REVISTA TÓPICOS
ISSN 2965-6672 | Qualis A2

O USO DA INTELIGÊNCIA ARTIFICIAL NO SUPORTE EMOCIONAL: UMA ANÁLISE DE RISCOS ÉTICOS, TÉCNICOS E REQUISITOS DE SEGURANÇA NA ENGENHARIA DE SOFTWARE

THE USE OF ARTIFICIAL INTELLIGENCE IN EMOTIONAL SUPPORT: AN ANALYSIS OF ETHICAL AND TECHNICAL RISKS AND SECURITY REQUIREMENTS IN SOFTWARE ENGINEERING

Ciências Exatas e da Terra, Engenharias, Ciências da Saúde •

26/11/2026

REGISTRO DOI: 10.70773/revistatopicos/795585165

Arielton Moura Pontes

Lucas Hermenegildo do Nascimento

RESUMO

O uso de inteligência artificial em aplicações de suporte emocional amplia o acesso a orientações de baixa intensidade, mas apresenta limitações técnicas, riscos de respostas inadequadas e desafios de privacidade e governança. O estudo analisa essas possibilidades e limites por meio de revisão bibliográfica narrativa e análise documental, mapeando riscos e propondo requisitos de software (R01 a R11) com critérios de verificação. Os resultados indicam que ferramentas conversacionais possuem potencial de apoio, mas exigem validação e supervisão. Evidenciou-se que questões de confiabilidade, privacidade e transparência tornam inadequada a substituição de profissionais de saúde mental por sistemas automatizados. Conclui-se que, na perspectiva da Engenharia de Software, essas ferramentas devem operar como sistemas sociotécnicos estruturados com limites de finalidade, segurança desde a concepção, minimização de dados, transparência, avaliação contínua e protocolos de encaminhamento para atendimento humano em situações críticas.

Palavras-chave: Inteligência artificial; Saúde mental; Chatbots; Ética em tecnologia; Engenharia de Software.

ABSTRACT

The use of artificial intelligence in emotional support applications expands access to low-intensity guidance, yet it presents technical limitations, risks of inappropriate responses, and challenges related to privacy and governance. The study examines these possibilities and limitations through a narrative literature review and document analysis, mapping risks and proposing software requirements (R01–R11) along with verification criteria. The findings indicate that conversational tools have the potential to provide support, but they require validation and oversight. It is evident that concerns

regarding reliability, privacy, and transparency make it inappropriate to replace mental health professionals with automated systems. The study concludes that, from a software engineering perspective, such tools should operate as structured sociotechnical systems—incorporating defined purposes, security by design, data minimization, transparency, continuous evaluation, and protocols for referring users to human care in critical situations.
Keywords: Artificial intelligence; Mental health; Chatbots; Technology ethics; Software Engineering.

1. INTRODUÇÃO

Sistemas de inteligência artificial capazes de conversar em linguagem natural já são utilizados em atividades que vão além da busca por informações. Uma pesquisa do Pew Research Center, realizada com adultos dos Estados Unidos de 17 a 23 de fevereiro de 2026, identificou que uma em cada dez pessoas utilizam chatbots para suporte emocional, embora as buscas mais comuns sejam pesquisas e trabalho (Pew Research Center, 2026). Esse dado ajuda a dimensionar a presença dessas ferramentas em interações relacionadas a questões pessoais, embora não represente necessariamente a realidade brasileira. Estudos apontam que a distribuição de profissionais é extremamente desigual, atingindo cerca de 9 psiquiatras a cada 100 mil habitantes em nações desenvolvidas, índice que cai para apenas 0,01 psiquiatra por 100 mil de habitantes em regiões em desenvolvimento (Vaidyam *et al.*, 2019). A facilidade de acesso e a possibilidade de conversar a qualquer momento ajudam a explicar o interesse por esses recursos. Entretanto, quando uma aplicação passa a ser utilizada em situações de sofrimento psicológico, não basta avaliar se ela responde de forma coerente. Também é necessário considerar a

qualidade das respostas, os limites do sistema e os riscos que podem surgir durante a interação.

Os resultados das pesquisas sobre agentes conversacionais para saúde mental indicam possibilidades de uso, mas não permitem tratar todas essas ferramentas como se tivessem o mesmo desempenho. He *et al.* (2023), em uma revisão sistemática com metanálise de ensaios clínicos randomizados, identificaram efeitos favoráveis em diferentes indicadores de saúde mental, principalmente no curto prazo. Os autores também observaram diferenças de eficácia entre os tipos de agentes e destacaram a necessidade de estudos de maior qualidade antes de uma adoção clínica mais ampla. Em uma análise voltada a jovens, Feng *et al.* (2025) encontraram evidências mais consistentes para a redução de sintomas depressivos do que para outros resultados, como ansiedade, estresse e bem-estar. Esses achados mostram que a avaliação precisa considerar o público, a finalidade e o resultado observado em cada estudo, em vez de atribuir uma eficácia geral à inteligência artificial.

A situação fica mais complexa com a expansão da inteligência artificial generativa. Diferentemente de sistemas limitados a respostas previamente definidas, modelos generativos conseguem produzir textos variados e adaptar a conversa às informações recebidas. Essa capacidade pode tornar a interação mais fluida, mas não comprova que o sistema compreenda a experiência de quem conversa com ele. Peixoto (2025) discute como a IA generativa pode simular aspectos da subjetividade por meio da linguagem sem possuir consciência ou vivência emocional. No contexto do suporte emocional, essa diferença merece atenção porque a forma da resposta pode influenciar a confiança depositada pelo usuário. Uma

conversa acolhedora não equivale, por si só, a uma avaliação psicológica ou a um acompanhamento profissional.

As limitações das evidências disponíveis reforçam essa preocupação. Wang, Zhou e Zhou (2025), ao revisarem aplicações de IA generativa na saúde mental, identificaram resultados promissores em algumas áreas, mas também destacaram a heterogeneidade dos estudos e a escassez de avaliações em condições reais de uso e ao longo do tempo. A diferença entre resultados obtidos em experimentos ou simulações e o comportamento de uma aplicação utilizada continuamente por pessoas precisa ser considerada. Um sistema pode apresentar desempenho satisfatório em uma avaliação específica sem que isso demonstre sua segurança em todas as situações em que venha a ser utilizado. Por esse motivo, benefícios potenciais e limitações devem ser examinados em conjunto.

No Brasil, as discussões sobre inteligência artificial na saúde também envolvem questões éticas, jurídicas e de responsabilidade. Elias *et al.* (2023), em uma revisão sistemática sobre as implicações bioéticas da IA na saúde, abordam temas como autonomia, explicabilidade, confiança, privacidade e segurança dos dados. Bentes, Sanches e Fonseca (2024), por sua vez, discutem os desafios regulatórios dos assistentes virtuais inteligentes aplicados à saúde mental. Essas contribuições ajudam a compreender que os riscos não se restringem à qualidade do conteúdo produzido pelo sistema. Eles também envolvem a forma como a ferramenta é apresentada, os limites informados ao usuário e as responsabilidades relacionadas à sua utilização.

A proteção das informações é parte desse problema. Durante uma conversa sobre sofrimento psicológico, o usuário pode fornecer

relatos íntimos e dados pessoais sensíveis. No Brasil, a Lei Geral de Proteção de Dados Pessoais (LGPD) estabelece regras para o tratamento dessas informações (Brasil, 2018). Na Engenharia de Software, isso exige decisões sobre coleta, armazenamento, acesso, compartilhamento e exclusão de dados. Também é necessário considerar o que será informado ao usuário e quais medidas poderão reduzir os riscos de acesso indevido ou utilização incompatível com a finalidade declarada. Essas questões não devem ser tratadas apenas como ajustes posteriores à implementação, pois influenciam a própria concepção do sistema.

Apesar da relevância dessas questões, identificar riscos à privacidade não é suficiente para orientar as decisões de desenvolvimento. É necessário traduzir essas preocupações em requisitos que possam ser verificados, como limitar a coleta aos dados necessários, definir prazos de retenção, restringir acessos e estabelecer procedimentos para exclusão das informações. Essa distinção também se aplica à avaliação de sistemas de suporte emocional: demonstrar que uma intervenção apresenta resultados favoráveis não significa que o software seja seguro diante de respostas inadequadas, exposição de dados ou situações que ultrapassem sua finalidade. O presente trabalho procura contribuir para essa lacuna ao relacionar os riscos discutidos na literatura a requisitos e mecanismos de segurança presentes na Engenharia de Software, sem presumir que essa proposta já tenha sido implementada ou validada empiricamente.

Diante desses problemas, este trabalho parte de uma questão voltada à Engenharia de Software: **Quais riscos técnicos, éticos e jurídicos associados ao uso de sistemas conversacionais de inteligência artificial para suporte emocional podem ser traduzidos em requisitos de software e mecanismos de**

segurança verificáveis? O objetivo é analisar as possibilidades e os limites dessas aplicações e relacionar os riscos identificados na literatura a requisitos que possam orientar seu desenvolvimento e sua avaliação. A discussão abrange confiabilidade, privacidade, transparência sobre as limitações do sistema, supervisão humana e encaminhamento para atendimento profissional quando necessário.

A contribuição pretendida não é demonstrar a eficácia clínica de um chatbot específico, mas examinar como os problemas identificados na literatura podem ser traduzidos em decisões de engenharia, requisitos de software e mecanismos de segurança.

Para desenvolver a análise, foi adotada uma abordagem bibliográfica e documental, baseada em publicações acadêmicas, documentos públicos e legislação. Não foram realizadas entrevistas, questionários, experimentos, testes com usuários ou coleta de relatos pessoais. Os resultados correspondem à análise das fontes selecionadas e não a uma avaliação empírica do desempenho de ferramentas em situações reais.

Depois desta introdução, a seção de trabalhos relacionados apresenta os procedimentos adotados na pesquisa. Já as seções seguintes discutem os sistemas conversacionais e seu uso no suporte emocional, examinam os riscos técnicos, éticos e jurídicos envolvidos e analisam suas implicações para a Engenharia de Software. As considerações finais apresentam as conclusões da análise e suas limitações.

2. FUNDAMENTAÇÃO TEÓRICA OU REVISÃO DA LITERATURA

Cho *et al.* (2026) realizaram uma revisão sistemática sobre chatbots baseados em grandes modelos de linguagem e sistemas de

inteligência artificial com capacidades de atuação autônoma voltados ao aconselhamento em saúde mental. Os autores examinaram as metodologias de desenvolvimento, as formas de avaliação e as salvaguardas éticas descritas nos estudos selecionados. Entre os problemas identificados, destacam-se a ausência de validação clínica independente em contextos reais de atendimento, a inconsistência na apresentação dos procedimentos de segurança e as limitações na documentação necessária para reproduzir e avaliar os sistemas. Os resultados indicam que, embora essas aplicações apresentem possibilidades de uso, ainda existem lacunas importantes na avaliação de sua segurança e confiabilidade
.

Em uma perspectiva mais diretamente voltada à mitigação de riscos, Olisaeloka *et al*. (2026) conduziram uma revisão de escopo sobre mecanismos de segurança empregados em chatbots generativos para saúde mental. A análise reuniu 21 estudos e organizou as estratégias identificadas em três domínios: medidas técnicas relacionadas aos sistemas de inteligência artificial, salvaguardas anteriores à implantação e ações de mitigação durante a utilização. Os autores identificaram problemas como a dificuldade de reconhecer situações de risco, a presença de informações clínicas incorretas, a limitação da supervisão humana e a insuficiência de procedimentos para situações de crise e para o acompanhamento de eventos adversos.

Os dois trabalhos contribuem para compreender os desafios relacionados à segurança dos sistemas conversacionais de inteligência artificial aplicados à saúde mental. Enquanto Cho *et al*. (2026) enfatizam as metodologias de desenvolvimento, os critérios de avaliação e a documentação das salvaguardas, Olisaeloka *et al*.

(2026) concentram-se na identificação e organização dos mecanismos de segurança e das estratégias de mitigação de riscos.

Diferentemente desses estudos, este artigo direciona a análise à Engenharia de Software, relacionando os riscos técnicos, éticos e jurídicos identificados na literatura a uma matriz de riscos e a onze requisitos (R01 a R11) com critérios e métodos de verificação, constituindo uma proposta conceitual fundamentada no contexto regulatório brasileiro.

3. METODOLOGIA

Este trabalho foi desenvolvido por meio de uma revisão bibliográfica narrativa, complementada pela análise de documentos oficiais e normas relacionadas ao uso da inteligência artificial na saúde mental. O objetivo foi reunir e discutir estudos sobre agentes conversacionais, suporte emocional, ética, segurança da informação e proteção de dados, buscando relacionar os riscos identificados a possíveis requisitos de software.

O levantamento considerou artigos científicos, revisões da literatura, publicações acadêmicas e documentos institucionais disponíveis gratuitamente. Foram consultadas fontes nacionais e internacionais, incluindo SciELO, PePSIC, PubMed, JMIR, Fiocruz, eduCAPES, PGE-SP e portais institucionais. Também foram considerados documentos de órgãos públicos relacionados à proteção de dados e à regulamentação da inteligência artificial.

A seleção do material levou em conta a relação entre cada publicação e o problema investigado. Foram priorizados trabalhos que abordassem agentes conversacionais aplicados à saúde mental, inteligência artificial generativa, avaliação de resultados, riscos

éticos, privacidade, segurança e governança. Também foram considerados estudos que contribuíssem para compreender como esses riscos podem ser tratados no desenvolvimento de software.

Foram excluídos materiais promocionais, publicações sem autoria identificável e trabalhos sem relação direta com os objetivos do artigo. Estudos sobre outras aplicações da inteligência artificial na saúde foram considerados apenas quando contribuíam para a discussão de aspectos relevantes, como segurança, avaliação, privacidade ou responsabilidade. A exclusão de uma publicação, portanto, não significa necessariamente que ela tenha baixa qualidade, mas que seu conteúdo não atendia ao foco deste trabalho.

A análise do material selecionado foi conduzida de forma qualitativa. Os conteúdos foram organizados em torno dos riscos técnicos, éticos e jurídicos associados ao uso de sistemas conversacionais para suporte emocional. A partir dessa organização, foram discutidas possíveis relações entre os problemas identificados na literatura e requisitos de software ou mecanismos de segurança que possam ser especificados e verificados durante o desenvolvimento desses sistemas.

A pesquisa não envolveu entrevistas, questionários, experimentos com participantes ou coleta de dados pessoais. Seu escopo se limita à análise de publicações e documentos existentes. Assim, as propostas apresentadas constituem uma discussão conceitual fundamentada nas fontes consultadas, e não uma validação experimental de ferramentas ou de intervenções em saúde mental.

3.1. Busca e Seleção do Material Bibliográfico

A busca bibliográfica foi orientada por combinações de termos relacionados à inteligência artificial, agentes conversacionais, saúde mental, suporte emocional, modelos de linguagem, ética, segurança e privacidade.

As buscas foram realizadas em bases científicas e fontes institucionais, de acordo com os objetivos indicados na Tabela 1. A seleção considerou a pertinência dos títulos e resumos em relação ao tema. Em seguida, os materiais considerados relevantes foram examinados para verificar sua contribuição à discussão dos riscos e à identificação de medidas aplicáveis à Engenharia de Software.

A análise foi organizada em cinco eixos: (a) agentes conversacionais utilizados em contextos de saúde mental; (b) aplicações de inteligência artificial generativa nesse campo; (c) efetividade e limitações das intervenções estudadas; (d) riscos éticos, jurídicos e relacionados à segurança e à privacidade; e (e) requisitos técnicos e mecanismos de governança aplicáveis ao desenvolvimento de sistemas conversacionais. As buscas foram realizadas com variações das combinações de termos apresentadas na Tabela 1:

Tabela 1. Resultados das buscas bibliográficas e processo de triagem.

Base de busca	Estratégia de busca aplicada	Artigos encontrados	Lidos	Aproveitados
PubMed	conversational agents AND mental health	422	4	3
PubMed	large language model AND	1285	4	1

	mental health			
PubMed	artificial intelligence AND mental health AND ethics	1519	2	1
PubMed Central (PMC)	chatbot AND mental health AND artificial intelligence	840	3	1
PubMed Central (PMC)	large language model AND mental health	1285	3	1
PubMed Central (PMC)	generative AI AND mental health	858	3	1
JMIR	chatbot AND mental health AND artificial intelligence AND conversational agents	782	3	3
SciELO	inteligência artificial e bioética e saúde	17	2	1
SciELO	ética e inteligência artificial	219	2	1
SciELO	inteligência artificial e psicologia	33	1	1
Fiocruz	assistentes virtuais e saúde mental e regulação	1	1	1

Fiocruz	inteligência artificial e autonomia do paciente	1	1	1
eduCAPES	chatbots e saúde mental	5454	1	1
PePSIC	tecnologias digitais AND psicanálise	1	1	1
PGE-SP	proteção de dados e saúde	2	1	1
GOVBR	ANPD e inteligência artificial e dados pessoais de saúde	105	3	3
GOVBR	Resolução 510	1	1	1

Fonte: Elaborado pelo autor (2026).

Critério temporal: As buscas foram realizadas no primeiro semestre de 2026, no período de março a junho. O artigo trabalha principalmente com conteúdos relativos aos anos entre 2019 e 2026 para trabalhos científicos, 2016 e 2018 para legislações e normas.

Critério linguístico: Foram incluídos artigos em português e inglês. Publicações em outros idiomas foram excluídas quando não havia tradução disponível.

Critério de pertinência: O estudo precisa contribuir para pelo menos uma das dimensões centrais do artigo: eficácia/uso, arquitetura de agentes, risco, ética, privacidade, segurança, supervisão ou requisitos de software.

Foram selecionadas e analisadas 23 referências, sendo 18 publicações acadêmicas e 5 documentos normativos e institucionais. A seleção considerou a contribuição de cada material para o desenvolvimento deste artigo, priorizando aqueles que abordam agentes conversacionais e aplicações de inteligência artificial na saúde mental, a efetividade dessas ferramentas e as limitações dos resultados encontrados, além de questões éticas, fatores humanos, privacidade, segurança da informação e regulamentação dessas aplicações.

3.2. Organização e Desenvolvimento da Pesquisa

A análise do material selecionado foi conduzida qualitativamente, por meio de uma síntese temática das publicações e dos documentos consultados. Inicialmente, foram identificadas as contribuições de cada fonte para a compreensão dos sistemas conversacionais, de suas aplicações em saúde mental e dos problemas técnicos, éticos e jurídicos associados ao seu uso. Em seguida, os problemas discutidos nas fontes foram agrupados conforme sua natureza e as possíveis consequências para os usuários e para o funcionamento do sistema.

A partir desse agrupamento, foram estabelecidas relações entre cada risco identificado, sua consequência potencial e as medidas que poderiam ser incorporadas ao desenvolvimento do software. Esse procedimento permitiu formular requisitos propostos, como transparência sobre os limites do sistema, proteção de dados pessoais, avaliação das respostas, monitoramento de alterações e encaminhamento para apoio humano em situações críticas. As relações entre riscos, consequências, requisitos e mecanismos de

segurança foram organizadas nas tabelas apresentadas nas seções seguintes.

O procedimento metodológico que orientou o desenvolvimento da pesquisa está sintetizado na Tabela 2:

Tabela 2. Etapas do procedimento metodológico da pesquisa.

Etapa	Denominação	Atividades e Questões Orientadoras
1	Levantamento Bibliográfico e Documental	Seleção de publicações brasileiras e internacionais de acesso livre (2019–2026), legislação nacional (LGPD, Resolução CNS nº 510/2016) e documentos sobre saúde digital e IA.
2	Análise Crítica e Multidisciplinar	Análise conceitual e crítica da literatura já existente, distinção entre geração estatística da IA e diagnóstico clínico, e mapeamento de riscos de privacidade e dependência.
3	Mapeamento de Requisitos de Software	Identificação dos riscos na literatura e sua tradução em requisitos de software para prevenir ou reduzir seus impactos, como proteção de dados, transparência e definição dos limites de atuação da IA.
4	Sistematização e Considerações Finais	Consolidação da visão sociotécnica e proposição de diretrizes para o desenvolvimento responsável de sistemas de suporte emocional.

Fonte: Elaborado pelo autor (2026).

3.3. Limitações da Pesquisa

Como limitações deste estudo, destaca-se o caráter bibliográfico e documental da pesquisa, baseado em revisão narrativa da literatura. A seleção e a interpretação das publicações foram realizadas por um único pesquisador, sem avaliação independente por outros revisores, o que pode influenciar a escolha das fontes e a interpretação dos resultados. Além disso, a análise depende das informações disponíveis nos trabalhos consultados e não incluiu testes empíricos, implementação prática ou a aplicação dos critérios de verificação a nenhum chatbot ou sistema conversacional existente.

Os métodos e requisitos propostos destinam-se exclusivamente à verificação conceitual da especificação de projeto, não tendo sido testados em ambiente de produção ou em cenários reais de atendimento. Dessa forma, os resultados permitem discutir riscos e formular recomendações conceituais para a Engenharia de Software, mas não permitem afirmar que os requisitos apresentados sejam suficientes para eliminar os riscos identificados ou que sua eficácia tenha sido comprovada empiricamente.

4. RESULTADOS E DISCUSSÕES

4.1. Sistemas Conversacionais e a Aproximação com a Linguagem Humana

O cenário global de saúde mental enfrenta um problema crítico na oferta de atendimento presencial especializado. Segundo Vaidyam *et al.* (2019), como descrito na introdução deste artigo, existe grande desigualdade de acesso a profissionais de saúde mental. Como consequência dessa escassez estrutural, a busca por alternativas tecnológicas ganhou força, impulsionando a Engenharia de

Software no desenvolvimento de sistemas conversacionais e chatbots projetados para interagir em linguagem natural e prover suporte emocional preliminar. Eles podem utilizar regras previamente definidas, modelos estatísticos, técnicas de aprendizado de máquina ou modelos de linguagem mais recentes. Meurer (2020) descreve os chatbots como agentes de conversação capazes de interagir por diferentes meios e discute seu potencial no apoio à saúde mental.

Embora os sistemas generativos consigam produzir respostas coerentes e interagir em linguagem natural, essa capacidade não representa compreensão humana ou experiência emocional. Suas respostas são produzidas a partir de padrões computacionais, não de uma compreensão clínica ou subjetiva do usuário (Zaidhaft; Reis; Winograd, 2024; Peixoto, 2025).

4.2. O Que a Evidência Permite Dizer Sobre Efetividade

A literatura experimental recente exige uma distinção entre a existência de resultados positivos e a demonstração de eficácia geral. Uma revisão sistemática e meta-análise publicada em 2023 reuniu 32 estudos, envolvendo 6.089 participantes, e encontrou efeitos favoráveis de curto prazo dos agentes conversacionais para diferentes desfechos, incluindo sintomas depressivos, ansiedade, estresse e bem-estar. Entretanto, os efeitos de longo prazo foram menos consistentes para diversos desfechos, o que exige cautela ao interpretar esses resultados como evidência de eficácia duradoura ou de adequação para qualquer situação clínica (He *et al.*, 2023).

Uma revisão sistemática publicada em 2025 analisou 29 estudos provenientes de ensaios clínicos randomizados sobre agentes

conversacionais utilizados no manejo de sintomas físicos e psicológicos. Entre os estudos incluídos, 17 avaliaram sintomas psicológicos, como depressão e ansiedade, 4 investigaram sintomas físicos e 8 abordaram ambas as categorias. Embora a maioria dos estudos tenha relatado experiências positivas e melhorias no manejo dos sintomas, 17 apresentaram alto risco de viés e outros 9 levantaram preocupações metodológicas. Esses resultados indicam que existem evidências favoráveis ao uso dessas ferramentas, mas a qualidade dos estudos limita a segurança das conclusões (Yang *et al.*, 2025).

Em jovens, uma meta-análise publicada em 2025 reuniu 14 artigos, correspondentes a 15 ensaios, com 1.974 participantes. Após o ajuste para viés de publicação, os agentes conversacionais baseados em inteligência artificial apresentaram efeito moderado a grande sobre sintomas depressivos. Contudo, os efeitos ajustados para ansiedade generalizada, estresse, afetos positivos e negativos e bem-estar mental não foram estatisticamente significativos. Esses resultados reforçam a necessidade de evitar generalizações entre diferentes problemas de saúde mental (Feng *et al.*, 2025).

A revisão sistemática de Wang, Zhou e Zhou (2025) destaca que os resultados obtidos com agentes conversacionais tradicionais não devem ser generalizados para sistemas generativos baseados em modelos de linguagem de grande escala. Por isso, a avaliação dessas ferramentas também deve considerar suas implicações éticas, a confiabilidade das respostas e as limitações das evidências disponíveis.

Contudo, neste artigo, a eficácia deve ser entendida como uma propriedade dependente da tarefa, do contexto, do público e da

qualidade das evidências disponíveis, enquanto a segurança deve ser tratada como requisito independente. A avaliação de sistemas de suporte emocional precisa considerar não apenas os resultados observados, mas também a privacidade, a transparência, os mecanismos de encaminhamento e os limites de atuação.

4.3. O Potencial de Apoio e Seus Limites

A literatura brasileira disponível não permite sustentar uma conclusão absoluta de que os chatbots são eficazes para qualquer necessidade de saúde mental. O estudo de Meurer (2020), elaborado como revisão de publicações sobre chatbots e saúde mental, encontrou resultados favoráveis na maior parte dos trabalhos analisados, mas também registrou ressalvas quanto à necessidade de pesquisas adicionais para validar protocolos e resultados. A própria forma como o estudo apresenta suas conclusões recomenda cautela.

Também é importante distinguir ferramentas de apoio de psicoterapia. Um sistema pode oferecer informação, exercícios estruturados, organização de rotinas, encaminhamento para serviços ou uma forma inicial de interação sem realizar, por exemplo, uma avaliação psicológica completa. A discussão brasileira sobre IA generativa na Psicologia reforça que a expansão desses recursos traz novas possibilidades, mas também exige competências, critérios éticos e regulação adequados (Peixoto, 2025).

Essa distinção evita dois extremos. O primeiro seria tratar qualquer chatbot como uma forma de tratamento psicológico. O segundo seria concluir que toda utilização é inútil. A posição mais coerente com a literatura analisada é intermediária: determinadas aplicações

podem funcionar como recursos complementares, mas seus limites precisam ser explicitados e sua segurança precisa ser tratada como requisito do sistema.

4.4. Proximidade, Linguagem e Dependência

Um sistema conversacional pode utilizar linguagem acolhedora, manter informações do contexto e adaptar suas respostas. Esses recursos podem aumentar a sensação de proximidade. O problema aparece quando a experiência do usuário passa a ocultar a natureza artificial da interação ou cria expectativas que o sistema não tem condições de cumprir.

Bentes, Sanches e Fonseca (2024), ao analisarem assistentes virtuais inteligentes e saúde mental no contexto brasileiro, destacam que essas tecnologias já são apropriadas para finalidades de cuidado e que a legislação brasileira ainda apresenta insuficiências para lidar com todos os riscos envolvidos. A questão, portanto, não se resume à qualidade da resposta: envolve também o desenho da relação entre usuário e sistema.

Do ponto de vista da Engenharia de Software, elementos da interface e regras de negócio podem influenciar a relação entre o usuário e o chatbot. Vaidyam *et al.* (2019) apontam que algumas pessoas podem desenvolver apego excessivo a esses sistemas, com o risco de desconforto quando não conseguem acessá-los ou de afastamento das relações presenciais. Embora essa possibilidade não signifique que todo uso frequente leve à dependência, ela precisa ser considerada nas decisões de projeto. Recursos que apresentam a IA como uma companhia permanente ou dificultam o acesso a apoio humano podem agravar esse risco. Por isso, a

aplicação deve deixar claros seus limites e facilitar o encaminhamento para ajuda profissional quando a situação relatada exigir acompanhamento especializado.

4.5. Riscos Técnicos, Éticos e Jurídicos

Com base na síntese integrativa descrita na metodologia, os riscos identificados foram organizados em seis grupos, detalhando suas consequências, requisitos de mitigação e referências de fundamentação conforme mostra a Tabela 3:

Tabela 3. Riscos, consequências, requisitos e referências de fundamentação

Grupo	Exemplo	Consequência	Requisitos	Autor(es)
Saída	Alucinação ou resposta incorreta	Orientação incorreta e indução ao erro	Avaliação automática e humana das respostas; mecanismo de fallback.	Hua *et al.* (2025); Cho *et al.* (2026); Olisaeloka *et al.* (2026).
Interação	Antropomorfização e falsa empatia	Confiança excessiva ou dependência afetiva	Identificação explícita da natureza artificial; controle do tom da interação.	Vaidyam *et al.* (2019); Bentes *et al.* (2024).
Dependência	Uso prolongado como substituto de apoio profissional	Possível atraso na busca de suporte especializado	Limites de uso; avisos sobre as limitações; encaminhamento para	Vaidyam *et al.* (2019); Bentes *et al.* (2024).

			apoio humano.	
Dados	Armazenamento de conversas sensíveis	Exposição de dados ou violação de privacidade	Minimização e mascaramento de dados; retenção limitada e descarte programado.	Brasil (2018); ANPD (2024); Olisaeloka *et al.* (2026).
Governança	Atualização de modelo sem validação formal	Regressão de desempenho ou falhas não detectadas	Testes regressivos; auditoria de versões; possibilidade de reversão (rollback).	Cho *et al.* (2026); Singhal *et al.* (2024); Shankar *et al.* (2026).
Escalonamento	Não reconhecer situações críticas	Agravamento do risco ao usuário	Classificação de situações críticas; protocolo de transição para apoio humano.	Hua *et al.* (2025); Cho *et al.* (2026); Olisaeloka *et al.* (2026).

Fonte: Elaborado pelo autor (2026).

4.6. Confiabilidade das Respostas

Um dos principais riscos de sistemas generativos é a produção de respostas que parecem adequadas, mas apresentam informações incorretas ou inadequadas. Em uma conversa comum, um erro pode ser corrigido sem grandes consequências. Em situações relacionadas à saúde mental, a consequência pode ser maior, especialmente quando a pessoa procura orientação em um momento de sofrimento intenso.

A revisão brasileira de inteligência artificial em saúde e bioética destaca a relação da tecnologia com questões de imprevisibilidade, confiança, segurança de dados, responsabilidade e desenvolvimento de sistemas (Elias *et al.*, 2023). A literatura sobre IA generativa na Psicologia também chama atenção para os desafios éticos decorrentes da capacidade de produzir linguagem e interagir de forma aparentemente personalizada (Peixoto, 2025).

Na Engenharia de Software, a resposta para esse problema não deve ser apenas escrever um aviso de que a IA pode errar. É necessário pensar em mecanismos de segurança, testes, limites de escopo, monitoramento e encaminhamento. Quanto maior o risco associado à decisão ou à orientação, menor deve ser a autonomia concedida ao sistema.

4.7. Privacidade e LGPD

Conversas de suporte emocional podem revelar informações íntimas e, dependendo do conteúdo, dados relacionados à saúde, classificados pela Lei Geral de Proteção de Dados Pessoais (LGPD) como dados pessoais sensíveis (Brasil, 2018). Isso torna inadequada a premissa de que mensagens digitadas em um chatbot configuram mero texto sem consequências jurídicas.

O tratamento dessas informações exige a observância estrita aos princípios da LGPD, destacando-se a finalidade, a adequação, a necessidade e a segurança (Brasil, 2018). A aplicação prática do princípio da minimização implica que o sistema não deve armazenar automaticamente todo o fluxo conversacional; se uma funcionalidade não exige memória persistente, a retenção deve ser questionada. Quando a persistência for necessária, devem ser

definidos prazos claros de retenção e procedimentos de expurgo, alinhando-se às hipóteses legais de conservação (Brasil, 2018; Autoridade Nacional de Proteção de Dados, s.d.).

Nesse sentido, a Autoridade Nacional de Proteção de Dados (ANPD) reforça que o tratamento de dados de saúde exige medidas técnicas e administrativas rigorosas para mitigar riscos de acesso não autorizado. Essa agenda regulatória ganha ainda mais relevância frente ao tratamento de dados de alto risco (Autoridade Nacional de Proteção de Dados, 2024).

Para o desenvolvimento de software, recomenda-se a adoção de uma abordagem de Privacy by Design, respondendo a indagações fundamentais desde a concepção: quais dados são realmente necessários, quem poderá acessá-los, por quanto tempo serão mantidos e como o titular poderá exercer seus direitos. Dessa forma, a privacidade deixa de ser tratada apenas como um termo de uso extensivo e passa a integrar diretamente a arquitetura do sistema (Simões, 2025).

4.8. Autonomia, Supervisão e Responsabilidade

A autonomia do usuário é outro ponto central. Um sistema que apresenta suas respostas como certezas pode induzir a pessoa a atribuir ao algoritmo uma autoridade maior do que aquela que ele possui. Malta e Lamy (2025) destacam, na discussão brasileira sobre inteligência artificial em saúde, os riscos para a autonomia quando existe baixa transparência ou ausência de supervisão humana.

Supervisão humana não deve ser tratada como um botão de emergência acrescentado no final do projeto. Ela precisa ser definida como parte da arquitetura e do fluxo operacional. Isso envolve

determinar em quais situações o sistema deixa de responder de maneira autônoma, quem recebe o encaminhamento, quais informações podem ser compartilhadas e como o incidente será registrado (Singhal *et al.*, 2024).

Também é necessário distinguir supervisão de transferência de responsabilidade. O fato de existir um profissional disponível não torna automaticamente segura uma aplicação. A equipe responsável pelo software continua tomando decisões sobre modelo, interface, memória, dados, limites de atuação e mecanismos de escalonamento. A governança precisa, portanto, acompanhar o ciclo de vida da aplicação (Singhal *et al.*, 2024; Shankar; Lim; Xu, 2026).

Uma consequência prática é a necessidade de rastreabilidade. Quando uma resposta problemática ocorrer, deve ser possível reconstruir, dentro dos limites de privacidade, qual versão do modelo estava ativa, quais componentes participaram da geração, quais regras foram acionadas e quais informações foram utilizadas. Essa capacidade é necessária para investigação de incidentes e para melhoria contínua (Singhal *et al.*, 2024; Shankar; Lim; Xu, 2026).

A responsabilidade não pode ser deslocada integralmente para o usuário, pois as decisões de projeto influenciam diretamente os riscos da aplicação. É essencial projetar interfaces transparentes e travas operacionais para que o sistema atue como ferramenta complementar, reconhecendo seus limites e encaminhando o usuário para redes de apoio.

Como aponta Santaella (2025), a ética é fundamental para orientar o desenvolvimento da inteligência artificial e definir com clareza os

seus limites de atuação. Por essa razão, a governança deve acompanhar todo o ciclo de vida do software. A avaliação não termina quando o sistema é publicado. Mudanças no modelo, nos dados, na infraestrutura ou no comportamento dos usuários podem alterar o perfil de risco. Uma aplicação responsável precisa de mecanismos de acompanhamento, registro de incidentes e revisão periódica.

4.9. Arquitetura Conceitual de Segurança

A partir dos riscos identificados na seção anterior, é proposta uma arquitetura conceitual em camadas. A proposta não representa uma implementação testada, logo, não deve ser apresentada como arquitetura validada experimentalmente. Seu objetivo é organizar os mecanismos que deveriam existir em um sistema de suporte emocional baseado em IA antes que a interação da IA generativa seja tratada como componente central da aplicação.

Na interface, o sistema deve informar de maneira transparente e persistente que a comunicação ocorre com um agente artificial e qual é sua finalidade. A camada da classificação de risco deve analisar a conversa segundo critérios previamente definidos e encaminhar situações que ultrapassem os limites de risco do projeto. A camada de geração pode utilizar um LLM, mas não deve possuir autonomia irrestrita: suas entradas e saídas precisam passar por mecanismos de controle. A camada de conhecimento da IA deve permitir verificar quando a resposta depende de fontes externas. Por último, a camada de governança deve registrar eventos necessários à auditoria, controlar retenção e acesso e permitir revisão periódica (Singhal *et al.*, 2024; Shankar; Lim; Xu, 2026).

4.10. Matriz de Risco e Prioridade de Mitigação

Para traduzir a discussão qualitativa em uma ferramenta de engenharia, a Tabela 4 organiza uma matriz dos riscos identificados na literatura. Os níveis e critérios de mitigação propostos baseiam-se na frequência de falhas relatadas em sistemas conversacionais.

Tabela 4. Matriz de riscos e prioridades de mitigação.

Risco	Probabilidade conceitual	Impacto	Prioridade	Mitigação
Resposta inadequada em situação de sofrimento	Alta	Muito alto	Crítica	Classificação de risco, guardrails, fallback e encaminhamento.
Alucinação da IA	Alta	Alto	Alta	RAG quando aplicável, avaliação de saída e fontes.
Dependência do usuário à IA	Média	Alto	Alta	Limites de finalidade, transparência e estímulo à rede humana.
Exposição de dados sensíveis	Média	Muito alto	Crítica	Minimização, criptografia, controle de acesso e retenção.
Mudança de comportame	Média	Alto	Alta	Testes regressivos e

nto após atualização				monitoramento.
Viés de linguagem e cultura	Média	Alto	Alta	Testes com diferentes grupos culturais e revisão de dados.
Falha de encaminhamento	Baixa/Média	Muito alto	Crítica	Protocolo de escalonamento e testes de cenários.

Fonte: Elaborado pelo autor (2026).

4.11. Autonomia da IA por Nível de Risco

A Tabela 5 sistematiza os níveis conceituais de risco e estabelece as ações esperadas do sistema conforme o grau de autonomia da IA.

Tabela 5. Níveis de classificação de risco e ações do sistema.

Nível	**Situação conceitual**	**Autonomia da IA**	**Ação do sistema**
1 — informativo	Perguntas gerais, educação, organização	Alta	Responder dentro do escopo declarado.
2 — apoio	Estresse, solidão, sofrimento sem sinais de crise	Moderada	Responder + orientar recursos humanos.
3 — atenção	Sofrimento relevante ou persistente	Baixa/moderada	Resposta restrita + incentivo a atendimento.

4 — alto risco	Indícios de situação crítica	Baixa	Interromper fluxo normal + protocolo de encaminhamento.
5 — emergência	Risco imediato	Mínima	Não tentar substituir intervenção humana; orientar busca de ajuda imediata.

Fonte: Elaborado pelo autor (2026).

4.12. Implicações para Engenharia de Software

A Tabela 6 relaciona os riscos mapeados aos requisitos de software e aos respectivos mecanismos de segurança e arquitetura.

Tabela 6. Mapeamento de riscos, requisitos de software e mecanismos de segurança.

Risco Mapeado	Requisito de Software	Mecanismo de Segurança / Arquitetural
Atribuição de autoridade clínica / Ilusão de presença humana	Transparência e Limites de Finalidade: O sistema deve informar de forma clara sua natureza automatizada e seu escopo não clínico.	Interface e Disclaimers: Exibição contínua de aviso explicitando tratar-se de um agente computacional sem capacidade diagnóstica ou terapêutica.
Dependência emocional e isolamento social	Indução ao Encaminhamento Humano: Requisito de negócio para limitar interações nocivas e promover a busca por ajuda especializada.	Botão de Emergência / Atravessamento de Rede: Exibição visível e acessível de contatos da rede de apoio (ex.: CVV 188 e CAPS) e envio de alertas em conversas prolongadas.

Respostas inadequadas ou alucinações em momentos de crise	Segurança por Design e Limites de Atuação: Garantir a contenção de conteúdos nocivos e interrupção imediata em quadros de alto risco.	Guardiões de Conteúdo e Travas Operacionais: Filtros de palavras-chave e classificadores de intenção que bloqueiam a IA e acionam protocolos rígidos em menções a situações críticas ou emergências.
Vazamento ou exposição de dados pessoais sensíveis	Privacy by Design e Minimização: Coletar e reter estritamente o necessário para o funcionamento do chat, atendendo à LGPD.	Criptografia e Controles de Acesso: Criptografia em trânsito e em repouso, pseudonimização e políticas rígidas de retenção e expurgo de dados.
Comportamento imprevisto após atualizações	Governança e Avaliação Contínua: O software deve ser monitorado continuamente ao longo de todo o seu ciclo de vida.	Auditoria de Logs e Testes Regressivos: Monitoramento de saídas do modelo, registro de incidentes e revisões periódicas de conformidade.

Fonte: Elaborado pelo autor (2026).

Essa análise permite identificar alguns requisitos que deveriam ser considerados desde as primeiras etapas de um projeto de suporte emocional baseado em IA. A Engenharia de Software deixa de tratar ética e privacidade como assuntos externos ao desenvolvimento e passa a incorporá-los ao próprio ciclo de vida do produto.

A partir das dimensões levantadas, a Tabela 7 detalha os requisitos de software propostos (R01 a R11), associando cada diretriz ao risco correspondente, ao critério e ao método de verificação.

Tabela 7. Requisitos do sistema, riscos associados e critérios de verificação.

ID	Dimensão	Risco	Requisito	Critério de verificação	Método de verificação
R01	Finalidade	Atribuição clínica / Falsa presença humana	Declarar escopo não clínico e ausência de finalidade terapêutic	Apresentação explícita da finalidade e dos limites	Análise das mensag s iniciai dos por de avis na

⚠ Esta tabela possui muitas colunas e foi cortada para impressão. Para visualizá-la completa, acesse o artigo original em: https://revistatopicos.com.br/artigos/o-uso-da-inteligencia-artificial-no-suporte-emocional-uma-analise-de-riscos-eticos-tecnicos-e-requisitos-de-seguranca-na-engenharia-de-software?noblockage

Fonte: Elaborado pelo autor (2026).

Os estudos analisados convergem quanto ao potencial dos agentes conversacionais para ampliar acesso e oferecer determinadas formas de apoio, mas não sustentam uma conclusão uniforme sobre eficácia. A diferença entre resultados de curto e longo prazo, a heterogeneidade das intervenções e a presença de risco de viés em parte considerável dos estudos indicam que o desempenho do sistema depende de características que não podem ser reduzidas ao fato de utilizar ou não inteligência artificial.

Essa constatação muda a forma como os resultados devem ser traduzidos para a Engenharia de Software. Se a eficácia depende da finalidade, do público e da forma de interação, então a especificação

do sistema precisa declarar esses elementos. Do mesmo modo, se os riscos aumentam quando o sistema possui maior autonomia e capacidade generativa, os mecanismos de controle precisam acompanhar essa autonomia.

A proposta deste trabalho procura fazer essa ponte. Em vez de apresentar ética e privacidade como tópicos externos ao desenvolvimento, os riscos são convertidos em requisitos que podem ser discutidos durante elicitação, arquitetura, implementação, testes e operação. O valor da proposta está menos em afirmar que esses riscos existem e mais em organizar como eles podem influenciar decisões de projeto.

Essa abordagem também permite identificar uma limitação importante da literatura: ainda existe diferença entre demonstrar que uma intervenção é aceitável ou capaz de produzir determinado efeito e demonstrar que o sistema permanece seguro quando utilizado em condições abertas e imprevisíveis. Para sistemas generativos, a avaliação precisa considerar tanto métricas de resultado quanto cenários de falha, privacidade, robustez e comportamento fora do escopo.

5. CONSIDERAÇÕES FINAIS

A análise realizada mostra que a inteligência artificial pode desempenhar um papel de apoio em determinadas situações relacionadas à saúde mental, mas os resultados analisados não permitem considerá-la substituta de profissionais qualificados, especialmente em situações que exigem avaliação clínica ou intervenção diante de situações de risco. Por essa razão, é importante que os usuários sejam informados de forma clara e

transparente sobre as limitações desses sistemas. O uso contínuo e a acomodação às orientações automatizadas fornecidas pela aplicação também podem distorcer a percepção dos limites da IA e favorecer uma confiança excessiva em suas respostas, levando o usuário a adiar a busca por atendimento profissional. A literatura consultada reconhece o potencial dos chatbots e de outras aplicações de IA, ao mesmo tempo em que destaca a necessidade de validação, supervisão, regulação e cuidados com os dados.

Em resposta à questão sobre quais riscos técnicos, éticos e jurídicos associados ao uso de sistemas conversacionais de inteligência artificial para suporte emocional podem ser traduzidos em requisitos de software e mecanismos de segurança verificáveis, a análise identificou riscos relacionados a respostas inadequadas, atribuição de autoridade clínica ao sistema, dependência emocional, exposição de dados pessoais sensíveis, falhas de encaminhamento e vieses linguísticos e culturais. Esses riscos podem orientar requisitos de transparência sobre a natureza e os limites da aplicação, proteção de dados desde a concepção, controle de acesso, limitação da coleta e da retenção de informações, testes de segurança em diferentes cenários e definição de procedimentos de encaminhamento para atendimento humano. A proposta organiza essas medidas em requisitos e mecanismos que podem ser examinados ao longo do ciclo de vida do software. Sua definição, contudo, não garante isoladamente a segurança da aplicação, pois a adequação desses mecanismos depende de avaliação, acompanhamento e revisão.

Como o estudo é bibliográfico e documental, os requisitos apresentados constituem uma proposta fundamentada na literatura analisada, e não resultados de implementação ou validação empírica.

O principal resultado desta revisão é, portanto, uma conclusão de equilíbrio. O problema não está em utilizar inteligência artificial para qualquer forma de suporte, nem em aceitar automaticamente toda aplicação como benéfica. O que importa é a finalidade do sistema, o contexto de uso, o nível de risco e a forma como o software foi projetado. Todos os objetivos da aplicação devem ser explícitos e transparentes, contribuindo para que o usuário compreenda com clareza os limites de atuação da inteligência artificial.

Para a Engenharia de Software, isso significa incorporar segurança, privacidade, transparência, limites de escopo e mecanismos de encaminhamento desde o início do desenvolvimento. Uma aplicação que conversa de maneira convincente não deve ser avaliada apenas pela fluidez de suas respostas. É necessário considerar também o que acontece quando ela erra, quando recebe uma informação sensível, quando o usuário passa a depender da interação ou quando a situação exige uma competência que o sistema não possui. Nesses casos, é necessário impor limites para a aplicação para garantir a integridade do usuário e recomendar ajuda profissional humana.

No Brasil, a LGPD oferece um conjunto importante de princípios e regras para proteção dos dados pessoais, enquanto a literatura mostra que a discussão sobre IA em saúde e saúde mental ainda está em construção. Esse cenário reforça a importância de soluções tecnológicas que sejam desenvolvidas com responsabilidade e de pesquisas acadêmicas que não antecipem conclusões além do que as evidências permitem.

Este estudo não realizou pesquisa de campo, avaliação clínica ou teste de usabilidade. Essa ausência foi deliberada e está de acordo

com o objetivo de realizar uma investigação exclusivamente bibliográfica e documental. Como desdobramento direto desta pesquisa, sugere-se que estudos futuros apliquem o conjunto de requisitos e critérios de verificação propostos (R01 a R11) a chatbots ou agentes conversacionais já existentes ou em fase de prototipagem, avaliando sua aplicabilidade prática e eficácia na mitigação dos riscos mapeados.

Pesquisas que pretendam avaliar usuários, profissionais ou pacientes deverão adotar metodologia própria e observar os procedimentos éticos aplicáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

Autoridade Nacional de Proteção de Dados (ANPD). Agenda Regulatória 2025–2026. Brasília, DF: ANPD, 2024. Aprovada pela Resolução CD/ANPD nº 23, de 9 de dezembro de 2024, com alterações posteriores. Disponível em: https://www.gov.br/anpd/pt-br/assuntos/regulacao/agenda-regulatoria-1/agenda-regulatoria-2025-2026. Acesso em: 10 mar. 2026.

Autoridade Nacional de Proteção de Dados (ANPD). Perguntas Frequentes. Seção sobre dados pessoais, dados pessoais sensíveis e hipóteses legais de tratamento. [s.d.]. Disponível em: https://www.gov.br/anpd/pt-br/acesso-a-informacao/perguntas-frequentes. Acesso em: 10 mar. 2026.

BENTES, Anna; SANCHES, Danielle; FONSECA, Paulo. Assistentes Virtuais Inteligentes e saúde mental: debates regulatórios no Brasil. Revista Eletrônica de Comunicação, Informação e Inovação em Saúde, Rio de Janeiro, v. 18, n. 3, 2024. DOI: https://doi.org/10.29397/reciis.v18i3.4310. Disponível em:

https://periodicos.fiocruz.br/index.php/pt-br/node/28099. Acesso em: 18 mar. 2026.

BRASIL. Conselho Nacional de Saúde. Resolução nº 510, de 7 de abril de 2016. Dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais cujos procedimentos metodológicos envolvam a utilização de dados diretamente obtidos com os participantes ou de informações identificáveis. Brasília, DF: Ministério da Saúde, 2016. Disponível em: https://www.gov.br/conselho-nacional-de-saude/pt-br/atos-normativos/resolucoes/2016/resolucao-no-510.pdf/view. Acesso em: 5 mar. 2026.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 5 mar. 2026.

CHO, Ha Na; WANG, Jiayuan; HU, Di; ZHENG, Kai. Large Language Model-Based Chatbots and Agentic AI for Mental Health Counseling: Systematic Review of Methodologies, Evaluation Frameworks, and Ethical Safeguards. JMIR AI, v. 5, e80348, 2026. DOI: https://doi.org/10.2196/80348. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC13032092/. Acesso em: 18 jun. 2026.

ELIAS, Mariele Abadia; FAVERSANI, Luciana Arruda; MOREIRA, Josiane Aparecida Vieira; MASIERO, Anelise Viapiana; CUNHA, Natalia Veronez da. Inteligência artificial em saúde e implicações bioéticas: uma revisão sistemática. Revista Bioética, Brasília, v. 31, 2023. DOI: https://doi.org/10.1590/1983-803420233542EN. Disponível

em: https://www.scielo.br/j/bioet/a/d9bswmTrshnRQSN6ff9WLkD/?lang=pt. Acesso em: 22 mar. 2026.

FENG, Yi; HANG, Yaming; WU, Wenzhi; SONG, Xiaohang; XIAO, Xiyao; DONG, Fangbai; QIAO, Zhihong. Effectiveness of AI-Driven Conversational Agents in Improving Mental Health Among Young People: Systematic Review and Meta- Analysis. Journal of Medical Internet Research, v. 27, e69639, 2025. DOI: https://doi.org/10.2196/69639. Disponível em: https://www.jmir.org/2025/1/e69639. Acesso em: 29 mar. 2026.

HE, Yuhao; YANG, Li; QIAN, Chunlian; LI, Tong; SU, Zhengyuan; ZHANG, Qiang; HOU, Xiangqing. Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. Journal of Medical Internet Research, v. 25, e43862, 2023. DOI: https://doi.org/10.2196/43862. Disponível em: https://www.jmir.org/2023/1/e43862. Acesso em: 25 mar. 2026.

HUA, Yining; SIDDALS, Steve; MA, Zilin; GALATZER-LEVY, Isaac; XIA, Winna; HAU, Christine; NA, Hongbin; FLATHERS, Matthew; LINARDON, Jake; AYUBCHA, Cyrus; TOROUS, John. Charting the evolution of artificial intelligence mental health chatbots from rule-based systems to large language models: a systematic review. World Psychiatry, v. 24, n. 3, p. 383–394, 2025. DOI: https://doi.org/10.1002/wps.21352. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC12434366/. Acesso em: 14 abr. 2026.

MALTA, Klauss Carvalho de; LAMY, Marcelo. Os desafios da autonomia do paciente frente ao uso da inteligência artificial na saúde. Ciência & Saúde Coletiva / Portal de Periódicos da Fiocruz,

2025. DOI: https://doi.org/10.17566/ciads.v14i4.1348. Disponível em: https://periodicos.fiocruz.br/pt-br/os-desafios-da-autonomia-do-paciente-frente-ao-uso-da-inteligencia-artificial-na-saude. Acesso em: 20 abr. 2026.

MEURER, Fábio. Chatbots: a efetividade da inteligência artificial no apoio à saúde mental. In: Ciências da Saúde: Teoria e Intervenção 3. Maringá: Unicesumar, 2020. cap. 7. DOI: https://10.0.88.5/at.ed.0942023047. Disponível em: https://educapes.capes.gov.br/bitstream/capes/569287/1/E-book%20Ci%C3%AAncias%20da%20Sa%C3%BAde%20Teoria%20e%20Interven%C3%A7%C3%A3o%203.pdf. Acesso em: 15 mar. 2026.

OLISAELOKA, Lotenna; RICHARDSON, Chris G.; WANG, Angel Y.; MUNTHALI, Richard J.; VIGO, Daniel V. Safety Mechanisms and Risk Mitigation in Generative AI Mental Health Chatbots: A Systematic Scoping Review. Healthcare, v. 14, n. 10, art. 1395, 2026. DOI: https://doi.org/10.3390/healthcare14101395. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC13205439/. Acesso em: 10 jun. 2026.

PEIXOTO, Adriano de Lemos Alves. Inteligência Artificial Generativa e seu Impacto na Formação e no Exercício Profissional da Psicologia. Psicologia: Ciência e Profissão, v. 45, n. spe1, e298224, 2025. DOI: https://doi.org/10.1590/1982-3703003298224. Disponível em: https://www.scielo.br/j/pcp/a/p3TgdLYJJTkLhzYy8K5T7wf/?lang=pt. Acesso em: 28 abr. 2026.

PEW RESEARCH CENTER. Search and work are the most common uses for chatbots, 1 in 10 use these tools for emotional support. 11 jun. 2026. Disponível em: https://www.pewresearch.org/chart/search-and-

work-are-the-most-common-uses-for-chatbots-1-in-10-use-these-tools-for-emotional-support/. Acesso em: 16 jun. 2026.

SANTAELLA, Lucia. A ética como guia para o uso da inteligência artificial generativa. Educação & Sociedade, v. 46, e296469, 2025. DOI: https://doi.org/10.1590/ES.296469. Disponível em: https://www.scielo.br/j/es/a/hnhzhvnQpRvvMwWts6z4bVs/?lang=pt. Acesso em: 2 mai. 2026.

SHANKAR, Ravi; LIM, Amaevia; XU, Qian. Barriers and Facilitators to the Use of Large Language Model-Based Conversational Agents in Mental Healthcare: A Systematic Review. Healthcare, v. 14, n. 10, 1267, 2026. DOI: https://doi.org/10.3390/healthcare14101267. Disponível em: https://pubmed.ncbi.nlm.nih.gov/42194358/. Acesso em: 5 jun. 2026.

SIMÕES, Izadora Rodrigues Normando. Tecnologia e saúde: a proteção de dados no Sistema Único de Saúde e a telemedicina. Revista da Escola Superior da Procuradoria-Geral do Estado de São Paulo, v. 16, n. 1, p. 135–166, 2025. DOI: https://doi.org/10.22491/respge.v16i1.1735. Disponível em: https://revistas.pge.sp.gov.br/revistaespgesp/article/view/1735. Acesso em: 8 mai. 2026.

SINGHAL, Sorabh; COOKE, Danielle L.; VILLAREAL, Ricardo I.; STODDARD, Joel J.; LIN, Chen-Tan; DEMPSEY, Allison G. Machine Learning for Mental Health: Applications, Challenges, and the Clinician's Role. Current Psychiatry Reports, v. 26, n. 12, p. 694–702, 2024. DOI: https://doi.org/10.1007/s11920-024-01561-w. Disponível em: https://pubmed.ncbi.nlm.nih.gov/39523249/. Acesso em: 15 mai. 2026.

VAIDYAM, Aditya Nrusimha; WISNIEWSKI, Hannah; HALAMKA, John David; KASHAVAN, Matcheri S.; TOROUS, John Blake. Chatbots and

Conversational Agents in Mental Health: A Review of the Psychiatric Landscape. Canadian Journal of Psychiatry. Revue canadienne de psychiatrie, v. 64, n. 7, p. 456-464, mar. 2019. DOI: https://doi.org/10.1177/0706743719828977. Disponível em: https://pubmed.ncbi.nlm.nih.gov/30897957/. Acesso em: 12 mar. 2026.

WANG, Xi; ZHOU, Yujia; ZHOU, Guangyu. The Application and Ethical Implication of Generative AI in Mental Health: Systematic Review. JMIR Mental Health, v. 12, e70610, 2025. DOI: https://doi.org/10.2196/70610. Disponível em: https://mental.jmir.org/2025/1/e70610. Acesso em: 22 mai. 2026.

YANG, Qingling; CHEUNG, Kin; ZHANG, Yan; ZHANG, Yazhou; QIN, Jing; XIE, Yao Jie. Conversational agents in physical and psychological symptom management: A systematic review of randomized controlled trials. International Journal of Nursing Studies, v. 163, 104991, 2025. DOI: https://doi.org/10.1016/j.ijnurstu.2024.104991. Disponível em: https://pubmed.ncbi.nlm.nih.gov/39799832/. Acesso em: 28 mai. 2026.

ZAIDHAFT, Eduardo; REIS, Ramon; WINOGRAD, Monah. A inteligência artificial pode substituir o psicanalista?: A crítica de Terrence Deacon à teoria computacional sobre a mente. Revista Brasileira de Psicanálise, v. 58, n. 4, p. 119–133, 2024. DOI: https://doi.org/10.69904/0486-641x.v58n4.08. Disponível em: https://pepsic.bvsalud.org/scielo.php?pid=S0486-641X2024000400119. Acesso em: 5 abr. 2026.