

Usabilidade de Interfaces Conversacionais com Inteligência Artificial Generativa em Aplicações mHealth: Uma revisão sistemática

Diego E. da Silva Lima¹, Noeli A. Pimentel Vaz^{1,2}, Sérgio T. Carvalho¹,
Luciana de O. Berretta¹

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – CEP 74001-970 – Goiânia – GO – Brasil

²Universidade Estadual de Goiás (UEG)
Caixa Postal 459 – CEP 75132-903 – Anápolis – GO – Brasil

{diegoenrique, noelivaz}@discente.ufg.br,

{sergiocarvalho, luciana.berretta}@ufg.br

Abstract. *This systematic review aims primarily to identify and analyze the methods and strategies employed to evaluate the usability and user experience of Artificial Intelligence-based chat interfaces in the context of mHealth applications. The study determined that evaluation strategies, both quantitative and qualitative, focus on criteria such as ease of use, utility, satisfaction, and information quality. Aspects such as interactivity, personalization, engagement, and data privacy were also taken into consideration. It is expected that this systematic review will contribute to the development of more accessible, culturally sensitive, and inclusive interfaces, as well as support studies focusing on mHealth applications.*

Resumo. *Esta revisão sistemática tem como objetivo principal identificar e analisar os métodos e estratégias empregados para avaliar a usabilidade e a experiência do usuário em interfaces conversacionais de Inteligência Artificial Generativa no contexto de aplicações mHealth. O estudo identificou que as estratégias de avaliação, tanto quantitativas quanto qualitativas, focam em critérios como facilidade de uso, utilidade, satisfação e qualidade da informação. Também foram considerados aspectos como interatividade, personalização, engajamento e privacidade dos dados. Espera-se que esta revisão sistemática contribua para o desenvolvimento de interfaces conversacionais mais acessíveis, culturalmente sensíveis e inclusivas, bem como para o apoio a estudos que foquem em aplicações em mHealth.*

1. Introdução

Devido à crescente demanda por saúde e ao foco em doenças crônicas, a telemedicina e as intervenções digitais tornaram-se fundamentais [Zhang and Song 2024]. Nesse contexto, os aplicativos de saúde móvel (*mobile health* - mHealth) estão revolucionando o setor ao oferecer assistência acessível e ao capacitar os pacientes para o autogerenciamento de sua saúde, promovendo melhores resultados [Raj et al. 2025].

Nesse sentido, as interfaces conversacionais são ferramentas transformadoras no ecossistema mHealth, pois simulam a interação humana por meio de linguagem natural em diversas plataformas [van Heerden et al. 2023]. Existem os tradicionais sistemas baseados em regras e os mais avançados, que empregam Inteligência Artificial (IA). Os baseados em regras dependem de estruturas predefinidas, como máquinas de estados e roteiros fixos de perguntas e respostas, para guiar a conversação. Por outro lado, as interfaces conversacionais de IA utilizam Processamento e Geração de Linguagem Natural para formular respostas coerentes e flexíveis, adaptando-se de forma mais eficaz às entradas do usuário. A evolução desses sistemas é dada pelos Grandes Modelos de Linguagem (*Large Language Models* - LLMs), como o ChatGPT, que oferecem capacidades avançadas de compreensão e geração de texto, similares às humanas [van Heerden et al. 2023].

Para garantir a retenção e uma experiência positiva, a interação com essas interfaces conversacionais deve ser natural, significativa e contextual [Golden and Aboujaoude 2024]. De forma geral, a usabilidade e a experiência do usuário (UX) são fatores essenciais para a aceitação, adoção e eficácia de qualquer solução mHealth baseada em IA. Portanto, um design de interface conversacional clara, intuitiva e de fácil uso é fundamental para minimizar as dificuldades e guiar o usuário. [D et al. 2024].

Apesar de seu grande potencial, as interfaces conversacionais em aplicações mHealth ainda enfrentam desafios significativos. As principais limitações incluem erros e imprecisões no reconhecimento da fala, dificuldade em manter a continuidade de um diálogo, fornecer informações com profundidade insuficiente e a necessidade de evoluir para interações multimodais, que vão além do texto [Choi et al. 2024].

Nesse contexto, o presente estudo analisa a literatura sobre as métricas de avaliação de usabilidade de interfaces conversacionais de IA generativa em aplicações mHealth, apresentando a seguinte estrutura: na Seção II, apresenta-se a metodologia utilizada para a busca, seleção e análise dos estudos. Na Seção III, apresentam-se os resultados obtidos e a discussão. E, por fim, na Seção IV, a conclusão do estudo.

2. Metodologia

A pesquisa utilizou um método empírico, implementado por meio de Revisão Sistemática da Literatura. A metodologia seguiu as diretrizes do Protocolo para Revisões Sistemáticas da Literatura (RSL) em Engenharia de Software [Kitchenham and Charters 2007], organizando o processo em três fases principais: planejamento, execução e apresentação dos resultados.

2.1. Planejamento

Durante a fase de planejamento, foram definidos os elementos centrais da revisão: os objetivos, as questões de pesquisa, as bases de dados, as *strings* de busca e os critérios de inclusão e exclusão dos artigos e a estratégia para seleção e extração de dados. Além disso, como apoio ao processo de documentação, foi utilizada a ferramenta Parsifal¹, uma ferramenta online desenvolvida para apoiar pesquisadores na realização de RSL.

O objetivo desta revisão sistemática é identificar e analisar os métodos empregados para avaliar a usabilidade e a experiência do usuário em interfaces conversacionais de IA

¹<https://parsif.al/>

generativa no contexto de aplicações mHealth. Busca-se compreender as principais barreiras enfrentadas pelos usuários, as estratégias de avaliação mais utilizadas e as soluções propostas na literatura para promover maior acessibilidade, engajamento e eficácia desses sistemas.

Para analisar os estudos sobre os métodos de avaliação de usabilidade dentro de aplicações mHealth, estabeleceu-se a seguinte pergunta de pesquisa que guiou esta revisão:

”Quais são as estratégias de avaliação em termos de usabilidade e experiência do usuário em agentes conversacionais baseados em IA generativa de aplicações mHealth?”

A resposta a essa questão visa contribuir para a compreensão abrangente dos métodos de avaliação existentes, orientando o desenvolvimento de interfaces mais acessíveis e inclusivas e, ao mesmo tempo, apoiando futuros estudos na área de mHealth.

A pesquisa foi realizada em quatro bases de dados científicas: IEEE Xplore, PubMed, Scopus e SpringerLink. Essas plataformas foram selecionadas por seu amplo reconhecimento acadêmico e por sua relevância e abrangência nas áreas de tecnologia e saúde, garantindo uma base de literatura sólida para a revisão.

A *string* de busca foi elaborada com base nas palavras-chave descritas na Tabela 1 para direcionar a pesquisa e capturar estudos relacionados ao tema da revisão. A seguinte *string* de busca foi adaptada conforme a sintaxe de cada base selecionada no estudo:

(“Usability”OR “User Experience”OR “Human-computer interaction”OR “User Satisfaction”) AND (“Chatbots”OR “Conversational AI”OR “Conversational Agents”OR “Virtual Assistants”OR “Conversational Interface”) AND (“LLM”OR “Large Language Model”OR “LLMs”OR “Large Language Models”OR “AI-based”OR “LLM-based”OR “AI-powered”) AND (“Methods”OR “Approaches”OR “Methodologies”OR “Metrics”OR “Techniques”OR “Evaluation”OR “Assessment”OR “Heuristic Evaluation”) AND (“mHealth”OR “Digital Health”OR “Mobile Health”)

Tabela 1. Palavras-chaves e Sinônimos

Palavra-Chave	Sinônimos
<i>Usability</i>	<i>User Experience, Human-computer interaction, User satisfaction</i>
<i>Chatbots</i>	<i>Conversational AI, Conversational Agents, Virtual Assistants, Conversational Interface</i>
<i>LLM</i>	<i>Large Language Model, LLMs, Large Language Models, Ai-based, LLM-based, AI-powered</i>
<i>Methods</i>	<i>Approaches, Methodologies, Metrics, Techniques, Evaluation, Assessments, Heuristic Evaluation</i>
<i>mHealth</i>	<i>Digital Health, Mobile hHealth</i>

A definição rigorosa dos critérios de inclusão e exclusão é uma etapa fundamental no protocolo de uma Revisão Sistemática da Literatura (RSL), para minimizar vieses e garantir que a seleção dos artigos esteja diretamente alinhada à pergunta de pesquisa [Galvão and Ricarte 2019]. Assim, os critérios de seleção dos artigos são apresentados na Tabela 2.

Tabela 2. Critérios de Seleção dos Artigos

Critérios de Inclusão	
CI.1	Trabalhos que abordam avaliação de usabilidade e experiência do usuário em agentes conversacionais baseados em IA de aplicações mHealth
Critérios de Exclusão	
CE.1	Não estão publicados em periódicos
CE.2	Nãos sejam artigos completos
CE.3	Não sejam revisões e/ou levantamentos
CE.4	O texto completo do artigo não está disponível para acesso
CE.5	Que não seja em inglês
CE.6	Que não seja publicado nos últimos 5 anos (2020 - 2025)
CE.7	Trabalhos que não abordam avaliação de usabilidade e experiência do usuário em agentes conversacionais baseados em IA de aplicações mHealth

A estratégia de seleção e extração de dados desta revisão sistemática adotou a abordagem PRISMA [Page et al. 2022], o que garante um processo rigoroso e transparente. As etapas de triagem e inclusão dos estudos estão apresentadas na Tabela 3, enquanto os dados extraídos estão consolidados na Tabela 4.

Tabela 3. Estratégia de seleção dos artigos

ES.1	Remoção de duplicatas
ES.2	Leitura dos títulos e resumos
ES.3	Leitura do texto completo dos artigos eleíveis
ES.4	Avaliação da qualidade
ES.5	Extração de dados

Tabela 4. Dados extraídos

DE.1	Tipo de aplicação mHealth
DE.2	Tipo de agente conversacional
DE.3	Critérios e métricas de usabilidade avaliados
DE.4	Instrumento de avaliação usado
DE.5	Método de coleta de dados
D5.6	Conclusões e recomendações dos autores

2.2. Execução

Na fase de execução, o planejamento da pesquisa foi posto em prática. As buscas nas bases de dados e os artigos encontrados foram filtrados e selecionados conforme os critérios de seleção e de avaliação de qualidade. Em seguida, os dados relevantes dos estudos selecionados foram extraídos e, por fim, sintetizados para a análise final dos achados da pesquisa.

A etapa inicial consistiu em uma pesquisa preliminar para identificar revisões sistemáticas que abordassem métodos e técnicas para avaliar a usabilidade de interfaces conversacionais baseadas em inteligência artificial generativa em aplicações mHealth. Essa análise inicial revelou uma lacuna na literatura, pois nenhum trabalho recuperado se alinhava diretamente à questão de pesquisa proposta neste estudo.

A string de busca foi elaborada com base em termos-chave de uma análise preliminar e aprimorada iterativamente por meio de pesquisas piloto, sendo projetada para localizar os termos em todo o conteúdo relevante (título, resumo, palavras-chave e texto) de artigos revisados por pares. A aplicação desta estratégia retornou 173 artigos, distribuídos da seguinte forma: IEEE Xplore - 7, PubMed - 6, Scopus - 9 e SpringerLink - 151 artigos.

O processo de seleção e triagem foi realizado em três etapas principais. A partir do conjunto inicial de 173 artigos, a primeira etapa consistiu na remoção de 3 estudos duplicados. Na segunda etapa, a análise de títulos e resumos excluiu 98 artigos. A terceira etapa envolveu a leitura completa dos 72 artigos restantes, resultando no descarte de 50 trabalhos que não se alinhavam aos objetivos da pesquisa. Desse modo, foram consolidados em 22 artigos, que foram submetidos à avaliação de qualidade.

Para a avaliação de qualidade dos artigos, foram estabelecidas três Questões de Qualidade (QQs) e uma respectiva escala de respostas, detalhadas nas Tabelas 5 e 6, respectivamente. A pontuação de cada artigo, com um máximo de 3 pontos, foi calculada pela soma das notas de cada item. Para garantir a relevância dos estudos, foi aplicado um limiar de corte, sendo selecionados apenas os artigos com pontuação superior a 1,5, resultando na seleção de 21 dos 22 artigos avaliados.

Na próxima seção, serão apresentados os resultados da avaliação e análise de qualidade dos artigos selecionados, bem como as informações que foram extraídas.

Tabela 5. Questões de Qualidade

QQ.1	O estudo apresenta uma avaliação em termos de usabilidade e experiência de usuário?
QQ.2	O estudo leva em consideração aplicações de interfaces conversacionais usando IA?
DE.3	O estudo leva em consideração aplicações mHealth?

Tabela 6. Escala de respostas às Questões de Qualidade

Avaliação	Sim	Parcialmente	Não
Pontuação	1	0.5	0

3. Resultados e Discussões

Considerando o período de seleção dos artigos, evidencia-se um forte crescimento no interesse pelo tema, com um salto de 9 artigos em 2023 para 42 em 2024 — um aumento superior a 350%. A aparente queda para 15 publicações em 2025 se explica pela natureza parcial da coleta de dados, que, para esta revisão, foi encerrada em julho de 2025. Este volume, alcançado em apenas sete meses, sinaliza uma forte tendência de crescimento, indicando que o total de publicações ao final do ano tem potencial para superar o de 2024 e, assim, consolidar a expansão da área de pesquisa.

Adicionalmente, foi realizado um levantamento dos instrumentos de avaliação empregados nos artigos selecionados. A Tabela 7 sintetiza essas informações, detalhando, para cada estudo, a abordagem metodológica (quantitativa, qualitativa ou mista) e os instrumentos utilizados.

Tabela 7. Instrumentos de avaliação encontrados nos artigos

Referência	Abordagem	Instrumento utilizado
[Lewis et al. 2024]	Quantitativa	Questionário de Linguagem
[Raj et al. 2025]	Mista	Não apresenta instrumentos formais de avaliação de usabilidade
[De la Puente et al. 2024]	Quantitativa	SUS
[D et al. 2024]	Quantitativa	Não apresenta instrumentos formais de avaliação de usabilidade
[Gabarron et al. 2024]	Qualitativa	Combinações de métodos avaliativos
[Görtz et al. 2023]	Mista	Questionários Estruturados
[Ma et al. 2024b]	Mista	UMUX-lite, AES e TAM
[Ma et al. 2024a]	Mista	UMUX-Lite, AES e TAM
[Golden and Aboujaoude 2024]	Quantitativa	FAITA-Mental Health
[Esmailzadeh 2025]	Mista	Framework de avaliação
[Azam et al. 2024]	Mista	CUQ e Comentários dos usuários
[Wang et al. 2025]	Quantitativa	MAUQ e DSSQ
[Zhang and Song 2024]	Quantitativa	CUQ
[Anisha et al. 2025]	Qualitativa	TAM, UTAUT, MARS e HBM
[Arigo et al. 2025]	Mista	DCBI e TWEETS
[Cevasco et al. 2024]	Mista	SUS, CSQ-8, NPS e Questionários personalizados
[van Heerden et al. 2023]	Qualitativa	TAM
[Choi et al. 2024]	Mista	Questionário estruturado
[Colakoglu et al. 2025]	Quantitativa	Análise de logs
[Nguyen et al. 2024]	Misto	SUS, CUQ, análise de logs e Comentários dos usuários
[Ashton et al. 2023]	Qualitativa	Entrevistas semiestruturadas

A análise dos estudos revela que as interfaces conversacionais em mHealth são projetadas para um público-alvo vasto e diversificado, que abrange desde a população em geral até pacientes com doenças crônicas, condições de saúde mental [De la Puente et al. 2024] e profissionais da área [D et al. 2024].

As aplicações de saúde móvel abrangidas nos artigos abrangem uma vasta gama de soluções digitais, predominantemente entregues por meio de *smartphones*, *tablets* e dispositivos vestíveis [Choi et al. 2024]. Uma categoria proeminente são os chatbots e agentes conversacionais (*Conversational Agents* - CAs), que empregam IA, incluindo os LLMs. Estas aplicações são desenvolvidas para o autogerenciamento da saúde e a prevenção de doenças crônicas, como hipertensão [Wang et al. 2025], diabetes tipo 2, insuficiência cardíaca, doença pulmonar obstrutiva crônica (DPOC) e condições cognitivas, como demência [Anisha et al. 2025].

Outras áreas de intervenção incluem a saúde mental (abordando estresse, ansiedade, transtorno obsessivo-compulsivo e avaliação do risco de suicídio) [Esmailzadeh 2025], promoção da atividade física [Gabarron et al. 2024], hábitos alimentares saudáveis [Ashton et al. 2023], saúde sexual e reprodutiva, com destaque para a prevenção do HIV e da gravidez na adolescência [van Heerden et al. 2023].

As funcionalidades vão desde a coleta de histórico de saúde familiar e fornecimento de informações sobre medicamentos até a sumarização de relatórios médicos [Raj et al. 2025] e, também, o apoio a profissionais de saúde para a otimização de fluxos de trabalho [Wang et al. 2025].

Muitos destes sistemas são projetados com recursos de acessibilidade para grupos específicos, como indivíduos com deficiência visual [Choi et al. 2024], e são dis-

ponibilizados por meio de aplicativos móveis dedicados, plataformas web e, em alguns casos, por meio de serviços de mensagens, como SMS ou Facebook Messenger [van Heerden et al. 2023].

A avaliação da usabilidade em aplicações mHealth e interfaces conversacionais concentra-se em critérios essenciais para garantir que as ferramentas sejam eficazes e satisfatórias [Ma et al. 2024a]. A análise dos artigos revelou que os principais critérios investigados são: a facilidade de uso, a utilidade, a eficácia percebida e a qualidade da informação. A facilidade de uso (ou intuitividade) refere-se à simplicidade com que os usuários operam o sistema [Anisha et al. 2025]. Já a utilidade e a eficácia percebidas avaliam se a aplicação auxilia no autogerenciamento da saúde e alcança seus objetivos, como mudanças comportamentais ou diagnósticos auxiliares [Anisha et al. 2025].

A qualidade da informação constitui um critério fundamental na avaliação de agentes conversacionais, sendo examinada por meio de três abordagens principais: Métricas Técnicas de Desempenho, Avaliação Subjetiva dos Usuários e Validação por Especialistas/Revisores Humanos [Azam et al. 2024]. A satisfação do utilizador é consistentemente avaliada por meio de questionários padronizados, como o *System Usability Scale* (SUS), *Chatbot Usability Questionnaire* (CUQ), *mHealth App Usability Questionnaire* (MAUQ), *UMUX-Lite e Acceptability E-scale* (AES), refletindo as reações subjetivas e o sentimento geral de convivência, segurança emocional e ausência de julgamento [Colakoglu et al. 2025].

Além disso, consideram-se critérios relacionados com a interatividade e a personalização, como a capacidade das interfaces conversacionais de compreender a entrada do utilizador (incluindo o reconhecimento de fala e a interpretação de homófonos), manter a continuidade do diálogo e adaptar suas respostas às necessidades individuais [Ashton et al. 2023]. O engajamento é medido pela intenção de uso contínuo, a taxa de retenção e a frequência e duração das interações [Gabarron et al. 2024]. A acessibilidade e a inclusão são avaliadas pela adaptabilidade a grupos com necessidades específicas, como pessoas com deficiência visual, e pela mitigação de vieses culturais e linguísticos. Por fim, a gestão de crises (com protocolos de escalonamento de risco) e a privacidade dos dados são aspectos éticos e de segurança cruciais que influenciam a confiança e a aceitação do utilizador [Golden and Aboujaoude 2024]. Os métodos e instrumentos de avaliação empregados nos artigos são diversos e abrangem tanto abordagens quantitativas quanto qualitativas, frequentemente em estudos de métodos mistos [Nguyen et al. 2024]. A usabilidade e a aceitabilidade são avaliadas de forma proeminente, utilizando escalas padronizadas como o *System Usability Scale*, *Chatbot Usability Questionnaire* [Nguyen et al. 2024], *mHealth App Usability Questionnaire* (MAUQ) [Wang et al. 2025], *UMUX-Lite e Acceptability E-scale* (AES) [Ma et al. 2024a], além de questionários customizados e entrevistas semiestruturadas [Anisha et al. 2025].

A avaliação da percepção do usuário frequentemente se baseia no Modelo de Aceitação da Tecnologia (TAM) e suas subestruturas, incluindo facilidade de uso percebida (*Perceived Ease of Use* - PEU), utilidade percebida (*Perceived Usefulness* - PU), atitude em relação ao uso (*Attitude Towards Use* - ATU) e intenção comportamental de uso (*Behavioral Intention to Use* - BIU) [Ma et al. 2024a]. A satisfação é um critério-chave, por vezes medida pelo *Net Promoter Score* (NPS) [Cevasco et al. 2024], e é explorada através de *feedback* qualitativo sobre a experiência do usuário [Wang et al. 2025].

Modelos específicos são validados por especialistas, que avaliam a utilidade, concisão, completude e clareza das respostas [Wang et al. 2025]. A qualidade da informação é verificada pela acurácia, coerência e adequação do conteúdo [Wang et al. 2025], por vezes com análise de legibilidade (*Flesch-Kincaid Grade Level Test*) [Golden and Aboujaoude 2024], e, em alguns casos, pela comparação com informações validadas por profissionais de saúde.

O engajamento do usuário é avaliado por dados de *log* (frequência de uso, duração da interação, cliques em botões, perguntas feitas) [Nguyen et al. 2024], e por escalas de autoavaliação (*Digital Behavior Change Intervention Engagement Scale*, *Twente Engagement with eHealth Technologies Scale*) [Arigo et al. 2025]. Há também uma preocupação com aspectos como a acessibilidade (para usuários com deficiência visual) [Choi et al. 2024] e a mitigação de vieses (culturais, linguísticos, demográficos) [van Heerden et al. 2023]. A avaliação pode incluir o impacto na eficiência do fluxo de trabalho dos clínicos e na redução da carga de trabalho [Wang et al. 2025]. A validade e a confiabilidade dos dados são asseguradas por meio de revisões independentes, discussões em grupo, triangulação de dados e uso de pseudônimos para proteger a identidade dos participantes [Ma et al. 2024a].

4. Conclusão

Esta Revisão Sistemática da Literatura analisou 21 artigos para identificar estratégias de avaliação de usabilidade em agentes conversacionais com IA generativa para aplicações mHealth. A análise revelou que o campo emprega um robusto e diversificado arsenal de métricas, combinando abordagens quantitativas e qualitativas. A usabilidade é predominantemente aferida por meio de escalas padronizadas, com destaque para o SUS [De la Puente et al. 2024] e questionários específicos, como o Chatbot CUQ [Zhang and Song 2024]. A aceitação da tecnologia é frequentemente investigada por meio de modelos teóricos, como o TAM [Ma et al. 2024a] e seus construtos de utilidade e facilidade de uso percebida. Complementarmente, o engajamento do usuário é medido por meio de métricas comportamentais (dados de *log* de uso) [Nguyen et al. 2024]. Desse modo, espera-se que esta síntese metodológica contribua diretamente para orientar o desenvolvimento de interfaces conversacionais mais acessíveis e inclusivas, oferecendo apoio robusto a futuros estudos focados em aplicações mHealth.

Referências

- Anisha, S. A., Sen, A., Ahmad, B., and Bain, C. (2025). Exploring acceptance of digital health technologies for managing non-communicable diseases among older adults: A systematic scoping review. *Journal of Medical Systems*, 49(1):35.
- Arigo, D., Jake-Schoffman, D. E., and Pagoto, S. L. (2025). The recent history and near future of digital health in the field of behavioral medicine: an update on progress from 2019 to 2024. *Journal of Behavioral Medicine*, 48(1):120–136.
- Ashton, L. M., Adam, M. T., Whatnall, M., Rollo, M. E., Burrows, T. L., Hansen, V., and Collins, C. E. (2023). Exploring the design and utility of an integrated web-based chatbot for young adults to support healthy eating: a qualitative study. *International Journal of Behavioral Nutrition and Physical Activity*, 20(1):119.

- Azam, A., Naz, Z., and Khan, M. U. G. (2024). Pharmallm: A medicine prescriber chatbot exploiting open-source large language models. *Human-Centric Intelligent Systems*, 4(4):527–544.
- Cevasco, K. E., Morrison Brown, R. E., Woldeselassie, R., and Kaplan, S. (2024). Patient engagement with conversational agents in health applications 2016–2022: A systematic review and meta-analysis. *Journal of Medical Systems*, 48(1):40.
- Choi, S., Seo, J., Hernandez, M., and Kitsiou, S. (2024). Conversational agents in mhealth: use patterns, challenges, and design opportunities for individuals with visual impairments. *Journal of Technology in Behavioral Science*, 9(4):912–923.
- Colakoglu, S., Durmus, M., Polat, Z. P., Yildiz, A., and Sezgin, E. (2025). User engagement with a multimodal conversational agent for self-care and chronic disease management: A retrospective analysis. *Journal of Medical Systems*, 49(1):76.
- D, U., S, K., Chakradhar, V., Teja, M. S., and Teja, M. S. (2024). Enabling mental wellness through interactive platforms. In *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, pages 1970–1975.
- De la Puente, G., Silva, A., and Felix, R. (2024). Development of a chatbot powered by artificial intelligence to diagnose and improve stress and anxiety levels in university students. In *2024 IEEE XXXI International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pages 1–8.
- Esmailzadeh, P. (2025). Decoding the cry for help: Ai’s emerging role in suicide risk assessment. *AI and Ethics*.
- Gabarron, E., Larbi, D., Rivera-Romero, O., and Denecke, K. (2024). Human factors in AI-Driven digital solutions for increasing physical activity: Scoping review. *JMIR Hum Factors*, 11:e55964.
- Galvão, M. C. B. and Ricarte, I. L. M. (2019). Revisão sistemática da literatura: Conceituação, produção e publicação. *Logeion: Filosofia da Informação*, 6(1):57–73.
- Golden, A. and Aboujaoude, E. (2024). Describing the framework for AI tool assessment in mental health and applying it to a generative AI Obsessive-Compulsive disorder platform: Tutorial. *JMIR Form Res*, 8:e62963.
- Görtz, M., Baumgärtner, K., Schmid, T., Muschko, M., Woessner, P., Gerlach, A., Byczkowski, M., Sülthmann, H., Duensing, S., and Hohenfellner, M. (2023). An artificial intelligence-based chatbot for prostate cancer education: Design and patient evaluation study. *Digit Health*, 9:20552076231173304.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. 2.
- Lewis, D.-M., DeRenzi, B., Misomali, A., Nyirenda, T., Phiri, E., Chifisi, L., Makwenda, C., and Lesh, N. (2024). Human review for post-training improvement of low-resource language performance in large language models. In *2024 IEEE 12th International Conference on Healthcare Informatics (ICHI)*, pages 592–597.
- Ma, Y., Achiche, S., Pomey, M.-P., Paquette, J., Adjoutah, N., Vicente, S., Engler, K., MARVIN chatbots Patient Expert Committee, Laymouna, M., Lessard, D., Lemire,

- B., Asselah, J., Therrien, R., Osmanliu, E., Zawati, M. H., Joly, Y., and Lebouché, B. (2024a). Adapting and evaluating an AI-Based chatbot through patient and stakeholder engagement to provide information for different health conditions: Master protocol for an adaptive platform trial (the MARVIN chatbots study). *JMIR Res Protoc*, 13:e54668.
- Ma, Y., Achiche, S., Tu, G., Vicente, S., Lessard, D., Engler, K., Lemire, B., MARVIN chatbots Patient Expert Committee, Laymouna, M., de Pokomandy, A., Cox, J., and Lebouché, B. (2024b). The first AI-based chatbot to promote HIV self-management: A mixed methods usability study. *HIV Med*, 26(2):184–206.
- Nguyen, M. H., Sedoc, J., and Taylor, C. O. (2024). Usability, engagement, and report usefulness of Chatbot-Based family health history data collection: Mixed methods analysis. *J Med Internet Res*, 26:e55164.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., and Moher, D. (2022). [the PRISMA 2020 statement: an updated guideline for reporting systematic reviewsdeclaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas]. *Rev Panam Salud Publica*, 46:e112.
- Raj, S., Shekhar, S., and P, B. (2025). Medihealth: An ai-driven mobile solution for enhanced health care decision-making and accessibility. In *2025 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, pages 874–880.
- van Heerden, A., Bosman, S., Swendeman, D., and Comulada, W. S. (2023). Chatbots for hiv prevention and care: a narrative review. *Current HIV/AIDS Reports*, 20(6):481–486.
- Wang, Y., Zhu, T., Zhou, T., Wu, B., Tan, W., Ma, K., Yao, Z., Wang, J., Li, S., Qin, F., Xu, Y., Tan, L., Liu, J., and Wang, J. (2025). Hyper-dream, a multimodal digital transformation hypertension management platform integrating large language model and digital phenotyping: Multicenter development and initial validation study. *Journal of Medical Systems*, 49(1):42.
- Zhang, S. and Song, J. (2024). A chatbot based question and answer system for the auxiliary diagnosis of chronic diseases based on large language model. *Scientific Reports*, 14(1):17118.