

REVISÃO DE LITERATURA IDENTIFICADO

INTELIGÊNCIA ARTIFICIAL GENERATIVA NO TREINAMENTO DA
COMUNICAÇÃO DE MÁS NOTÍCIAS: REVISÃO DE ESCOPO

GENERATIVE ARTIFICIAL INTELLIGENCE IN BREAKING BAD NEWS
COMMUNICATION TRAINING: A SCOPING REVIEW

Vitor Comassetto Paes¹; Gabriel Henrique Carvalho Leite Romeiro¹; Tacy Santana Machado¹;
Jessyca Lane Fausto Lira¹; Luciano Timbo Barbosa¹; Reinaldo dos Santos Moura²

1 Acadêmico(a) de Medicina, Centro Universitário de Maceió (UNIMA/Afya), Maceió/AL

2 Docente de Medicina, Centro Universitário de Maceió (UNIMA/Afya), Maceió/AL

E-mail do autor principal: reinaldo.santos@afya.com.br

Eixo Temático: Eixo 03 - Tecnologia e Inovação

RESUMO

A comunicação de más notícias é competência clínica essencial, frequentemente malconduzida por falta de treinamento estruturado, e a disseminação de IA generativa, como os grandes modelos de linguagem (LLM), abriu via inédita de simulação escalável para essa habilidade, ainda que a extensão e a qualidade dessa produção científica não houvessem sido mapeadas. Objetivou-se mapear a extensão, a natureza e as lacunas da literatura sobre o uso de IA generativa no treinamento de comunicação de más notícias em estudantes e profissionais de saúde. Trata-se de revisão de escopo conduzida segundo a metodologia do Joanna Briggs Institute (JBI) e reportada conforme o checklist PRISMA-ScR, com buscas sem restrição de data em PubMed/MEDLINE, LILACS/BVS, ERIC e Semantic Scholar, critérios de elegibilidade explícitos e estrutura População-Conceito-Contexto, resultando em 457 registros identificados, 434 após remoção de duplicatas e 11 estudos incluídos após triagem em três etapas. Os 11 estudos incluídos foram publicados entre 2023 e 2026, predominantemente estudos piloto ou quase-experimentais com estudantes de medicina utilizando ChatGPT/GPT-4 em cenários de más notícias oncológicas, mamográficas ou de emergência, com autoconfiança autorreferida como desfecho mais relatado e melhora estatisticamente significativa na maioria dos estudos, mas ausência quase completa de desfechos objetivos e de grupos-controle. Conclui-se que a literatura sobre IA generativa aplicada a esse treinamento é recente, numericamente pequena e concentrada em estudos piloto sem desfechos objetivos, havendo lacuna clara para estudos controlados com avaliação de transferência da habilidade para a prática clínica real.

Palavras-chave: Inteligência Artificial. Revelação da Verdade. Educação Médica. Comunicação. Simulação de Paciente.

ABSTRACT

Breaking bad news is an essential clinical competency, frequently poorly conducted due to lack of structured training, and the spread of generative artificial intelligence (AI) tools, such as large language models (LLM), has opened an unprecedented avenue of scalable simulation for this skill, even though the extent and quality of this scientific production had not yet been mapped. The objective was to map the extent, nature, and gaps in the

literature on the use of generative AI in training breaking bad news communication among health students and professionals. This is a scoping review conducted according to the Joanna Briggs Institute (JBI) methodology and reported according to the PRISMA-ScR checklist, with searches without date restriction in PubMed/MEDLINE, LILACS/BVS, ERIC, and Semantic Scholar, explicit eligibility criteria, and a Population-Concept-Context structure, resulting in 457 records identified, 434 after duplicate removal, and 11 studies included after three-stage screening. The 11 included studies were published between 2023 and 2026, predominantly pilot or quasi-experimental studies with medical students using ChatGPT/GPT-4 in oncology, mammography, or emergency bad-news scenarios, with self-reported confidence as the most frequently reported outcome and statistically significant improvement in most studies, but an almost complete absence of objective outcomes and control groups. It is concluded that the literature on generative AI applied to this training is recent, numerically small, and concentrated in pilot studies without objective outcomes, with a clear gap for controlled studies assessing skill transfer to real clinical practice.

Keywords: Artificial Intelligence. Truth Disclosure. Education, Medical. Communication. Patient Simulation.

INTRODUÇÃO

A comunicação de más notícias, definida como qualquer informação que altere de forma grave e negativa a perspectiva do paciente sobre seu futuro, é uma das competências clínicas mais desafiadoras da prática em saúde. O protocolo SPIKES, formalizado há mais de duas décadas, tornou-se referência internacional para estruturá-la¹. A literatura mostra déficit persistente de treinamento formal, com impacto na experiência do paciente e no desgaste emocional do profissional, incluindo fadiga de compaixão². O treinamento tradicional depende de pacientes simulados, estratégia eficaz, porém onerosa e pouco escalável³.

A disseminação de IA generativa, sobretudo após o lançamento público de LLM como o ChatGPT em novembro de 2022, abriu via alternativa de simulação conversacional de baixo custo, disponível a qualquer momento e potencialmente escalável para treino de comunicação sensível. Diferentemente dos pacientes virtuais tradicionais, roteirizados e de interação limitada, os LLM geram respostas dinâmicas e contextualizadas, ampliando o realismo percebido da simulação.

Revisões de escopo já mapearam o uso de IA no treinamento de comunicação em saúde de forma geral⁴ e pacientes virtuais baseados em LLM na educação médica^{5,6}, além de uma revisão rápida sobre pacientes virtuais para más notícias não necessariamente baseados em IA generativa⁷. Nenhuma isola a interseção entre IA generativa e o treinamento específico da comunicação de más notícias, cenário singular pela carga emocional e pelo protocolo próprio consolidado. Esta revisão teve por objetivo mapear a extensão, a natureza e as lacunas da

literatura sobre IA generativa no treinamento de comunicação de más notícias em estudantes e profissionais de saúde.

MÉTODO

Trata-se de revisão de escopo conduzida segundo a metodologia da JBI e reportada conforme o PRISMA-ScR. A pergunta foi estruturada em formato PCC: estudantes, residentes e profissionais de saúde em formação ou prática; uso de IA generativa no treinamento de comunicação de más notícias; educação em saúde, sem restrição geográfica ou de data.

A busca foi realizada em quatro bases — PubMed/MEDLINE, LILACS/BVS, ERIC e Semantic Scholar —, cobrindo a produção biomédica geral, regional/latino-americana, em educação em saúde e interdisciplinar entre saúde e ciência da computação. As estratégias combinaram termos MeSH/DeCS e livres de IA generativa e más notícias/diagnóstico difícil (Tabela 1).

Critérios de elegibilidade a priori. Inclusão: ferramenta de IA generativa aplicada ao treinamento de comunicação de más notícias, diagnóstico difícil ou prognóstico grave, em estudantes/profissionais de saúde, qualquer desenho primário e idioma. Exclusão: IA não generativa; comunicação genérica sem menção explícita a más notícias; IA como ferramenta de análise; população de pacientes; revisões já publicadas (tratadas como literatura relacionada).

A seleção ocorreu em três etapas: triagem por título, por resumo e elegibilidade em texto completo, com critérios registrados e aplicados por revisor único, limitação discutida adiante. Cada estudo elegível foi avaliado também por critérios metodológicos básicos (grupo de comparação, cegamento, instrumento validado, tamanho amostral), reportados na síntese, prática recomendada pela JBI ainda que não obrigatória em revisão de escopo.

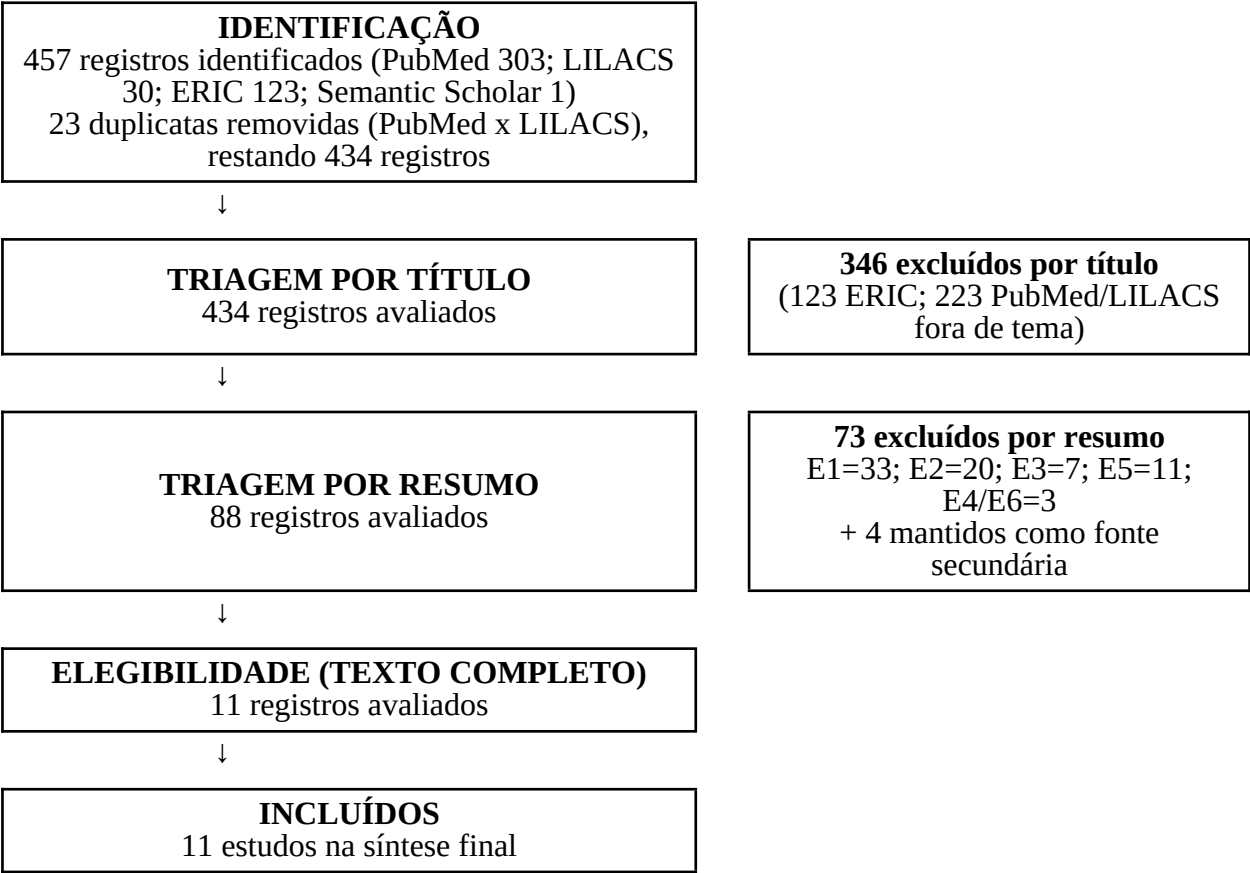
A busca identificou 457 registros (303 PubMed/MEDLINE, 30 LILACS/BVS, 123 ERIC, 1 Semantic Scholar); após remoção de 23 duplicatas, 434 avançaram à triagem por título, que excluiu 346 (123 do ERIC, majoritariamente sobre IA no ensino de idiomas). Avançaram 88 registros à triagem por resumo, incluindo 6 capturados apenas pelo LILACS, não identificados pelo PubMed, evidenciando o valor da busca em múltiplas bases.

Na triagem por resumo, 73 registros foram excluídos (E1=33; E2=20; E5=11; E3=7; E4/E6=3); 4 foram mantidos como fonte secundária de contexto, incluindo um estudo em que a IA generativa foi usada como ferramenta de análise de dados sobre más notícias, não como intervenção de treinamento⁸. Os 11 restantes preencheram todos os critérios em texto completo e compuseram a síntese final (Figura 1).

Base	String de busca (síntese)
PubMed/MEDLINE	(IA generativa OR LLM OR ChatGPT OR chatbot OR paciente virtual) AND (revelação da verdade OR más notícias OR conversa difícil OR comunicação de doença grave)
LILACS/BVS	(inteligência artificial OR LLM OR ChatGPT OR paciente virtual) AND (más notícias OR revelação da verdade OR conversa difícil)
ERIC	(artificial intelligence OR LLM OR ChatGPT OR chatbot OR virtual patient) AND (breaking bad news OR bad news OR difficult conversation OR communication skills)
Semantic Scholar	(generative artificial intelligence OR LLM OR ChatGPT OR chatbot) AND (breaking bad news OR bad news OR difficult conversation)

Tabela 1. Estratégias de busca por base de dados, sem restrição de data. Strings completas com operadores MeSH/DeCS na versão de protocolo.

Figura 1. Fluxograma PRISMA-ScR.



Nota: critérios de exclusão por resumo: E1 IA não generativa; E2 comunicação genérica sem más notícias como foco; E3 IA como ferramenta de análise, não de treinamento; E4 população incorreta; E5 revisão já publicada; E6 domínio fora da saúde.

RESULTADOS

Os 11 estudos incluídos na síntese (Figura 1, Tabela 2) foram publicados entre 2023 e 2026, com concentração em 2025-2026 (8 dos 11). A maioria teve origem nos EUA e Reino Unido, além de um estudo do México e um de origem não identificada. Os desenhos predominantes foram piloto ou pré-pós de braço único, sem grupo controle, amostras pequenas

(1-162 participantes), nenhum ensaio randomizado. A ferramenta mais usada foi o ChatGPT, presente em 9 dos 11 estudos, incluindo variante por voz.

Quanto ao cenário, tratou-se de más notícias em emergência⁹, anestesiologia¹⁰, simulação dedicada (BRAVE)¹¹, comparação IA x humano¹², mamografia anormal^{13,14} e piloto dedicado¹⁵; um, de oncologia, não foi localizado em bases indexadas, exigindo verificação¹⁶; dois trataram más notícias entre múltiplos cenários^{17,18}; e um ampliou a interpretação para diagnóstico de condição crônica grave (diabetes tipo 2)¹⁹.

O desfecho mais relatado foi autoconfiança/autoeficácia autorreferida, com melhora significativa em todos os estudos com avaliação pré-pós. Nenhum estudo utilizou avaliação por observador cego com instrumento validado, nem mediu transferência da habilidade para encontros clínicos reais.

Autor/Ano	País	Desenho	N	Ferramenta	Principal desfecho
Webb JJ, 2023	EUA	Prova de conceito	1	ChatGPT-3.5	Viabilidade de role-play e feedback via SPIKES.
Phillips M et al., 2024	EUA	Relato breve	NI	ChatGPT	Discussão qualitativa sobre potencial em anestesiologia.
Karnieli-Miller O, Elyoseph Z, 2026	Israel	Pré-pós braço único	99	IA generativa (BRAVE)	Autoeficácia p<0,001; disposição p<0,001.
Armstrong ME et al., 2026	Reino Unido	Comparativo IA x humano	155	ChatGPT	61% preferiu conteúdo gerado por IA.
Weisman D et al., 2025	EUA	Desenvolvimento multifásico	NI	GPT-4	Desenvolvimento e usabilidade da plataforma.
Comulada WS et al., 2025	EUA	Piloto braço único	10	IA por voz (GPT)	SUS médio 91; alta usabilidade percebida.
Müller Y et al., 2026	Alemanha	Quase-experimental	162	GPT-4o	Aumento de competência percebida, d=0,58; p<0,001.
Mukadam A et al., 2025	Reino Unido	Métodos mistos	27	ChatGPT-4o (voz)	Confiança em más notícias: mediana 2 para 4; p<0,001.
Chiu J et al., 2025	EUA	Piloto pré-pós	NI	ChatGPT	Confiança 3 para 4,17; p=0,002.
Kotov M, 2026	NI	Implementação real	NI	Chatbot (ChatGPT)	Potencial como recurso escalável; fonte não indexada, checar idoneidade.
Suárez-García RX et al., 2025	México	Pré-pós simulação	NI	IA generativa	Melhora da comunicação diagnóstica (diabetes tipo 2).

Tabela 2. Síntese dos 11 estudos incluídos. NI = não informado.

CONCLUSÃO

O achado central não é que a IA generativa funciona para treinar más notícias, é que a evidência ainda não permite afirmar isso com segurança: todos os estudos com desfecho pré-pós relataram melhora significativa, mas essa uniformidade, somada à ausência de grupo-controle e de cegamento, é compatível tanto com efeito real quanto com efeito de novidade e viés de publicação. Nenhuma revisão prévia havia sintetizado essa interseção, o que sustenta a contribuição desta revisão para a organização inicial de um campo jovem e concentrado nos últimos dois anos.

A literatura atual não sustenta a adoção da IA generativa como substituto do treinamento tradicional, apenas como ferramenta complementar de baixo custo e alta disponibilidade. As perspectivas futuras passam por estudos controlados, com avaliação por avaliador cego e instrumento validado, e mensuração de transferência para o encontro clínico real; a seleção por revisor único e a ausência de bases pagas reforçam a necessidade de replicação futura.

REFERÊNCIAS

1. Baile WF, Buckman R, Lenzi R, Glober G, Beale EA, Kudelka AP. SPIKES-A six-step protocol for delivering bad news: application to the patient with cancer. *Oncologist*. 2000;5(4):302-11.
2. Tan CY, Thum NWH, Mohan T, Cheng AKJ, Govindasamy R, Abdul Hamid NAB, et al. Breaking bad news and the cost of caring on physicians: a systematic scoping review. *BMJ Support Palliat Care*. 2026;16(3):509-18.
3. Lee J, Kim H, Kim KH, Jung D, Jowsey T, Webster CS. Effective virtual patient simulators for medical communication training: a systematic review. *Med Educ*. 2020;54(9):786-95.
4. Stamer T, Steinhäuser J, Flägel K. Artificial intelligence supporting the training of communication skills in the education of health care professions: scoping review. *J Med Internet Res*. 2023;25:e43311.
5. Zeng J, Qi W, Shen S, Liu X, Li S, Wang B, et al. Embracing the future of medical education with large language model-based virtual patients: scoping review. *J Med Internet Res*. 2025;27:e79091.
6. Jiang J, Ye MZ, Kwok TT, Wong JYH. GenAI-supported virtual patients in health care education: systematic review. *J Med Internet Res*. 2026;28:e82756.
7. Kang W. The use of virtual patients for breaking bad news: a rapid review. *Clin Teach*. 2024;21(2):e13681.
8. Vitale E, Conte L, Cioce M, Cornacchione P, Capuano A, Massafra R, et al. Communicating bad news in oncology and hematology settings: a statistic and large language model for interpreting nurses' difficulties and emotions. *Palliat Support Care*. 2025;23:e203.
9. Webb JJ. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. *Cureus*. 2023;15(5):e38755.
10. Phillips M, Vermeylen J, Ballard H. From bytes to empathy: can ChatGPT teach anesthesiologists how to deliver bad news? *Acad Med*. 2024;99(4):347.
11. Karnieli-Miller O, Elyoseph Z. Breaking bad news with reflective AI virtual experience (BRAVE): development and evaluation of a generative AI-based simulation tool. *Adv Simul (Lond)*. 2026;11(1):30.
12. Armstrong ME, Bright P, Glen S, Marshall D, Jones MC. Can artificial intelligence help with the development of generic clinical skills when breaking bad news? A quantitative evaluation. *Clin Med (Lond)*. 2026;26(3):100562.
13. Weisman D, Sugarman A, Huang YM, Gelberg L, Ganz PA, Comulada WS. Development of a GPT-4-powered virtual simulated patient and communication training platform for medical students to practice discussing abnormal mammogram results with patients: multiphase study. *JMIR Form Res*. 2025;9:e65670.
14. Comulada WS, Ganz PA, Huang YM, Weisman D, Sugarman A, Lee D, et al. A pilot test of an AI voice-driven simulation with feedback for medical students to practice discussing diagnostic mammogram results with patients. *Cureus*. 2025;17(10):e95606.
15. Chiu J, Castro B, Ballard I, Nelson K, Zarutskie P, Olaiya OK, et al. Exploration of the role of ChatGPT in teaching communication skills for medical students: a pilot study. *Med Sci Educ*. 2025;35(4):1871-82.
16. Kotov M. Ferramenta de IA conversacional baseada em ChatGPT para treinamento de oncologistas na comunicação com pacientes: um estudo de implementação no mundo real. *Rev Oncol Clin*. 2026.
17. Müller Y, Lemberg H, Gellert P, Zoellick JC. Training empathetic communication skills in medical students with a role-prompted GPT-4o chatbot: quasi-experimental intervention study. *JMIR Med Educ*. 2026;12:e88092.
18. Mukadam A, Suresh S, Jacobs C. Beyond traditional simulation: an exploratory study on the effectiveness and acceptability of ChatGPT-4o advanced voice mode for communication skills practice among medical students. *Cureus*. 2025;17(5):e84381.
19. Suárez-García RX, Chavez-Castañeda Q, Orrico-Pérez R, Valencia-Marin S, Castañeda-Ramírez AE, Quiñones-Lara E, et al. DIALOGUE: a generative AI-based pre-post simulation study to enhance diagnostic communication in medical students through virtual type 2 diabetes scenarios. *Eur J Investig Health Psychol Educ*. 2025;15(8):152.