**Trilha 1 - Teorias e Desenvolvimento**

**GOVERNANÇA DE IA NO SETOR HOSPITALAR: UMA REVISÃO BIBLIOMÉTRICA SOBRE RISCOS CONTEXTUAIS, AMEAÇAS CIBERNÉTICAS E INTEGRAÇÃO CLÍNICO-REGULATÓRIA**

*AI GOVERNANCE IN THE HOSPITAL SECTOR: A BIBLIOMETRIC REVIEW ON CONTEXTUAL RISKS, CYBERSECURITY THREATS, AND CLINICAL-REGULATORY INTEGRATION*

**Wagner de Paula Rodrigues**

Mestre. UNINOVE – Brazil.

wprodriguesbr@uni9.edu.br.

**Ivanir Costa**

Doutor. UNINOVE – Brazil.

ivanirc@uni9.pro.br.

**Luiz Fernando Rodrigues Pinto**

Doutor. UNINOVE – Brazil.

luiz.rodrigues@uni9.pro.br.

**RESUMO**

A inteligência artificial (IA) tem impulsionado a digitalização da saúde, promovendo avanços em diagnóstico por imagem, apoio à decisão clínica e monitoramento remoto. Contudo, sua adoção em hospitais introduz riscos complexos, como viés algorítmico, ameaças cibernéticas e lacunas regulatórias, que desafiam a governança da IA, especialmente no Brasil, onde coexistem marcos como a Lei Geral de Proteção de Dados (LGPD), as diretrizes da Anvisa e os padrões da Organização Nacional de Acreditação (ONA). Embora metodologicamente robustos, frameworks como o NIST AI RMF e o AI Act da União Europeia mostram-se pouco sensíveis às especificidades clínicas e operacionais do setor hospitalar. Este estudo parte da hipótese de que tais referenciais não oferecem diretrizes acionáveis para o contexto brasileiro. Para investigar essa lacuna, foi realizada uma revisão bibliométrica com 252 publicações indexadas nas bases Scopus e Compendex (1989–2025), analisadas por meio do software VOSviewer e curadoria semântica com thesaurus. A análise de coocorrência de palavras-chave revelou quatro clusters temáticos: ecossistemas de cuidado, governança de dados, riscos técnicos em IA e estratégias decisórias. Os resultados apontam fragmentação entre os domínios técnico, ético e assistencial, com baixa atenção à explicabilidade algorítmica e à validação clínica. Tais lacunas reforçam a necessidade de um modelo de governança setorial que integre práticas regulatórias, conhecimento clínico e segurança cibernética, traduzindo diretrizes internacionais em soluções operacionais viáveis no contexto hospitalar brasileiro.

**PALAVRAS-CHAVE:** Inteligência Artificial, Saúde, Governança de IA, Cibersegurança, Inteligência Artificial Explicável (XAI), Conformidade Regulatória.

**ABSTRACT**

Artificial intelligence (AI) has driven the digital transformation of healthcare, fostering advances in medical imaging, clinical decision support, and remote monitoring. However, its adoption in hospitals introduces complex risks, such as algorithmic bias, cybersecurity threats, and regulatory gaps, which challenge AI governance—particularly in Brazil, where legal and institutional frameworks coexist, including the General Data Protection Law (LGPD), Anvisa guidelines, and the standards of the National Accreditation Organization (ONA). Although methodologically robust, frameworks such as the NIST AI RMF and the European Union AI Act are insufficiently sensitive to the clinical and operational specificities of the hospital sector. This study is based on the hypothesis that such frameworks do not offer actionable guidance for the Brazilian context. To investigate this gap, a bibliometric review was conducted based on 252 publications indexed in the Scopus and Compendex databases (1989–2025), analyzed using VOSviewer and a custom-built semantic thesaurus. The keyword co-occurrence analysis revealed four thematic clusters: care ecosystems, data governance, technical risks in AI, and decision-making strategies. The results show fragmentation across technical, ethical, and clinical domains, with limited attention to algorithmic explainability and clinical validation. These gaps underscore the need for a sector-specific AI governance model that integrates regulatory practices, clinical expertise, and cybersecurity, translating international guidelines into operationally viable solutions for the Brazilian hospital context.

**KEYWORDS:** *Artificial Intelligence, Healthcare, AI Governance, Cybersecurity, Explainable Artificial Intelligence (XAI), Regulatory Compliance.*

1. **INTRODUÇÃO**

A inteligência artificial (IA) é uma das tecnologias mais transformadoras da atualidade, com impactos crescentes em setores estratégicos como saúde, transporte e educação. Sua adoção tem promovido ganhos em eficiência e personalização, mas também impõe desafios éticos, regulatórios e socioeconômicos. A Organização para a Cooperação e Desenvolvimento Econômico (OECD, 2023) destaca a importância de frameworks normativos que mitiguem riscos e maximizem benefícios, assegurando alinhamento com valores democráticos e direitos fundamentais.

No setor da saúde, a IA tem promovido avanços em diagnóstico por imagem, apoio à decisão clínica e telemedicina (Bohr; Memarzadeh, 2020). Contudo, esses benefícios são acompanhados por riscos emergentes, como viés algorítmico, que pode acentuar desigualdades no acesso ao cuidado (Obermeyer et al., 2019), e falhas na proteção de dados sensíveis, que comprometem a segurança do paciente e a confiança institucional (Borycki; Kushniruk, 2021). Esses riscos são críticos em contextos clínicos complexos, onde decisões assistenciais exigem julgamento profissional e validação contextual (Al-Maamari, 2025; Boonstra; Laven, 2022)

Diversas iniciativas regulatórias internacionais vêm buscando estruturar os riscos associados à IA, como o NIST AI RMF (2023), o AI Act da União Europeia e a ISO 42001 (2024). Contudo, embora representem avanços relevantes, esses referenciais adotam abordagens de caráter geral e, segundo a OECD (2023), sua efetiva implementação requer adaptações específicas ao contexto de cada setor, de modo a considerar características técnicas, regulatórias e operacionais próprias. No Brasil, a fragmentação entre normativas, como a LGPD, regulamentações da Anvisa e padrões da ONA, destaca a necessidade de abordagens mais integradas e contextualizadas (Radanliev *et al.*, 2021). Nesse sentido, a governança da IA pode ser compreendida como o conjunto de processos, estruturas, ferramentas e atores responsáveis por assegurar o uso responsável da IA, assegurando que seu desenvolvimento e aplicação estejam alinhados a princípios éticos, legais e sociais (OECD, 2023).

Diante desse cenário, este estudo realiza uma revisão bibliométrica da literatura científica sobre governança e riscos associados ao uso da IA no setor hospitalar. A análise foi conduzida com base em dados das bases Scopus e Compendex, utilizando o software VOSviewer e orientações metodológicas de Donthu et al. (2021). O objetivo é avaliar se o conhecimento disponível sustenta a formulação de abordagens de governança sensíveis às especificidades clínicas, técnicas e regulatórias da saúde.

1. **FUNDAMENTOS DE GESTÃO DE RISCO NA SAÚDE**

A gestão de riscos na saúde envolve processos estruturados para identificar, avaliar e monitorar eventos adversos que comprometam a segurança do paciente e a sustentabilidade institucional. Inicialmente voltada a incidentes clínicos, como erros de medicação e infecções hospitalares, essa abordagem evoluiu diante da digitalização dos serviços e da crescente complexidade dos sistemas assistenciais (COSO, 2017; ISO 42001, 2024).

A adoção da IA tem ampliado a acurácia diagnóstica e a eficiência operacional, com aplicações em visão computacional, prontuários eletrônicos e automação administrativa (Bohr; Memarzadeh, 2020). Contudo, esses avanços são acompanhados por riscos que desafiam os modelos tradicionais de controle, como vieses algorítmicos, baixa qualidade dos dados e falta de explicabilidade clínica (Borycki; Kushniruk, 2021). Além disso, vulnerabilidades cibernéticas e dificuldades de integração entre sistemas heterogêneos elevam a exposição a incidentes de segurança (Jalali; Kaiser, 2018).

Frameworks como o NIST AI RMF (2023), o EU AI Act e a ISO 42001 (2024) fornecem diretrizes relevantes, mas ainda carecem de sensibilidade às especificidades clínicas e operacionais dos ambientes hospitalares (Kritharidou et al., 2023). A ausência de integração com fluxos de trabalho e a baixa aderência a normativas setoriais, como as diretrizes da FDA para Software as a Medical Device (SaMD), revelam limitações na governança da IA em saúde (Reddy et al., 2019).

Diante desse cenário, a literatura aponta a necessidade de abordagens mais responsivas, que combinem inovação tecnológica com segurança do paciente, transparência e validação contextual (Borycki; Kushniruk, 2021). Inspirado por metodologias consolidadas (Donthu et al., 2021; Van Eck; Waltman, 2010), este estudo busca oferecer uma base empírica para interpretar tais desafios e fundamentar propostas de gestão de risco adaptadas ao setor. Esta seção, portanto, estrutura o referencial teórico que orienta a análise, articulando fundamentos clássicos com os imperativos contemporâneos da transformação digital e da adoção segura da IA.

* 1. MODELOS CLÁSSICOS DE GESTÃO DE RISCO

A gestão de riscos em saúde apoia-se historicamente em frameworks normativos que organizam o processo em etapas iterativas de identificação, análise, tratamento, monitoramento e comunicação (ISO 31000, 2018). Embora eficazes em contextos organizacionais diversos, tais modelos enfrentam limitações em ambientes clínicos, caracterizados por alta variabilidade operacional, interdependência técnico-assistencial e exigências éticas específicas (Vincent; Amalberti, 2016).

Dentre os modelos setoriais, destaca-se o Healthcare Failure Mode and Effect Analysis (HFMEA), voltado à avaliação proativa de falhas em processos como administração de medicamentos e cirurgias. Contudo, a introdução da IA exige abordagens complementares, dado o surgimento de riscos tecnológicos, falhas algorítmicas e desafios à segurança da informação (Borycki; Kushniruk, 2021).

O modelo Swiss Cheese, de Reason, permanece relevante ao abordar falhas latentes que atravessam múltiplas barreiras organizacionais. Em ambientes digitalizados, sua eficácia depende da capacidade institucional de antecipar vulnerabilidades e integrar equipes clínicas e técnicas (Perneger, 2005; Vincent; Amalberti, 2016).

Mais recentemente, o NIST AI RMF (NIST AI RMF, 2023) introduziu diretrizes voltadas à governança da IA, com foco em transparência, confiabilidade e mitigação de viés. Apesar de seu avanço metodológico, ainda carece de alinhamento com as especificidades do setor saúde, como validação contextual de algoritmos, interoperabilidade clínica e aderência a normas como LGPD, Anvisa e ONA (NIST AI RMF, 2023; Al-Maamari, 2025).

A análise bibliométrica deste estudo reforça que riscos associados à IA em saúde, como os identificados nos clusters de governança, cibersegurança, ética regulatória e infraestrutura assistencial não são plenamente contemplados pelos modelos clássicos. Embora setores como defesa e finanças adotem frameworks mais dinâmicos, como o *Cybersecurity Maturity Model Certification* 2.0 (CMMC) e o *Committee of Sponsoring Organizations of the Treadway Commission - Enterprise Risk Management* (COSO ERM), no setor saúde persiste uma lacuna estrutural entre risco técnico, prática clínica e regulação ética.

* 1. RISCOS ESPECÍFICOS DA INTELIGÊNCIA ARTIFICIAL NA SAÚDE

A IA tem transformado práticas clínicas e administrativas na saúde, promovendo ganhos em precisão diagnóstica, eficiência operacional e personalização terapêutica (Bohr; Memarzadeh, 2020). No entanto, sua adoção introduz riscos complexos, ampliados pela sensibilidade dos dados, pela criticidade dos serviços e pela assimetria de poder entre desenvolvedores, gestores e profissionais clínicos (Borycki; Kushniruk, 2021).

Diferentemente de setores como o financeiro ou o varejo, onde os riscos são majoritariamente operacionais ou reputacionais, na saúde os impactos da IA afetam diretamente princípios como equidade, autonomia, dignidade e segurança do paciente. Isso exige uma governança que una rigor técnico, responsabilidade ética e conformidade regulatória (Reddy *et al.*, 2020).

O Quadro 1 sintetiza as fontes de risco associadas à adoção de IA em ambientes clínicos, com base na ISO 42001 (2024). Embora essas categorias sejam analiticamente úteis, suas fronteiras frequentemente se sobrepõem na prática assistencial:

Quadro 1: Fontes de Risco associadas à AI em saúde, com base na ISO 42001 (2024)

|  |  |  |  |
| --- | --- | --- | --- |
| Fonte de Risco | Risco | Descrição | Impacto |
| Aprendizado de máquina. Anexo C.3.4 (ISO 42001, 2024) | Viés algorítmico | Decisões enviesadas devido a dados de treinamento historicamente enviesados (ex.: raça, gênero), perpetuando desigualdades. | Ameaça à equidade e justiça distributiva (Obermeyer *et al.*, 2019) |
| Aprendizado de máquina; Anexo C.3.4 (ISO 42001, 2024)  Hardware do sistema. Anexo C.3.5 (ISO 42001, 2024) | *Data poisoning* e ataques adversariais | Inserção de dados maliciosos ou ruído visando manipular previsões do modelo, explorando vulnerabilidades técnicas. | Pode levar a erros clínicos graves (Kritharidou *et al.*, 2023) |
| Transparência e explicabilidade. Anexo C.3.2 (ISO 42001, 2024) | Falta de interpretabilidade (XAI) | Modelos opacos dificultam auditoria e compreensão pelas partes interessadas. | Fragiliza a confiança e a responsabilização. (Reddy *et al.*, 2020; Borycki; Kushniruk, 2021) |
| Ciclo de vida e complexidade ambiental. Anexos C.3.6 e C.3.1 (ISO 42001, 2024) | Conformidade regulatória e bioética | Dificuldade de aderência simultânea a diferentes marcos legais e éticos (LGPD, Anvisa, princípios da bioética). | Risco jurídico e ético (Al-Maamari, 2025) |
| Nível de automação. Anexo C.3.3 (ISO 42001, 2024) | Dependência excessiva da automação | Redução da supervisão humana em decisões clínicas críticas, gerando confiança desmedida na IA. | Risco sistêmico e perda do julgamento clínico (Vincent; Amalberti, 2016) |
| Integração sistêmica. Anexos C.3.6 e C.3.5 (ISO 42001, 2024) | Integração deficiente com sistemas legados | Incompatibilidades entre sistemas de IA e infraestrutura existente, comprometendo a interoperabilidade. | Perda de dados e atrasos no cuidado (Jalali; Kaiser, 2018) |

Fonte: o Autor

No caso brasileiro, tais riscos são agravados por disparidades de maturidade digital entre instituições públicas e privadas, baixa capacitação em cibersegurança e fragmentação regulatória. Essa heterogeneidade compromete a resiliência do sistema e amplia a vulnerabilidade operacional (Radanliev et al., 2021). A articulação entre riscos técnicos, éticos e operacionais ressalta a inadequação de frameworks genéricos para orientar a governança da IA no setor saúde. Isso torna imprescindível uma avaliação crítica das limitações dos modelos atualmente disponíveis e justifica a adoção de uma abordagem bibliométrica para mapear soluções adaptadas à realidade setorial.

* 1. LIMITAÇÕES DOS FRAMEWORKS EXISTENTES E JUSTIFICATIVAS PARA A PROPOSTA

Embora frameworks como a ISO 31000 (2018), o NIST AI RMF (2023), a ISO 42001 (2024) e a ISO 27001 (2022) promovam resiliência organizacional em diversos setores, sua aplicação ao contexto hospitalar revela limitações relevantes. Tais modelos multissetoriais não contemplam adequadamente exigências clínicas, como a transparência em diagnósticos algorítmicos, a validação contextual de sistemas de IA e a proteção de dados sensíveis de pacientes (Al-Maamari, 2025; Reddy et al., 2020).

Além disso, observa-se uma dissociação entre referenciais técnicos e o conhecimento prático em ambientes clínicos críticos, como UTIs e triagens automatizadas. A ausência de integração entre requisitos técnicos, éticos e operacionais compromete a legitimidade das soluções baseadas em IA e dificulta sua aceitação por profissionais da saúde (Borycki; Kushniruk, 2021).

Outros setores, como defesa e aviação, já dispõem de frameworks maduros para lidar com riscos automatizados. O CMMC define níveis de maturidade em cibersegurança, enquanto o COSO ERM estrutura princípios compatíveis com governança baseada em risco (COSO, 2017). Na aviação, confiança e previsibilidade de sistemas são requisitos para decisões críticas. No setor saúde, entretanto, a alta variabilidade clínica e os riscos contextuais dificultam a adoção direta desses modelos, exigindo abordagens mais adaptativas (Vincent; Amalberti, 2016). Conforme argumenta Al-Maamari (2025), a ausência de estruturas regulatórias harmonizadas em países com arcabouços fragmentados, como o Brasil, amplia o descompasso entre os riscos associados à IA e os mecanismos de mitigação disponíveis. A simultaneidade de regulamentações como a LGPD, as diretrizes da Anvisa e os padrões da ONA impõe desafios adicionais à implementação de soluções integradas e operacionais.

Diante dos riscos analisados, torna-se evidente que frameworks genéricos não são suficientes para orientar de forma eficaz a governança da IA em saúde. Justifica-se, portanto, a adoção de uma abordagem bibliométrica capaz de mapear, de forma sistemática, os principais desafios e lacunas conceituais do campo. Tal etapa é essencial para fundamentar estratégias regulatórias mais sensíveis às especificidades clínicas e institucionais do setor.

1. **METODOLOGIA**

Dada a fragmentação entre os domínios clínico, técnico e regulatório na governança da IA em saúde, conforme discutido na Seção 1, optou-se por uma revisão bibliométrica para mapear a estrutura temática e as lacunas da literatura sobre riscos associados à IA no setor hospitalar. Essa abordagem, fundamentada em *science mapping*, permite analisar sistematicamente a produção científica, fornecendo evidências para frameworks de governança adaptados (Donthu *et al.*, 2021).

A metodologia adotada segue as diretrizes de Donthu et al. (2021), estruturada em quatro etapas complementares: (i) definição do objetivo e escopo da pesquisa; (ii) construção da estratégia de busca e seleção das bases de dados, utilizando palavras-chave livres e operadores booleanos; (iii) extração, normalização e pré-processamento dos metadados bibliográficos, com padronização semântica por meio de *thesaurus*; e (iv) aplicação da análise de coocorrência no software VOSviewer (versão 1.6.20), gerando visualizações estruturais e temporais de padrões temáticos.

Todas as análises foram realizadas exclusivamente com o VOSviewer, reconhecido por sua robustez na construção de mapas bibliométricos (Van Eck & Waltman, 2010). Essa escolha garante consistência e replicabilidade, alinhando-se a boas práticas em estudos de governança (D’Ascenzo et al., 2024).

* 1. DEFINIÇÃO DO ESCOPO E OBJETIVOS DA PESQUISA

A presente revisão bibliométrica tem como objetivo identificar e analisar os principais eixos temáticos, lacunas conceituais e tendências emergentes na literatura científica sobre governança da IA no setor hospitalar, com foco em riscos contextuais, cibersegurança e conformidade regulatória. O escopo foi delimitado por três domínios críticos(i) gestão de riscos aplicados à saúde, com atenção especial a ameaças cibernéticas e falhas operacionais relacionadas a IA; (ii) estruturação de práticas de governança clínica e digital frente às novas tecnologias; e (iii) conformidade com exigências regulatórias nacionais e internacionais, como a LGPD, diretrizes da Anvisa e normas hospitalares. Optou-se por uma revisão bibliométrica operacionalizada por meio do *science mapping*, conforme orientações de Donthu *et al.* (2021), com o intuito de explorar o estado atual da produção científica e oferecer subsídios empíricos para futuras propostas de modelos setoriais de governança em saúde.

* 1. ESTRATÉGIA DE BUSCA E BASES DE DADOS

A estratégia de busca foi estruturada com o objetivo de identificar publicações científicas que abordassem a interseção entre gestão de riscos, IA e saúde. O recorte temático concentrou-se nas dimensões de segurança clínica, cibersegurança e conformidade regulatória, reconhecidas como essenciais para compreender os desafios contemporâneos da adoção de sistemas algorítmicos em ambientes assistenciais. A seleção das bases seguiu o critério de complementaridade temática: optou-se pela *Scopus* e pela *Compendex* (*Engineering Village*). Essa abordagem capturou perspectivas clínicas, tecnológicas e regulatórias, essenciais para estudos de risco sistêmico em IA aplicada à saúde.

* + 1. *STRING* DE BUSCA

As buscas foram realizadas entre os dias 20 e 23 de maio de 2025, com *strings* adaptadas às particularidades sintáticas de cada base de dados. Na *Scopus*, os termos foram aplicados nos campos de título, resumo e palavras-chave por meio da expressão *TITLE-ABS-KEY ("****risk management****" AND "****artificial intelligence****" AND* ***("healthcare"*** *OR* ***"medical"****)) AND (LIMIT-TO (DOCTYPE,* ***"ar"****) OR LIMIT-TO (DOCTYPE,* ***"re"****) OR LIMIT-TO (DOCTYPE,* ***"cp"****) OR LIMIT-TO (DOCTYPE,* ***"cr"****))*. Já na *Compendex* (*Engineering Village*), utilizou-se a string *((((****"risk management"*** *AND* ***"artificial intelligence"*** *AND (****"healthcare"*** *OR* ***"medical"*** *)) WN ALL)) AND ((****{ca}*** *OR* ***{ja}*** *OR* ***{cp}*** *OR* ***{ip})*** *WN DT))*, aplicada em todos os campos indexados.

Adotou-se o filtro de idioma para o inglês, priorizando publicações de maior impacto internacional, conforme recomendação metodológica de Donthu *et al.* (2021). Embora isso exclua parte da produção regional em português e espanhol, a escolha favorece a homogeneidade e a comparabilidade do corpus. Nenhum filtro temporal foi aplicado, visando capturar a evolução natural da literatura sobre o tema.

* + 1. CRITÉRIOS DE INCLUSÃO E EXCLUSÃO

A seleção documental seguiu os princípios recomendados para revisões bibliométricas robustas (Donthu *et al.*, 2021), garantindo coerência temática, qualidade dos metadados e reprodutibilidade analítica. O Quadro 2 sintetiza os critérios de elegibilidade adotados.

Quadro 2: Critérios de inclusão e exclusão

|  |  |  |
| --- | --- | --- |
| Critério | Inclusão | Exclusão |
| Tipo de documento | Artigos, revisões, artigos de periódico (*Article*, *Journal Article*, *Review*) revisados por pares | Artigos em pré-publicação (*Article in Press*), artigos de conferência (*Conference Article*, *Conference Paper*), proceedings (*Conference Proceeding*), revisões de conferência (*Conference Review*) |
| Período | 1989–2024 |  |
| Idioma | Todos |  |
| Metadados | Completos (título, autores, palavras-chave, fonte, citações) | Metadados ausentes ou incompletos |
| Relevância temática | IA, Gestão de risco e saúde (*healthcare*) | Publicações sem aderência conceitual |

Fonte: o Autor

* + 1. PROCESSO DE SELEÇÃO

A seleção dos documentos seguiu uma abordagem sistemática voltada à construção de um *corpus* bibliográfico coerente com os objetivos do *science mapping*. O processo foi conduzido de forma a assegurar a representatividade temática, a qualidade bibliográfica e a integridade dos metadados, conforme diretrizes propostas por Donthu *et al.* (2021). A seleção final incluiu apenas documentos com metadados completos e alinhamento conceitual, preservando a integridade e a representatividade do *corpus* para as análises subsequentes.

* 1. EXTRAÇÃO E PRÉ-PROCESSAMENTO DOS DADOS

A preparação do *corpus* para a análise bibliométrica envolveu a aplicação de procedimentos técnicos articulados em três subetapas sequenciais: (i) exportação dos metadados e análise de afiliações institucionais; (ii) normalização de formatos; e (iii) filtragem e limpeza semântica para assegurar a robustez do *corpus*. Cada uma dessas etapas foi conduzida para garantir a integridade estrutural dos metadados, a uniformização dos registros e a coerência semântica necessária às análises no VOSviewer, conforme orientações metodológicas de Van Eck e Waltman (2010).

* + 1. EXPORTAÇÃO DOS METADADOS E ANÁLISE DE AFILIAÇÕES

Os registros bibliográficos foram extraídos das bases *Scopus* e *Compendex* em formato CSV, preservando os campos essenciais à análise de coocorrência.

* + 1. Normalização de formatos

Realizou-se a padronização sintática dos arquivos, com ajuste dos delimitadores e harmonização estrutural dos campos, a fim de viabilizar sua ingestão técnica no VOSviewer e garantir consistência nos dados processados.

* + 1. LIMPEZA SEMÂNTICA

A consolidação semântica do *corpus* envolveu a construção de um *thesaurus* personalizado, desenvolvido com base em análise lexical iterativa. A curadoria do *thesaurus* foi realizada em colaboração entre autor e orientador, com validação em amostras parciais do *corpus* para assegurar estabilidade e coerência terminológica.

1. **RESULTADOS E DISCUSSÃO**

Esta seção apresenta os resultados da análise bibliométrica, detalhando o percurso metodológico desde a construção do *corpus* documental até a geração dos mapas de coocorrência semântica.

* 1. CONSTRUÇÃO DO CORPUS DOCUMENTAL

A construção do corpus bibliográfico baseou-se nos critérios metodológicos descritos na Seção 3, com foco na rastreabilidade e na qualidade da base empírica. O objetivo foi selecionar publicações relevantes à governança de riscos associados à IA na saúde, com ênfase em segurança clínica, cibersegurança e conformidade regulatória. As subseções a seguir descrevem esse processo de forma detalhada, assegurando transparência e rastreabilidade na formação da base empírica utilizada.

* + 1. ESTRATÉGIA DE BUSCA E COLETA

A busca foi realizada entre 20 e 23 de maio de 2025 nas bases Scopus (321 registros) e Compendex (587 registros), totalizando 908 documentos. Adotou-se como filtro a língua inglesa, priorizando publicações de maior visibilidade e impacto internacional, conforme recomendações metodológicas de Donthu *et al.* (2021. Os registros foram exportados em formato CSV com os principais metadados bibliográficos.

* + 1. DEDUPLICAÇÃO E CRITÉRIOS DE ELEGIBILIDADE

A partir dos 908 registros inicialmente coletados nas bases *Scopus* e *Compendex*, foi realizada a deduplicação manual com auxílio do Microsoft Excel. Em seguida, aplicaram-se os critérios de elegibilidade descritos na Seção 3.2.2, resultando em um *corpus* final de 252 documentos, sendo 203 provenientes da *Scopus* e 49 da *Compendex*.

Seguindo as recomendações de Donthu *et al.* (2021) para revisões bibliométricas, toda a trajetória de construção do *corpus* foi documentada de forma transparente e rastreável. O processo de filtragem adotou quatro fases: identificação, triagem, elegibilidade e inclusão, tal como preconizado pelos autores para assegurar reprodutibilidade e clareza metodológica em estudos de *science mapping*. A Tabela 1 sintetiza as etapas do processo de triagem e seleção documental.

Tabela 1 – Etapas da seleção

|  |  |  |
| --- | --- | --- |
| Etapa | n | Descrição |
| Registros identificados | 908 | Total recuperado entre 20 e 23/05/2025 |
| Registros duplicados removidos | 146 | Remoção manual e por planilha Excel |
| Registros triados | 762 | Títulos e resumos avaliados conforme os critérios de inclusão |
| Registros excluídos | 510 | Por tipo documental. (*Articles in Press*, *Conference Papers*, *Proceedings*, *Reviews* de conferência.) |
| Relatórios avaliados para elegibilidade | 252 | Documentos elegíveis segundo critérios de elegibilidade. (*Article*, *Journal Article*, *Review)* |
| Relatórios excluídos | 0 | Todos os documentos atendiam aos critérios definidos |
| Estudos incluídos na revisão | 252 | *Corpus* final: 203 documentos da *Scopus* e 49 da *Compendex* |

Fonte: Adaptado pelo autor com base em Donthu *et al.* (2021)

A Figura 1 ilustra a distribuição dos documentos por tipo e base, destacando a predominância de artigos revisados por pares. Essa escolha metodológica visou assegurar a robustez e a confiabilidade do *corpus*.

Gráfico, Gráfico de explosão solar

O conteúdo gerado por IA pode estar incorreto.

Figura 1: Distribuição dos documentos por tipo e base de dados

Fonte: o Autor

* + 1. PANORAMA QUANTITATIVO E QUALITATIVO DA PRODUÇÃO CIENTÍFICA

O corpus final, composto por 252 documentos, apresentou distribuição predominante de *journal articles* (76%) e *reviews* (24%), conforme a Tabela 2. A expressiva presença (41%) de revisões sugere um campo em consolidação conceitual, com esforços voltados à sistematização do conhecimento (Donthu *et al.*, 2021).

Tabela 2: Distribuição dos documentos por base de dados e tipo de publicação

|  |  |  |  |
| --- | --- | --- | --- |
| Base de Dados | *Journal Article* | *Review* | Total |
| *Scopus* | 144 | 59 | 203 |
| *Compendex* | 49 | - | 49 |

Fonte: o Autor

Não houve restrição temporal na coleta, permitindo observar um crescimento acentuado a partir de 2020, com destaque para os anos de 2023 a 2025, como ilustrado na Figura 2. Tal tendência sinaliza a crescente atenção internacional às questões de risco, regulação e segurança da IA em saúde.

Gráfico, Gráfico de barras

O conteúdo gerado por IA pode estar incorreto.

Figura 2: Distribuição de publicações por ano de publicação

Fonte: o Autor

Além da dimensão temporal, a análise geográfica da produção científica, conforme observa-se na Figura 3, indica uma concentração significativa nos Estados Unidos (12,9%) e na China (12,3%), que juntas respondem por 25,5% do total de publicações sobre governança de riscos e IA em saúde. Este protagonismo pode ser atribuído ao volume de investimentos em pesquisa, desenvolvimento tecnológico e políticas regulatórias específicas para IA nesses países. Outras nações com participação expressiva incluem Japão, Itália, Índia e Alemanha, todas com percentuais superiores a 5%. O Brasil figura com apenas 2,7% das publicações, o que revela uma contribuição ainda incipiente nesse campo estratégico e ressalta a necessidade de fortalecer iniciativas institucionais e acadêmicas nacionais, especialmente voltadas à governança da IA em saúde.

Gráfico, Gráfico de mapa de árvore

O conteúdo gerado por IA pode estar incorreto.

Figura 3: Distribuição de publicações por país de origem

Fonte: o Autor

Esses resultados evidenciam um campo em expansão, dominado por centros de pesquisa consolidados. A etapa seguinte da análise concentrou-se na padronização lexical do *corpus*, por meio de *thesaurus* personalizado, resultando na identificação de 43 termos-chave com frequência mínima de 10 ocorrências. Esses termos fundamentaram a geração dos mapas de coocorrência, detalhados na próxima seção.

* 1. PROCEDIMENTOS ANALÍTICOS DE COOCORRÊNCIA

A análise semântica do *corpus* foi conduzida no software VOSviewer (Van Eck; Waltman, 2010), reconhecido por sua robustez na construção e interpretação de redes de coocorrência. O estudo concentrou-se exclusivamente nas palavras-chave (*all keywords*), com o objetivo de identificar núcleos conceituais, lacunas temáticas e padrões evolutivos da literatura sobre governança de riscos relacionados à IA no setor da saúde.

Para assegurar coesão terminológica e precisão lexical, foi construído um *thesaurus* personalizado, desenvolvido iterativamente pelo autor em conjunto com o orientador. Esse processo incluiu:

* Unificação de sinônimos e variantes ortográficas, como *cybersecurity* e *cyber security*;
* Remoção de termos genéricos e não analíticos, como *study*, *article*, *male*, *adult*;
* Preservação seletiva de termos amplos, como *disease*, mantidos apenas em expressões compostas relevantes (ex.: *disease prediction*).

A aplicação do *thesaurus* garantiu a consistência semântica necessária para as etapas posteriores de análise de coocorrência, densidade e agrupamento temático. A comparação entre os mapas gerados com e sem padronização é apresentada nas seções seguintes.

A análise de coocorrência considerou todas as palavras-chave com frequência mínima de 10 ocorrências, resultando em 43 termos qualificados. O método de normalização adotado foi o *Association Strength*, por sua capacidade de evidenciar relações proporcionais mesmo em distribuições assimétricas.

* 1. RESULTADOS ESTRUTURAIS DA ANÁLISE SEMÂNTICA

A partir do *corpus* refinado e da padronização semântica realizada, foi possível aplicar a análise de coocorrência de palavras-chave no VOSviewer, resultando em uma representação estruturada dos principais conceitos que compõem a literatura sobre governança de riscos em IA na saúde. Esta seção apresenta os resultados obtidos em três níveis complementares: (i) os termos mais frequentes e sua centralidade no campo; (ii) os clusters temáticos resultantes da análise de agrupamento; e (iii) as visualizações estruturais e temporais que permitem observar a organização conceitual e a evolução temática ao longo do tempo.

* + 1. TERMOS MAIS FREQUENTES

A aplicação do limiar mínimo de 10 ocorrências resultou em 43 termos qualificados. A Tabela 3 apresenta os dez mais frequentes, com destaque para *risk management* e *artificial intelligence*, que evidenciam a centralidade da discussão sobre governança e tecnologia em ambientes clínicos. Termos como *cybersecurity*, *ethics* e *privacy* refletem preocupações crescentes com responsabilidade algorítmica e proteção de dados sensíveis (Khalifa; Househ, 2021; Borycki; Kushniruk, 2021). Já a presença de expressões como *decision support system* e *electronic health record* sinaliza o papel prático da IA no suporte à decisão clínica e na automação de fluxos assistenciais.

Tabela 3: Termos mais frequentes no *corpus* analisado

|  |  |
| --- | --- |
| Termo | Ocorrências |
| risk management | 2.020 |
| artificial intelligence | 1.860 |
| machine learning | 800 |
| healthcare | 720 |
| decision making | 420 |
| decision support system | 340 |
| electronic health record | 300 |
| Ethics | 200 |
| cybersecurity | 130 |
| Privacy | 120 |

Fonte: o Autor

* + 1. CLUSTERS TEMÁTICOS

A análise de coocorrência revelou quatro clusters temáticos principais, apresentados na Tabela 4, a partir dos termos extraídos com uso do *thesaurus* e organizados pelo método de *association strength* no VOSviewer (Van Eck; Waltman, 2010). Esses agrupamentos representam quatro clusters temáticos conceituais inter-relacionados na literatura.

Tabela 4: Tópicos centrais por cluster temático identificado

|  |  |  |  |
| --- | --- | --- | --- |
| Cluster | Foco | Termos-Chave | Frequência |
| 1 | Ecossistema clínico e sistemas de saúde | *healthcare*, *patient safety*, *patient care*, *telemedicine* | 2710 |
| 2 | Práticas clínicas e governança de dados | *electronic health record*, *clinical practice*, *ethics* | 1850 |
| 3 | IA e riscos técnicos | *machine learning*, *deep learning*, *cybersecurity* | 2150 |
| 4 | Estratégias de governança e decisão em IA | *risk management*, *artificial intelligence*, *decision making* | 5160 |

Fonte: o Autor

O Cluster 1 agrupa termos ligados ao ecossistema clínico e aos sistemas de cuidado, como *healthcare* e *patient safety*, refletindo preocupações com a infraestrutura assistencial e a segurança do paciente.

O Cluster 2 enfatiza práticas clínicas e governança de dados, reunindo conceitos como *electronic health record*, *privacy* e *ethics*, que expressam a tensão entre inovação tecnológica e exigências regulatórias.

O Cluster 3 concentra os aspectos técnicos e computacionais da IA, com ênfase em ferramentas como *machine learning*, *prediction* e *cybersecurity*. A coocorrência desses termos revela uma preocupação emergente com os riscos algorítmicos, como ataques adversariais, falhas de explicabilidade e vulnerabilidades de segurança que afetam diretamente a confiabilidade das aplicações clínicas baseadas em IA.

Por fim, o Cluster 4 representa o eixo estratégico da literatura, estruturado em torno de conceitos como *risk management* e *decision support system*. Este agrupamento sintetiza os esforços voltados à formulação de modelos de governança que conciliem inovação tecnológica, requisitos clínicos e demandas regulatórias, estando diretamente alinhado ao escopo deste estudo.

Essa organização conceitual fornece uma visão integrada do campo, permitindo visualizar tanto convergências temáticas quanto lacunas estruturais ainda não resolvidas. As visualizações a seguir complementam essa análise com representações estruturais e temporais das redes semânticas geradas.

* + 1. VISUALIZAÇÕES BIBLIOMÉTRICAS

Com o objetivo de ilustrar a estrutura conceitual e a evolução temática da literatura, foram geradas duas visualizações bibliométricas no VOSviewer: um mapa estrutural de coocorrência e um mapa temporal de palavras-chave, ambos baseados no corpus refinado, com aplicação de *thesaurus* e limiar mínimo de 10 ocorrências.

Testes preliminares sem padronização semântica revelaram dispersão conceitual, sobreposição temática e predominância de termos genéricos, como *article* e *review* (Figura 4).

Gráfico, Gráfico de dispersão

O conteúdo gerado por IA pode estar incorreto.

Figura 4: Mapa de coocorrência sem aplicação de *thesaurus*

Fonte: o Autor

A Figura 5 complementa essa análise ao mostrar a distribuição temporal de termos com base no ano médio de ocorrência, destacando a transição de temas estruturais (ex.: *clinical decision support system*) para tópicos emergentes como *chatGPT*, *adversarial machine learning* e *medical device regulation*, a partir de 2022.

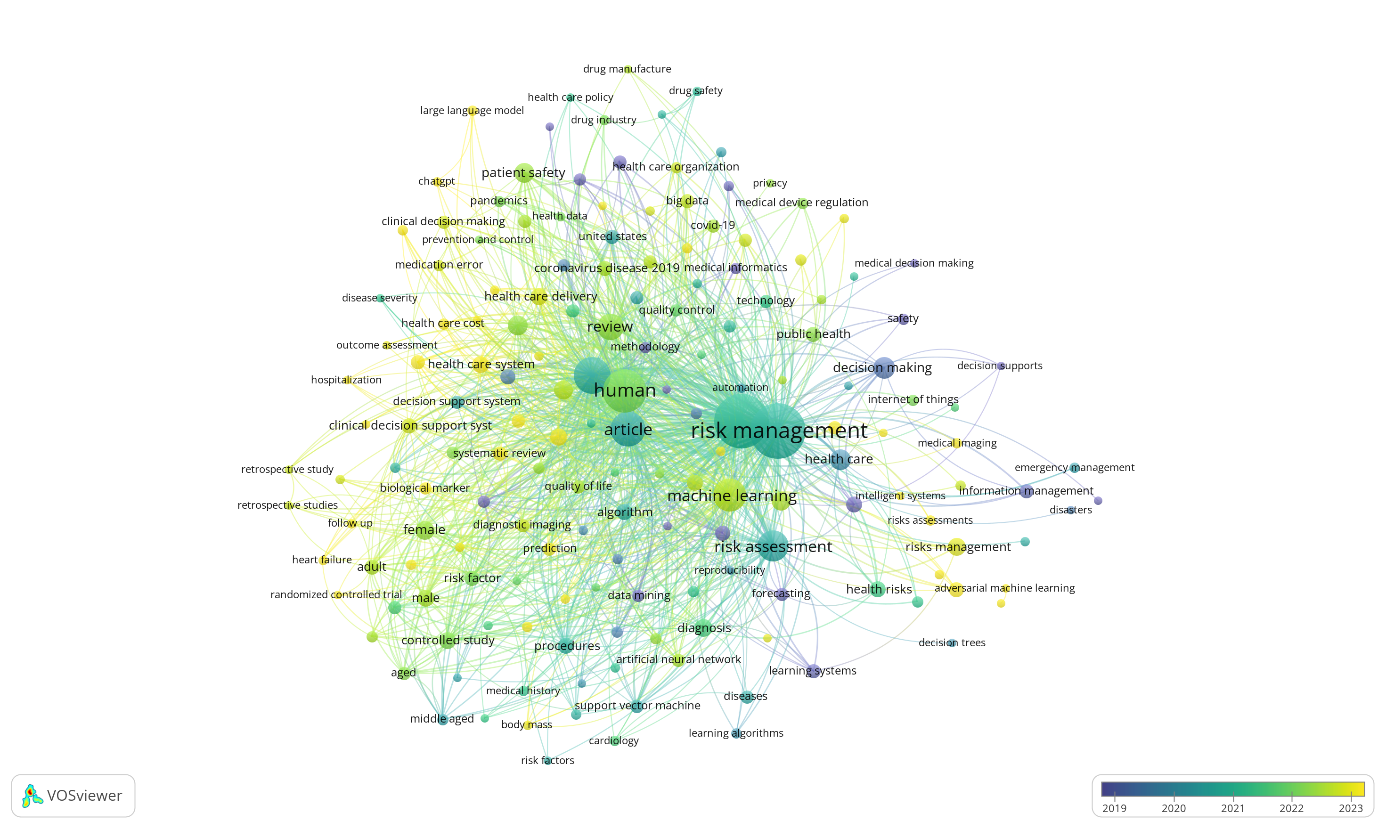
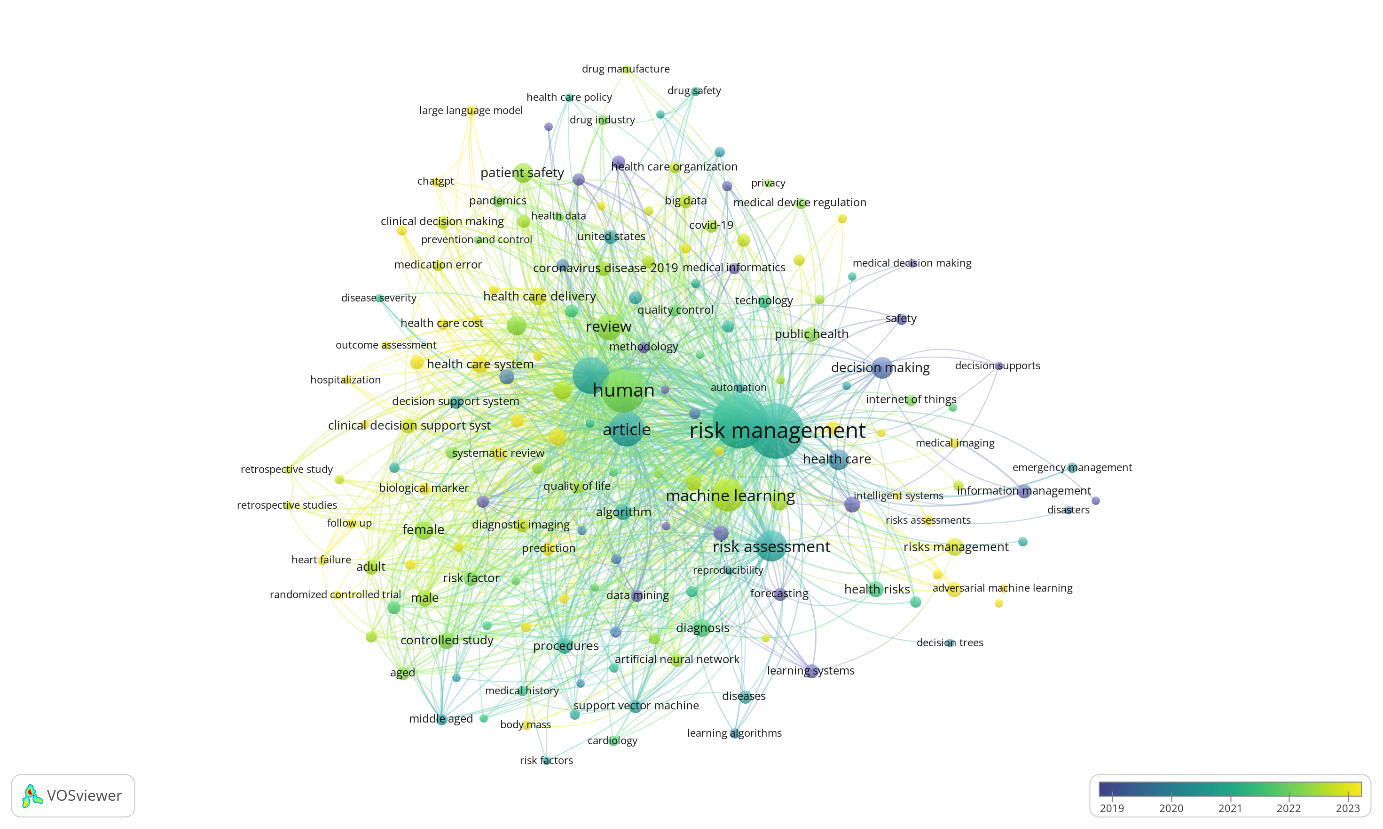


Figura 5: Mapa de temporal sem aplicação de *thesaurus*

Fonte: o Autor

Essa evolução sugere um amadurecimento do campo em direção à integração de tecnologias de IA, segurança da informação e padrões regulatórios em contextos clínicos, alinhando-se às transformações regulatórias recentes e à crescente complexidade dos riscos associados. A curadoria semântica permitiu consolidar 43 termos representativos e gerar visualizações mais coesas. Com a aplicação do *thesaurus* e a elevação do limiar para 10 ocorrências por termo, obteve-se uma base mais estável e refinada, resultando em 43 termos qualificados. O mapa de coocorrência estrutural (Figura 6) passou a evidenciar quatro agrupamentos conceituais distintos, refletindo com clareza os domínios temáticos identificados: clínico, técnico, regulatório e estratégico. A dimensão dos nós é proporcional à frequência dos termos, e as cores indicam a associação de cada termo ao seu respectivo cluster.

Diagrama

O conteúdo gerado por IA pode estar incorreto.

Figura 6: Mapa estrutural de coocorrência: clusters temáticos

Fonte: o Autor

Complementarmente, o mapa temporal de coocorrência (Figura 7) apresenta a média de ano de publicação dos documentos em que cada termo aparece, permitindo identificar tópicos consolidados e temas emergentes ao longo do tempo. Termos como *risk management* e *patient safety* predominam em tons azul-esverdeados, indicando maior recorrência em anos anteriores (2019–2021), enquanto expressões como *explainable AI*, *cybersecurity* e *data privacy* aparecem em tons amarelados, sinalizando sua ascensão mais recente (2022–2025).

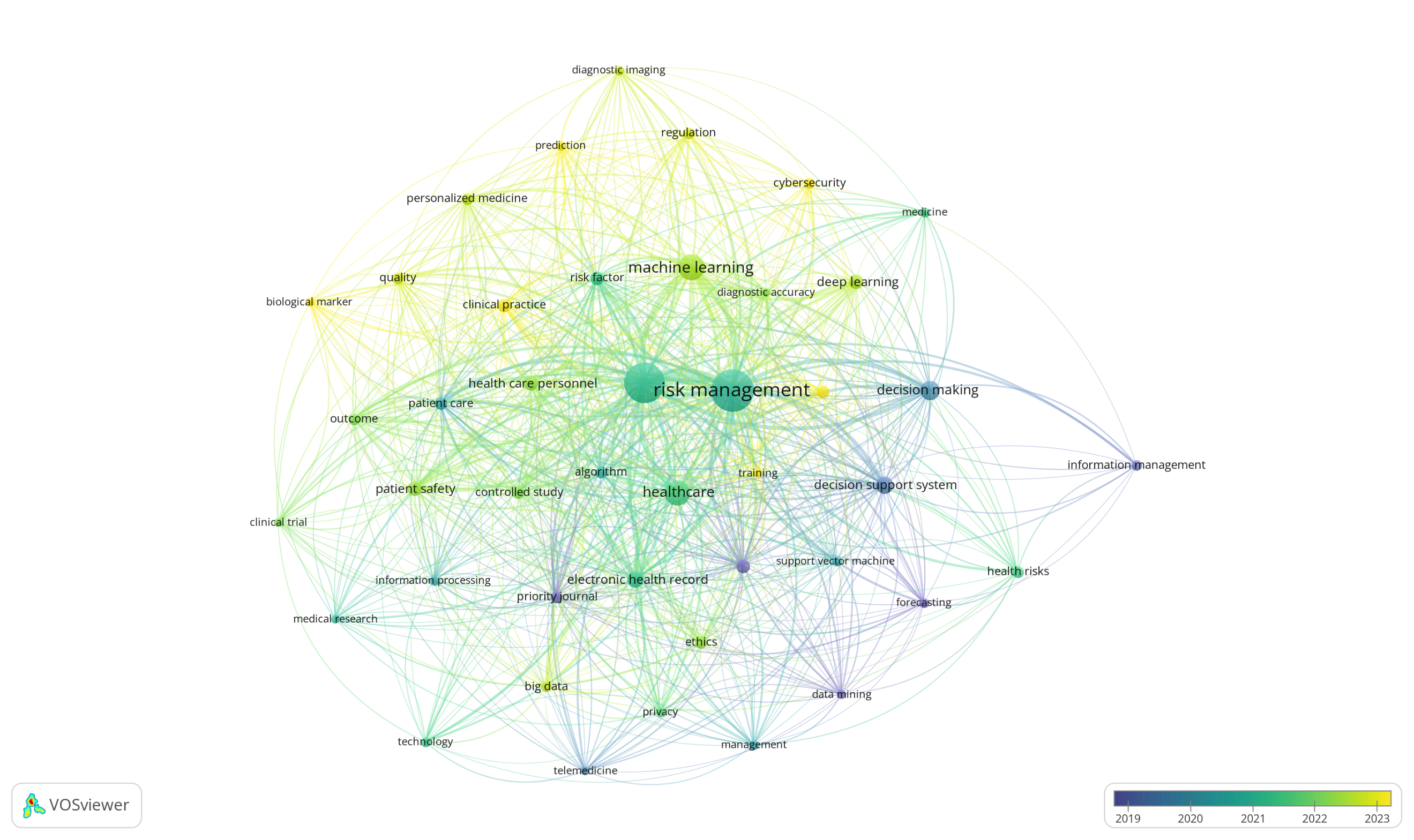
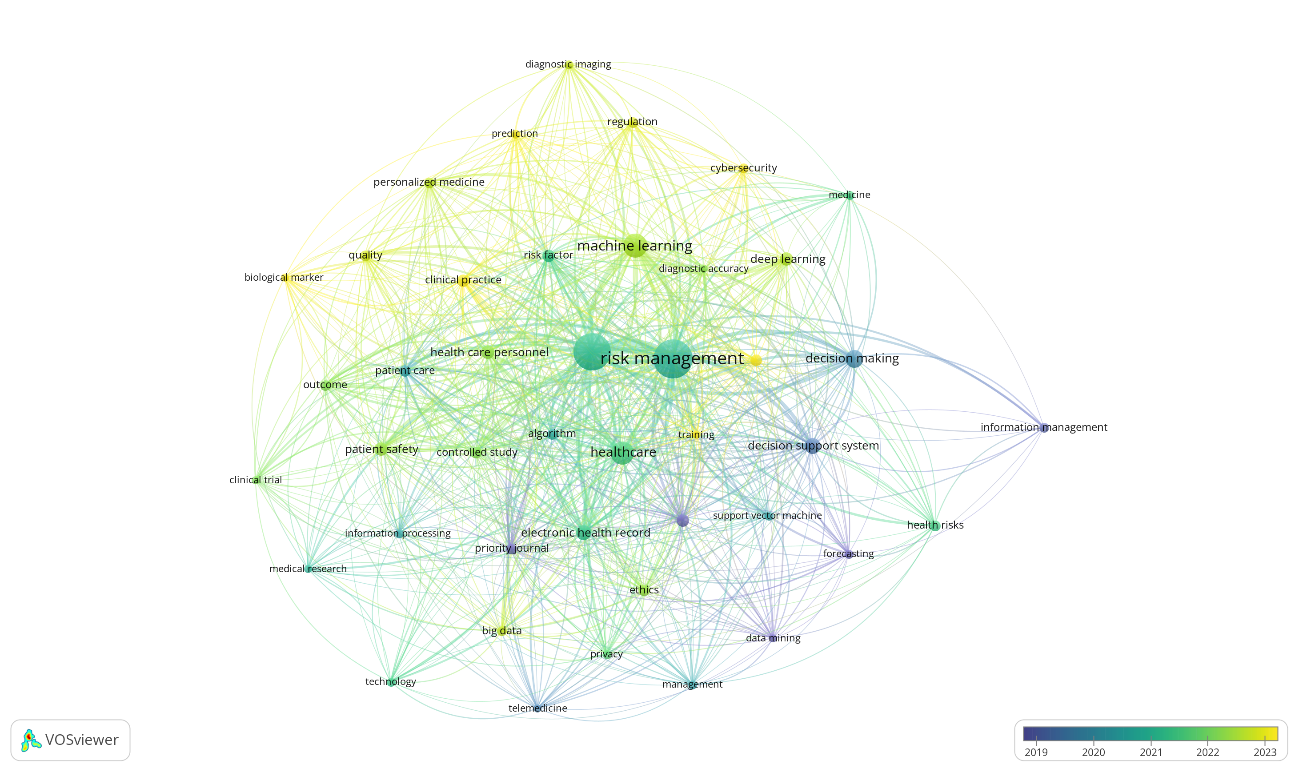


Figura 7: Mapa temporal de coocorrência: tendências emergentes

Fonte: o Autor

Essas representações reforçam o amadurecimento do campo em direção à integração entre inovação algorítmica, governança regulatória e práticas clínicas, com destaque para temas emergentes em segurança, transparência e conformidade.

* + 1. SÍNTESE ANALÍTICA DOS CLUSTERS

A análise semântica permitiu identificar quatro clusters temáticos que articulam dimensões clínicas, técnicas, regulatórias e estratégicas da governança da IA na saúde. O Quadro 3 resume os focos de cada agrupamento e seus termos representativos.

Quadro 3: Caracterização dos clusters temáticos

|  |  |  |
| --- | --- | --- |
| Cluster | Foco | Exemplos de termos |
| 1 | Ecossistema clínico e sistemas de saúde | *healthcare, patient safety, telemedicine, training* |
| 2 | Práticas clínicas e governança de dados | *electronic health record, clinical practice, privacy, ethics* |
| 3 | IA e riscos técnicos | *machine learning, cybersecurity, regulation, prediction* |
| 4 | Estratégias de governança e decisão em IA | *risk management, decision making, artificial intelligence, information management* |

Fonte: o Autor

O Cluster 1 refere-se ao ecossistema clínico e à infraestrutura assistencial, com destaque para *healthcare*, *patient safety* e *telemedicine*. Esses termos refletem os desafios da digitalização em ambientes hospitalares marcados por variabilidade tecnológica e maturidade cibernética desigual (Jalali; Kaiser, 2018).

O Cluster 2 aborda práticas clínicas e governança de dados, reunindo *electronic health record*, *privacy* e *ethics*. As tensões entre inovação, responsabilidade e conformidade regulatória são centrais nesse agrupamento, especialmente diante da crescente exigência por sistemas transparentes e auditáveis (Kritharidou et al., 2023; Al-Maamari, 2025).

O Cluster 3 enfoca os riscos técnicos da IA, como ataques adversariais, falhas de explicabilidade e desafios de interoperabilidade, refletidos em termos como *machine learning*, *cybersecurity* e *prediction* (Radanliev *et al.*, 2021; Borycki; Kushniruk, 2021).

O Cluster 4 representa o núcleo estratégico da literatura, centrado em *risk management* e *artificial intelligence*, funcionando como eixo transversal entre os demais domínios. Ainda assim, modelos como o NIST AI RMF (NIST AI RMF, 2023) revelam limitações quando aplicados a ambientes clínicos de baixa maturidade digital. Essa limitação, discutida de forma ampliada na Seção 2.3, reforça a necessidade de abordagens setoriais mais sensíveis à complexidade da saúde carecem de sensibilidade setorial, sobretudo para aplicações críticas (Al-Maamari, 2025).

A configuração desses clusters revela que, apesar dos avanços, a literatura segue fragmentada entre os domínios técnico, ético e assistencial. Persistem lacunas na integração entre esses campos, o que reforça a necessidade de modelos mais contextualizados e adaptativos à realidade operacional da saúde. Tais achados fundamentam a discussão apresentada nas considerações finais deste estudo. A seguir, as conclusões do estudo sintetizam esses achados e apontam direções para aprofundamentos conceituais e futuras investigações aplicadas sobre a governança da IA em saúde.

1. **CONSIDERAÇÕES FINAIS E IMPLICAÇÕES**

Esta revisão bibliométrica mapeou a produção científica recente sobre governança e gestão de riscos associados à inteligência artificial (IA) no setor da saúde. A partir da análise de 252 publicações indexadas nas bases Scopus e Compendex, com apoio do software VOSviewer e de um *thesaurus* semântico, identificaram-se quatro clusters conceituais que estruturam o campo: (i) ecossistema clínico e sistemas de cuidado; (ii) práticas clínicas e governança de dados; (iii) IA e riscos técnicos; e (iv) estratégias de governança e decisão em IA. Esses agrupamentos refletem os principais eixos temáticos da literatura, como segurança clínica, cibersegurança, transparência algorítmica e conformidade regulatória.

A contribuição central deste trabalho reside na organização empírica da literatura por meio de núcleos conceituais interligados, revelando lacunas estruturantes que dificultam a consolidação de modelos de governança eficazes no setor. A análise demonstrou que o conhecimento científico permanece fragmentado entre os domínios clínico, técnico e normativo. Termos como *risk management*, *cybersecurity* e *ethics* são recorrentes, enquanto conceitos fundamentais para a confiabilidade algorítmica, como *explainable AI*, *human-in-the-loop* e *clinical validation*, aparecem com baixa frequência. Essa assimetria compromete a articulação entre inovação e responsabilidade, evidenciando a ausência de abordagens integradas capazes de responder à complexidade dos ambientes hospitalares. Como indicam estudos anteriores, como o de D’Ascenzo *et al.* (2024), a falta de articulação entre regulação, prática clínica e riscos algorítmicos configura uma limitação persistente.

No plano prático, constata-se que os frameworks mais utilizados, como o NIST AI RMF, o AI Act da União Europeia e a ISO 42001, ainda carecem de mecanismos sensíveis ao contexto assistencial, sendo pouco eficazes em ambientes com infraestrutura heterogênea e baixa maturidade digital, como os encontrados em muitos hospitais no Brasil. Essa limitação tem implicações diretas para a gestão organizacional: a adoção de IA sem diretrizes robustas pode amplificar riscos operacionais, favorecer decisões clínicas automatizadas sem supervisão adequada e comprometer a segurança do paciente. Nesse contexto, os achados deste estudo oferecem subsídios relevantes para o desenvolvimento de políticas institucionais voltadas à transparência, validação clínica e auditoria contínua, contribuindo para a qualificação da tomada de decisão e para o fortalecimento da governança interna em organizações de saúde.

Entre as limitações da pesquisa, destaca-se seu caráter teórico e exploratório, baseado exclusivamente em análise bibliométrica de publicações indexadas. Não foram conduzidas investigações empíricas, estudos de caso ou validações em campo. Frente a esse desafio, propõe-se o desenvolvimento de um modelo modular e setorial de governança da IA, que traduza diretrizes globais em práticas clínicas viáveis, incluindo mecanismos como validação contextual, auditorias algorítmicas e protocolos de responsabilização. Como agenda de pesquisa futura, recomenda-se a aplicação de métodos mistos, como *topic modeling*, estudos qualitativos e abordagens participativas, para aprofundar os eixos aqui mapeados. Também se sugere o desenvolvimento e teste de frameworks setoriais validados em contextos hospitalares reais, com a participação ativa de profissionais da saúde, gestores, especialistas em IA e reguladores.

Os achados desta revisão reforçam a urgência de modelos de governança que não apenas assegurem conformidade normativa, mas também promovam equidade, segurança e viabilidade operacional. Diferentemente de estudos anteriores, esta pesquisa contribui ao sistematizar o campo com base em evidências empíricas e ao propor caminhos concretos para a construção de soluções regulatórias adaptadas à realidade da saúde digital.

1. **REFERÊNCIAS**

AL-MAAMARI, A. **Between innovation and oversight: a cross-regional study of AI risk management frameworks in the EU, U.S., UK, and China**. arXiv preprint, 2025. Disponível em: <https://arxiv.org/abs/2503.05773>.

BOONSTRA, A., LAVEN, M. **Influence of artificial intelligence on the work design of emergency department clinicians a systematic literature review.** BMC Health Serv Res 22, 669 (2022). <https://doi.org/10.1186/s12913-022-08070-7>.

BORYCKI, E. M.; KUSHNIRUK, A. W. **The safety of AI in healthcare: emerging issues and considerations for healthcare.** In: HOUSEH, M.; BORYCKI, E.; KUSHNIRUK, A. (ed.). Multiple perspectives on artificial intelligence in healthcare. Cham: Springer, 2021. p. 15–30. DOI: <https://doi.org/10.1007/978-3-030-67303-1_2>.

BOHR, A.; MEMARZADEH, K. **The rise of artificial intelligence in healthcare applications.** In: BOHR, A.; MEMARZADEH, K. (ed.). Artificial intelligence in healthcare. Cambridge, MA: Academic Press, 2020. p. 25–60. DOI: <https://doi.org/10.1016/B978-0-12-818438-7.00002-2>.

COSO (Committee of Sponsoring Organizations of the Treadway Commission). (2017). **Enterprise Risk Management** - Integrating with Strategy and Performance.

D’ASCENZO, F. *et al.* **Artificial intelligence in healthcare: a bibliometric and topic modeling analysis**. Discover Artificial Intelligence, [S. l.], v. 4, n. 1, p. 1–15, 2024. DOI: <https://doi.org/10.1016/j.sftr.2024.100198>.

DONTHU, N. *et al.* **How to conduct a bibliometric analysis: An overview and guidelines**. Journal of Business Research, v. 133, p. 285–296, 2021. DOI: <https://doi.org/10.1016/j.jbusres.2021.04.070>.

ISO 27001. **ABNT NBR ISO/IEC 27001:2022** – Tecnologia da informação — Segurança da informação, cibersegurança e proteção da privacidade — Sistemas de gestão de segurança da informação — Requisitos. Rio de Janeiro: Associação Brasileira de Normas Técnicas, 2023.

ISO 31000. **ABNT NBR ISO 31000:2018** – Gestão de riscos — Diretrizes. Rio de Janeiro: Associação Brasileira de Normas Técnicas, 2018.

ISO 42001. **ABNT NBR ISO/IEC 42001:2024** – Tecnologia da informação — Inteligência artificial — Sistema de gestão. Rio de Janeiro: Associação Brasileira de Normas Técnicas, 2024.

JALALI, M. S.; KAISER, J. P. **Cybersecurity in hospitals: a systematic, organizational perspective.** Journal of Medical Internet Research, v. 20, n. 5, e10059, 2018. Disponível em: <https://doi.org/10.2196/10059>.

KHALIFA, M.; HOUSEH, M. **Utilizing health analytics in improving the performance of hospitals and healthcare services: promises and challenges.** In: HOUSEH, M.; BORYCKI, E.; KUSHNIRUK, A. (ed.). Multiple perspectives on artificial intelligence in healthcare. Cham: Springer, 2021. p. 31–48. DOI: <https://doi.org/10.1007/978-3-030-67303-1_3>.

KRITHARIDOU, M. *et al.* **Ethicara for Responsible AI in Healthcare: A System for Bias Detection and AI Risk Management**. [S.l.: s.n.], 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/39435256/>.

LGPD. **Lei nº 13.709**, de 14 de agosto de 2018. Dispõe sobre a proteção de dados pessoais e altera a Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet). Diário Oficial da União: seção 1, Brasília, DF, ano CLVI, n. 157, p. 1–5, 15 ago. 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>.

NIST AI RMF. **Artificial Intelligence Risk Management Framework** (AI RMF 1.0). Gaithersburg, MD: NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 2023. Disponível em: <https://doi.org/10.6028/NIST.AI.100-1>.

NIST. **Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile**. Gaithersburg, MD: NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY, 2024. (NIST AI 600-1). Disponível em: <https://doi.org/10.6028/NIST.AI.600-1>.

OECD (2023), **The state of implementation of the OECD AI Principles four years on**, OECD Artificial Intelligence Papers, No. 3, OECD Publishing, Paris, <https://doi.org/10.1787/835641c9-en>.

OECD (2024), **Governing with Artificial Intelligence: Are governments ready?**, OECD Artificial Intelligence Papers, No. 20, OECD Publishing, Paris, <https://doi.org/10.1787/26324bc2-en>.

OBERMEYER, Z.; POWERS, B.; VOGELI, C.; MULLAINATHAN, S. **Dissecting racial bias in an algorithm used to manage the health of populations**. Science, v. 366, n. 6464, p. 447-453, 2019. Disponível em: <https://doi.org/10.1126/science.aax2342>.

PERNEGER, T. V. **The Swiss cheese model of safety incidents: are there holes in the metaphor?** BMC Health Services Research, v. 5, n. 71, 2005. DOI: <https://doi.org/10.1186/1472-6963-5-71>.

RADANLIEV, P. *et al.* **Artificial intelligence in cyber-physical systems.** AI & Society, v. 36, n. 3, p. 783–796, 2021. DOI: <https://doi.org/10.1007/s00146-020-01049-0>.

REDDY, S. *et al.* **A Governance Model for the Application of AI in Healthcare.** Journal of the American Medical Informatics Association, v. 27, n. 3, p. 491-497, 2020. DOI: <https://doi.org/10.1093/jamia/ocz192>.

VAN ECK, N. J.; WALTMAN, L. **Software survey: VOSviewer, a computer program for bibliometric mapping.** Scientometrics, v. 84, p. 523–538, 2010. DOI: <https://doi.org/10.1007/s11192-009-0146-3>.

VINCENT, C.; AMALBERTI, R. **Safer healthcare: strategies for the real world**. Cham: Springer, 2016. Disponível em: <https://doi.org/10.1007/978-3-319-25559-0>.