Revista Ibero-Americana de Humanidades, Ciências e Educação — REASE

doi.org/10.51891/rease.v12i11.31860

INTELIGÊNCIA ARTIFICIAL NA PRÁTICA FARMACÊUTICA ENTRE EVIDÊNCIAS DE EFICÁCIA, LIMITAÇÕES METODOLÓGICAS E DESAFIOS DE IMPLEMENTAÇÃO RESPONSÁVEL

ARTIFICIAL INTELLIGENCE IN PHARMACEUTICAL PRACTICE: EVIDENCE OF EFFICACY, METHODOLOGICAL LIMITATIONS, AND CHALLENGES OF RESPONSIBLE IMPLEMENTATION

INTELIGENCIA ARTIFICIAL EN LA PRÁCTICA FARMACÉUTICA: EVIDENCIAS DE EFICACIA, LIMITACIONES METODOLÓGICAS Y DESAFÍOS DE IMPLEMENTACIÓN RESPONSABLE

Maria Eduarda Magalhães da Silva[1]
Maria Vitória Guimarães da Cunha[2]
Millena Pereira Xavier[3]
Kelly Mayanny Inácio Silva[4]
Karin Anne Margaridi Gonçalves[5]
Erika Carolina Vieira Almeida[6]

RESUMO: Foram analisadas aplicações da inteligência artificial na prática farmacêutica com base em estudos originais publicados entre 2021 e 2026. A metodologia consistiu em revisão sistemática em seis bases de dados com quatro estratégias de busca, resultando em 8.622 registros identificados, triagem de 140 artigos em texto completo e inclusão de 31 estudos organizados em quatro eixos temáticos: suporte à decisão clínica e adesão terapêutica, farmacovigilância e predição de reações adversas, interações medicamentosas e farmacocinética, e descoberta de fármacos. Os resultados demonstraram eficácia clínica em ensaios clínicos randomizados, com adesão medicamentosa de 77,4% versus 58,6% no controle e retorno sobre investimento de 12,4:1 em programa de gestão farmacêutica com suporte de IA. Modelos preditivos de farmacovigilância alcançaram AUC entre 0,876 e 0,921, e modelos de linguagem de grande escala apresentaram taxas de alucinação factual entre 6% e 23%, com erros clinicamente perigosos em até 7,3% das recomendações geradas. A análise concluiu que a barreira central à adoção dessas tecnologias não é de desempenho computacional, mas de translação clínica, evidenciada pela queda de F1-score de 0,82 para 0,34 na transferência inter-hospitalar e pela ausência de validação prospectiva multicêntrica na maioria dos estudos.

Palavras-chave: Inteligência artificial. Farmácia clínica. Farmacovigilância.

[1] Discente do curso de Farmácia da Universidade de Gurupi (UnirG).
[2] Discente do curso de Farmácia da Universidade de Gurupi (UnirG).
[3] Mestre em Gestão de Políticas Públicas. Docente do curso de Farmácia da Universidade de Gurupi (UnirG).
[4] Especialista em Farmácia Clínica e Prescrição Farmacêutica. Docente do curso de Farmácia da Universidade de Gurupi (UnirG).
[5] Mestre em Desenvolvimento Regional. Docente do curso de Farmácia da Universidade de Gurupi (UnirG).
[6] Mestre em Biotecnologia. Docente do curso de Farmácia da Universidade de Gurupi (UnirG).

ABSTRACT: This paper analyzed the applications of artificial intelligence in pharmaceutical practice based on original studies published between 2021 and 2026. The methodology consisted of a systematic literature review in six databases with four search strategies, resulting in 8,622 identified records, screening of 140 full-text articles, and inclusion of 31 studies organized into four thematic areas: clinical decision support and medication adherence, pharmacovigilance and adverse drug reaction prediction, drug interactions and pharmacokinetics, and drug discovery. The results demonstrated clinical efficacy in randomized clinical trials, with medication adherence of 77.4% versus 58.6% in the control group and a return on investment of 12.4:1 in an AI-supported pharmaceutical management program. Pharmacovigilance predictive models achieved AUC values between 0.876 and 0.921, while large language models showed factual hallucination rates between 6% and 23%, with clinically dangerous errors in up to 7.3% of generated recommendations. The analysis concluded that the central barrier to adopting these technologies is not computational performance but clinical translation, evidenced by a drop in F1-score from 0.82 to 0.34 in inter-hospital model transfer and by the absence of prospective multicenter validation in most studies.

Keywords: Artificial intelligence. Clinical pharmacy. Pharmacovigilance.

RESUMEN: Este artículo analizó las aplicaciones de la inteligencia artificial en la práctica farmacéutica con base en estudios originales publicados entre 2021 y 2026. La metodología consistió en una revisión sistemática en seis bases de datos con cuatro estrategias de búsqueda, resultando en 8.622 registros identificados, evaluación de 140 artículos en texto completo e inclusión de 31 estudios en cuatro ejes temáticos: soporte a la decisión clínica y adherencia, farmacovigilancia y predicción de reacciones adversas, interacciones medicamentosas y farmacocinética, y descubrimiento de fármacos. Los resultados demostraron eficacia clínica en ensayos aleatorizados, con adherencia medicamentosa del 77,4% frente al 58,6% en el control y retorno de 12,4:1 en gestión farmacéutica con IA. Los modelos predictivos de farmacovigilancia alcanzaron AUC entre 0,876 y 0,921, y los modelos de lenguaje de gran escala presentaron tasas de alucinación factual entre el 6% y el 23%, con errores clínicamente peligrosos en hasta el 7,3% de las recomendaciones generadas. El análisis concluyó que la barrera central para la adopción de estas tecnologías no es el rendimiento computacional sino la traslación clínica, evidenciada por caída del F1-score de 0,82 a 0,34 en transferencia inter-hospitalaria y por la ausencia de validación prospectiva multicéntrica en la mayoría de los estudios.

Palabras clave: Inteligencia artificial. Farmacia clínica. Farmacovigilancia.

1 INTRODUÇÃO

A inteligência artificial (IA) tem assumido papel crescente nas ciências da saúde, impulsionada pela expansão dos dados clínicos, pelo desenvolvimento de recursos computacionais e pelo aprimoramento de algoritmos de aprendizado de máquina (GANGWAL A, et al., 2024). Na prática farmacêutica, a complexidade dos regimes terapêuticos, a ocorrência de polifarmácia, a variabilidade interindividual das respostas farmacocinéticas e farmacodinâmicas e a expansão do arsenal terapêutico têm favorecido o desenvolvimento de sistemas capazes de integrar e analisar grandes volumes de dados de forma sistemática

(GANGWAL A, et al., 2024). Nesse contexto, algoritmos preditivos têm sido investigados em diferentes domínios, incluindo farmacovigilância, gestão da adesão terapêutica, modelagem farmacocinética individualizada, detecção de interações medicamentosas (DDI) e triagem computacional de candidatos a fármacos (KHA QH, et al., 2024; KOMURA H, et al., 2021).

Mais recentemente, o desenvolvimento de modelos de linguagem de grande escala (MLGE), como GPT-4, Claude e Gemini, ampliou as possibilidades de utilização de sistemas capazes de processar consultas farmacológicas em linguagem natural (ZHANG H, et al., 2026). Paralelamente, aumentaram as preocupações relacionadas à ocorrência de alucinações factuais, à desatualização das informações e à necessidade de supervisão especializada durante a utilização desses sistemas (YANG SS, et al., 2026). Embora o número de aplicações investigadas tenha aumentado, permanecem desafios relacionados à validação dos modelos em diferentes contextos e à translação de seus resultados para situações de prática clínica verificável (DIL-NAHLIELI D, et al., 2024; GANGWAL A, et al., 2024).

O presente estudo teve como objetivo analisar as aplicações da inteligência artificial (IA) na prática farmacêutica, com base em estudos originais publicados entre 2021 e 2026, considerando as evidências disponíveis, os principais contextos de aplicação, as limitações metodológicas e as perspectivas para sua implementação na prática profissional.

2 MÉTODOS

Trata-se de uma revisão sistemática da literatura, relatada segundo as recomendações do Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) (PAGE MJ, et al., 2021), por meio de quatro estratégias de busca independentes nas bases PubMed, Scopus, Web of Science, ScienceDirect, SciELO e Google Acadêmico, com descritores MeSH e termos livres combinados por operadores booleanos AND e OR, filtradas para publicações de 1° de janeiro de 2021 a 1° de agosto de 2026.

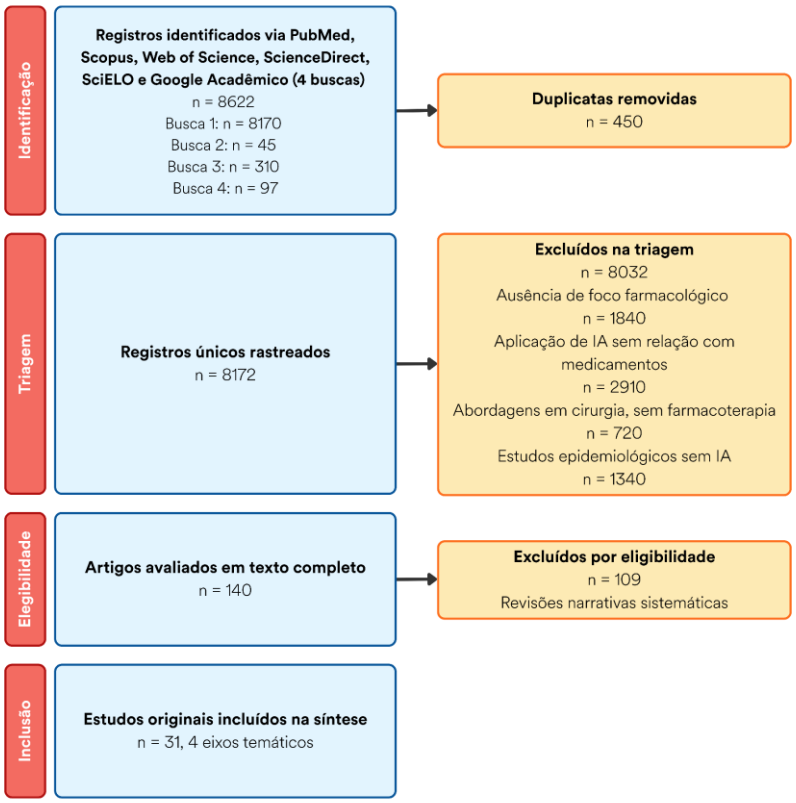
Foram organizados quatro blocos de busca principais em todas as plataformas, assim descritos: a Busca 1 empregou os termos ("Artificial Intelligence"[Mesh] OR artificial intelligence OR machine learning OR deep learning) AND (pharmac* OR pharmacist* OR pharmacy OR pharmaceutical*); a Busca 2, ("Artificial Intelligence"[Mesh] OR artificial intelligence OR machine learning) AND ("clinical pharmacy" OR "pharmaceutical care" OR "medication therapy management"); a Busca 3, ("Artificial Intelligence"[Mesh] OR machine learning OR deep learning) AND ("drug discovery" OR "drug development" OR "drug

design"); e a Busca 4, ("Artificial Intelligence"[Mesh] OR machine learning) AND (pharmacovigilance OR "adverse drug reaction*" OR "medication safety").

Foram incluídos estudos originais, incluindo ensaios clínicos randomizados, estudos observacionais e de coorte, estudos quase-experimentais, estudos de validação tecnológica, estudos de desenvolvimento e validação de modelos preditivos e estudos comparativos, publicados entre 1º de janeiro de 2021 e 1º de agosto de 2026 e relacionados à aplicação de IA em contextos farmacêuticos ou farmacoterapêuticos. Foram excluídos estudos de revisão narrativa ou sistemática, meta-análises, editoriais, cartas ao editor, protocolos, resumos de congresso, estudos exclusivamente relacionados à IA diagnóstica sem componente farmacêutico, estudos sem texto completo disponível e duplicatas.

Dos 8.622 registros brutos, 450 duplicatas foram removidas, resultando em 8.172 registros únicos triados por título e resumo. Os 8.032 registros excluídos nessa etapa distribuíram-se entre: diagnóstico por imagem sem enfoque farmacológico (n = 1.840), IA em saúde sem relação com medicamentos (n = 2.910), estudos cirúrgicos sem componente farmacoterapêutico (n = 720), estudos epidemiológicos sem componente de IA (n = 1.340) e outros critérios de escopo não relacionados à prática farmacêutica (n = 1.222). Dos 140 artigos avaliados em texto completo, 109 foram excluídos por se tratar de revisões narrativas (n ≈ 70) ou revisões sistemáticas e meta-análises (n ≈ 39), por não atenderem ao critério de inclusão de estudos originais, resultando em 31 estudos selecionados para a síntese. O fluxo completo de identificação, triagem e inclusão dos estudos está representado na Figura 1, elaborada conforme as recomendações do PRISMA 2020 (PAGE MJ, et al., 2021).

Figura 1 – Fluxograma do processo de identificação, triagem, elegibilidade e inclusão dos estudos, elaborado conforme as recomendações do PRISMA 2020 (PAGE et al., 2021)



Fonte: SILVA MEM et al., (2026).

A extração de dados utilizou instrumento estruturado contemplando autores, tipo de estudo, metodologia, tecnologia de IA, área de aplicação, métricas de desempenho, desfechos clínicos, limitações e perspectivas. Devido à heterogeneidade dos desenhos metodológicos e dos desfechos avaliados, os resultados foram sintetizados de forma narrativa e organizada por eixos temáticos. Não foi realizada metanálise quantitativa.

3 RESULTADOS

3.1 Características gerais dos estudos

Dos 140 artigos avaliados em texto completo, 31 estudos originais foram selecionados para esta síntese, compreendendo ensaios clínicos randomizados (RCT), estudos observacionais e de coorte, estudos quasi-experimentais, estudos de validação tecnológica, estudos de desenvolvimento e validação de modelos preditivos e estudos comparativos (Tabela 1).

Os 31 estudos selecionados foram publicados entre 2021 e 2026, com concentração nos anos de 2024 a 2026 (n = 22; 71%). Os países mais representados foram Estados Unidos (n = 6), China (n = 4) e Japão (n = 5), além de consórcios internacionais (n = 9). O perfil metodológico predominante compreendeu estudos de desenvolvimento ou validação de modelos (35,5%), seguidos por ensaios clínicos randomizados (19,4%). A América Latina foi representada por um único estudo original. A distribuição completa dos estudos por eixo temático, tipo de estudo e país de origem está apresentada na Figura 2.

3.2 Apoio à decisão clínica farmacêutica e gestão da adesão terapêutica

Wang L, et al. (2024) conduziram um ensaio clínico randomizado multicêntrico em cinco hospitais de cardiologia na China, com 52 semanas de seguimento e análise por intenção de tratar. A intervenção farmacêutica mediada por aplicativo WeChat com suporte de IA resultou em maior adesão medicamentosa no grupo intervenção (77,4% versus 58,6%; p<0,001), além de redução das hospitalizações por insuficiência cardíaca (HR = 0,72; IC95% 0,56–0,93; p=0,011) e melhora de 12,3 pontos na qualidade de vida pelo instrumento MLHFQ (p<0,001).

Tabela 1 – Caracterização dos 31 estudos originais incluídos na análise, organizados por eixo temático. Siglas: RCT = ensaio clínico randomizado; ML = machine learning; DL = deep learning; PLN = processamento de linguagem natural; MLGE = modelo de linguagem de grande escala; PK = farmacocinética; PopPK = farmacocinética de população; DDI = interação medicamentosa; AUC = área sob a curva ROC; F1 = F1-score; R^2 = coeficiente de determinação; κ = Kappa de Cohen; ROI = retorno sobre investimento; PDC = proporção de dias cobertos; RMSE = raiz do erro quadrático médio

Estudo	Autoria e ano	País	Tipo de estudo	Tecnologia de IA	Principal desfecho
EIXO I — IA no suporte à decisão clínica farmacêutica e adesão terapêutica					
E1	WANG et al., 2024	China	RCT multicêntrico	ML / app WeChat	Adesão 77,4% vs 58,6% (p<0,001); HR hospitalização=0,72 (IC95% 0,56-0,93)
E2	GUTIERREZ; SAKULBUMRUNGSIL, 2023	Tailândia	RCT	Sistema especialista (ML)	Controle PA 68,3% vs 42,9% (p=0,01); redução PAS -11,4 vs -5,2 mmHg
E3	WORRALL et al., 2025	EUA	Quasi-experimental	ML / plataforma IA	PDC +5,9% (HAS), +7,9% (colesterol), +6,4% (DM); 2.762 lacunas identificadas
E4	OHNO et al., 2025	Japão	Validação tecnológica	BERT-CRF (PLN)	F1-score=0,82; queda para 0,34-0,41 na transferência inter-hospitalar
E5	TSAI et al., 2025	EUA	RCT crossover	ML / eye-tracking	Deslocamento visual 19-26% para interface IA; dwell time 2,3 s vs 1,4 s (p<0,01)
E6	LESTER et al., 2025	EUA	RCT crossover	ML uncertainty-aware	Acurácia 91,3% vs 85,1% (p=0,024); tempo de reação -18,4%
E7	KESSLER et al., 2021	EUA	Observacional retrospectivo	ML / CMM-Wrap	ROI=12,4:1; redução custos totais -19,3% (p<0,001)

Estudo	Autoria e ano	País	Tipo de estudo	Tecnologia de IA	Principal desfecho
E8	DIL-NAHLIELI et al., 2024	Israel	Validação tecnológica	ML / API	Acurácia 88%; VPP 83%; VPN 91%; 5,04 PRM/paciente
E9	HUANG et al., 2025	Alemanha	Desenvolvimento de modelo	XGBoost / AiCure	AUC=0,92; HR recaída em aderentes=0,485
E10	OLENDER; ROY; NISHTALA, 2025	Reino Unido	Coorte prospectiva	XGBoost / DBI	AUC-ROC=0,863 (vs RF=0,780 e RL=0,608; p<0,001)
E11	KORB-SAVOLDELLI et al., 2023	França	Observacional transversal	CART (ML)	κ=0,62; instrumento de 5 itens aplicável em 3-4 min
EIXO II — IA na farmacovigilância e predição de reações adversas a medicamentos					
E12	KIRYU, 2023	Japão	Desenvolvimento de modelo	Random Forest / JADER	AUC=0,891; acurácia 87,4%; 3 RAM raras previamente subnotificadas identificadas
E13	PRATT et al., 2024	EUA	Consenso Delphi	Farmacogenômica + IA	5 variantes DPYD Tier 1 definidas; redução esperada de toxicidade grave de 30-40%
E14	IMAI, 2023	Japão	Análise de banco de dados	Rede bayesiana / JADER	Especificidade 0,91 vs PRR=0,74; 12 clusters de co-ocorrência RAM identificados
E15	AMBE, 2026	China	Desenvolvimento de modelo	XGBoost + SHAP	AUC-ROC=0,876; validação temporal prospectiva AUC=0,841
E16	VARGAS-NERI et al., 2026	América Latina	Descritivo / simpósio	PLN / NLP	Subnotificação RAM 94-98% na região; apenas 2 projetos PLN em pt/es identificados
E17	CHAN et al., 2025	Internacional	Validação tecnológica	XGBoost + SHAP (XAI) / FAERS	AUC=0,884; 23 sinais novos de RAM; SHAP concordou com mecanismo esperado em 79%
E18	LEE; CHEN, 2021	Japão	Desenvolvimento de modelo	DNN (SMILES + fingerprints)	AUC médio=0,876 (27 endpoints); AUC=0,901 para eventos adversos graves
EIXO III — IA em interações medicamentosas, farmacocinética e dosagem de precisão					
E19	KHA et al., 2024	Internacional	Validação prospectiva	GraphSAGE (GNN)	AUC=0,921; 15 DDI novas confirmadas prospectivamente em conjunto independente
E20	KOMURA et al., 2021	EUA / Internacional	Colaborativa público-privada	PBPK (SimCYP / GastroPlus)	AUC dentro de 2x do observado em 78% (SimCYP); banco com >10.000 parâmetros gerado
E21	GUO; DONG; HAO, 2024	China	Desenvolvimento de modelo	ChemBERTa / transfer learning	R^2=0,79 para clearance hepático; 90% do desempenho com apenas 500 moléculas
E22	UNO; NAKAMARU; YAMASHITA, 2024	China	Desenvolvimento de modelo	LASSO / ML-PopPK	Redução de 35% no tempo vs stepwise; detecção de 8,3% de concentrações atípicas
E23	FOCHESATO et al., 2026	Internacional	Desenvolvimento de modelo	NLME + ML (híbrido)	R^2=0,87 vs NONMEM=0,79; simulações 500x mais rápidas com >10.000 pacientes
EIXO IV — IA em descoberta de fármacos e avaliação crítica de sistemas de IA					
E24	ZHANG et al., 2026	França	Validação tecnológica	MLGEs (GPT-4 / Gemini / Claude)	Sensibilidade 40-82%; alucinação factual 6-23%; nenhum modelo atingiu sensib. >85% sem alucinações
E25	YANG et al., 2026	Japão	Estudo comparativo	MLGEs (4 modelos)	Acurácia GPT-4o=84,7%; menor alucinação Claude 3.5=6,1%; chain-of-thought +12-18%

Estudo	Autoria e ano	País	Tipo de estudo	Tecnologia de IA	Principal desfecho
E26	NGUYEN et al., 2026	EUA	Survey / análise curricular	ML (avaliação)	0/139 programas Pharm.D. com disciplina de IA; ~50% dos docentes com familiaridade limitada
E27	STEWART et al., 2026	EUA	Estudo piloto	Ambient AI (PLN)	Documentação 8,2→3,1 min (-62,2%); completude de campos 74%→89%
E28	ALAGARSAMY et al., 2026	China / Internacional	Desenvolvimento de modelo	ChemBERTa (transformer)	AUC médio=0,892 (12 benchmarks); Tox21=0,912; penetração BHE=0,936
E29	BITENCOURT-FERREIRA; SILVA; AZEVEDO, 2021	China / EUA	Desenvolvimento de modelo	GNN (afinidade de ligação)	RMSE=0,932 (CDK2); melhora 20-30% vs QSAR clássicos
E30	NISHIDA et al., 2026	EUA / Internacional	Avaliação comparativa	MLGEs (3 modelos) / oncologia	Concordância 74-81% com NCCN/ESMO; erros clinicamente perigosos em 2,7-7,3%
E31	PARKES; ZIAEE; O'REILLY, 2024	Internacional	Validação tecnológica	RF + DNN + DFT (híbrido)	Acurácia=91,4%; AUC=0,921 para seleção de co-cristais farmacêuticos

Fonte: SILVA MEM et al., 2026.

Figura 2 – Distribuição dos 31 estudos originais incluídos na revisão segundo critérios de eixo temático, perfil metodológico agrupado e a distribuição geográfica dos estudos por país ou região de origem



Fonte: SILVA MEM et al., 2026.

Gutierrez MM e Sakulbumrungsil R (2023) avaliaram, em ensaio clínico randomizado realizado na Tailândia, um sistema especialista baseado em regras derivadas de diretrizes para apoiar a orientação farmacêutica no manejo da hipertensão arterial. O grupo intervenção apresentou redução da pressão arterial sistólica de 11,4 mmHg versus 5,2 mmHg no grupo controle (p=0,02), com controle pressórico alcançado por 68,3% versus 42,9% dos participantes (p=0,01), além de melhora nos escores de adesão pelo Morisky-8.

Worrall C, et al. (2025) avaliaram uma plataforma de IA para gestão de adesão em estudo quasi-experimental multicêntrico com 10.477 pacientes nos EUA. A proporção de dias cobertos (PDC), métrica de adesão baseada em dispensação farmacêutica, melhorou para hipertensão

(+5,9%), colesterol (+7,9%) e diabetes (+6,4%), com p<0,01 para todos os indicadores. O sistema identificou automaticamente 2.762 lacunas de adesão acionáveis, com 60,6% dos pacientes beneficiados em pelo menos uma medida de qualidade.

Kessler S, et al. (2021) estimaram o impacto econômico de uma plataforma de IA para gestão farmacêutica em estudo observacional retrospectivo com 2.150 beneficiários do Medicaid da Califórnia. Foi estimado retorno sobre investimento (ROI) de 12,4:1, associado à redução de 19,3% nos custos totais de cuidado (p<0,001), de 15,1% nas visitas à emergência (p=0,002) e de 9,4% nas internações hospitalares (p=0,008), com economia estimada de 14 milhões de dólares por ano.

Dois RCTs crossover conduzidos na University of Michigan, nos Estados Unidos, investigaram a interação cognitiva entre farmacêuticos e sistemas de IA durante a verificação de medicamentos. Tsai CC, et al. (2025) avaliaram a interação entre farmacêuticos e sistemas de IA durante tarefas de verificação de medicamentos por meio de rastreamento ocular. A exposição a uma condição de baixa utilidade clínica deslocou 19% a 26% das fixações visuais para a interface, sem melhora na acurácia. Lester C et al. (2025) demonstraram que modelos uncertainty-aware, que exibem ao profissional a probabilidade de confiança da predição em vez de apenas fornecer a recomendação final, reduziram o tempo de reação em 18,4% (p=0,003; d de Cohen=0,71) e elevaram a acurácia de verificação de 85,1% para 91,3% (p=0,024) em relação a modelos caixa-preta, com redução adicional de 12,1 pontos no escore NASA-TLX de carga cognitiva.

Ohno Y, et al. (2025) validaram um modelo BERT-CRF (Bidirectional Encoder Representations from Transformers - Conditional Random Field), arquitetura de processamento de linguagem natural (PLN) que combina representação bidirecional de texto com identificação de entidades nomeadas como medicamentos, diagnósticos e sintomas, em 12.004 registros de cuidado farmacêutico hospitalar no Japão. O F1-score, métrica que pondera sensibilidade e precisão do modelo, variou de 0,67 com 1.200 registros de treinamento a 0,82 com o corpus completo, atingindo um patamar de eficiência com apenas 30% dos dados. A transferência do modelo para outro hospital resultou em queda do F1-score para 0,34 - 0,41.

Dil-Nahlieli D, et al. (2024) validaram o sistema Artificial Pharmacology Intelligence (API; MDI Health) em 49 pacientes polimedicados com registros eletrônicos ambulatoriais, tendo como padrão-ouro a avaliação por comitê de farmacêuticos especialistas. O sistema identificou uma média de 5,04 problemas relacionados a medicamentos (PRM) por paciente,

com valor preditivo positivo (VPP) de 83%, valor preditivo negativo (VPN) de 91%, sensibilidade de 79%, especificidade de 94% e acurácia global de 88%.

Olender RT, et al. (2025) demonstraram, em coorte prospectiva com 86.870 adultos com 65 anos ou mais do UK Biobank, que XGBoost combinado ao Drug Burden Index (DBI), índice que quantifica a carga farmacológica anticolinérgica e sedativa de um esquema medicamentoso, alcançou AUC-ROC de 0,863 para readmissão hospitalar em 30 dias, superando o Random Forest (AUC=0,780; p<0,001) e a regressão logística (AUC=0,608; p<0,001). O DBI foi o principal preditor pelo método SHAP (SHapley Additive exPlanations).

Korb-Savoldelli V, et al. (2023) desenvolveram instrumento reduzido de autorrelato de adesão com cinco itens derivados por árvore de decisão CART e consenso Delphi, demonstrando concordância substancial com avaliação multidisciplinar completa (coeficiente kappa = 0,62) e aplicabilidade em 3 a 4 minutos de preenchimento ambulatorial.

3.3 Farmacovigilância e predição de reações adversas a medicamentos

Os estudos de farmacovigilância apresentaram desempenho discriminativo variável conforme a base de dados, o algoritmo e o desfecho analisado. Kiryu Y (2023) comparou quatro algoritmos de machine learning em validação cruzada de dez partições, com Random Forest obtendo acurácia de 87,4% e AUC de 0,891, além de identificar três RAM raras previamente subnotificadas para antidiabéticos orais. Imai S (2023) combinou análise de correspondência múltipla, clustering k-means e redes bayesianas ao mesmo banco, identificando 12 clusters de co-ocorrência de RAM não capturados pelo método convencional de desproporcionalidade (PRR, Proportional Reporting Ratio). A rede bayesiana apresentou especificidade de 0,91 para causalidade farmacológica versus 0,74 do PRR.

Ambe K (2026) desenvolveu modelo preditivo de RAM hospitalares na China, utilizando dados retrospectivos com variáveis demográficas, histórico medicamentoso, resultados laboratoriais e comorbidades. O algoritmo XGBoost obteve AUC-ROC de 0,876, acurácia de 83,4%, sensibilidade de 79,2% e especificidade de 85,9%. Na validação temporal em coorte prospectiva independente, o AUC se manteve em 0,841.

Lee CY e Chen YPP (2021) desenvolveram rede neural profunda (DNN) alimentada por representações estruturais moleculares (SMILES e fingerprints de Morgan) para predição de eventos adversos, com base nos bancos FAERS (FDA Adverse Event Reporting System),

SIDER e OFFSIDES. O modelo alcançou AUC médio de 0,876 em 27 endpoints avaliados e de 0,901 especificamente para eventos adversos graves na validação temporal.

Chan FY, et al. (2025) desenvolveram sistema de inteligência artificial explicável (XAI) em farmacovigilância, baseado em XGBoost com análise SHAP, aplicado a mais de 18 milhões de relatórios do FAERS. O sistema alcançou AUC de 0,884 e identificou 23 sinais novos de RAM confirmados por comparação com a base OFFSIDES. As explicações SHAP geradas concordaram com o mecanismo farmacológico esperado em 79% dos casos avaliados por especialistas.

Pratt VM, et al. (2024) conduziram estudo de consenso Delphi modificado com oito organizações científicas internacionais, incluindo o CPIC (Clinical Pharmacogenetics Implementation Consortium) e o PharmGKB (Pharmacogenomics Knowledgebase), para definição de recomendações de genotipagem do gene DPYD, responsável pelo metabolismo de quimioterápicos fluoropirimidínicos como 5-fluorouracil e capecitabina. Foram definidas cinco variantes de rastreamento obrigatório (Tier 1), com estimativa de redução de 30% a 40% na toxicidade grave com implementação prospectiva. Vargas-Neri JL, et al. (2026) documentaram, a partir de dois simpósios da ISoP LATAM com especialistas de 12 países, subnotificação de RAM entre 94% e 98% na América Latina e identificação de apenas dois projetos de PLN em espanhol ou português voltados à farmacovigilância em toda a região.

3.4 Interações medicamentosas, farmacocinética e dosagem individualizada

Kha QH, et al. (2024) desenvolveram e validaram prospectivamente modelo de predição de DDI em antidiabéticos orais, comparando as arquiteturas LightGBM, GraphSAGE e MolFormer. O GraphSAGE (Graph Sample and Aggregate), tipo de rede neural de grafo (GNN) que representa moléculas codificando átomos como nós e ligações químicas como arestas, alcançou AUC de 0,921. A validação em conjunto independente de DDI documentadas em 2023 e não presentes no treinamento confirmou prospectivamente 15 novas interações clinicamente relevantes (KHA QH, et al., 2024).

Komura H, et al. (2021) reportaram parceria público-privada entre FDA (Food and Drug Administration) e PhRMA (Pharmaceutical Research and Manufacturers of America) para padronização de modelos PBPK (Physiologically Based Pharmacokinetics), abordagem que estima parâmetros farmacocinéticos a partir de dados fisiológicos mensuráveis. Os modelos demonstraram predição da área sob a curva (AUC) dentro de duas vezes o valor observado em

78% dos casos no SimCYP, com banco de mais de 10.000 parâmetros de alta qualidade gerado pela iniciativa.

Guo W, et al. (2024) demonstraram que o transfer learning, estratégia em que modelo pré-treinado em grande volume de dados moleculares é adaptado para tarefas específicas com dados reduzidos, aqui aplicado ao preditor ChemBERTa, alcançou coeficiente de determinação R^2 de 0,79 para clearance hepático, superando modelos como de Random Forest (R^2=0,67) e redes de passagem de mensagem (message passing) moleculares (R^2=0,73), de modo que o ajuste fino com apenas 500 moléculas recuperou 90% do desempenho obtido com 5.000 exemplos. Fochesato A, et al. (2026) propuseram modelos híbridos integrando modelagem não linear de efeitos mistos (NLME), abordagem farmacométrica padrão para estimativa de parâmetros farmacocinéticos populacionais e sua variabilidade interindividual, com machine learning para inibidores de CDK4/6 (Cyclin-Dependent Kinase 4/6) no câncer de mama. O modelo híbrido NLME+ML alcançou R^2=0,87 contra 0,79 do NONMEM convencional, com modelos substitutos 500 vezes mais rápidos em simulações com mais de 10.000 pacientes.

3.5 Descoberta de fármacos e avaliação crítica de modelos de linguagem de grande escala

Alagarsamy S, et al. (2026) avaliaram o desempenho do modelo ChemBERTa em 12 benchmarks padronizados de propriedades farmacológicas. O modelo apresentou AUC média de 0,892, superior aos modelos comparadores nas condições avaliadas, com AUC de 0,912 para toxicidade celular e de 0,936 para penetração da barreira hematoencefálica.

Zhang H, et al. (2026), na França, compararam três MLGE (GPT-4, Google Gemini e Anthropic Claude) na detecção de DDI que geraram notificações na base nacional de farmacovigilância (BNPV). A sensibilidade para detecção de DDI graves variou de 40% a 82% conforme o modelo, e as taxas de alucinação factual variaram de 6% a 23%. Nenhum dos três modelos atingiu simultaneamente sensibilidade superior a 85% e ausência de alucinações.

Yang SS, et al. (2026), no Japão, avaliaram quatro MLGE (GPT-4o, Claude 3.5 Sonnet, Gemini 1.5 Pro e Llama 3.1-70B) com quatro estratégias de prompt engineering em 200 questões padronizadas de consulta farmacêutica. A estratégia chain-of-thought, que instrui o modelo a raciocinar explicitamente passo a passo antes de fornecer a resposta, melhorou a acurácia em 12% a 18% em relação ao zero-shot em todos os modelos. O GPT-4o com *structured prompting* obteve 84,7% de acurácia, e o Claude 3.5 apresentou a menor taxa de alucinação (6,1%).

Nishida M, et al. (2026) verificaram que recomendações geradas por GPT-4, Claude 3 e Gemini 1.5 para 150 cenários de câncer de mama avançado apresentaram concordância de 74,0% a 81,3% com os guidelines NCCN/ESMO, mas com taxas de erros clinicamente perigosos entre 2,7% e 7,3%, sem incorporação automática de atualizações dos guidelines.

Nguyen KA, et al. (2026) realizaram análise curricular de 139 programas Pharm.D. (Doctor of Pharmacy) acreditados pela ACPE (Accreditation Council for Pharmacy Education) nos EUA, identificando ausência de disciplina dedicada à IA em todos os programas avaliados em 2024, com aproximadamente 50% do corpo docente reportando familiaridade limitada com o tema. Stewart MP, et al. (2026) avaliaram, em estudo piloto, implementação de Ambient AI, ferramenta que captura e estrutura automaticamente as informações das interações farmacêutico-paciente por transcrição de áudio ambiental em tempo real. O tempo de documentação por consulta reduziu de 8,2 para 3,1 minutos (queda de 62,2%), com completude dos campos obrigatórios passando de 74% para 89% e satisfação de 4,2/5,0 na escala Likert.

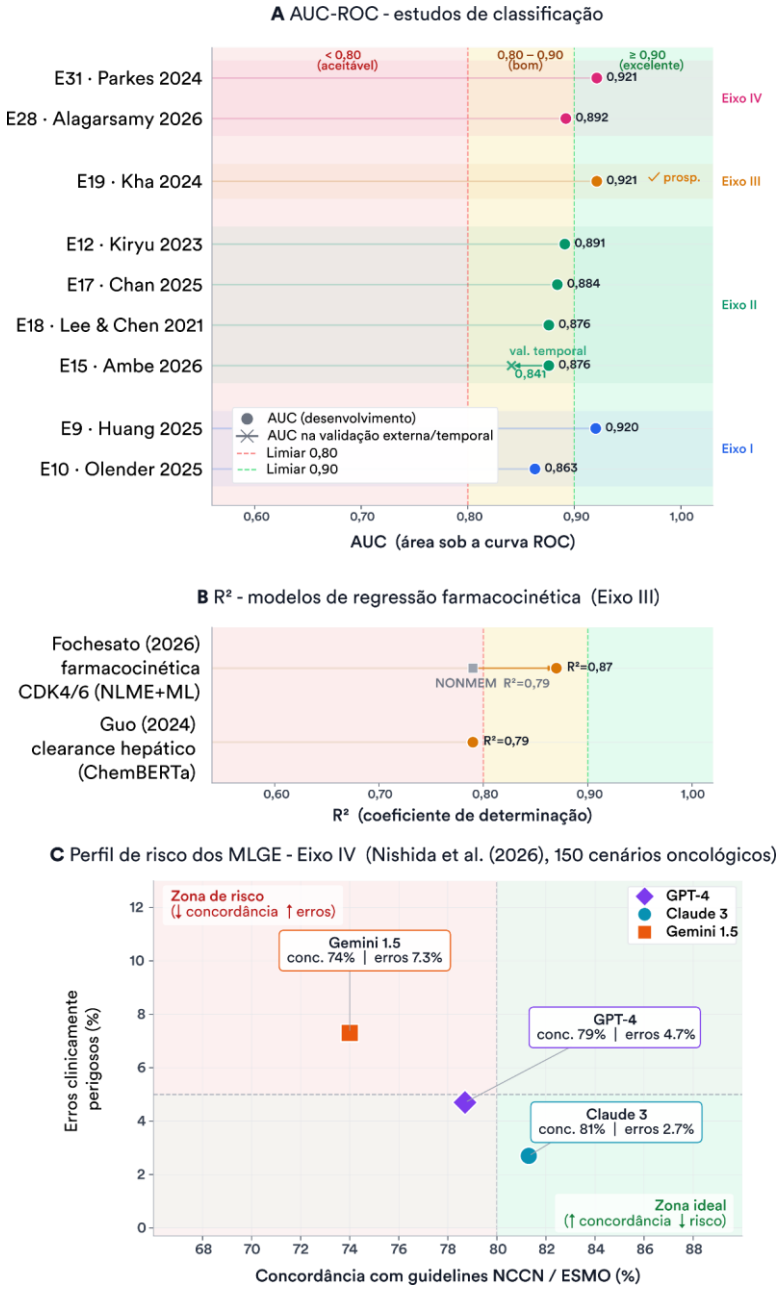
3.6 Síntese das métricas de desempenho por tipo e domínio de aplicação

A Figura 3 organiza o desempenho dos modelos de IA em três painéis independentes, cada um correspondendo a uma classe de métrica com natureza e escala de interpretação próprias: o Painel A apresenta a AUC-ROC dos estudos de classificação organizados por eixo temático, com faixas de cor delimitando os intervalos aceitável (< 0,80), bom (0,80–0,90) e excelente (≥ 0,90), sendo que a seta indica queda de desempenho na validação externa ou temporal; o Painel B apresenta o coeficiente de determinação R^2 dos modelos de regressão farmacocinética do Eixo III, com o quadrado cinza indicando o desempenho do modelo convencional de referência; e o Painel C apresenta o perfil de risco dos MLGE avaliados por Nishida et al. (2026) em 150 cenários oncológicos, com o eixo horizontal indicando concordância com os guidelines NCCN/ESMO (%) e o eixo vertical a taxa de erros clinicamente perigosos (%), sendo a região inferior direita o perfil ideal de desempenho.

Na Figura 3, o Painel A apresenta nove estudos com AUC-ROC entre 0,876 e 0,921, destacando E19 (Kha et al., 2024), único com validação prospectiva independente, e E15 (Ambe et al., 2026), cuja queda temporal da AUC para 0,841 sugere degradação do desempenho. No Painel B, os modelos farmacocinéticos superaram as referências convencionais, com R^2 de 0,87 para E23 e 0,79 para E21, métricas não comparáveis às AUC. No Painel C, nenhum modelo de MLGE combinou elevada concordância com as diretrizes NCCN/ESMO e baixa taxa de erros

perigosos, evidenciando que desempenho técnico não garante segurança clínica na decisão terapêutica oncológica.

Figura 3 – Desempenho dos modelos de IA por tipo de métrica e domínio de aplicação, organizado em três painéis com métricas distintas e não intercambiáveis: AUC-ROC para modelos de classificação (A), coeficiente de determinação R² para modelos de regressão farmacocinética (B) e perfil de risco dos modelos de linguagem de grande escala em cenários oncológicos (C)



Fonte: SILVA MEM et al., 2026.

4 DISCUSSÃO

O achado central desta revisão é que a IA apresenta potencial para apoiar diferentes dimensões da prática farmacêutica, embora a força das evidências varie substancialmente conforme o desenho do estudo, o contexto de validação e o domínio de aplicação. Os ensaios clínicos randomizados de Wang L, et al. (2024) e Gutierrez MM e Sakulbumrungsil R (2023) fornecem evidências de efeito causal das intervenções avaliadas sobre desfechos clínicos relacionados à adesão terapêutica e ao controle pressórico. Entretanto, esses resultados não podem ser diretamente extrapolados para estudos de desenvolvimento e validação de modelos preditivos, que constituíram parcela expressiva dos estudos incluídos e avaliaram predominantemente métricas de desempenho algorítmico.

A comparação entre os domínios analisados sugere diferentes graus de evidência para aplicação clínica. Na farmacovigilância, os estudos incluídos apresentaram desempenho discriminativo relevante, embora muitos tenham utilizado bases históricas de notificações espontâneas, limitando a avaliação de desempenho em condições prospectivas. Uma exceção é o estudo de Pratt VM, et al. (2024), que apresentou recomendações de implementação relacionadas à genotipagem DPYD. Na farmacocinética, os modelos híbridos e as estratégias de transfer learning apresentaram resultados promissores nas condições avaliadas, embora permaneçam limitadas as evidências provenientes de validações prospectivas multicêntricas.

A queda do F1-score de 0,82 para 0,34–0,41 observada por Ohno Y, et al. (2025) após a transferência do modelo para outro hospital evidencia o impacto potencial da mudança de contexto sobre o desempenho algorítmico. Esse resultado é compatível com problemas de generalização associados à diferença entre os dados utilizados no desenvolvimento do modelo e aqueles encontrados em ambientes externos. Kha QH, et al. (2024), por sua vez, demonstraram uma estratégia de validação prospectiva em conjunto independente, constituindo uma referência metodológica relevante entre os estudos incluídos.

O achado de Yang SS, et al. (2026), segundo o qual a estratégia chain-of-thought esteve associada a aumento de 12% a 18% na acurácia em comparação ao zero-shot, evidencia que a forma de interação com o modelo pode influenciar seu desempenho. Esse resultado reforça a importância de capacitar os profissionais para formular consultas, interpretar criticamente as respostas e reconhecer limitações dos sistemas generativos.

Entre os estudos incluídos nesta revisão, não foram identificados trabalhos que avaliassem de forma abrangente o impacto clínico de MLGE em condições rotineiras de prática

farmacêutica. Estudos prospectivos de implementação e impacto clínico permanecem, portanto, necessários para determinar a efetividade e a segurança dessas tecnologias em contextos reais.

No contexto latino-americano, os achados de Nguyen KA, et al. (2026) e Vargas-Neri JL, et al. (2026) apontam lacunas relacionadas à formação em IA e à utilização de ferramentas de processamento de linguagem natural em farmacovigilância. No contexto brasileiro, esses resultados sugerem oportunidades para investigação do uso de PLN em sistemas de notificação de eventos adversos, incluindo o NOTIVISA, especialmente considerando a necessidade de aprimorar a identificação e a análise de sinais de segurança. A regulamentação do papel do farmacêutico diante das tecnologias digitais e da inteligência artificial, mencionada no manuscrito por meio da Resolução CFF nº 10/2024, indica reconhecimento institucional da transformação tecnológica da prática profissional. Entretanto, a implementação segura dessas ferramentas também depende de capacitação profissional, infraestrutura de dados e mecanismos de validação compatíveis com os diferentes contextos de utilização.

Entre as limitações desta revisão estão a elevada heterogeneidade dos desenhos metodológicos, das populações, das tecnologias avaliadas e dos desfechos, o que inviabilizou a realização de uma síntese quantitativa comparável entre os estudos. Deve-se considerar também a possibilidade de viés de publicação, uma vez que estudos com resultados positivos podem apresentar maior probabilidade de publicação; contudo, essa possibilidade não foi formalmente mensurada nesta revisão.

5 CONSIDERAÇÕES FINAIS

Os estudos analisados indicam potencial da inteligência artificial para apoiar a segurança, a eficiência e a racionalidade do uso de medicamentos em diferentes contextos da prática farmacêutica. Entretanto, a heterogeneidade metodológica, as limitações de validação externa, a dependência dos dados de treinamento e os riscos associados aos modelos generativos ainda restringem a generalização dos resultados e sua adoção autônoma na prática clínica. Assim, o principal desafio da IA na farmácia não é o desempenho computacional, mas sua translação clínica, exigindo validações prospectivas multicêntricas, regulamentação específica, atualização da formação farmacêutica e tecnologias adaptadas aos contextos regionais. No Brasil, a integração de PLN em português ao NOTIVISA da ANVISA destaca-se como uma oportunidade de alto impacto para a segurança dos medicamentos. Os achados sustentam a utilização da IA como ferramenta de apoio ao trabalho farmacêutico, e não como substituta do

julgamento clínico profissional, desde que sua utilização seja acompanhada de validação adequada, supervisão humana, transparência e capacitação dos profissionais.

REFERÊNCIAS

ALAGARSAMY S, et al. Transformers in drug discovery: fine-tuning ChemBERTa for high-accuracy prediction of solubility, toxicity and binding affinity. **Drug Discovery Today**, 2026; 31(1): 104602.

AMBE K. Development of a Data-Driven Prediction Model of Adverse Drug Reactions Using Large-Scale Medical Information and Machine Learning. **Biological & Pharmaceutical Bulletin**, 2026; 49(2): 213-219.

CHAN FY, et al. Web-Based Explainable Machine Learning-Based Drug Surveillance for Predicting Sunitinib- and Sorafenib-Associated Thyroid Dysfunction: Model Development and Validation Study. JMIR Formative Research, 2025; 9: e67767.52

DIL-NAHLIELI D, et al. Validation of a novel Artificial Pharmacology Intelligence (API) system for the management of patients with polypharmacy. Research in Social & Administrative Pharmacy, 2024; 20(7): 633-639.

FOCHESATO A, et al. Building Hybrid Pharmacometric-Machine Learning Models in Oncology Drug Development: Current State and Recommendations. CPT: Pharmacometrics & Systems Pharmacology, 2026; 15(1): e70113.

GANGWAL A, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. Frontiers in Pharmacology, 2024; 15: 1331062.

GUO W, et al. Transfer learning empowers accurate pharmacokinetics prediction of small samples. Drug Discovery Today, 2024; 29(4): 103946.

GUTIERREZ MM, SAKULBUMRUNGSIL R. Effectiveness of a pharmacist-led expert system intervention for medication adherence and blood pressure control of adults with hypertension: A randomized controlled trial. Research in Social & Administrative Pharmacy, 2023; 19(6): 931-943.

IMAI S. A Novel Approach to Analyze the Factors Affecting Adverse Drug Reactions by Combination of Electronic Medical Record Database and Machine Learning Method. Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan, 2023; 143(6): 485-489.

KESSLER S, et al. Economic and utilization outcomes of medication management at a large Medicaid plan with disease management pharmacists using a novel artificial intelligence platform from 2018 to 2019. Journal of Managed Care & Specialty Pharmacy, 2021; 27(9): 1186-1196.

KHA QH, et al. Development and validation of a machine learning model for predicting drug-drug interactions with oral diabetes medications. Methods, 2024; 232: 81-88.

KIRYU Y. Development of a Medical Big Data Analysis System Utilizing Artificial Intelligence Analytics in Clinical Pharmacy. Yakugaku Zasshi: Journal of the Pharmaceutical Society of Japan, 2023; 143(6): 501-505.

KOMURA H, et al. A public-private partnership to enrich the development of in silico predictive models for pharmacokinetic and cardiotoxic properties. Drug Discovery Today, 2021; 26(5): 1275-1283.

KORB-SAVOLDELLI V, et al. Psychometric Properties of a Machine Learning-Based Patient-Reported Outcome Measure on Medication Adherence: Single-Center, Cross-Sectional, Observational Study. Journal of Medical Internet Research, 2023; 25: e42384.

LEE CY, CHEN YPP. Prediction of drug adverse events using deep learning in pharmaceutical discovery. Briefings in Bioinformatics, 2021; 22(2): 1884-1901.

LESTER C, et al. Effect of Uncertainty-Aware AI Models on Pharmacists' Reaction Time and Decision-Making in a Web-Based Mock Medication Verification Task: Randomized Controlled Trial. JMIR Medical Informatics, 2025; 13: e64902.

NGUYEN KA, et al. Preparing Future Pharmacists for an AI-Enhanced Health Care Environment: Educational Strategies and Implementation Challenges. Journal of the American College of Clinical Pharmacy, 2026; 9(2): e70162.

NISHIDA M, et al. Comparative evaluation of treatment recommendations generated by generative AI and breast cancer specialists for advanced and recurrent breast cancer. Breast Cancer, 2026; 33(4): 841-852.

OHNO Y, et al. Performance Improvement of a Natural Language Processing Tool for Extracting Patient Narratives Related to Medical States From Japanese Pharmaceutical Care Records by Increasing the Amount of Training Data. JMIR Medical Informatics, 2025; 13: e68863.

OLENDER RT, et al. Potentially inappropriate polypharmacy is an important predictor of 30-day emergency hospitalisation in older adults: a machine learning feature validation study. Age and Ageing, 2025; 54(6): afaf156.

PAGE MJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, 2021; n71.

PRATT VM, et al. DPYD Genotyping Recommendations: A Joint Consensus Recommendation of the Association for Molecular Pathology and collaborating organizations. The Journal of Molecular Diagnostics, 2024; 26(10): 851-863.

ROTHBARTH R. CFF regulamenta atuação farmacêutica em saúde digital e inteligência artificial. São Paulo: Machado Meyer Advogados, 2024.

STEWART MP, et al. Using an Ambient Artificial Intelligence Tool to Document Clinical Pharmacist Patient Encounters: Real-World Evidence From a Federally Qualified Health Center. Journal of the American College of Clinical Pharmacy, 2026; 9(7): e70242.

TSAI CC, et al. Effect of Artificial Intelligence Helpfulness and Uncertainty on Cognitive Interactions with Pharmacists: Randomized Controlled Trial. Journal of Medical Internet Research, 2025; 27: e59946.

VARGAS-NERI JL, et al. Strengthening Pharmacovigilance in Latin America: Multidisciplinary Insights from Two ISoP Latin American Chapter Symposia. Drug Safety, 2026; 49(8): 871-882.

WANG L, et al. Pharmacist-Led Management Model and Medication Adherence Among Patients With Chronic Heart Failure: A Randomized Clinical Trial. JAMA Network Open, 2024; 7(12): e2453976.

WORRALL C, et al. Impact of a clinical pharmacist-led, artificial intelligence-supported medication adherence program on medication adherence performance, chronic disease control measures, and cost savings. Journal of the American Pharmacists Association, 2025; 65(1): 102271.

YANG SS, et al. Comparison of Large Language Model-Based Systems and Prompt Engineering for Internal Medicine Clinical Pharmacy Cases. Journal of the American College of Clinical Pharmacy, 2026; 9(7): e70235.

ZHANG H, et al. Performance validation of large language models for drug-drug interaction consultations: Potential pharmacist assistant in outpatient pharmaceutical services. Journal of the American Pharmacists Association, 2026; 66(3): 103077.