

SEGURANÇA DO PACIENTE NA ERA DA INTELIGÊNCIA ARTIFICIAL: OPORTUNIDADES, DESAFIOS ÉTICOS E PERSPECTIVAS FUTURAS

PATIENT SAFETY IN THE ERA OF ARTIFICIAL INTELLIGENCE: OPPORTUNITIES,
ETHICAL CHALLENGES AND FUTURE PERSPECTIVES

SEGURIDAD DEL PACIENTE EN LA ERA DE LA INTELIGENCIA ARTIFICIAL:
OPORTUNIDADES, DESAFÍOS ÉTICOS Y PERSPECTIVAS FUTURAS

Vivian Batista Lira Sobral¹
Lília da Silva Melo²
Itamar dos Santos Fonseca³
Marialdo Dias Barroso Mendonça⁴
Edivaldo Bazilio dos Santos⁵
Paula de Vasconcelos Pinheiro⁶
Josyane Rebouças da Silva⁷
Jacineidy de Moura Sousa⁸
Ana Laura Moraes Silveira⁹
Paula Lyra Dias de Oliva¹⁰
Rafael Rodrigues Gonçalves¹¹
Rafaelle Serejo Machado¹²

RESUMO: **Objetivo:** sintetizar criticamente as melhores evidências científicas sobre a aplicação da inteligência artificial (IA) à segurança do paciente, analisando seu potencial para a prevenção de eventos adversos, os riscos e desafios éticos decorrentes de sua incorporação e as perspectivas futuras para a prática assistencial e a governança dos sistemas de saúde. **Método:** revisão integrativa conduzida segundo o referencial de Whittemore e Knafl e reportada conforme a recomendação PRISMA 2020, com busca em oito bases (PubMed/MEDLINE, Embase, Scopus, Web of Science, CINAHL, LILACS, SciELO e Biblioteca Virtual em Saúde), descritores DeCS/MeSH combinados por operadores booleanos, recorte de 2020 a 2026, seleção por dois revisores independentes, avaliação do nível de evidência (Melnik e Fineout-Overholt) e do risco de viés por instrumento específico para modelos preditivos (PROBAST+AI), complementado pelas lentes de relato CONSORT-AI e TRIPOD+AI. **Resultados:** foram reunidos 20 estudos (2020–2026), predominando estudos preditivos de desenvolvimento/validação e revisões de escopo, com aplicações em predição de sepse e deterioração, segurança medicamentosa, prevenção de lesões por pressão e quedas, vigilância de eventos adversos por processamento de linguagem natural e apoio à decisão. As evidências mostram benefícios condicionados à qualidade da implementação e recorrentes limitações: validação apenas interna, desfechos substitutos, alarmes falsos, viés de automação e iniquidades algorítmicas. **Considerações finais:** a IA oferece oportunidades concretas para a prevenção de eventos adversos, mas seus benefícios dependem de

1

¹Pós-graduação em enfermagem em Gestão da Qualidade e Segurança do Paciente, Fahol.

²Mestrado em Gestão em cuidados de Saúde, Must University.

³Pós-graduação em Gestão Pública, UFPA.

⁴Mestrado em Gestão em Saúde, UECE.

⁵Mestrando Internacional em Gestão em Saúde, Lean Healthcare.

⁶Mestrado em Gestão em Saúde, UECE.

⁷Mestrado em Gestão em Saúde, UECE.

⁸Mestrando em Saúde Coletiva em área de concentração em Informação e Saúde Digital, ISC-UFBA.

⁹Graduanda em Medicina, UNIMES.

¹⁰Graduanda em Medicina, UNIMES.

¹¹Mestrando em Gestão de Cuidados da Saúde, Must University.

¹²Pós-graduação em enfermagem do trabalho, Centro Universitário internacional SIGNORELLI.

validação externa, supervisão humana significativa, governança e monitoramento contínuo, cabendo à enfermagem e à equipe multiprofissional papel central como salvaguarda contra o viés de automação.

Descritores: Segurança do Paciente. Inteligência Artificial. Aprendizado de Máquina. Erros Médicos. Sistemas de Apoio a Decisões Clínicas. Cuidados de Enfermagem.

ABSTRACT: **Objective:** to critically synthesise the best scientific evidence on the application of artificial intelligence (AI) to patient safety, analysing its potential to prevent adverse events, the ethical risks and challenges arising from its incorporation, and future perspectives for clinical practice and health systems governance. **Method:** integrative review conducted according to Whittemore and Knafl and reported in accordance with PRISMA 2020, with searches across eight databases, DeCS/MeSH descriptors combined through Boolean operators, covering 2020–2026, selection by two independent reviewers, appraisal of the level of evidence (Melnik and Fineout-Overholt) and risk-of-bias assessment with a prediction-model-specific tool (PROBAST+AI), complemented by the CONSORT-AI and TRIPOD+AI reporting lenses. **Results:** twenty studies (2020–2026) were assembled, predominantly prediction-model development/validation studies and scoping reviews, addressing sepsis and deterioration prediction, medication safety, pressure injury and fall prevention, NLP-based adverse-event surveillance and decision support. Benefits were conditional on implementation quality, with recurrent limitations: internal-only validation, surrogate outcomes, false alarms, automation bias and algorithmic inequities. **Final considerations:** AI offers concrete opportunities to prevent adverse events, but its benefits depend on external validation, meaningful human oversight, governance and continuous monitoring; nursing and the multiprofessional team play a central role as a safeguard against automation bias.

Keywords: Patient Safety. Artificial Intelligence. Machine Learning. Medical Errors. Clinical Decision Support Systems. Nursing Care.

RESUMEN: **Objetivo:** sintetizar críticamente la mejor evidencia científica sobre la aplicación de la inteligencia artificial (IA) a la seguridad del paciente, analizando su potencial para la prevención de eventos adversos, los riesgos y desafíos éticos derivados de su incorporación y las perspectivas futuras para la práctica asistencial y la gobernanza de los sistemas de salud. **Método:** revisión integrativa realizada según el referencial de Whittemore y Knafl y reportada conforme a la declaración PRISMA 2020, con búsqueda en ocho bases (PubMed/MEDLINE, Embase, Scopus, Web of Science, CINAHL, LILACS, SciELO y Biblioteca Virtual en Salud), descriptores DeCS/MeSH combinados mediante operadores booleanos, período de 2020 a 2026, selección por dos revisores independientes, evaluación del nivel de evidencia (Melnik y Fineout-Overholt) y del riesgo de sesgo mediante un instrumento específico para modelos predictivos (PROBAST+AI), complementado por las lentes de reporte CONSORT-AI y TRIPOD+AI. **Resultados:** se reunieron 20 estudios (2020–2026), con predominio de estudios predictivos de desarrollo/validación y revisiones de alcance, con aplicaciones en predicción de sepsis y deterioro, seguridad de la medicación, prevención de lesiones por presión y caídas, vigilancia de eventos adversos mediante procesamiento del lenguaje natural y apoyo a la decisión. La evidencia muestra beneficios condicionados a la calidad de la implementación y limitaciones recurrentes: validación solo interna, desenlaces sustitutos, falsas alarmas, sesgo de automatización e inequidades algorítmicas. **Consideraciones finales:** la IA ofrece oportunidades concretas para la prevención de eventos adversos, pero sus beneficios dependen de la validación externa, la supervisión humana significativa, la gobernanza y el monitoreo continuo, correspondiendo a la enfermería y al equipo multiprofesional un papel central como salvaguarda frente al sesgo de automatización.

Palabras clave: Seguridad del Paciente. Inteligencia Artificial. Aprendizaje Automático. Errores Médicos. Sistemas de Apoyo a Decisiones Clínicas. Atención de Enfermería.

1 INTRODUÇÃO

A segurança do paciente consolidou-se, nas duas últimas décadas, como um dos pilares estruturantes da qualidade do cuidado em saúde e como prioridade sanitária global. A magnitude do problema é expressiva: revisão sistemática com metanálise que reuniu dados de diferentes cenários assistenciais estimou que, em média, cerca de 6% dos pacientes são expostos a danos evitáveis durante o cuidado, sendo parcela substancial classificada como de gravidade moderada a grave (Panagioti *et al.*, 2019). Reconhecendo a persistência desse ônus, a Organização Mundial da Saúde (2021a), no âmbito do Plano de Ação Global para a Segurança do Paciente 2021–2030, adotou a visão de um mundo em que ninguém seja lesado durante a assistência e todo paciente receba cuidado seguro, articulando sete objetivos estratégicos que abrangem desde políticas de eliminação do dano evitável até a construção de sistemas de alta confiabilidade.

O movimento contemporâneo de segurança do paciente tem marco simbólico na publicação, ainda na virada do milênio, do relatório *To Err Is Human: Building a Safer Health System*, que deslocou a compreensão do erro assistencial da culpabilização individual para a análise sistêmica das condições latentes que o favorecem (Kohn; Corrigan; Donaldson, 2000). Dessa inflexão derivou o entendimento de que a maioria dos eventos adversos resulta de falhas de processo e de organização, e não de incompetência isolada, o que fundamenta a noção de cultura de segurança e a abordagem sistêmica hoje incorporada às políticas internacionais (Organização Mundial da Saúde, 2021a).

No Brasil, a institucionalização da agenda deu-se sobretudo a partir de 2013, com a criação do Programa Nacional de Segurança do Paciente por meio da Portaria GM/MS nº 529/2013 (Brasil, 2013a) e com a definição, pela Agência Nacional de Vigilância Sanitária, das ações para a segurança do paciente em serviços de saúde e da obrigatoriedade dos Núcleos de Segurança do Paciente (Brasil, 2013b). Esse arcabouço, aliado aos protocolos básicos de segurança: identificação do paciente, segurança na prescrição e administração de medicamentos, prevenção de quedas e de lesões por pressão, entre outros, estabeleceu a base normativa e operacional sobre a qual as tecnologias emergentes passam agora a ser avaliadas.

É nesse cenário que a inteligência artificial se apresenta como tecnologia potencialmente transformadora. Entende-se por IA o conjunto de métodos computacionais capazes de executar tarefas que tradicionalmente exigiriam cognição humana; entre suas vertentes destacam-se o aprendizado de máquina, em que modelos aprendem padrões a partir de dados; o aprendizado

profundo, baseado em redes neurais de múltiplas camadas; e, mais recentemente, os grandes modelos de linguagem e os modelos multimodais, cuja adoção tem sido a mais veloz já observada em aplicações de saúde (Rajpurkar *et al.*, 2022; Organização Mundial da Saúde, 2024). A promessa de tais ferramentas para a segurança do paciente reside em sua capacidade de processar, em tempo quase real, volumes de dados clínicos incompatíveis com a cognição humana desassistida.

As aplicações potenciais são amplas e recobrem praticamente todo o espectro da segurança assistencial: predição de deterioração clínica e detecção precoce de sepse; sistemas de apoio à decisão clínica; segurança medicamentosa e prevenção de erros de prescrição, dispensação e administração; prevenção de quedas e de lesões por pressão; vigilância e notificação de eventos adversos apoiadas por processamento de linguagem natural; e análise preditiva de risco em nível individual e populacional. Ilustrativo do estado da arte e de suas armadilhas, é o caso de um modelo proprietário de predição de sepse amplamente implementado em centenas de hospitais, cuja validação externa independente revelou desempenho discriminatório substancialmente inferior ao anunciado pelo fabricante, com elevada taxa de alarmes e capacidade limitada de antecipar o desfecho (Wong *et al.*, 2021).

Com efeito, a incorporação da IA à assistência não é isenta de riscos. A literatura documenta a ocorrência de vieses algorítmicos capazes de reproduzir e amplificar desigualdades preexistentes como no algoritmo de gestão populacional que, ao utilizar o custo assistencial como variável substituta da necessidade de saúde, subestimou sistematicamente a gravidade de pacientes negros (Obermeyer *et al.*, 2019). Somam-se a esse risco os alarmes falsos e a consequente fadiga de alertas, o viés de automação, tendência do profissional a confiar acriticamente na saída do sistema, a opacidade dos modelos de aprendizado profundo, sintetizada na metáfora da caixa-preta, e os erros de generalização decorrentes de validação restrita ao ambiente de origem. Não por acaso, o exemplo da sepse evidencia que desempenho satisfatório em validação interna não garante utilidade clínica em contextos distintos (Wong *et al.*, 2021).

Essas fragilidades projetam-se sobre dimensões éticas e regulatórias incontornáveis. A Organização Mundial da Saúde (2021b) propôs seis princípios consensuais para a governança ética da IA em saúde: proteção da autonomia; promoção do bem-estar e da segurança; transparência e explicabilidade; responsabilização; inclusão e equidade; e sustentabilidade e responsividade, posteriormente ampliados por orientação específica sobre modelos

multimodais (Organização Mundial da Saúde, 2024). No plano nacional, a proteção de dados pessoais sensíveis é regida pela Lei Geral de Proteção de Dados (Brasil, 2018), enquanto o marco regulatório setorial avança rapidamente: o Conselho Federal de Medicina publicou, em 2026, resolução que disciplina o uso de sistemas de IA na prática médica, reafirmando a IA como ferramenta de apoio e mantendo a responsabilidade clínica sob o profissional (Conselho Federal de Medicina, 2026), ao passo que tramita no Congresso Nacional o projeto de marco legal da IA e a Agência Nacional de Vigilância Sanitária regula os softwares como dispositivos médicos (Brasil, 2022; Brasil, 2023). Trata-se, portanto, de um cenário regulatório em rápida transformação, o que reforça a necessidade de análises atualizadas.

Nesse conjunto de tensões, a enfermagem e a equipe multiprofissional ocupam posição estratégica. Como categorias que operam na interface entre o dado, o sistema e o leito, cabe-lhes exercer a supervisão humana significativa, não a mera presença formal do profissional no circuito decisório, mas o julgamento clínico ativo capaz de reconhecer, contestar e corrigir saídas algorítmicas inadequadas. É precisamente nessa função de vigilância que reside a principal salvaguarda contra o viés de automação e a delegação acrítica de decisões, o que torna imprescindível compreender de que modo a IA se articula às práticas de segurança do paciente.

Apesar do crescimento exponencial de publicações sobre IA em saúde, persistem lacunas relevantes: a literatura é marcada por estudos com validação apenas interna e desfechos substitutos, com escassa avaliação de impacto sobre desfechos assistenciais reais e de suas implicações para a prática da enfermagem e a governança dos serviços. Sínteses críticas que articulem, simultaneamente, oportunidades, riscos éticos e perspectivas futuras com apreciação metodológica rigorosa do risco de viés típico dos estudos de IA, permanecem escassas. Diante disso, o presente estudo tem por objetivo sintetizar as melhores evidências científicas sobre a aplicação da inteligência artificial à segurança do paciente, respondendo à pergunta norteadora: quais evidências científicas demonstram como a inteligência artificial influencia a segurança do paciente e quais são as oportunidades, os desafios éticos e as perspectivas futuras decorrentes de sua incorporação?

2 MÉTODO

2.1 Tipo de estudo

Trata-se de revisão integrativa da literatura, método que permite reunir e sintetizar achados de estudos com delineamentos diversos, experimentais e não experimentais em torno

de um mesmo problema. Adotou-se o referencial de Whitemore e Knafl (2005), organizado em seis etapas: identificação do problema; busca na literatura; avaliação dos dados; análise dos dados; apresentação; e síntese interpretativa. O fluxo de reporte PRISMA 2020 foi adotado como padrão de transparência e rastreabilidade do processo de seleção (Page *et al.*, 2021).

2.2 Pergunta de pesquisa

A pergunta norteadora foi estruturada pela estratégia PICO: População: pacientes assistidos em serviços de saúde e profissionais/equipes que utilizam sistemas de IA; Interesse: aplicação da IA às práticas e aos processos de segurança do paciente; Contexto: serviços hospitalares, unidades de terapia intensiva, atenção primária e sistemas de saúde.

2.3 Fontes e descritores

Planejou-se busca em oito fontes: PubMed/MEDLINE, Embase, Scopus, Web of Science, CINAHL, LILACS, SciELO e BVS. Utilizaram-se descritores DeCS e MeSH e sinônimos livres — *Patient Safety, Artificial Intelligence, Machine Learning, Deep Learning, Medical Errors, Clinical Decision Support Systems, Ethics, Clinical e Nursing Care* —, combinados pelos operadores AND e OR. A estratégia principal (PubMed/MEDLINE) foi: ("Patient Safety"[Mesh] OR "Medical Errors"[Mesh] OR "patient safety"[tiab] OR "adverse event*"[tiab]) AND ("Artificial Intelligence"[Mesh] OR "Machine Learning"[Mesh] OR "Deep Learning"[Mesh] OR "clinical decision support"[tiab]), adaptada à sintaxe de cada base.

6

2.4 Critérios de elegibilidade

Inclusão: estudos primários e secundários sobre a aplicação da IA a desfechos ou processos de segurança do paciente, publicados entre 2020 e 2026, em português, inglês ou espanhol, com texto completo. Exclusão: editoriais, cartas e resumos sem dados primários; estudos com foco exclusivamente computacional sem vínculo com segurança assistencial; duplicatas; e trabalhos sem texto completo recuperável.

2.5 Seleção, extração e síntese

A seleção ocorrerá em duas fases (títulos/resumos e textos completos), por dois revisores independentes, com concordância aferida pelo coeficiente Kappa e apresentação no fluxograma PRISMA 2020. A extração empregará instrumento estruturado (autor/ano, país, objetivo,

delineamento, amostra/contexto, aplicação de IA, principais resultados e nível de evidência), seguida de análise temática.

2.6 Nível de evidência e risco de viés

O nível de evidência será classificado pela hierarquia de sete níveis de Melnyk e Fineout-Overholt (2019). Para o risco de viés, considerando o predomínio de estudos de modelos preditivos, adotar-se-á o PROBAST+AI (Moons *et al.*, 2025), atualização do PROBAST (Wolff *et al.*, 2019); a completude do relato será apreciada à luz de CONSORT-AI (Liu *et al.*, 2020) e TRIPOD+AI (Collins *et al.*, 2024), com exame explícito de validação apenas interna, ausência de validação externa/prospectiva, desfechos substitutos e relato incompleto de equidade e calibração.

3 RESULTADOS

Esta etapa consolidou uma base de 20 estudos reais, indexados e verificáveis, publicados entre 2020 e 2026. Predominaram estudos de desenvolvimento e/ou validação de modelos preditivos (Wong *et al.*, 2021; Adams *et al.*, 2022; Escobar *et al.*, 2020; Song *et al.*, 2021; Nakagami *et al.*, 2021; Lados-Martin *et al.*, 2022; Mamlook; Wells; Sawyer, 2023) e revisões de escopo ou sistemáticas (Bates *et al.*, 2021; Syrowatka *et al.*, 2022; Bedi *et al.*, 2025; Murphy *et al.*, 2023), acompanhados de estudos experimentais sobre viés de automação (Dratsch *et al.*, 2023; Gaube *et al.*, 2021), estudos de modelagem e avaliação (Gichoya *et al.*, 2022; Hager *et al.*, 2024; Ozonoff *et al.*, 2022) e contribuições da enfermagem brasileira (Gonçalves *et al.*, 2020; Scherer, 2022; Vitorino; Yoshinari Júnior, 2023). Os estudos distribuem-se por diferentes países, com destaque para Estados Unidos, Brasil, países europeus e asiáticos, e recobrem os principais domínios de dano evitável.

As aplicações concentram-se na predição de sepse e deterioração clínica, na segurança medicamentosa, na prevenção de lesões por pressão e quedas, na vigilância de eventos adversos por processamento de linguagem natural e no apoio à decisão clínica, incluindo grandes modelos de linguagem. De modo transversal, os achados evidenciam que os benefícios são fortemente condicionados à qualidade da implementação e à integração ao fluxo de trabalho, e que persistem limitações metodológicas recorrentes — validação apenas interna, desfechos substitutos, elevada taxa de falsos alarmes, viés de automação e iniquidades algorítmicas, detalhadas na Discussão. O Quadro 1 sintetiza os estudos incluídos.

Figura 1 — Fluxograma PRISMA 2020 (estrutura para preenchimento após a busca completa)

IDENTIFICAÇÃO: registros identificados nas bases (PubMed/MEDLINE, Embase, Scopus, Web of Science, CINAHL, LILACS, SciELO, BVS) [n = ____] → registros removidos antes da triagem: duplicatas [n = ____]; outros [n = ____]

TRIAGEM: registros triados por título/resumo [n = ____] → excluídos [n = ____]

ELEGIBILIDADE: relatórios buscados para recuperação [n = ____] → não recuperados [n = ____]; avaliados em texto completo [n = ____] → excluídos, com motivos [n = ____]

INCLUÍDOS: estudos incluídos na síntese [n = 20]

Nota: a estrutura segue o PRISMA 2020 (Page et al., 2021). Os quantitativos de identificação, triagem e elegibilidade devem ser preenchidos após a execução completa e reprodutível das buscas em cada base, na data da submissão. O número de estudos incluídos (20) corresponde à base verificada nesta etapa.

Quadro 1 — Caracterização dos estudos incluídos (n = 20), 2020–2026

Autor (ano)	País	Objetivo	Delineamento	Amostra/Con texto	Aplicação de IA	Principais resultados	N E
Wong et al. (2021)	EUA	Validar externamente modelo proprietário de sepse	Coorte retrospectiva (validação externa)	27.697 pacientes; sistema hospitalar	Predição de sepse (ESM)	AUROC 0,63 (vs 0,76–0,83 do fabricante); baixa sensibilidade; alto volume de alertas	I V
Adams et al. (2022)	EUA	Avaliar desfechos após implantação do TREWS	Coorte prospectiva multicêntrica	590.736 monitorados; 6.877 com sepse; 5 hospitais	Alerta de sepse por ML (TREWS)	Interação com o alerta associada a menor tempo até antibiótico e menor mortalidade	I V
Escobar et al. (2020)	EUA	Avaliar detecção automatizada de deterioração	Quase-experimental (19 hospitais)	Internações; enfermeiros de monitoramento remoto	Escore de deterioração (AAM) + resposta rápida	Redução de mortalidade e de internação em UTI no grupo intervenção	III
Song et al. (2021)	EUA	Predizer LP a partir de dados de enfermagem	Desenvolvimento de modelo (coorte retrospectiva)	5 hospitais; fenótipos de avaliação de enfermagem	Predição de lesão por pressão (RF)	RF AUC 0,92–0,94; nível de consciência (Glasgow) como preditor-chave	I V
Nakagami et al. (2021)	Japão	Predizer LP hospitalar via prontuário	Coorte observacional retrospectiva	Hospital universitário	Predição de LP (ML/EHR)	Desempenho superior a escalas tradicionais; relato conforme TRIPOD	I V
Ladíos-Martin et al. (2022)	Espanha	Modelo preditivo de quedas para a enfermagem	Desenvolvimento/v alidação de modelo	Pacientes internados; prontuário eletrônico	Predição de risco de queda (ML)	Apoio à decisão que reduz coleta manual; desempenho moderado-bom	I V
Mamlook; Wells; Sawyer (2023)	EUA	Predizer infecção de sítio cirúrgico (ISC)	Desenvolvimento/v alidação de modelo	Fatores pré-operatórios e do procedimento	Predição de ISC (ML)	Modelos de ML superaram abordagens estatísticas tradicionais	I V

Autor (ano)	País	Objetivo	Delineamento	Amostra/Con texto	Aplicação de IA	Principais resultados	N E
Syrowatka et al. (2022)	EUA/Canadá	Mapear casos de uso de IA para reduzir EAM	Revisão de escopo	78 artigos	IA/PLN para predição e detecção de EAM	Casos de uso promissores; evidência inicial; faltam ensaios prospectivos	V
Bates et al. (2021)	EUA	Sintetizar potencial da IA em 8 domínios de dano	Revisão de escopo	392 estudos	IA para predição/detecção de EA	Maior potencial em EAM, deterioração e erro diagnóstico	V
Bedi et al. (2025)	EUA	Revisar testes/avaliação de LLM em saúde	Revisão sistemática	Estudos de aplicações de LLM	LLMs em tarefas clínicas	Avaliação heterogênea; poucos avaliam segurança e uso real	V
Murphy et al. (2023)	Países Baixos	Revisar PLN para detecção de EAM	Revisão de escopo	Métodos de aprendizado supervisionado	PLN para detecção de EAM	Heterogeneidade; desafios de generalização e validação	V
Dratsch et al. (2023)	Alemanha/P. Baixos	Avaliar viés de automação na mamografia	Experimento prospectivo (reader study)	27 radiologistas; 50 mamografias	Sugestão BI-RADS por IA	Sugestões incorretas reduziram a acurácia; menos experientes mais afetados	III
Gaube et al. (2021)	Alemanha/EUA	Avaliar suscetibilidade a conselhos de IA	Estudo experimental	Profissionais; radiografias de tórax	Apoio à decisão por IA	Conselhos incorretos reduziram o desempenho, sobretudo dos menos experientes	III
Gichoya et al. (2022)	EUA/multinac.	Verificar reconhecimento de raça por IA em imagens	Estudo de modelagem	Múltiplas bases de imagem	Aprendizado profundo	IA prediz raça a partir de imagens — risco de iniquidade oculta	V I
Hager et al. (2024)	Alemanha/R. Unido	Avaliar limitações de LLMs na decisão clínica	Estudo de avaliação (benchmark)	Casos clínicos reais	LLMs como apoio à decisão	Desempenho inferior ao humano e inconsistente; risco de segurança	V I
Ozonoff et al. (2022)	EUA	Vigilância de eventos de segurança via PLN	Estudo de avaliação/validação	Notas clínicas hospitalares	PLN (classificador de texto)	Boa sensibilidade; complementa notificação voluntária subutilizada	V I
Ghassemi; Oakden-Rayner; Beam (2021)	Multinac.	Analisar limites da IA explicável em saúde	Análise crítica (viewpoint)	—	Explicabilidade de (XAI)	Métodos atuais de XAI não garantem confiança/transparência ao nível do paciente	V II

Autor (ano)	País	Objetivo	Delineamento	Amostra/Con texto	Aplicação de IA	Principais resultados	N E
Gonçalves et al. (2020)	Brasil	Relatar implantação de algoritmo de IA para sepse	Relato de experiência (antes/depois)	Hospital filantrópico; enfermeiros	Algoritmo de IA para sepse	Descreve papel do enfermeiro na criação/uso e efeitos no processo de trabalho	V I
Scherer (2022)	Brasil	Refletir sobre IA no apoio à decisão na sepse	Reflexão	—	IA na predição de sepse	Discute potencial e limites da IA como apoio à decisão de enfermagem	V II
Vitorino; Yoshinari Júnior (2023)	Brasil	Discutir IA como aliada na enfermagem	Reflexão	—	IA na prática de enfermagem	Aponta desafios, oportunidades e responsabilidade profissional	V II

NE: nível de evidência (Melnik e Fineout-Overholt, 2019). LP: lesão por pressão; EAM: evento adverso a medicamento; EA: evento adverso; PLN: processamento de linguagem natural; LLM: grande modelo de linguagem; ML: aprendizado de máquina; RF: random forest; ESM: Epic Sepsis Model; AAM: Advance Alert Monitor; TREWS: Targeted Real-time Early Warning System.

Quadro 2 — Síntese das principais aplicações de IA voltadas à segurança do paciente

Aplicação	Finalidade	Evidência de efetividade	Limitações
Predição de sepse/deterioração	Antecipar deterioração para intervenção precoce	Mista: implantação bem integrada associada a melhores desfechos (Adams; Escobar); validação externa pode reduzir muito o desempenho (Wong)	Alarmes falsos, fadiga de alertas, perda de desempenho fora do ambiente de origem
Segurança medicamentosa	Predizer/detectar EAM e otimizar alertas	Promissora na interceptação de erros e redução da carga de alertas (Syrowatka; Murphy)	Poucos estudos prospectivos; heterogeneidade; risco de subdetecção
Prevenção de lesão por pressão	Estratificar risco de LP	Modelos superam escalas tradicionais em discriminação (Song; Nakagami)	Predomínio de validação interna; foco em predição, não em dano evitado
Prevenção de quedas	Estratificar risco de queda	Apoio à decisão de enfermagem com desempenho moderado-bom (Ladinos-Martin)	Validação externa limitada; sensibilidade variável
Vigilância de EA (PLN)	Detectar e classificar eventos e incidentes	Boa sensibilidade; complementa notificação subutilizada (Ozonoff)	Especificidade variável; necessidade de revisão humana
Apoio à decisão (CDSS/LLM)	Auxiliar o julgamento clínico	LLMs ainda inferiores ao humano e inconsistentes (Hager; Bedi)	Alucinação, viés de automação, avaliação insuficiente de segurança
Equidade e auditoria	Detectar e mitigar viés algorítmico	Evidência de viés relevante (Obermeyer; Gichoya)	Métricas de equidade pouco reportadas; risco de ampliar desigualdades

Quadro 3 — Atribuições por categoria profissional na supervisão e no uso seguro de sistemas de IA

Categoria	Atribuições
Enfermeiro	Validar alertas à beira do leito; exercer supervisão humana significativa; documentar e notificar discrepâncias; educar a equipe; participar da avaliação e da governança local
Médico	Decisão clínica final; validar recomendações; assumir a responsabilidade clínica (CFM, 2026); integrar a IA ao raciocínio diagnóstico
Farmacêutico	Revisar alertas de segurança medicamentosa; apoiar a conciliação medicamentosa com ML; avaliar a relevância clínica dos alertas
Núcleo de Segurança do Paciente / gestão	Monitoramento pós-implantação; auditoria de desempenho e equidade; gestão de risco; definição de protocolos de uso
Informática em saúde / TI	Curadoria e qualidade dos dados; integração e manutenção; monitoramento de deriva do modelo
Todas as categorias	Formação e competências digitais; cultura de segurança; reporte de eventos relacionados à IA

Tabela 1 — Indicadores de avaliação de sistemas de IA aplicados à segurança do paciente

Indicador	Como avaliar
Acurácia/discriminação	AUROC, sensibilidade, especificidade e calibração
Validação externa	Desempenho em ambiente independente do de treinamento (idealmente prospectivo)
Taxa de falsos alarmes	Proporção de alertas não acionáveis e carga total de alerta
Impacto em desfechos	Efeito sobre mortalidade, tempo até tratamento e eventos evitados
Equidade entre subgrupos	Desempenho estratificado por sexo, raça/cor, idade e comorbidade
Monitoramento pós-implantação	Vigilância de deriva, reavaliação periódica e auditoria contínua

Quadro 4 — Princípios éticos e marcos regulatórios: OMS, marcos internacionais e Brasil

Dimensão	OMS	Marcos internacionais	Brasil
Transparência/explicabilidade	Princípio da transparência e explicabilidade	Guias de relato (TRIPOD+AI; CONSORT-AI)	CFM 2.454/2026: avaliação de risco e rastreabilidade
Responsabilização	Princípio da responsabilização	EU AI Act (abordagem baseada em risco)	PL 2338/2023; responsabilidade clínica do profissional (CFM)
Equidade/inclusão	Princípio da inclusão e equidade	Literatura sobre vies (Obermeyer; Gichoya)	SUS e universalidade; atenção a iniquidades
Privacidade e dados	Proteção de dados pessoais	GDPR	LGPD (Lei 13.709/2018); COFEN 696/2022
Autonomia/supervisão humana	Autonomia; humano no controle	"Human-in-the-loop"	IA como ferramenta de apoio (CFM 2.454/2026)

Figura 2 — Ciclo de uso seguro da IA na assistência (descrição conceitual)

O uso seguro da IA na assistência pode ser representado como um ciclo contínuo: (1) dado clínico de qualidade, curado e representativo → (2) modelo/algoritmo treinado e validado → (3) recomendação ou alerta gerado ao ponto de cuidado → (4) supervisão humana significativa, exercida pela enfermagem e pela equipe multiprofissional, que interpreta, contesta e contextualiza a saída → (5) decisão clínica e ação assistencial → (6) desfecho do paciente →

(7) monitoramento pós-implantação e auditoria (deriva do modelo, taxa de falsos alarmes, equidade entre subgrupos) → retroalimentação que corrige os dados e o modelo, reiniciando o ciclo. A supervisão humana (etapa 4) é o principal ponto de controle contra o viés de automação, e o monitoramento contínuo (etapa 7) é a salvaguarda contra a degradação silenciosa do desempenho ao longo do tempo.

4 DISCUSSÃO

4.1 Segurança do paciente e inteligência artificial: convergências conceituais e promessas

A convergência entre segurança do paciente e IA repousa em uma afinidade conceitual: ambas deslocam o foco do erro individual para a análise sistêmica dos processos (Kohn; Corrigan; Donaldson, 2000). Revisões de escopo abrangentes indicam que a IA tem potencial de atuação em todos os grandes domínios de dano evitável, com alvos iniciais mais promissores em eventos adversos a medicamentos, deterioração clínica e erro diagnóstico (Bates *et al.*, 2021; Rajpurkar *et al.*, 2022). Contudo, a promessa deve ser lida com cautela: a maior parte da evidência disponível demonstra capacidade preditiva, e não redução comprovada de dano, de modo que a passagem da predição ao desfecho assistencial permanece o principal teste a ser vencido.

12

4.2 Predição de deterioração clínica e detecção precoce de sepse

A sepse é o campo que melhor ilustra tanto o potencial quanto as armadilhas da IA. A validação externa independente de um modelo proprietário amplamente disseminado revelou desempenho muito inferior ao anunciado, com baixa sensibilidade e elevada carga de alarmes (Wong *et al.*, 2021), ao passo que um sistema de alerta desenvolvido e avaliado prospectivamente em múltiplos hospitais associou-se a menor tempo até o antibiótico e a melhores desfechos quando os profissionais interagiram com o alerta (Adams *et al.*, 2022). O contraste evidencia que a utilidade clínica não reside no algoritmo isolado, mas em sua integração ao fluxo de trabalho e na resposta humana que o sucede, como também demonstrou a redução de mortalidade obtida por um programa de deterioração mediado por enfermeiros de monitoramento remoto (Escobar *et al.*, 2020). No Brasil, relatos de implantação e reflexões da enfermagem corroboram a centralidade do enfermeiro nesse processo (Gonçalves *et al.*, 2020; Scherer, 2022).

4.3 Segurança medicamentosa e prevenção de erros de medicação

Na segurança medicamentosa, a IA tem sido explorada para prever e detectar eventos adversos a medicamentos e para otimizar alertas, tornando-os mais específicos e acionáveis (Syrowatka *et al.*, 2022). O processamento de linguagem natural amplia a capacidade de identificar eventos descritos em texto livre, embora com desafios de generalização e validação (Murphy *et al.*, 2023). A evidência, porém, é majoritariamente retrospectiva e heterogênea, com escassez de estudos prospectivos que demonstrem redução efetiva de dano ao paciente, limitação que recomenda prudência antes da incorporação acrítica dessas ferramentas à prescrição e à dispensação.

4.4 Prevenção de quedas, lesões por pressão e eventos adversos assistenciais

Modelos de aprendizado de máquina para predição de lesão por pressão, quedas e infecção de sítio cirúrgico tendem a superar, em discriminação, as escalas tradicionais de avaliação de risco (Song *et al.*, 2021; Nakagami *et al.*, 2021; Ladios-Martin *et al.*, 2022; Mamlook; Wells; Sawyer, 2023). Entretanto, dois limites são recorrentes: o predomínio de validação apenas interna e o uso de desfechos substitutos, a maioria dos estudos prediz o risco, mas não demonstra a redução do evento evitado. É sintomático que preditores de enfermagem, como o nível de consciência, figurem entre as variáveis mais importantes (Song *et al.*, 2021), o que reforça que o valor desses modelos depende da qualidade dos dados assistenciais gerados à beira do leito.

4.5 Sistemas de apoio à decisão clínica: benefícios, fadiga de alertas e viés de automação

Os sistemas de apoio à decisão podem qualificar o cuidado, mas introduzem riscos próprios. O excesso de alertas pouco específicos conduz à fadiga de alertas, enquanto a confiança acrítica na saída do sistema caracteriza o viés de automação. Estudos experimentais demonstram que sugestões incorretas de IA degradam o desempenho de profissionais inclusive de radiologistas experientes, e sobretudo dos menos experientes (Dratsch *et al.*, 2023; Gaube *et al.*, 2021). Esses achados impõem que a IA seja desenhada para preservar a autonomia e o escrutínio do profissional, e não para substituí-lo.

4.6 Vigilância, notificação e análise de eventos adversos apoiadas por IA

A vigilância de eventos adversos beneficia-se do processamento de linguagem natural aplicado a notas clínicas e relatos de incidentes, que recupera eventos subnotificados pelos sistemas voluntários, com boa sensibilidade (Ozonoff *et al.*, 2022; Murphy *et al.*, 2023). A especificidade variável, contudo, impõe a manutenção da revisão humana, o que posiciona a IA como instrumento de triagem e priorização e não de substituição do julgamento das equipes e dos Núcleos de Segurança do Paciente.

4.7 Vieses algorítmicos, equidade e risco de ampliação de desigualdades

A equidade é uma das fronteiras éticas mais sensíveis. Algoritmos podem reproduzir e amplificar desigualdades, como no caso em que o uso do custo assistencial como variável substituta da necessidade subestimou a gravidade de pacientes negros (Obermeyer *et al.*, 2019); modelos de aprendizado profundo chegam a inferir a raça autorrelatada a partir de imagens médicas, mesmo quando não há intenção de fazê-lo, o que cria um vetor de iniquidade difícil de detectar (Gichoya *et al.*, 2022). Em um sistema universal como o SUS, a avaliação de desempenho estratificada por subgrupos torna-se requisito ético e sanitário, e não mera formalidade técnica.

4.8 Transparência, explicabilidade e o problema da "caixa-preta"

A opacidade dos modelos alimenta a expectativa de que a explicabilidade resolveria os problemas de confiança e viés. Argumenta-se, porém, que os métodos atuais de IA explicável oferecem descrições aproximadas, mas não garantem transparência real ao nível da decisão individual, podendo inclusive induzir confiança indevida (Ghassemi; Oakden-Rayner; Beam, 2021). Disso decorre que a confiança não deve derivar de explicações pós-hoc, mas de validação rigorosa, monitoramento e governança — deslocando o problema da "caixa-preta" do campo técnico para o da responsabilização institucional.

4.9 Responsabilização profissional e jurídica diante de decisões mediadas por algoritmos

A responsabilização é princípio consensual da governança ética da IA em saúde (Organização Mundial da Saúde, 2021b) e foi incorporada ao ordenamento brasileiro pela resolução que disciplina o uso da IA na prática médica, a qual reafirma a IA como ferramenta de apoio e mantém a responsabilidade clínica sob o profissional (Conselho Federal de Medicina,

2026). A tramitação do marco legal da IA (Brasil, 2023) tende a consolidar uma abordagem baseada em risco. Permanece, todavia, a questão da responsabilidade distribuída entre desenvolvedor, instituição e profissional — tema que exige clareza normativa para que a supervisão humana não se reduza a um mecanismo de transferência de culpa.

4.10 Privacidade, proteção de dados e consentimento

O funcionamento da IA depende de grandes volumes de dados sensíveis, o que coloca a privacidade no centro do debate. No Brasil, a Lei Geral de Proteção de Dados (Brasil, 2018) estabelece as bases legais do tratamento, e a regulação da atuação da enfermagem na saúde digital exige plataformas seguras, confidencialidade e consentimento (Conselho Federal de Enfermagem, 2022), em consonância com os princípios internacionais de governança (Organização Mundial da Saúde, 2021b). A tensão entre a utilidade dos dados e a proteção do paciente recomenda minimização, segurança e transparência como padrões inegociáveis.

4.11 Atuação da enfermagem e da equipe multiprofissional na supervisão humana significativa

A supervisão humana significativa é o principal ponto de controle do uso seguro da IA, e a enfermagem ocupa posição privilegiada para exercê-la, por sua presença contínua à beira do leito e por gerar boa parte dos dados que alimentam os modelos (Escobar *et al.*, 2020; Gonçalves *et al.*, 2020; Vitorino; Yoshinari Júnior, 2023). Supervisão significativa não é presença formal no circuito, mas julgamento ativo capaz de reconhecer e contestar saídas inadequadas, justamente a competência que mitiga o viés de automação evidenciado em contextos de menor experiência (Dratsch *et al.*, 2023; Gaube *et al.*, 2021).

15

4.12 Formação, competências digitais e educação permanente para o uso seguro da IA

A eficácia da supervisão humana depende de competências digitais que ainda não estão consolidadas na formação em saúde. A literatura sobre LLMs demonstra desempenho inconsistente e inferior ao humano em decisão clínica, reforçando a necessidade de letramento em IA para interpretar criticamente suas saídas (Hager *et al.*, 2024; Bedi *et al.*, 2025). A educação permanente e a normatização da atuação na saúde digital (Conselho Federal de Enfermagem, 2022) são, portanto, condições para que a incorporação tecnológica se traduza em segurança, e não em novos riscos (Vitorino; Yoshinari Júnior, 2023).

4.13 Governança, validação, regulação e avaliação de tecnologias em saúde

A governança responsável articula validação rigorosa, regulação e avaliação de tecnologias em saúde. Instrumentos específicos, PROBAST+AI para o risco de viés (Moons *et al.*, 2025) e guias de relato como TRIPOD+AI e CONSORT-AI (Collins *et al.*, 2024; Liu *et al.*, 2020) oferecem parâmetros para distinguir modelos robustos de promessas infundadas, enquanto a avaliação sistemática de aplicações de LLM ainda é heterogênea e pouco voltada à segurança (Bedi *et al.*, 2025). A orientação internacional (Organização Mundial da Saúde, 2021b; 2024) e o marco nacional (Conselho Federal de Medicina, 2026) convergem na exigência de validação externa, idealmente prospectiva, e de monitoramento pós-implantação — etapas frequentemente ausentes nos estudos disponíveis.

4.14 Barreiras à implementação e perspectivas futuras baseadas em evidências

As principais barreiras à implementação segura incluem a qualidade e a representatividade dos dados, a integração ao fluxo de trabalho, a deriva do modelo ao longo do tempo, as iniquidades algorítmicas e as lacunas de formação. As perspectivas futuras, baseadas em evidências, apontam para a priorização de ensaios prospectivos com desfechos assistenciais reais, o relato sistemático de equidade entre subgrupos, a cautela na adoção de LLMs em decisões de alto risco (Hager *et al.*, 2024; Bedi *et al.*, 2025) e o fortalecimento da validação externa como requisito mínimo (Wong *et al.*, 2021). No Brasil, o cenário regulatório em rápida transformação (Conselho Federal de Medicina, 2026; Brasil, 2023) exige acompanhamento contínuo para alinhar a inovação às salvaguardas de segurança.

5 CONSIDERAÇÕES FINAIS

As evidências reunidas indicam que a inteligência artificial influencia a segurança do paciente ao longo de todo o continuum assistencial da predição de deterioração e sepse à segurança medicamentosa, à prevenção de lesões por pressão e quedas, à vigilância de eventos adversos e ao apoio à decisão. Respondendo à pergunta norteadora, as oportunidades são concretas, mas seus benefícios são condicionais: dependem da qualidade da implementação, da integração ao fluxo de trabalho e, sobretudo, da supervisão humana significativa. Os desafios éticos vieses algorítmicos e iniquidades, opacidade, responsabilização e proteção de dados são incontornáveis, e as perspectivas futuras passam por validação externa e prospectiva, relato de equidade, governança e monitoramento contínuo.

Para a prática clínica e a gestão, impõe-se tratar a IA como ferramenta de apoio sob julgamento humano ativo, fortalecendo o papel da enfermagem e dos Núcleos de Segurança do Paciente na validação de alertas, no monitoramento pós-implantação e na auditoria de desempenho e equidade. Para a formação multiprofissional, é urgente desenvolver competências digitais e letramento em IA por meio da educação permanente. Para as políticas públicas, o avanço regulatório brasileiro deve assegurar transparência, responsabilização e proteção de dados sem obstruir a inovação.

Como limitações, registra-se que a base de estudos aqui consolidada deverá ter seus quantitativos de identificação e triagem preenchidos após a execução completa e reprodutível das buscas, conforme o fluxograma PRISMA 2020; soma-se a isso a heterogeneidade dos delineamentos e a restrição idiomática. Recomendam-se, como agenda de pesquisa, estudos prospectivos com desfechos assistenciais reais, avaliações de impacto sobre a prática de enfermagem e investigações sobre supervisão humana significativa e mitigação do viés de automação em contextos brasileiros.

REFERÊNCIAS

- ADAMS, R. *et al.* Prospective, multi-site study of patient outcomes after implementation of the TREWS machine learning-based early warning system for sepsis. *Nature Medicine*, [s. l.], v. 28, n. 7, p. 1455-1460, 2022. DOI: 10.1038/s41591-022-01894-0.
- BATES, D. W. *et al.* The potential of artificial intelligence to improve patient safety: a scoping review. *npj Digital Medicine*, [s. l.], v. 4, art. 54, 2021. DOI: 10.1038/s41746-021-00423-6.
- BEDI, S. *et al.* Testing and evaluation of health care applications of large language models: a systematic review. *JAMA*, [s. l.], v. 333, n. 4, p. 319-328, 2025. DOI: 10.1001/jama.2024.21700.
- BRASIL. Ministério da Saúde. Portaria GM/MS nº 529, de 1º de abril de 2013. Institui o Programa Nacional de Segurança do Paciente (PNSP). Brasília, DF: Ministério da Saúde, 2013a.
- BRASIL. Agência Nacional de Vigilância Sanitária. Resolução da Diretoria Colegiada – RDC nº 36, de 25 de julho de 2013. Institui ações para a segurança do paciente em serviços de saúde e dá outras providências. Brasília, DF: Anvisa, 2013b.
- BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018.
- BRASIL. Agência Nacional de Vigilância Sanitária. Resolução da Diretoria Colegiada – RDC nº 657, de 24 de março de 2022. Dispõe sobre a regularização de software como dispositivo médico (SaMD). Brasília, DF: Anvisa, 2022.

BRASIL. Congresso Nacional. Projeto de Lei nº 2.338, de 2023. Dispõe sobre o uso da inteligência artificial. Brasília, DF, 2023.

COLLINS, G. S. *et al.* TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, [s. l.], v. 385, p. e078378, 2024. DOI: 10.1136/bmj-2023-078378.

CONSELHO FEDERAL DE ENFERMAGEM (Brasil). Resolução COFEN nº 696, de 17 de maio de 2022. Dispõe sobre a atuação da Enfermagem na Saúde Digital, normatizando a Telenfermagem. Brasília, DF: Cofen, 2022.

CONSELHO FEDERAL DE MEDICINA (Brasil). Resolução CFM nº 2.454, de 27 de fevereiro de 2026. Define normas para a utilização de sistemas de inteligência artificial na prática médica. Brasília, DF: CFM, 2026.

DRATSCH, T. *et al.* Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*, [s. l.], v. 307, n. 4, p. e222176, 2023. DOI: 10.1148/radiol.222176.

ESCOBAR, G. J. *et al.* Automated identification of adults at risk for in-hospital clinical deterioration. *New England Journal of Medicine*, [s. l.], v. 383, n. 20, p. 1951-1960, 2020. DOI: 10.1056/NEJMsa2001090.

GAUBE, S. *et al.* Do as AI say: susceptibility in deployment of clinical decision-aids. *npj Digital Medicine*, [s. l.], v. 4, art. 31, 2021. DOI: 10.1038/s41746-021-00385-9.

GHASSEMI, M.; OAKDEN-RAYNER, L.; BEAM, A. L. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, [s. l.], v. 3, n. 11, p. e745-e750, 2021. DOI: 10.1016/S2589-7500(21)00208-9.

GICHOYA, J. W. *et al.* AI recognition of patient race in medical imaging: a modelling study. *The Lancet Digital Health*, [s. l.], v. 4, n. 6, p. e406-e414, 2022. DOI: 10.1016/S2589-7500(22)00063-2.

GONÇALVES, L. S. *et al.* Implementation of an artificial intelligence algorithm for sepsis detection. *Revista Brasileira de Enfermagem*, [s. l.], v. 73, n. 3, p. e20180421, 2020. DOI: 10.1590/0034-7167-2018-0421.

HAGER, P. *et al.* Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nature Medicine*, [s. l.], v. 30, n. 9, p. 2613-2622, 2024. DOI: 10.1038/s41591-024-03097-1.

KOHN, L. T.; CORRIGAN, J. M.; DONALDSON, M. S. (ed.). To err is human: building a safer health system. Washington, DC: National Academies Press, 2000.

LADIOS-MARTIN, M. *et al.* Development of a predictive inpatient falls risk model using machine learning. *Journal of Nursing Management*, [s. l.], v. 30, n. 8, p. 3777-3786, 2022. DOI: 10.1111/jonm.13760.

LIU, X. *et al.* Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ*, [s. l.], v. 370, p. m3164, 2020. DOI: 10.1136/bmj.m3164.

MAMLOOK, R. E. A.; WELLS, L. J.; SAWYER, R. Machine-learning models for predicting surgical site infections using patient pre-operative risk and surgical procedure factors. *American Journal of Infection Control*, [s. l.], v. 51, n. 5, p. 544-550, 2023.

MELNYK, B. M.; FINEOUT-OVERHOLT, E. Evidence-based practice in nursing & healthcare: a guide to best practice. 4. ed. Philadelphia: Wolters Kluwer, 2019.

MOONS, K. G. M. *et al.* PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ*, [s. l.], v. 388, p. e082505, 2025. DOI: 10.1136/bmj-2024-082505.

MURPHY, R. M. *et al.* Adverse drug event detection using natural language processing: a scoping review of supervised learning methods. *PLoS One*, [s. l.], v. 18, n. 1, p. e0279842, 2023. DOI: 10.1371/journal.pone.0279842.

NAKAGAMI, G. *et al.* Supervised machine learning-based prediction for in-hospital pressure injury development using electronic health records: a retrospective observational cohort study in a university hospital in Japan. *International Journal of Nursing Studies*, [s. l.], v. 119, p. 103932, 2021. DOI: 10.1016/j.ijnurstu.2021.103932.

OBERMEYER, Z. *et al.* Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, [s. l.], v. 366, n. 6464, p. 447-453, 2019. DOI: 10.1126/science.aax2342.

19

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Global patient safety action plan 2021–2030: towards eliminating avoidable harm in health care. Genebra: OMS, 2021a.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Ethics and governance of artificial intelligence for health: WHO guidance. Genebra: OMS, 2021b.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Genebra: OMS, 2024.

OZONOFF, A. *et al.* Electronic surveillance of patient safety events using natural language processing. *Health Informatics Journal*, [s. l.], v. 28, n. 4, p. 14604582221132429, 2022. DOI: 10.1177/14604582221132429.

PAGE, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, [s. l.], v. 372, p. n71, 2021. DOI: 10.1136/bmj.n71.

PANAGIOTI, M. *et al.* Prevalence, severity, and nature of preventable patient harm across medical care settings: systematic review and meta-analysis. *BMJ*, [s. l.], v. 366, p. l4185, 2019. DOI: 10.1136/bmj.l4185.

RAJPURKAR, P. *et al.* AI in health and medicine. *Nature Medicine*, [s. l.], v. 28, n. 1, p. 31-38, 2022. DOI: 10.1038/s41591-021-01614-0.

SCHERER, J. S. Para além da tecnologia: a inteligência artificial pode apoiar decisões clínicas na predição da sepse? *Revista Brasileira de Enfermagem*, [s. l.], v. 75, n. 5, p. e20210586, 2022. DOI: 10.1590/0034-7167-2021-0586.

SONG, W. *et al.* Predicting pressure injury using nursing assessment phenotypes and machine learning methods. *Journal of the American Medical Informatics Association*, [s. l.], v. 28, n. 4, p. 759-765, 2021. DOI: 10.1093/jamia/ocaa336.

SYROWATKA, A. *et al.* Key use cases for artificial intelligence to reduce the frequency of adverse drug events: a scoping review. *The Lancet Digital Health*, [s. l.], v. 4, n. 2, p. e137-e148, 2022. DOI: 10.1016/S2589-7500(21)00229-6.

VITORINO, L. M.; YOSHINARI JÚNIOR, G. H. A inteligência artificial como aliada na enfermagem brasileira: desafios, oportunidades e responsabilidade profissional. *Revista Brasileira de Enfermagem*, [s. l.], v. 76, n. 3, p. e760301, 2023. DOI: 10.1590/0034-7167.2023-760301.

WHITTEMORE, R.; KNAFL, K. The integrative review: updated methodology. *Journal of Advanced Nursing*, [s. l.], v. 52, n. 5, p. 546-553, 2005. DOI: 10.1111/j.1365-2648.2005.03621.x.

WOLFF, R. F. *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, [s. l.], v. 170, n. 1, p. 51-58, 2019. DOI: 10.7326/M18-1376.

WONG, A. *et al.* External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, [s. l.], v. 181, n. 8, p. 1065-1070, 2021. DOI: 10.1001/jamainternmed.2021.2626.