovation and precaution, transparency and privacy, standardization and local adaptation, supporting a continuous institutional cycle for responsible artificial intelligence governance.

**Theoretical / methodological contributions**
Theoretically, the study integrates governance, ethics, regulation, institutional capacity, and equity, overcoming fragmented approaches. Methodologically, it distinguishes the descriptive corpus from the analytical core, specifies selection criteria, and organizes convergences, tensions, and gaps, providing replicable structure for future reviews and empirical research.

**Social / management contributions**
The study offers healthcare managers a practical cycle to guide decisions on the adoption, validation, and monitoring of artificial intelligence systems. Socially, it promotes transparency, accountability, safety, participation, and equity, helping prevent harm, reduce inequalities, and strengthen public trust in healthcare institutions.

**Keywords:** Artificial Intelligence Governance, Algorithmic equity, Digital Health, Systematized Review, Health

# GOVERNANÇA RESPONSÁVEL DA INTELIGÊNCIA ARTIFICIAL EM SAÚDE NO BRASIL: REVISÃO SISTEMATIZADA E PROPOSIÇÃO DE UM CICLO INSTITUCIONAL

## 1 Introdução

A Inteligência Artificial deixou de ser um tema restrito à computação e passou a integrar debates centrais da Administração, da Saúde Coletiva, do Direito, da Bioética e das Políticas Públicas. Na saúde, sistemas algorítmicos podem influenciar diagnóstico, tratamento, vigilância epidemiológica, gestão de recursos e acesso a serviços. Essas aplicações ampliam a capacidade de reconhecer padrões e antecipar riscos, mas também redistribuem autoridade, responsabilidades e possibilidades de contestação (Dourado, 2023; Soares; Chiavegatto Filho, 2026; Torres; Wermelinger; Ferreira, 2025).

A governança da IA não se confunde com governança de dados, conformidade legal ou gestão de tecnologia, embora dependa das três. Neste artigo, ela é definida como o arranjo de princípios, atores, responsabilidades e mecanismos verificáveis que orienta a seleção, o desenvolvimento, a contratação, a implantação, o monitoramento, a contestação e a retirada de sistemas de IA. Essa definição processual e institucional difere de concepções predominantemente normativas, centradas em princípios, e de concepções técnico-gerenciais, centradas no gerenciamento de riscos. A posição adotada integra ambas: princípios só se tornam governança quando convertidos em decisões, responsáveis, evidências e mecanismos de correção.

No Brasil, a discussão ocorre sob condições específicas: organização federativa do Sistema Único de Saúde (SUS), desigualdade regional, heterogeneidade de infraestrutura, dependência de fornecedores, sensibilidade dos dados e assimetrias de capacidade entre organizações. Por isso, o artigo tem foco analítico no contexto brasileiro. A literatura e os marcos internacionais são empregados como contraste teórico-regulatório, não como ampliação artificial do corpus nacional.

A lacuna enfrentada reside na fragmentação entre ética, proteção de dados, regulação sanitária, validação técnica e capacidade organizacional. Estudos abordam esses temas, mas raramente mostram como convertê-los em um processo institucional integrado, rastreável e adaptativo. Pergunta-se: como a literatura acadêmica caracteriza a governança da IA em saúde no Brasil, quais tensões e lacunas permanecem e como seus achados podem fundamentar um ciclo institucional de decisão?

O objetivo é analisar criticamente a produção acadêmica recente sobre governança da IA em saúde, caracterizar o corpus, explicitar convergências e divergências e propor um ciclo institucional fundamentado. A contribuição não é apresentar um modelo universal, mas organizar pontos de decisão que possam ser posteriormente testados em hospitais, redes de atenção e órgãos públicos.

## 2 Posicionamento conceitual e interlocução internacional

Como fundamento seminal, Rhodes (1996) distingue governança de governo ao definir governança como uma forma de coordenação constituída por redes interorganizacionais auto-organizadas, caracterizadas pela interdependência, pela troca de recursos, por regras compartilhadas e por ajustes contínuos entre atores. Diferentemente do governo hierárquico, a governança distribui autoridade e capacidade de decisão entre organizações públicas, privadas e atores sociais que controlam recursos distintos. Essa concepção é especialmente pertinente à IA em saúde, pois o desenvolvimento, o fornecimento, a regulação, o uso clínico, a gestão de dados e o controle social envolvem múltiplas organizações. Contudo, a autonomia das redes também pode dificultar direção e

responsabilização, por isso, a governança da IA requer mecanismos explícitos de decisão, supervisão, contestação e prestação de contas.

As concepções de governança de IA podem ser agrupadas, analiticamente, em três ênfases. A primeira é principiológica: direitos humanos, autonomia, justiça, transparência e responsabilidade orientam usos legítimos. A segunda é regulatória: obrigações variam conforme finalidade e risco, com requisitos de documentação, supervisão e avaliação de conformidade. A terceira é organizacional: governança é uma capacidade contínua de mapear contexto, medir riscos, decidir respostas e aprender com resultados. A definição adotada neste artigo assume que nenhuma dessas ênfases é suficiente isoladamente.

A Organização Mundial da Saúde (OMS, 2021) coloca autonomia humana, bem-estar, transparência, responsabilização, inclusão e sustentabilidade no centro da IA em saúde. Os Princípios de IA da OCDE articulam valores humanos, transparência, robustez e accountability (OCDE, 2019/2024). O AI Risk Management Framework do NIST operacionaliza o gerenciamento por quatro funções: governar, mapear, medir e gerenciar, e trata a governança como transversal (NIST, 2023). Na União Europeia, o AI Act adota lógica baseada em risco e sujeita sistemas de alto risco, inclusive determinadas aplicações vinculadas a dispositivos médicos, a exigências reforçadas de gestão, documentação e supervisão (UNIÃO EUROPEIA, 2024). No campo de dispositivos médicos, orientações internacionais de Good Machine Learning Practice destacam qualidade dos dados, avaliação ao longo do ciclo de vida e monitoramento do desempenho (FDA; HEALTH CANADA; MHRA, 2021).

A literatura científica internacional aprofunda a operacionalização desses marcos. Reddy et al. (2020) propõem um modelo específico de governança da IA em saúde, articulando responsabilidade, transparência, confiança, regulação e supervisão durante a incorporação clínica. Wiens et al. (2019) defendem uma trajetória responsável para o aprendizado de máquina em saúde, desde a definição do problema e a qualidade dos dados até a validação, a implantação e o monitoramento. Morley et al. (2020) demonstram que os desafios éticos se distribuem entre questões epistêmicas, normativas e de rastreabilidade, exigindo respostas institucionais em diferentes níveis. Vayena, Blasimme e Cohen (2018), por sua vez, relacionam confiança à proteção de dados, à transparência algorítmica, à equidade e à responsabilização. Esses artigos não foram acrescentados retroativamente ao corpus nacional de 59 produções, são mobilizados como literatura internacional de contraste para situar teoricamente o ciclo proposto.

Esses marcos e estudos aproximam-se ao rejeitar avaliações pontuais e ao exigir atenção ao ciclo de vida. Divergem, porém, no grau de obrigatoriedade, no objeto regulado e na escala: a OMS enfatiza ética e direitos em saúde; a OCDE oferece princípios gerais; o NIST fornece estrutura voluntária de risco; o AI Act institui obrigações jurídicas proporcionais ao risco; e a GMLP se concentra em produtos médicos. O ciclo proposto adiante não substitui esses instrumentos. Ele os traduz para decisões organizacionais em saúde e acrescenta priorização pública e aprendizagem institucional como elos explícitos.

## 3 Metodologia

O estudo consiste em revisão sistematizada da literatura, de abordagem qualitativa, orientada à construção de um estado do conhecimento. Utiliza procedimentos transparentes e ordenados de busca, seleção, caracterização e síntese, sem reivindicar o escopo de uma revisão sistemática exaustiva. Essa denominação é mais adequada porque a busca se concentrou em três fontes, não houve registro prévio de protocolo nem dupla avaliação independente, e a extração disponível não permite reconstruir numericamente cada motivo de exclusão dentro da primeira triagem (Romanowski; Ens, 2006; Morosini; Fernandes, 2014).

A coleta ocorreu entre março e abril de 2026 na Biblioteca Digital Brasileira de Teses

e Dissertações (BDTD), no Portal de Periódicos CAPES e na SciELO. Foram combinados descritores em português e inglês referentes a IA ("inteligência artificial" OR "artificial intelligence"), saúde (saúde OR health) e governança/riscos (governança OR governance OR regulação OR ethics OR "proteção de dados" OR privacy OR explicabilidade OR bias). Consideraram-se os registros disponíveis até abril de 2026.

Os critérios de inclusão foram: texto integral disponível; relação substantiva com saúde; e abordagem de ao menos uma dimensão de governança, regulação, ética, proteção de dados, segurança, explicabilidade, responsabilização, capacidade institucional ou equidade. Excluíram-se duplicidades; estudos exclusivamente técnicos, sem discussão de governança ou impacto institucional; textos sem acesso integral; e produções cujo objeto não apresentava vínculo relevante com saúde.

**Etapa 1 – identificação.** As buscas nas três fontes recuperaram 312 registros: 219 na BDTD, 20 no Portal de Periódicos CAPES e 73 na SciELO.

**Etapa 2 – triagem e elegibilidade geral.** Após a remoção de duplicidades, examinaram-se títulos e resumos e verificaram-se disponibilidade do texto integral e aderência geral ao tema. Foram retirados 253 registros por duplicidade, ausência de texto completo, enfoque exclusivamente técnico ou falta de relação substantiva com governança e saúde. Como a matriz original registrou esses motivos de forma agregada, não foram atribuídas posteriormente quantidades específicas a cada motivo. Essa etapa resultou em 59 produções elegíveis.

**Etapa 3 – caracterização do corpus.** As 59 produções, e não 59 artigos, foram mantidas no mapeamento descritivo e classificadas por origem documental, tipo, período, escopo e tema predominante: 30 artigos científicos, 19 dissertações ou trabalhos aplicados e 10 teses. Portanto, os 59 estudos constituem o corpus de caracterização do estado do conhecimento. **Etapa 4 – seleção do núcleo analítico.** Os 59 textos foram examinados integralmente.

Para integrar o aprofundamento, o estudo deveria: (a) focalizar diretamente a governança da IA em saúde; (b) oferecer contribuição explícita para ao menos uma dimensão regulatória, ética, técnica ou político-institucional; (c) apresentar densidade analítica suficiente para comparação; e (d) contribuir para a diversidade de perspectivas do núcleo. Treze estudos atenderam conjuntamente a esses critérios. Os outros 46 não foram eliminados da revisão: permaneceram na caracterização do corpus, mas não foram mobilizados na análise qualitativa aprofundada.

**Etapa 5 – síntese qualitativa.** Os 13 estudos do núcleo foram submetidos à comparação temática de convergências, tensões, lacunas e implicações. É esse subconjunto analítico que fundamenta diretamente os cinco movimentos do ciclo institucional proposto.

A análise temática de conteúdo compreendeu: leitura e codificação de objetivos, métodos e achados; agrupamento de unidades de sentido por convergência, tensão e silêncio; comparação entre os estudos; e síntese interpretativa (Bardin, 2016). O corpus de 59 sustenta a caracterização descritiva, enquanto o núcleo de 13 sustenta a discussão aprofundada. Essa separação evita apresentar o subconjunto analítico como se fosse a totalidade da produção localizada.

**Tabela 1 – Fluxo de identificação, seleção e aprofundamento do corpus**

| Etapa | Procedimento | Resultado | n |
|---|---|---|---|
| Etapa 1 – Identificação | Busca nas três fontes | Registros localizados | 312 |
| Etapa 2 – Triagem geral | Duplicidades, título/resumo, texto integral e aderência geral | Registros retirados | 253 |
| Etapa 3 – Caracterização | Classificação por origem, tipo, período, escopo e tema predominante | Corpus de caracterização | 59 |
| Etapa 4 – Seleção analítica | Leitura integral e aplicação dos quatro critérios de aprofundamento | Estudos mantidos apenas no mapeamento | 46 |
| Etapa 5 – Síntese qualitativa | Comparação temática de convergências, tensões e lacunas | Núcleo analítico | 13 |

*Nota. Elaborada pela autora (2026) com base na matriz de leitura. Os motivos da triagem inicial foram registrados de forma agregada e, por isso, não recebem contagens retrospectivas por submotivo.*

## 4 Resultados e discussão

### 4.1 Caracterização do corpus

O corpus de 59 produções reúne 29 trabalhos acadêmicos: 10 teses e 19 dissertações ou trabalhos aplicados e 30 artigos. A composição é equilibrada entre produção acadêmica longa e artigos, o que amplia a diversidade metodológica, mas também exige cautela: teses, comentários, revisões e estudos empíricos não oferecem o mesmo nível ou tipo de evidência.

A produção é recente e acelerada. Quarenta e cinco estudos (76,3%) foram publicados entre 2023 e 2026; 12 (20,3%) entre 2020 e 2022; e apenas dois (3,4%) antes de 2020. O predomínio recente indica expansão do campo, mas também baixa sedimentação teórica e rápida obsolescência regulatória. Quanto ao foco, 48 obras apresentam relação direta com saúde; as demais contribuem por conceitos adjacentes, como governança colaborativa, transformação digital, capacidade organizacional e cultura de dados.

A classificação por tema predominante mostra maior concentração em regulação, políticas públicas e responsabilização (23,7%), seguida de capacidade institucional, adoção e gestão (22,0%), ética, equidade impactos humanos (20,3%), validação, explicabilidade e segurança técnica (18,6%) e dados, privacidade segurança e infraestrutura (15,3%). Quanto à origem, 29 trabalhos acadêmicos foram recuperados na BDTD. Os 30 artigos derivam das buscas em CAPES e SciELO, mas a matriz original não preservou a base individual de recuperação de cada artigo; por isso, esses registros são apresentados conjuntamente, sem divisão retrospectiva.

**Tabela 2 – Caracterização do corpus de 59 produções**

| Dimensão | Categoria | n | % |
|---|---|---|---|
| Origem documental | BDTD – teses e dissertações/trabalhos aplicados | 29 | 49,2 |
| Origem documental | CAPES/SciELO – artigos sem desagregação individual na matriz | 30 | 50,8 |
| Tipo | Artigos científicos | 30 | 50,8 |
| Tipo | Dissertações/trabalhos aplicados | 19 | 32,2 |
| Tipo | Teses | 10 | 16,9 |
| Período | 2010–2019 | 2 | 3,4 |
| Período | 2020–2022 | 12 | 20,3 |
| Período | 2023–2026 | 45 | 76,3 |
| Escopo | Foco direto em saúde | 48 | 81,4 |

| | | | |
|---|---|---|---|
| Escopo | Contribuição adjacente | 11 | 18,6 |
| Tema predominante | Regulação, políticas públicas e responsabilização | 14 | 23,7 |
| Tema predominante | Capacidade institucional, adoção e gestão | 13 | 22,0 |
| Tema predominante | Ética, equidade e impactos humanos | 12 | 20,3 |
| Tema predominante | Validação, explicabilidade e segurança técnica | 11 | 18,6 |
| Tema predominante | Dados, privacidade, segurança e infraestrutura | 9 | 15,3 |

Nota. Percentuais calculados sobre n = 59; pequenas diferenças decorrem de arredondamento. Cada produção foi classificada em apenas um tema predominante, definido a partir do título, objetivo, palavras-chave e foco analítico, de modo que as categorias temáticas totalizam 59 e 100%. A matriz permite comprovar 29 trabalhos acadêmicos provenientes da BDTD e 30 artigos recuperados no conjunto CAPES/SciELO, mas não permite desagregar retrospectivamente os 30 artigos entre essas duas bases.

## 4.2 Convergências, tensões e lacunas

A convergência mais forte é que a proteção de dados constitui condição necessária, mas não suficiente. Dourado e Aith (2022) situam a LGPD como ponto de partida, Gonçalo et al. (2025) reforçam normas de acesso, armazenamento e interoperabilidade, Silveira (2021) destaca segurança e responsabilidades, Sembay e Macedo (2024) acrescentam proveniência e rastreabilidade. Em conjunto, os estudos deslocam o debate de consentimento formal para governança contínua dos dados. A tensão surge quando compartilhamento e interoperabilidade, necessários à assistência e à pesquisa, ampliam superfícies de exposição e possibilidades de reutilização não antecipada.

A segunda convergência envolve supervisão humana, explicabilidade e responsabilização. Nunes, Guimarães e Dadalto (2022) defendem autonomia, dignidade e responsabilidade profissional, Dourado (2023) aponta limites de decisões automatizadas, Soares e Chiavegatto Filho (2026) relacionam essas exigências aos princípios do SUS. Contudo, “manter um humano no circuito” não resolve o problema se o profissional não tiver tempo, competência ou autoridade para contestar o sistema. A literatura oscila entre transparência como abertura técnica e explicação como informação útil ao usuário. A primeira pode colidir com segurança, privacidade e segredo industrial, a segunda pode ser insuficiente para auditoria.

A terceira convergência é a necessidade de validação contextual. Costa Filho (2021), Bernardi (2024) e Torres, Wermelinger e Ferreira (2025) mostram que o desempenho técnico não equivale automaticamente a um valor público ou a segurança clínica. A tensão está entre padronização e adaptação: padrões comuns facilitam avaliação e escala, mas diferenças populacionais, epidemiológicas, organizacionais e territoriais exigem validação local. Sem essa etapa, métricas médias podem ocultar desempenho desigual.

A quarta convergência vincula capacidade institucional e equidade. Machado (2025) identifica sub-representação da saúde, ausência de critérios técnicos, descontinuidade e opacidade em políticas brasileiras. Vadlamani e Wachira (2025) mostram que efeitos distributivos e desigualdades de gênero precisam ser avaliados explicitamente. O conflito central opõe eficiência agregada e justiça distributiva: um sistema pode melhorar um indicador médio e, simultaneamente, piorar resultados para grupos sub-representados. A literatura ainda oferece poucos indicadores operacionais para decidir quando essa diferença exige recalibração, restrição ou retirada.

Essas tensões revelam que a governança não deve buscar eliminar todos os conflitos por consenso abstrato. Sua função é tornar escolhas visíveis, justificar prioridades, atribuir responsabilidades e definir como divergências serão monitoradas e revistas.

**Tabela 3 – Tensões críticas e lacunas de pesquisa**

| Tensão | Posições em diálogo | Lacuna para a tese |
|---|---|---|
| Inovação × precaução | Dourado (2023); Soares e Chiavegatto Filho (2026) | Critérios empíricos para graduar risco e decidir implantação, suspensão ou retirada. |
| Interoperabilidade × privacidade | Silveira (2021); Sembay e Macedo (2024); Gonçalo et al. (2025) | Arranjos de acesso, finalidade e rastreabilidade em redes multi-institucionais. |
| Transparência × proteção de ativos | Dourado e Aith (2022); Nunes, Guimarães e Dadalto (2022) | Níveis de explicação úteis a pacientes, profissionais, gestores e auditores. |
| Padronização × adaptação local | Costa Filho (2021); Bernardi (2024); Torres, Wermelinger e Ferreira (2025) | Protocolos de validação externa e local por população, território e fluxo assistencial. |
| Eficiência × equidade | Corrêa (2022); Vadlamani e Wachira (2025); Soares e Chiavegatto Filho (2026) | Indicadores distributivos e limiares de ação para grupos sub-representados. |
| Responsabilidade compartilhada × diluição de responsabilidade | Dourado (2023); Machado (2025) | Matrizes de decisão e responsabilização entre fornecedores, gestores, clínicos e reguladores. |

## 4.3 Ciclo institucional de governança fundamentado na literatura

A síntese sustenta um ciclo de cinco movimentos. Cada movimento deriva de problemas e recomendações identificados no núcleo analítico e encontra correspondência parcial em referenciais internacionais. A proposição é original na articulação e na sequência adaptativa, não na criação isolada de seus componentes. O ciclo começa antes da aquisição e retorna ao início quando monitoramento ou prestação de contas alteram a justificativa do uso.

**Tabela 4 – Fundamentação dos cinco movimentos do ciclo institucional**

| Movimento | Decisão central | Fundamentação no corpus | Diálogo internacional |
|---|---|---|---|
| 1. Priorização pública | O problema justifica IA e gera valor público superior às alternativas? | Machado (2025); Dourado (2023); Soares e Chiavegatto Filho (2026) | OMS: finalidade pública e bem-estar; NIST: mapeamento do contexto. Literatura científica: Reddy et al. (2020). |
| 2. Diligência prévia | Fornecedor, evidências, dados, riscos e responsabilidades são aceitáveis? | Silveira (2021); Sembay e Macedo (2024); Dourado e Aith (2022); Gonçalo et al. (2025) | AI Act: gestão de risco e documentação; GMLP: qualidade e boas práticas no ciclo de vida. Literatura científica: Vayena, Blasimme e Cohen (2018). |
| 3. Validação local | O desempenho é seguro, útil e equitativo na população e nos fluxos reais? | Costa Filho (2021); Corrêa (2022); Bernardi (2024); Vadlamani e Wachira (2025) | GMLP: dados representativos e avaliação clínico-técnica; NIST: medir. Literatura científica: Wiens et al. (2019). |
| 4. Monitoramento contínuo | Há degradação, incidentes, efeitos no trabalho ou diferenças entre grupos? | Bernardi (2024); Torres, Wermelinger e Ferreira (2025); Soares e Chiavegatto Filho (2026) | NIST: medir e gerenciar; GMLP: monitoramento do produto; AI Act: pós-mercado. Literatura científica: Wiens et al. (2019) e Reddy et al. (2020). |

| | | | |
|---|---|---|---|
| 5. Prestação de contas | Quem explica, corrige, repara e decide continuidade ou retirada? | Dourado (2023); Nunes, Guimarães e Dadalto (2022); Machado (2025) | OMS/OCDE: accountability e contestabilidade; NIST: governança transversal. Literatura científica: Morley et al. (2020) e Vayena, Blasimme e Cohen (2018). |

A priorização pública impede que o apelo da inovação substitua a definição do problema. A organização deve explicitar benefício esperado, população afetada, alternativa não algorítmica, custo de oportunidade e alinhamento ao SUS. Esse movimento deriva da crítica de Machado (2025) à ausência de critérios públicos e da exigência de Dourado (2023) de regulação orientada ao risco.

A diligência prévia antecede contratação e implantação. Inclui avaliação do fornecedor, finalidade declarada, proveniência e representatividade dos dados, evidência de desempenho, segurança, interoperabilidade, condições de atualização, repartição de responsabilidades e possibilidade de auditoria. Silveira (2021), Sembay e Macedo (2024), Dourado e Aith (2022) e Gonçalo et al. (2025) sustentam esse movimento ao mostrar que proteção jurídica, segurança e rastreabilidade precisam operar em conjunto.

A validação local testa utilidade, segurança e equidade no contexto real. Deve comparar o desempenho global e por subgrupos, avaliar integração ao fluxo e registrar divergências entre recomendações algorítmicas e decisões humanas. Costa Filho (2021) e Bernardi (2024) apoiam a avaliação contextual; Corrêa (2022) e Vadlamani e Wachira (2025) justificam a inclusão explícita de justiça e efeitos distributivos.

O monitoramento contínuo acompanha degradação, mudanças de dados, incidentes, excesso de alertas, dependência da automação, sobrecarga profissional e diferenças entre grupos. Atualizações relevantes exigem controle de versão e nova validação. Bernardi (2024) aproxima implementação e avaliação, Torres, Wermelinger e Ferreira (2025) exigem contextualização e capacidade institucional.

A prestação de contas reúne documentação, explicação, contestação, correção e reparação. Deve definir quem autoriza, quem monitora, quem responde a incidentes e quem pode interromper o sistema. Canais acessíveis a profissionais e usuários fecham o ciclo: evidências de falha ou baixo valor retornam à priorização e podem justificar modificação ou retirada.

### 4.4 Agenda de pesquisa para a tese

A revisão aponta uma agenda centrada na pergunta: como organizações brasileiras de saúde constroem capacidades e distribuem responsabilidades para governar sistemas de IA durante seu ciclo de vida? Essa formulação desloca o problema de uma lista abstrata de princípios para práticas organizacionais observáveis.

Cinco lacunas sustentam essa agenda: (a) poucos estudos empíricos acompanham sistemas após a implantação; (b) faltam critérios comparáveis de validação local e equidade; (c) a distribuição de responsabilidades entre fornecedor, gestor, equipe clínica, controlador de dados e regulador é pouco observada; (d) há escassez de estudos sobre contestação, reparação e retirada de sistemas; e (e) os marcos internacionais são raramente traduzidos para capacidades e restrições de organizações brasileiras. Cada lacuna pode ser investigada por estudos de caso múltiplos, análise documental, entrevistas com stakeholders e comparação entre aplicações de diferentes níveis de risco.

## 5 Considerações finais

A literatura caracteriza a governança da IA em saúde como um arranjo multidimensional e contínuo que articula regulação, ética, dados, validação, supervisão, capacidade institucional e equidade. O foco brasileiro evidencia desafios próprios do SUS e de organizações com capacidades desiguais, enquanto o contraste internacional mostra que o

campo converge para abordagens baseadas em risco e ciclo de vida.

A revisão também demonstra que o consenso aparente encobre conflitos substantivos. Interoperabilidade pode ampliar exposição, transparência pode colidir com privacidade e proteção de ativos, padronização pode ignorar diferenças locais, supervisão humana pode ser apenas simbólica e ganhos médios de eficiência podem aprofundar desigualdades. Tornar essas tensões verificáveis é parte da governança, não um obstáculo externo a ela.

O ciclo de priorização pública, diligência prévia, validação local, monitoramento contínuo e prestação de contas constitui a contribuição central. Sua fundamentação foi explicitada estudo por estudo e relacionada a referenciais internacionais. Ainda assim, trata-se de proposição teórica: sua utilidade, sequência, indicadores e condições de implementação precisam ser avaliados empiricamente.

As limitações decorrem da concentração da busca em três fontes, da heterogeneidade das produções, do registro agregado dos motivos de exclusão inicial e da seleção intencional de 13 estudos para aprofundamento. Essas escolhas não invalidam a síntese, mas delimitam seu alcance.

Conclui-se que governar a IA em saúde significa organizar a capacidade coletiva de decidir antes, durante e depois da implantação. A legitimidade da inovação depende menos da promessa do algoritmo do que da possibilidade de demonstrar valor público, prevenir dano, identificar desigualdades, contestar decisões e aprender institucionalmente.

**Declaração de uso de inteligência artificial generativa**

Foi utilizada ferramenta de inteligência artificial generativa como apoio à revisão linguística, à reorganização textual, à padronização formal e à construção visual das tabelas. As decisões de escopo, seleção, categorização, interpretação, posicionamento conceitual e validação das referências permanecem sob responsabilidade das autoras.

**Referências**

BARDIN, Laurence. Análise de conteúdo. São Paulo: Edições 70, 2016.

BERNARDI, Filipe Andrade. Digital health research governance: from FAIR to RE-AIM. 2024. Tese (Doutorado) – Universidade de São Paulo, São Carlos, 2024.

CORRÊA, Leonardo Cambraia. Análise bioética do uso de Big Data em saúde. 2022. Dissertação (Mestrado em Bioética) – Universidade de Brasília, Brasília, 2022.

COSTA FILHO, Raimir Holanda. Smart-GISSA: um sistema para governança em saúde digital baseado em aprendizado de máquina. 2021. Tese (Doutorado) – Universidade Federal do Ceará, Fortaleza, 2021.

DOURADO, Daniel de Araujo. Regulação da inteligência artificial na saúde. 2023. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2023.

DOURADO, Daniel de Araujo; AITH, Fernando Mussa Abujamra. A regulação da inteligência artificial na saúde no Brasil começa com a Lei Geral de Proteção de Dados Pessoais. Revista de Saúde Pública, v. 56, 80, 2022. DOI: 10.11606/s1518-8787.2022056004461.

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. Official Journal of the European Union, 2024.

FDA; HEALTH CANADA; MHRA. Good Machine Learning Practice for Medical Device Development: Guiding Principles. 2021.

GONÇALO, Wemerson; SOUZA, Maria Clara de; SANTOS, Wellington Pinheiro dos; OLIVEIRA, Fábio Henrique Cavalcanti de. Abordagens regulatórias na proteção de

dados em saúde: uma revisão integrativa de 2018 a 2023. Physis: Revista de Saúde Coletiva, v. 35, n. 1, e350113, 2025. DOI: 10.1590/S0103-73312025350113pt.

MACHADO, Mariana. A inteligência artificial na saúde: um estudo da Política Nacional de IA no Brasil entre 2019 e 2025. 2025. 100 f. Dissertação (Mestrado em Ciências) – Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, 2025.

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

MOROSINI, Marília Costa; FERNANDES, Cleoni Maria Barboza. Estado do conhecimento: conceitos, finalidades e interlocuções. Educação Por Escrito, v. 5, n. 2, p. 154–164, 2014.

NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Gaithersburg: National Institute of Standards and Technology, 2023. DOI: 10.6028/NIST.AI.100-1.

NUNES, Heloá da Conceição; GUIMARÃES, Rita Miranda Coessens; DADALTO, Luciana. Desafios bioéticos do uso da inteligência artificial em hospitais. Revista Bioética, v. 30, n. 1, p. 82–93, 2022. DOI: 10.1590/1983-80422022301509PT.

OECD. Recommendation of the Council on Artificial Intelligence. Paris: Organisation for Economic Co-operation and Development, 2019, atualizada em 2024.

REDDY, Sandeep; ALLAN, Sonia; COGHLAN, Simon; COOPER, Paul. A governance model for the application of AI in health care. Journal of the American Medical Informatics Association, v. 27, n. 3, p. 491–497, 2020. DOI: 10.1093/jamia/ocz192.

RHODES, R. A. W. The new governance: governing without government. Political Studies, v. 44, n. 4, p. 652–667, 1996. DOI: 10.1111/j.1467-9248.1996.tb01747.x.

ROMANOWSKI, Joana Paulin; ENS, Romilda Teodora. As pesquisas denominadas do tipo "estado da arte" em educação. Revista Diálogo Educacional, v. 6, n. 19, p. 37–50, 2006.

SEMBAY, Márcio José; MACEDO, Douglas Dyllon Jeronimo de. Proveniência de dados e inteligência artificial em saúde: desafios e implicações éticas sob a ótica da ciência da informação. LOGEION: Filosofia da Informação, v. 11, ed. esp., p. 1–17, 2024. DOI: 10.21728/logeion.2024v11e-7388.

SILVEIRA, Suzana Aparecida. Segurança da informação e proteção de dados pessoais: estudo de caso e proposta de governança para serviços de saúde. 2021. 211 p. Dissertação (Mestrado Profissional em Inovação Tecnológica) – Universidade Federal de São Paulo, São José dos Campos, 2021.

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: 10.1590/1980-549720260006.2.

TORRES, Douglas Rodrigues; WERMELINGER, Eduardo Dias; FERREIRA, Aldo Pacheco. Aplicação da Inteligência Artificial na Atenção Primária à Saúde: revisão de escopo e avaliação crítica. Saúde em Debate, v. 49, n. 145, e10070, 2025. DOI: 10.1590/2358-2898202514510070P.

VAYENA, Effy; BLASIMME, Alessandro; COHEN, I. Glenn. Machine learning in medicine: addressing ethical challenges. PLOS Medicine, v. 15, n. 11, e1002689, 2018. DOI: 10.1371/journal.pmed.1002689.

VADLAMANI, Suman; WACHIRA, Elizabeth. AI's ongoing impact: implications of AI's effects on health equity for women's healthcare providers. Revista Panamericana de Salud Pública, v. 49, e19, 2025. DOI: 10.26633/RPSP.2025.19.

WIENS, Jenna et al. Do no harm: a roadmap for responsible machine learning for health care. Nature Medicine, v. 25, n. 9, p. 1337–1340, 2019. DOI: 10.1038/s41591-019-

0548-6.

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