

REVISÃO DE LITERATURA IDENTIFICADO

INTELIGÊNCIA ARTIFICIAL GENERATIVA APLICADA À DOCUMENTAÇÃO CLÍNICA E AO PRONTUÁRIO ELETRÔNICO: UMA REVISÃO INTEGRATIVA GENERATIVE ARTIFICIAL INTELLIGENCE APPLIED TO CLINICAL DOCUMENTATION AND ELECTRONIC HEALTH RECORDS: AN INTEGRATIVE REVIEW

Larissa Leite de Souza Leão^{1*}, Gabriel Henrique Carvalho Leite Romeiro¹, Vitor Comassetto Paes¹, Jussara Guimarães da Rocha Lima¹, Joana Santos Cesar de Carvalho¹, Reinaldo dos Santos Moura²

¹Acadêmicos de Medicina, Afya Centro Universitário de Maceió (UNIMA); ²Docente do Curso de Medicina, Afya Centro Universitário de Maceió (UNIMA).

*Autora principal. E-mail: larissaleao@alunos.afya.com.br

Eixo: 03 - Tecnologia e Inovação

RESUMO

O tempo dedicado à documentação clínica no prontuário eletrônico é apontado como um dos principais determinantes do esgotamento profissional entre médicos, superando com frequência o tempo de contato direto com o paciente. Assistentes de documentação baseados em inteligência artificial (IA) generativa, os chamados “scribes ambientais”, surgiram como estratégia para automatizar a geração de notas clínicas a partir da conversa médico-paciente, reduzindo a carga administrativa. Objetivo: sintetizar as evidências científicas disponíveis sobre o uso de IA generativa na documentação clínica e no prontuário eletrônico, considerando ganhos de eficiência, segurança e aceitação. Método: revisão integrativa de literatura, com busca em PubMed/MEDLINE, Europe PMC, JMIR e NEJM AI, priorizando estudos publicados entre 2024 e 2026. Resultados: foram incluídas 14 referências. Ensaio clínico randomizado com 238 médicos mostrou redução modesta no tempo de documentação e melhora em indicadores de esgotamento com uma das duas ferramentas testadas, sem diferença significativa com a outra. Estudos de simulação identificaram taxa média de erro de 26,3% nas notas geradas, incluindo alucinações com potencial de dano clínico. Aplicações em sumarização de alta hospitalar e em psiquiatria mostraram resultados mistos, com achado paradoxal de menor desfecho terapêutico apesar de notas mais detalhadas. Conclusão: a IA generativa aplicada à documentação clínica apresenta potencial real de redução de carga administrativa, mas a evidência de segurança ainda é limitada e heterogênea, exigindo supervisão humana e validação específica por especialidade antes de adoção em larga escala.

Palavras-chave: Inteligência Artificial; Registros Eletrônicos de Saúde; Documentação; Esgotamento Profissional; Processamento de Linguagem Natural.

ABSTRACT

The time spent on clinical documentation in electronic health records is identified as one of the main determinants of physician burnout, often exceeding time spent in direct contact with patients. Documentation assistants based on

generative artificial intelligence (AI), the so-called “ambient scribes,” have emerged as a strategy to automate the generation of clinical notes from physician-patient conversations, reducing administrative burden. Objective: to synthesize the available scientific evidence on the use of generative AI in clinical documentation and electronic health records, considering gains in efficiency, safety, and acceptance. Methods: integrative literature review, with searches in PubMed/MEDLINE, Europe PMC, JMIR, and NEJM AI, prioritizing studies published between 2024 and 2026. Results: 14 references were included. A randomized clinical trial with 238 physicians showed a modest reduction in documentation time and improvement in burnout indicators with one of the two tools tested, with no significant difference for the other. Simulation studies identified a mean error rate of 26.3% in generated notes, including hallucinations with potential for clinical harm. Applications in hospital discharge summarization and psychiatry showed mixed results, with a paradoxical finding of a lower rate of concrete therapeutic action despite more detailed notes. Conclusion: generative AI applied to clinical documentation shows real potential to reduce administrative burden, but safety evidence remains limited and heterogeneous, requiring human oversight and specialty-specific validation before large-scale adoption.

Keywords: Artificial Intelligence; Electronic Health Records; Documentation; Burnout, Professional; Natural Language Processing.

1. INTRODUÇÃO

A documentação clínica no prontuário eletrônico consome parcela expressiva da jornada de trabalho médica. Estudo de tempo e movimento com 57 médicos de quatro especialidades mostrou que praticamente metade do tempo assistencial é dedicada ao prontuário e a tarefas administrativas, enquanto o contato direto com o paciente ocupa pouco mais de um quarto desse tempo¹. Essa sobrecarga documental é reconhecida como determinante direto do esgotamento profissional, o que tem motivado a discussão explícita sobre se a inteligência artificial deveria assumir parte dessa tarefa².

Nesse cenário, surgiram os chamados “scribes ambientais”: ferramentas de IA generativa que, com consentimento do paciente, captam a conversa da consulta por áudio, transcrevem o conteúdo e utilizam um modelo de linguagem de grande porte para estruturar automaticamente uma nota clínica formal, extraíndo sintomas, histórico, avaliação e plano terapêutico. Diferentemente de um transcritor literal, o sistema reorganiza e sintetiza a informação em formato clínico, funcionando como assistente digital dentro da sala de consulta.

Apesar do rápido crescimento na adoção dessas ferramentas, a literatura sobre seu desempenho real permanece dispersa entre estudos de eficiência, de segurança e acurácia, e de aceitação por pacientes e profissionais, sem síntese recente que integre essas três dimensões de forma crítica. Diante disso, esta revisão busca responder: qual é o estado atual da evidência científica sobre o

uso de IA generativa na documentação clínica, considerando de forma conjunta ganhos de eficiência, riscos de segurança e aceitação por pacientes e profissionais?

2. OBJETIVO

Sintetizar, por meio de revisão integrativa, as evidências científicas disponíveis sobre o uso de inteligência artificial generativa na documentação clínica e no prontuário eletrônico, considerando ganhos de eficiência e impacto sobre o esgotamento profissional, segurança e acurácia das notas geradas, e aceitação por pacientes e profissionais.

3. MÉTODO

Trata-se de revisão integrativa de literatura, de caráter narrativo, sem metanálise, dada a heterogeneidade dos desenhos de estudo disponíveis. A busca foi realizada em agosto de 2026 nas bases PubMed/MEDLINE, Europe PMC, JMIR Publications e NEJM AI, combinando os descritores “artificial intelligence”, “generative AI”, “large language model”, “ambient scribe”, “clinical documentation”, “electronic health record” e “physician burnout”, priorizando publicações entre 2024 e 2026.

Foram incluídos estudos primários, ensaios clínicos, revisões sistemáticas e de escopo, e artigos de perspectiva revisados por pares que avaliassem diretamente o uso de IA generativa na geração ou no apoio à documentação clínica. Foram excluídos estudos de IA aplicada exclusivamente ao diagnóstico clínico sem relação com documentação, editoriais sem dado próprio e conteúdos sem revisão por pares. Todas as referências foram verificadas individualmente quanto a autoria, periódico, ano e identificador digital (DOI e/ou PMID) junto às bases PubMed e Europe PMC.

Como se trata de revisão de literatura, sem coleta direta de dados de pacientes, o presente trabalho está dispensado de submissão a Comitê de Ética em Pesquisa, conforme item 4.3 do edital do 6º AFYA Global Meeting, que exige parecer ético apenas para estudos com envolvimento direto ou indireto de seres humanos.

4. RESULTADOS

Eficiência e impacto sobre o esgotamento profissional

O estudo de maior rigor metodológico identificado é o ensaio clínico randomizado de Lukac et al., que distribuiu 238 médicos de 14 especialidades entre duas ferramentas de IA generativa (Microsoft Dragon Copilot e Nabla) e cuidado usual, acompanhando cerca de 72 mil consultas³. Uma das ferramentas reduziu o tempo de documentação em cerca de 10% em relação ao controle, enquanto a outra não mostrou diferença significativa; ambas se associaram a melhora em medidas

de experiência do médico e de carga de trabalho percebida, mas não necessariamente a redução consistente do tempo total. Revisão rápida de evidência do mundo real de Kanaparthi et al., reunindo estudos entre 2014 e 2024, converge com esse achado: ganhos de eficiência são consistentes, mas heterogêneos entre plataformas e contextos⁴. Revisão narrativa de Razaghi et al. sintetiza esse mesmo padrão, destacando que a maior parte da evidência ainda vem de estudos observacionais de curto prazo, não de ensaios controlados⁵. Análise de barreiras à escala de Ohde et al. identificou que a adoção depende fortemente de infraestrutura tecnológica local e de treinamento da equipe, não apenas da qualidade do modelo de IA em si⁶.

Segurança, precisão e alucinações

O valor mais consistentemente demonstrado nesses estudos é, ao mesmo tempo, sua maior fragilidade: a precisão do conteúdo gerado. Anderson et al. avaliaram quatro plataformas comerciais de scribe ambiental em consultas simuladas e encontraram taxa média de erro de 26,3% nas notas geradas, com média de três erros por caso com potencial de dano moderado a grave⁷. Topaz et al. descrevem riscos ainda pouco documentados dessas ferramentas, incluindo alucinações, omissões críticas e atribuição incorreta de falas entre paciente e médico⁸. Palm et al., em avaliação validada da qualidade de notas geradas por um scribe baseado em modelo de linguagem, confirmam variabilidade relevante de qualidade entre casos clínicos, reforçando a necessidade de revisão humana sistemática antes da incorporação definitiva ao prontuário⁹. Em conjunto, esses três estudos sustentam que a capacidade demonstrada de gerar texto clinicamente plausível não equivale, por si só, a garantia de exatidão factual, distinção especialmente relevante para conteúdo sensível como medicação, alergia e diagnóstico.

Aplicações além da consulta ambulatorial e diferenças por especialidade

O uso de IA generativa na documentação não se limita à nota de consulta. Hains et al. avaliaram a preparação de resumos de alta hospitalar a partir de dados reais de prontuário eletrônico, com resultados promissores quanto à viabilidade técnica, mas ainda sem validação prospectiva de impacto clínico¹⁰. Em psiquiatria, Castro et al. encontraram achado paradoxal ao comparar consultas de atenção primária documentadas com e sem scribe ambiental: as notas com IA continham descrições mais detalhadas de sintomas psiquiátricos, mas a probabilidade de uma intervenção concreta (encaminhamento, novo diagnóstico ou prescrição) foi menor nas consultas assistidas por IA¹¹. Esse achado ilustra que maior volume ou detalhamento textual não implica,

necessariamente, melhor desfecho assistencial, e que efeitos específicos por especialidade ainda são pouco compreendidos.

Aceitação, consentimento e considerações éticas

A aceitação por pacientes é heterogênea. Levantamento pré-implementação de Leiserowitz et al. em centro médico acadêmico encontrou atitudes mistas, com parcela relevante dos pacientes relatando preocupação com precisão e privacidade antes de qualquer exposição real à tecnologia¹². Lawrence et al. discutem o desafio do consentimento informado para documentação ambiental com IA generativa em contexto ambulatorial, destacando que a natureza passiva da captação de áudio cria tensão com marcos regulatórios de privacidade já estabelecidos¹³. Editorial de Leung et al. sintetiza esses desafios, defendendo integração responsável que combine ganho de eficiência com transparência ativa ao paciente e supervisão clínica contínua, não substituição do julgamento médico¹⁴.

5. CONCLUSÃO

A inteligência artificial generativa aplicada à documentação clínica apresenta potencial real de redução da carga administrativa, com evidência de ensaio clínico randomizado mostrando ganhos modestos de eficiência para pelo menos uma ferramenta avaliada. Entretanto, a evidência de segurança ainda é limitada e heterogênea, com taxas de erro e alucinação relatadas não desprezíveis em conteúdo clinicamente sensível, e efeitos específicos por especialidade, como o achado paradoxal em psiquiatria, ainda pouco compreendidos.

Conclui-se que essas ferramentas não substituem a responsabilidade do médico pela nota clínica final, mas funcionam como apoio condicionado à revisão humana sistemática, consentimento informado explícito do paciente e validação específica por contexto de uso, não apenas por desempenho técnico geral do modelo.

REFERÊNCIAS

1. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Ann Intern Med*. 2016;165(11):753-760.
2. Miao J, Thongprayoon C, Cheungpasitporn W. Should Artificial Intelligence Be Used for Physician Documentation to Reduce Burnout? *Kidney360*. 2024;5(5):765-767.
3. Lukac PJ, Turner W, Vangala S, Chin AT, Khalili J, Shih YT, et al. Ambient AI Scribes in Clinical Practice: A Randomized Trial. *NEJM AI*. 2025;2(12).

4. Kanaparthi NS, Villuendas-Rey Y, Bakare T, Diao Z, Iscoe M, Loza A, et al. Real-World Evidence Synthesis of Digital Scribes Using Ambient Listening and Generative Artificial Intelligence for Clinician Documentation Workflows: Rapid Review. *JMIR AI*. 2025;4:e76743.
5. Razaghi M, Hafez A, Farina JM, Scalia IG, Pereyra M, Abdelfattah FE, et al. Transforming clinical documentation with ambient artificial intelligence (AI) scribes: a narrative review of technology, impact, and implementation. *Cardiovasc Diagn Ther*. 2026;16(1).
6. Ohde JW, Thompson A, Liu Z, Rost LM, Overgaard JD, Tan JYE, et al. Barriers and opportunities of scaling ambient AI scribes for clinical documentation across diverse healthcare settings. *NPJ Digit Med*. 2026;9(1):369.
7. Anderson TN, Mohan V, Dorr DA, Ratwani RM, Biro JM, Gold JA. Evaluating the Quality and Safety of Ambient Digital Scribe Platforms Using Simulated Ambulatory Encounters. *Mayo Clin Proc Digit Health*. 2025;3(4):100292.
8. Topaz M, Peltonen LM, Zhang Z. Beyond human ears: navigating the uncharted risks of AI scribes in clinical practice. *NPJ Digit Med*. 2025;8(1):569.
9. Palm E, Manikantan A, Mahal H, Belwadi SS, Pepin ME. Assessing the quality of AI-generated clinical notes: validated evaluation of a large language model ambient scribe. *Front Artif Intell*. 2025;8:1691499.
10. Hains L, Kleinig O, Murugappa A, Gluck S, Marks J, Gilbert T, et al. Large language model discharge summary preparation using real-world electronic medical record data shows promise. *Intern Med J*. 2025.
11. Castro VM, McCoy TH, Verhaak P, Ramachandiran A, Perlis RH. Psychiatric Documentation and Management in Primary Care With Artificial Intelligence Scribe Use. *JAMA Psychiatry*. 2026;83(3):281-286.
12. Leiserowitz G, Mansfield J, MacDonald S, Jost M. Patient Attitudes Toward Ambient Voice Technology: Preimplementation Patient Survey in an Academic Medical Center. *JMIR Med Inform*. 2025;13:e77901.
13. Lawrence K, Kuram VS, Levine DL, Sharif S, Polet C, Malhotra K, et al. Informed Consent for Ambient Documentation Using Generative AI in Ambulatory Care. *JAMA Netw Open*. 2025;8(7):e2522400.
14. Leung TI, Coristine AJ, Benis A. AI Scribes in Health Care: Balancing Transformative Potential With Responsible Integration. *JMIR Med Inform*. 2025;13:e80898.