ReGeo ISSN: 2177-3246

INTELIGÊNCIA ARTIFICIAL CONVERSACIONAL EM SAÚDE MENTAL: LACUNAS CONCEITUAIS E REQUISITOS PARA UM FRAMEWORK INTEGRATIVO

CONVERSATIONAL ARTIFICIAL INTELLIGENCE IN MENTAL HEALTH: CONCEPTUAL GAPS AND REQUIREMENTS FOR AN INTEGRATIVE FRAMEWORK

INTELIGENCIA ARTIFICIAL CONVERSACIONAL EN SALUD MENTAL: LAGUNAS CONCEPTUALES Y REQUISITOS PARA UN MARCO INTEGRADOR

https://doi.org/10.56238/revgeov17n8-023

Data de submissão: 02/07/2026 **Data de publicação:** 02/08/2026

Izabela Souza da Silva
Doutora em Propriedade Intelectual
Instituição: Universidade Federal do Paraná (UFPR), Núcleo de Inovação Tecnológica (NIT/UFPR)
E-mail: izabellapsi100@gmail.com

Maria Goretti Fernandes
Doutora em Ciências da Saúde
Instituição: Universidade Federal do Paraná (UFPR), Departamento de Prevenção e Reabilitação em Fisioterapia
E-mail: gorettifernandes@ufpr.br

RESUMO

A inteligência artificial conversacional, que abrange agentes conversacionais, chatbots, assistentes baseados em grandes modelos de linguagem (LLMs) e aplicações de inteligência artificial generativa, expandiu-se rapidamente no campo da saúde mental. Entretanto, a literatura permanece fragmentada e carece de uma estrutura capaz de organizar os sistemas segundo sua autonomia clínica, seu estágio no ciclo de vida e os mecanismos terapêuticos mobilizados. Objetivo. Demonstrar, com base em evidências, a necessidade científica de um novo framework conceitual e especificar as condições que essa estrutura deverá satisfazer, sem propor o modelo nesta etapa. Método. Realizou-se um mapeamento sistemático de lacunas conceituais. Foram recuperados 758 registros desduplicados nas bases OpenAlex e PubMed, referentes ao período de 2015 a 2025. A triagem em duas etapas identificou 169 publicações com contribuição conceitual, das quais 51 compuseram o conjunto central de extração estruturada. Doze dimensões analíticas foram codificadas. Resultados. A cobertura mostrou-se desigual. Autonomia clínica (29%) e estágio do ciclo de vida (16%) foram as dimensões menos contempladas e coocorreram em apenas três modelos. Em média, cada modelo cobriu 6,7 das 12 dimensões; apenas dois apresentaram validação empírica e nenhum modelo de maturidade foi identificado. Somente dois modelos abordaram simultaneamente as três condições estruturais centrais, embora não se caracterizassem como frameworks integrativos. Conclusão. A literatura disponível descreve, regula e revisa a inteligência artificial conversacional em saúde mental, mas ainda não oferece uma arquitetura integradora. Um novo framework deverá articular autonomia clínica graduada, ciclo de vida, mecanismo terapêutico e evidência em uma estrutura única e testável.

Palavras-chave: Inteligência Artificial Conversacional. Saúde Mental. Grandes Modelos de Linguagem. Saúde Digital. Governança Algorítmica. Framework Conceitual.

ABSTRACT
Conversational artificial intelligence, including conversational agents, chatbots, assistants based on large language models (LLMs), and generative artificial intelligence applications, has rapidly expanded in mental health. However, the literature remains fragmented and lacks an integrative structure capable of organizing systems according to clinical autonomy, life-cycle stage, and therapeutic mechanisms. Objective. To demonstrate, based on evidence, the scientific need for a new conceptual framework and to specify the conditions that such a structure should meet, without proposing the model at this stage. Method. A systematic conceptual gap mapping was conducted. A total of 758 deduplicated records were retrieved from OpenAlex and PubMed for the period 2015-2025. Two-stage screening identified 169 publications with a conceptual contribution, of which 51 composed the core set for structured extraction. Twelve analytical dimensions were coded. Results. Coverage was uneven. Clinical autonomy (29%) and life-cycle stage (16%) were the least represented dimensions and co-occurred in only three models. On average, each model covered 6.7 of the 12 dimensions; only two had empirical validation, and no maturity model was identified. Only two models simultaneously addressed the three core structural conditions, although neither constituted an integrative framework. Conclusion. The available literature describes, regulates, and reviews conversational artificial intelligence in mental health, but it still does not provide an integrative architecture. A new framework should articulate graded clinical autonomy, life cycle, therapeutic mechanism, and evidence within a single, testable structure.

Keywords: Conversational Artificial Intelligence. Mental Health. Large Language Models. Digital Health. Algorithmic Governance. Conceptual Framework.

RESUMEN
La inteligencia artificial conversacional, que abarca agentes conversacionales, chatbots, asistentes basados en grandes modelos de lenguaje (LLM) y aplicaciones de inteligencia artificial generativa, se ha expandido rápidamente en el campo de la salud mental. Sin embargo, la literatura sigue fragmentada y carece de un marco capaz de organizar los sistemas según su autonomía clínica, su etapa en el ciclo de vida y los mecanismos terapéuticos movilizados. Objetivo. Demostrar, con base en la evidencia, la necesidad científica de un nuevo marco conceptual y especificar las condiciones que este marco debería satisfacer, sin proponer el modelo en esta etapa. Método. Se llevó a cabo un mapeo sistemático de brechas conceptuales. Se recuperaron 758 registros deduplicados de las bases de datos OpenAlex y PubMed, que cubren el período de 2015 a 2025. El cribado en dos etapas identificó 169 publicaciones con contribuciones conceptuales, de las cuales 51 conformaron el conjunto central para la extracción estructurada. Se codificaron doce dimensiones analíticas. Resultados. La cobertura resultó ser desigual. La autonomía clínica (29 %) y la etapa del ciclo de vida (16 %) fueron las dimensiones menos abordadas y coexistieron en solo tres modelos. En promedio, cada modelo cubrió 6,7 de las 12 dimensiones; solo dos presentaron validación empírica y no se identificó ningún modelo de madurez. Únicamente dos modelos abordaron simultáneamente las tres condiciones estructurales centrales, aunque no se caracterizaron como marcos integradores. Conclusión: La literatura disponible describe, regula y revisa la inteligencia artificial conversacional en salud mental, pero aún no ofrece una arquitectura integradora. Un nuevo marco debería articular la autonomía clínica gradual, el ciclo de vida, el mecanismo terapéutico y la evidencia en una estructura única y verificable.

Palabras clave: Inteligencia Artificial Conversacional. Salud Mental. Modelos de Lenguaje a Gran Escala. Salud Digital. Gobernanza Algorítmica. Marco Conceptual.

1 INTRODUÇÃO

A incorporação de sistemas de inteligência artificial capazes de sustentar diálogos em linguagem natural aos cuidados em saúde mental deixou de constituir uma possibilidade prospectiva e passou a integrar iniciativas de implementação em diferentes contextos assistenciais. Agentes conversacionais destinados à triagem, à psicoeducação, ao monitoramento de sintomas e ao apoio entre sessões evoluíram de protótipos acadêmicos a aplicações amplamente disponíveis. Concomitantemente, a difusão dos grandes modelos de linguagem (LLMs) generativos ampliou tanto a capacidade de interação quanto os riscos associados à segurança, à geração de conteúdo impreciso e à responsabilização. Revisões sistemáticas indicam efeitos favoráveis de agentes conversacionais sobre sintomas depressivos e sofrimento psicológico (Li et al., 2023), ao passo que sínteses recentes destacam riscos de alucinação, segurança e indefinição de responsabilidades (Lawrence et al., 2024).

A expressão “além dos chatbots” representa uma delimitação conceitual, e não apenas tecnológica. O termo chatbot designa predominantemente uma interface, embora o domínio compreenda sistemas heterogêneos: agentes baseados em regras ou modelos generativos; ferramentas de autoajuda ou sistemas de apoio à decisão clínica; aplicações informativas ou soluções que exercem diferentes níveis de autoridade terapêutica. A organização desse domínio requer critérios conceituais estáveis que permitam distinguir as aplicações quanto à função clínica, ao grau de autonomia, ao estágio de desenvolvimento e ao mecanismo de ação. Contudo, a literatura ainda ainda não oferece uma arquitetura capaz de integrar esses elementos de modo sistemático.

A proposição de novos frameworks sem demonstração prévia de sua necessidade pode ampliar a fragmentação conceitual já observada. Por essa razão, adotou-se uma estratégia de prova de necessidade: mapear sistematicamente as estruturas disponíveis, caracterizar suas contribuições e limitações e verificar se permanece uma lacuna não contemplada pelos referenciais existentes. Trata-se, portanto, de uma hipótese passível de verificação empírica no próprio corpus: um novo framework somente se justifica caso nenhuma estrutura atual integre, de forma articulada e testável, os eixos considerados essenciais ao uso clínico seguro.

O estudo teve quatro objetivos complementares: mapear frameworks, taxonomias, modelos teóricos, modelos de governança, modelos de maturidade e consensos ou diretrizes sobre inteligência artificial conversacional em saúde mental; extrair objetivos, dimensões conceituais, fundamentos teóricos, potencialidades, limitações e lacunas de um conjunto central de modelos; construir uma matriz comparativa e uma análise de cobertura dimensional; e determinar, com base nesse mapeamento, se uma nova estrutura conceitual se mostra cientificamente necessária. Nesta etapa, não se propõe um novo framework.

2 FUNDAMENTAÇÃO TEÓRICA

A análise foi organizada a partir de três tradições conceituais que estruturam o debate sobre inteligência artificial conversacional em saúde mental. Embora compartilhem objetos de interesse, essas tradições enfatizam dimensões distintas, o que favorece a fragmentação conceitual do domínio.

2.1 A TRADIÇÃO DA ALIANÇA TERAPÊUTICA

Na psicoterapia, a qualidade do vínculo entre paciente e terapeuta constitui um preditor consistente de desfechos. Quando transposta para agentes conversacionais, essa tradição suscita a questão de saber em que medida um sistema artificial pode estabelecer uma relação funcionalmente análoga à aliança de trabalho. Modelos de antropomorfismo em serviços digitais oferecem subsídios para compreender a atribuição de qualidades humanas a agentes artificiais (Blut et al., 2021). Paralelamente, estudos sobre agentes adaptativos examinam como personalidade e estilo conversacional influenciam o engajamento (Ahmad et al., 2022). Não obstante, permanece insuficientemente elucidado o mecanismo pelo qual a interação linguística produziria efeitos terapêuticos.

2.2 A TRADIÇÃO DA SAÚDE DIGITAL E DA CIÊNCIA DE IMPLEMENTAÇÃO

Na tradição da saúde digital e da ciência de implementação, o agente conversacional é concebido como uma intervenção que deve ser avaliada, regulada e incorporada aos fluxos assistenciais. Essa perspectiva prioriza métricas de eficácia, validação técnica e contexto de implementação, tendo produzido taxonomias de avaliação e modelos de governança (Abd-Alrazaq et al., 2020; Abbasian et al., 2024). Sua principal contribuição reside no rigor avaliativo. Entretanto, frequentemente pressupõe sistemas relativamente estáticos, o que limita sua aplicabilidade a modelos generativos que podem sofrer alterações de desempenho e comportamento ao longo do tempo.

2.3 A TRADIÇÃO DA ÉTICA E GOVERNANÇA DA IA

A tradição ética e de governança da inteligência artificial concentra-se em princípios como autonomia, não maleficência, justiça, transparência e responsabilização aplicados a chatbots de saúde mental (Coghlan et al., 2023; Vilaza; McCashin, 2021). Sua contribuição normativa é central para a adoção responsável dessas tecnologias. Todavia, essa tradição tende a pressupor que o objeto regulado já esteja conceitualmente definido. Sem a especificação do tipo de sistema, de seu grau de autonomia e de sua posição no ciclo de vida, os princípios de governança permanecem desvinculados do nível concreto de risco e responsabilidade.

2.4 A LACUNA ENTRE AS TRADIÇÕES

Em conjunto, as três tradições esclarecem dimensões complementares — mecanismo, avaliação e norma —, mas não fornecem uma ontologia que as integre. Desse modo, formulou-se a hipótese de que a fragmentação conceitual decorre da ausência de uma estrutura capaz de articular autonomia clínica, mecanismo terapêutico, evidência e ciclo de vida. O mapeamento realizado buscou verificar empiricamente essa hipótese.

3 MÉTODO

3.1 DELINEAMENTO

Conduziu-se um mapeamento sistemático de lacunas conceituais (conceptual gap mapping). Esse delineamento é apropriado quando o objetivo não consiste em estimar efeitos, mas em caracterizar a organização conceitual de um domínio e identificar dimensões sub-representadas ou ausentes. O protocolo foi estabelecido a priori e registrado como artefato da pesquisa, com definição de critérios de elegibilidade, procedimentos de triagem e esquema de extração.

3.2 CRITÉRIOS DE INCLUSÃO E EXCLUSÃO

Foram aplicados dois critérios cumulativos de elegibilidade. O primeiro, relativo ao objeto, exigiu que o estudo abordasse especificamente inteligência artificial conversacional, incluindo agentes conversacionais, chatbots, LLMs, inteligência artificial generativa, assistentes conversacionais e agentes relacionais no contexto da saúde mental. Foram excluídas aplicações restritas à visão computacional, ao aprendizado de máquina preditivo, ao processamento de sinais ou a outras modalidades sem interação dialógica. O segundo critério, referente à contribuição conceitual, exigiu a apresentação de framework, taxonomia, modelo teórico, modelo de governança, modelo de maturidade, consenso, diretriz ou revisão que sintetizasse uma estrutura organizadora. Estudos empíricos primários sem contribuição estrutural foram registrados, porém excluídos da extração aprofundada.

3.3 FONTES E BUSCA

As buscas foram conduzidas nas bases OpenAlex e PubMed, com combinações de termos relacionados ao objeto tecnológico, ao contexto de saúde mental e ao tipo de estrutura conceitual. O recorte temporal abrangeu publicações de 2015 a 2025. As estratégias de busca foram definidas previamente e mantidas constantes durante a recuperação dos registros.

Nota de transparência metodológica. A recuperação no repositório arXiv não foi concluída devido à indisponibilidade técnica transitória do serviço. Em conformidade com o protocolo, a análise prosseguiu com os registros recuperados nas bases OpenAlex e PubMed. Consequentemente, a cobertura corresponde à indexação dessas fontes e não deve ser interpretada como exaustiva.

3.4 SELEÇÃO E TRIAGEM

Os registros foram desduplicados sucessivamente por DOI, PMID e título normalizado. A triagem ocorreu em duas etapas. Na primeira, aplicou-se um filtro programático baseado na presença de indicadores de inteligência artificial conversacional e saúde mental, associado à ausência de modalidades excluídas. Na segunda, realizou-se classificação assistida para atribuição do tipo de estrutura e da decisão de inclusão. Posteriormente, selecionou-se um conjunto central para extração aprofundada, considerando relevância tipológica e impacto.

3.5 EXTRAÇÃO ESTRUTURADA

Cada item do conjunto central foi submetido a um protocolo de extração padronizado. Foram registrados o nome do modelo, o tipo de estrutura, os objetivos, as dimensões conceituais, os fundamentos teóricos, o contexto de uso, as potencialidades, as limitações, as lacunas não resolvidas, as oportunidades futuras, o estatuto evidencial e a presença ou ausência de 12 dimensões analíticas normalizadas. Sempre que uma dimensão não era abordada pela fonte, registrou-se “não informado”, sem inferência de atributos ausentes. Também se documentou a base textual de cada extração texto completo, resumo ou título para qualificar o nível de confiança.

3.6 AS DOZE DIMENSÕES ANALÍTICAS

As 12 dimensões analíticas foram definidas para viabilizar a comparação entre estruturas de natureza distinta: autonomia clínica; segurança e risco; governança e ética; mecanismo terapêutico; papel humano-IA; avaliação e validação; contexto de implementação; estágio do ciclo de vida; população e equidade; dados e privacidade; tecnologia baseada em LLMs ou inteligência artificial generativa; e evidência de eficácia.

3.7 SÍNTESE

Com base nas extrações, foram elaboradas uma matriz comparativa entre modelos e atributos, uma análise de cobertura dimensional, uma matriz de coocorrência entre dimensões e uma análise da

distribuição tipológica ao longo do tempo. Todas as frequências apresentadas nos resultados foram calculadas a partir dos artefatos da pesquisa, sem estimativas.

4 RESULTADOS

4.1 FLUXO DE SELEÇÃO

Após a desduplicação, foram identificados 758 registros. A primeira etapa de triagem reteve 261 estudos com indicadores de inteligência artificial conversacional em saúde mental. A segunda confirmou 169 publicações com contribuição conceitual. Desse conjunto, 51 modelos compuseram a amostra central de extração aprofundada. A extração baseou-se em texto completo em 32 casos, em resumo em 17 e apenas no título em dois.

4.2 COMPOSIÇÃO TIPOLÓGICA DO CAMPO

A distribuição dos tipos de estrutura no conjunto incluído (n = 169) e na amostra central (n = 51) é apresentada na Tabela 1.

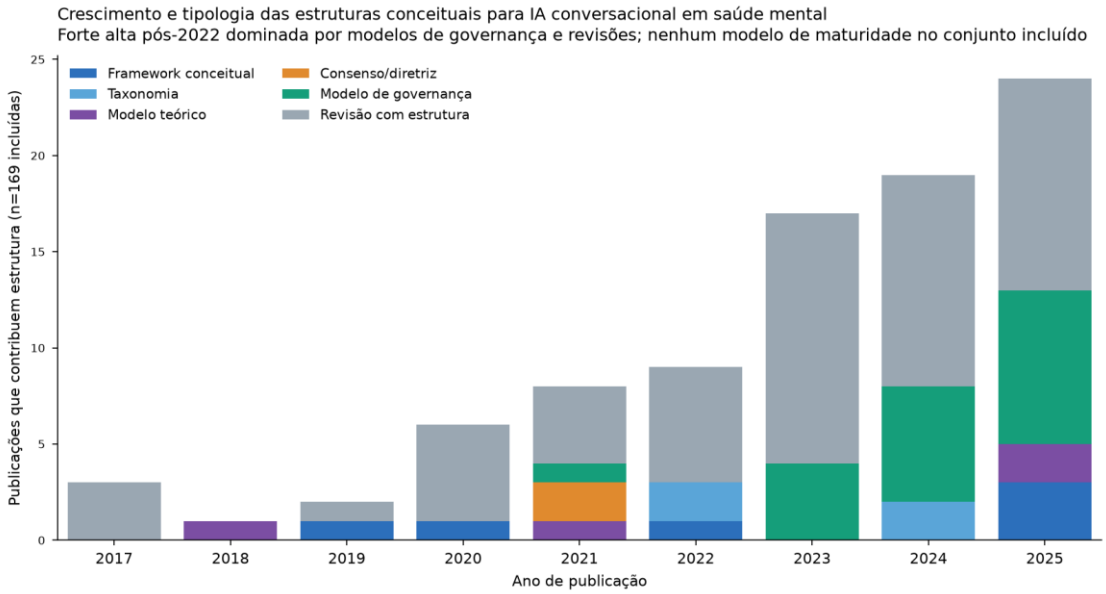
Tabela 1 – Distribuição dos tipos de estrutura conceitual.

Tipo de estrutura	Conjunto incluído (n=169)	Conjunto central (n=51)
Revisão com estrutura	99	10
Modelo de governança	36	16
Framework conceitual	11	11
Taxonomia	7	9
Modelo teórico	8	4
Consenso/diretriz	8	1
Modelo de maturidade	0	0
Total	169	51

Fonte: elaboração própria com base nos dados da pesquisa (2026).

Os resultados evidenciam duas tendências. A primeira é a predominância de revisões estruturadas e modelos de governança, que, em conjunto, representam a maior parcela das publicações incluídas. A segunda é a ausência de modelos de maturidade em todo o corpus de 169 estudos. Em contraste, frameworks conceituais, taxonomias e modelos teóricos — estruturas com maior potencial de organização ontológica — permanecem comparativamente menos frequentes.

Figura 1 – Crescimento e tipologia das estruturas conceituais ao longo do tempo



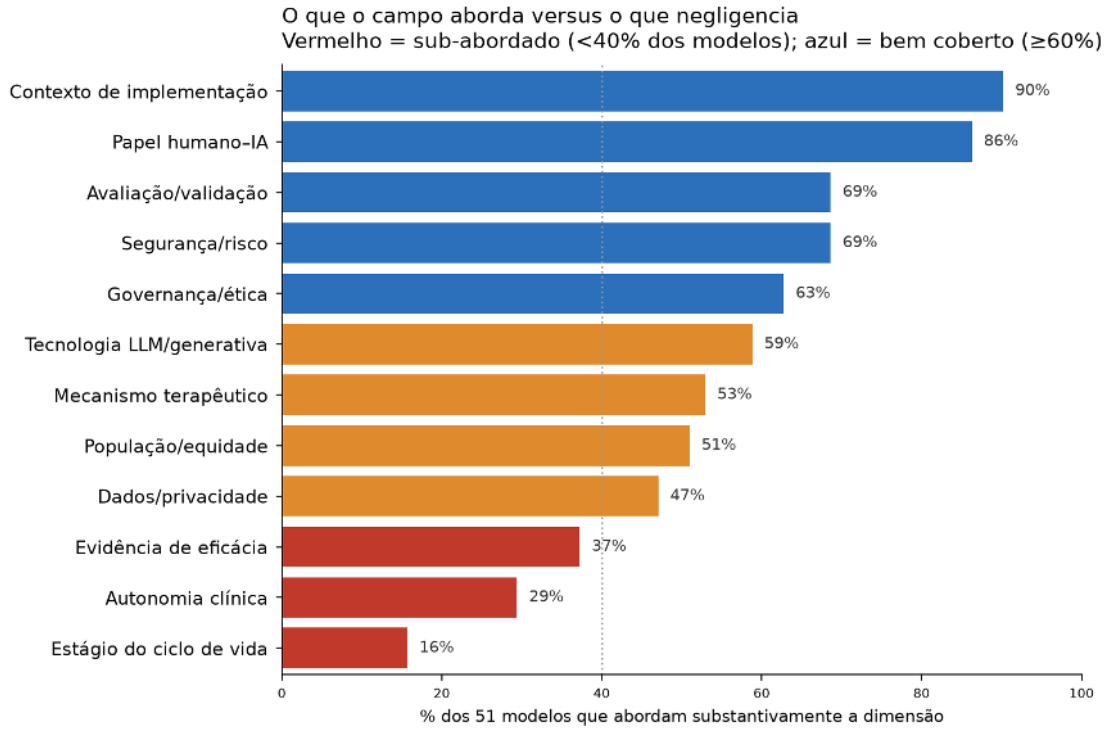
Fonte: elaboração própria com base nos dados da pesquisa (2026).

Conforme ilustrado na Figura 1, a produção de estruturas conceituais permaneceu limitada até 2020 e apresentou crescimento acentuado a partir de 2022, período coincidente com a popularização dos LLMs generativos. Esse crescimento, entretanto, foi impulsionado principalmente por revisões e modelos de governança, e não pelo desenvolvimento de frameworks integrativos ou modelos teóricos. A ausência de modelos de maturidade indica que o domínio ainda carece de instrumentos capazes de representar estágios de desenvolvimento de sistemas e organizações implementadoras. Depreende-se, portanto, que a produção normativa e de síntese avançou mais rapidamente do que a elaboração de arquiteturas conceituais.

4.3 COBERTURA DIMENSIONAL: O QUE O CAMPO ABORDA E O QUE NEGLIGENCIA

A proporção de modelos centrais que contemplam cada uma das 12 dimensões analíticas é apresentada na Figura 2.

Figura 2 – Cobertura dimensional — o que o campo aborda versus negligencia



Fonte: elaboração própria com base nos dados da pesquisa (2026).

A Figura 2 permite distinguir três níveis de cobertura. Contexto de implementação (90%) e papel humano-IA (86%) apresentaram elevada presença, indicando ampla atenção ao local de inserção dos sistemas e à distribuição de funções entre profissionais e tecnologias. Em nível intermediário, mecanismo terapêutico e dados ou privacidade foram abordados por aproximadamente metade dos modelos. Em contrapartida, evidência de eficácia (37%), autonomia clínica (29%) e estágio do ciclo de vida (16%) constituíram as dimensões menos contempladas. Esse padrão sugere que os referenciais existentes abordam com maior frequência aspectos descritivos e organizacionais do que relações que exigem integração clínica, técnica e temporal.

Tabela 2 – Cobertura das doze dimensões analíticas no conjunto central (n=51).

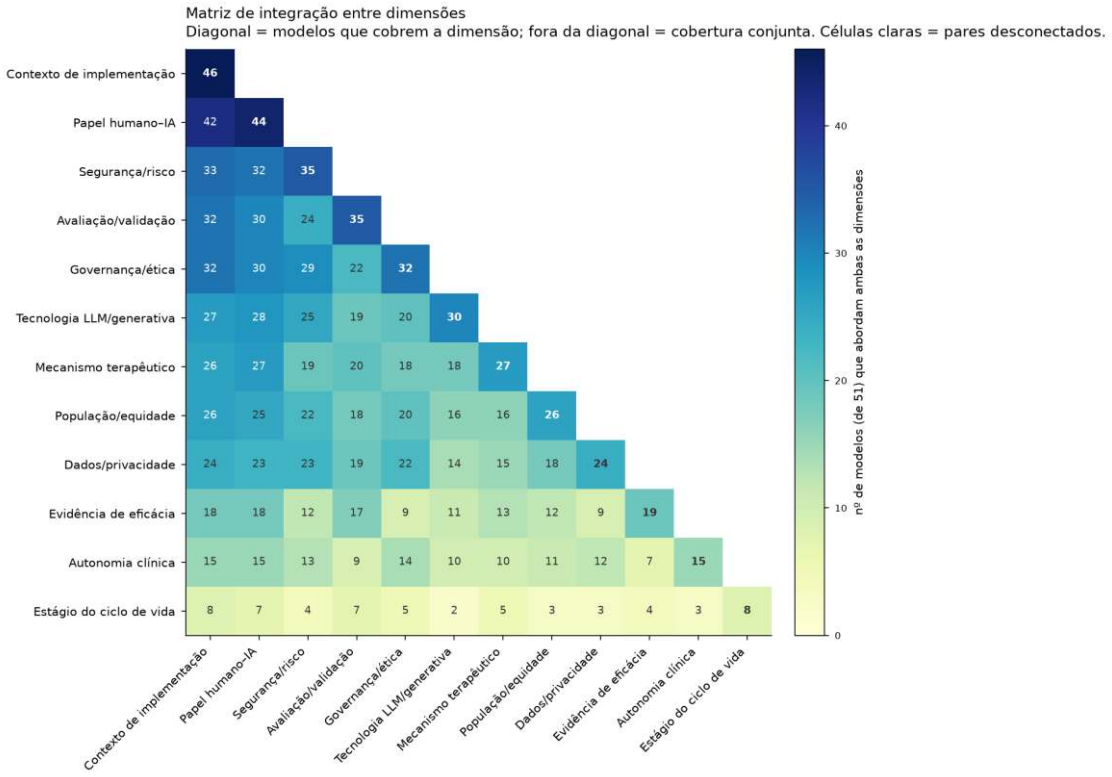
Dimensão	% de modelos
Contexto de implementação	90%
Papel humano–IA	86%
Segurança/risco	69%
Avaliação/validação	69%
Governança/ética	63%
Tecnologia LLM/generativa	59%
Mecanismo terapêutico	53%
População/equidade	51%
Dados/privacidade	47%
Evidência de eficácia	37%
Autonomia clínica	29%
Estágio do ciclo de vida	16%

Fonte: elaboração própria com base nos dados da pesquisa (2026).

4.4 INTEGRAÇÃO ENTRE DIMENSÕES

A cobertura isolada de dimensões não implica integração entre elas. Por essa razão, a Figura 3 apresenta a matriz de coocorrência, que expressa o número de modelos que abordam conjuntamente cada par de dimensões.

Figura 3 – Matriz de integração entre dimensões



Fonte: elaboração própria com base nos dados da pesquisa (2026).

Os dados da Figura 3 evidenciam baixa integração nos pares conceitualmente mais relevantes. O estágio do ciclo de vida foi associado à tecnologia generativa em apenas dois modelos e à segurança em quatro. Autonomia clínica e ciclo de vida coocorreram em três modelos. A relação entre mecanismo terapêutico e evidência de eficácia, necessária para articular plausibilidade teórica e demonstração empírica, esteve presente em 13 modelos; já a relação entre autonomia clínica e evidência foi identificada em sete. Em média, cada modelo contemplou 6,7 das 12 dimensões, enquanto nove cobriram quatro ou menos. Assim, amplitude dimensional e integração testável não se concentraram em uma única estrutura.

4.5 ESTATUTO EVIDENCIAL DAS PRÓPRIAS ESTRUTURAS

Entre os 51 modelos centrais, apenas dois apresentaram validação empírica; 23 foram classificados como puramente conceituais e 18 como sínteses de evidências. Esse perfil indica que o

domínio se organiza majoritariamente por meio de estruturas que ainda não foram submetidas a validação própria. Tal constatação não constitui uma avaliação negativa de estudos isolados, mas evidencia a inexistência de um arcabouço empiricamente validado que possa ser adotado como referência integradora.

4.6 LACUNAS AUTOIDENTIFICADAS PELOS AUTORES

Além da codificação realizada neste estudo, os 51 artigos centrais registraram 141 lacunas não resolvidas. O agrupamento temático apresentado na Tabela 3 indica convergência entre os autores quanto à insuficiência de integração entre efetividade clínica, regulação, segurança, aliança terapêutica, design e fundamentação mecanística.

Tabela 3 – Meta-lacunas recorrentes autoidentificadas pelos autores do conjunto central.

Meta-lacuna (autoidentificada)	**Nº de artigos**
Efetividade clínica e medição de desfechos	13
Estruturas regulatórias e jurídicas	11
Segurança, gestão de risco e diretrizes éticas	10
Aliança terapêutica e dinâmica relacional humano–IA	8
Orientação de design, desenvolvimento e implementação	8
Qualidade da evidência, método e padronização de métricas	8
Transparência, responsabilização e credibilidade dos dados	7
Compreensão de partes interessadas e preferências dos usuários	6
Compreensão mecanística e fundamentação teórica	6

Fonte: elaboração própria com base nos dados da pesquisa (2026).

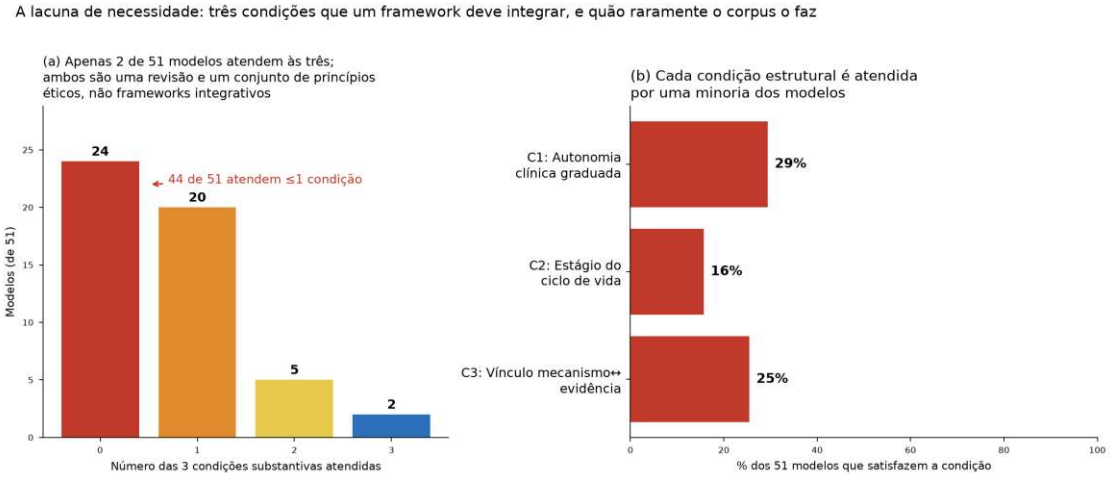
5 DISCUSSÃO

Os resultados sustentam uma proposição delimitada e passível de refutação: a literatura disponível oferece recursos para descrever, regular e revisar a inteligência artificial conversacional em saúde mental, mas permanece insuficiente para organizá-la em uma arquitetura integrada. A discussão a seguir examina por que essa insuficiência não pode ser superada apenas pela extensão de estruturas existentes.

5.1 TRÊS CONDIÇÕES ESTRUTURAIS QUE NENHUM MODELO SATISFAZ CONJUNTAMENTE

A necessidade de integração foi examinada a partir de três condições estruturais: autonomia clínica graduada, localização no ciclo de vida e vínculo entre mecanismo terapêutico e evidência. A distribuição dessas condições entre os modelos é apresentada na Figura 4.

Figura 4 – A lacuna de necessidade: três condições que um framework deve integrar



Fonte: elaboração própria com base nos dados da pesquisa (2026).

Conforme a Figura 4, 24 dos 51 modelos não atenderam a nenhuma das três condições e 44 contemplaram, no máximo, uma. Apenas dois modelos atenderam simultaneamente às três condições; contudo, ambos corresponderam a uma revisão de evidências e a um conjunto de princípios éticos, e não a frameworks conceituais integrativos. Consideradas isoladamente, as condições também foram minoritárias: autonomia clínica graduada ocorreu em 29% dos modelos, ciclo de vida em 16% e vínculo mecanismo-evidência em 26%. Esses resultados indicam que a lacuna não decorre da ausência de um componente pontual, mas da inexistência de uma estrutura que integre os três elementos.

5.2 POR QUE EXTENSÃO NÃO BASTA

A extensão de um framework existente revela-se insuficiente por três razões. Inicialmente, autonomia clínica e ciclo de vida são as dimensões menos cobertas e coocorrem em apenas três modelos, o que impede sua simples incorporação a uma arquitetura já consolidada. Além disso, os gêneros predominantes, revisões e modelos de governança, desempenham funções distintas da organização ontológica: regulam ou sintetizam um objeto previamente definido, mas não especificam o tipo de sistema, o nível de autonomia, o estágio de desenvolvimento e o mecanismo terapêutico. Por fim, as lacunas autoidentificadas convergem para problemas interdependentes de efetividade, regulação, segurança, aliança terapêutica e fundamentação mecanística. Essa convergência sugere a ausência de uma estrutura relacional, e não apenas de elementos isolados.

5.3 DIÁLOGO COM AS ESTRUTURAS DE MAIOR ALCANCE

As estruturas de maior alcance reforçam, por contraste, a natureza da lacuna. A tipologia de quatro formas de integração da inteligência artificial conversacional à psicoterapia proposta por Miner

et al. (2019) diferencia níveis de entrega humano-IA, mas não estabelece uma gradação de autonomia associada ao risco. Lawrence et al. (2024) sistematizam um amplo conjunto de oportunidades e riscos, porém sem convertê-los em uma lógica organizadora testável. De modo semelhante, propostas conceituais recentes avançam em abrangência, mas permanecem sem validação empírica (Bing; An-Yue, s.d.). Portanto, a amplitude temática observada em alguns trabalhos não se traduz em integração operacional.

5.4 IMPLICAÇÕES

Do ponto de vista científico, os achados deslocam a agenda de simplesmente propor um novo framework para especificar uma arquitetura cuja necessidade decorre das evidências. Na prática clínica e regulatória, os resultados demonstram que instrumentos normativos exigem um referente ontológico estável: sistemas de baixa autonomia em fase piloto e sistemas de alta autonomia em operação contínua apresentam perfis de risco distintos e não devem ser submetidos a princípios indiferenciados. No âmbito educacional, a taxonomia proposta pelas dimensões analisadas pode subsidiar a formação de profissionais capazes de avaliar ferramentas segundo autonomia, mecanismo, evidência e maturidade.

6 LIMITAÇÕES

Este estudo apresenta limitações relacionadas à cobertura das fontes, à heterogeneidade da base textual e à operacionalização analítica. A busca restringiu-se a OpenAlex e PubMed, uma vez que a recuperação no arXiv não foi concluída por indisponibilidade técnica. Assim, trabalhos em pré-publicação na ciência da computação podem estar sub-representados. Além disso, a extração baseou-se em texto completo para 32 dos 51 modelos centrais, em resumo para 17 e apenas em título para dois; consequentemente, dimensões presentes no corpo dos artigos podem ter sido subestimadas. A codificação das 12 dimensões e a classificação tipológica envolveram julgamento analítico, embora tenham sido orientadas por esquema fixo e pela regra de não inferir atributos ausentes. O recorte temporal de 2015 a 2025 e a predominância de publicações em inglês também podem limitar a generalização. Tais restrições recomendam cautela na interpretação das frequências, mas não anulam o padrão central de baixa cobertura de autonomia clínica, ciclo de vida e integração.

7 CONSIDERAÇÕES FINAIS

O mapeamento sistemático demonstrou que a literatura sobre inteligência artificial conversacional em saúde mental dispõe de instrumentos numerosos para descrição, regulação e síntese, mas ainda ainda carece de uma arquitetura conceitual integradora. Entre 169 estruturas identificadas e

51 modelos submetidos à extração aprofundada, autonomia clínica e ciclo de vida foram as dimensões menos contempladas e raramente integradas. Também não foram identificados modelos de maturidade, e apenas duas estruturas abordaram simultaneamente as três condições centrais, sem se configurarem como frameworks integrativos.

A necessidade de um novo framework resulta, portanto, de uma especificação empírica. Essa estrutura deverá posicionar cada sistema em um eixo graduado de autonomia clínica, situá-lo em um ciclo de vida com perfil de risco variável e relacionar o mecanismo terapêutico à evidência disponível e ao nível de autonomia. Além disso, deverá operar como uma arquitetura integrada e testável, e não como uma enumeração de princípios. A construção e a validação dessa proposta constituem a etapa subsequente desta linha de pesquisa.

FINANCIAMENTO

Este trabalho foi desenvolvido no âmbito do projeto Consolidação da Agência de Inovação da Universidade Federal do Paraná (UFPR) no Ecossistema de Inovação do Paraná, financiado pela Fundação Araucária, por meio do Edital CP 24/2024 – Programa Institucional Bolsa-Técnico Agente Regional de Inovação.

REFERÊNCIAS

1. ABBASIAN, M. et al. Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. npj Digital Medicine, 2024. DOI: 10.1038/s41746-024-01074-z.

2. ABD-ALRAZAQ, A. et al. Perceptions and Opinions of Patients About Mental Health Chatbots: Scoping Review. Journal of Medical Internet Research, 2020. DOI: 10.2196/17828.

3. ABD-ALRAZAQ, A. et al. Technical Metrics Used to Evaluate Health Care Chatbots: Scoping Review. Journal of Medical Internet Research, 2020. DOI: 10.2196/18301.

4. AHMAD, R. et al. Designing Personality-Adaptive Conversational Agents for Mental Health Care. Information Systems Frontiers, 2022. DOI: 10.1007/s10796-022-10254-9.

5. ASHRAFUL, I.; MOALLA, C.B.; AMINUL, I. RACares: a conceptual design to guide mHealth relational agent development based on a systematic review. mHealth, s.d. DOI: 10.21037/mhealth-23-46.

6. BING, C.; AN-YUE, J.. Generative AI as interactional infrastructure for meaning-centered care in later life. Frontiers in Psychiatry, s.d. DOI: 10.3389/fpsyt.2026.1883962.

7. BLUT, M. et al. Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI. Journal of the Academy of Marketing Science, 2021. DOI: 10.1007/s11747-020-00762-y.

8. CHEW, H.S.J.; ACHANANUPARP, P. Perceptions and Needs of Artificial Intelligence in Health Care to Increase Adoption: Scoping Review. Journal of Medical Internet Research, 2022. DOI: 10.2196/32939.

9. COGHLAN, S. et al. To chat or bot to chat: Ethical issues with using chatbots in mental health. Digital Health, 2023. DOI: 10.1177/20552076231183542.

10. CRUZ-GONZALEZ, P. et al. Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications. Psychological Medicine, 2025. DOI: 10.1017/s0033291724003295.

11. DEKKER, I. et al. Optimizing Students' Mental Health and Academic Performance: AI-Enhanced Life Crafting. Frontiers in Psychology, 2020. DOI: 10.3389/fpsyg.2020.01063.

12. ERMOLINA, A.; TIBERIUS, V.. Voice-Controlled Intelligent Personal Assistants in Health Care: International Delphi Study. Journal of Medical Internet Research, 2021. DOI: 10.2196/25312.

13. GUO, Z. et al. Large Language Models for Mental Health Applications: Systematic Review. JMIR Mental Health, 2024. DOI: 10.2196/57400.

14. HOERMANN, S. et al. Application of Synchronous Text-Based Dialogue Systems in Mental Health Interventions: Systematic Review. Journal of Medical Internet Research, 2017. DOI: 10.2196/jmir.7023.

15. LAWRENCE, H.R. et al. The Opportunities and Risks of Large Language Models in Mental Health. JMIR Mental Health, 2024. DOI: 10.2196/59479.

16. LAYMOUNA, M. et al. Roles, Users, Benefits, and Limitations of Chatbots in Health Care: Rapid Review. Journal of Medical Internet Research, 2024. DOI: 10.2196/56930.

17. LI, H. et al. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. npj Digital Medicine, 2023. DOI: 10.1038/s41746-023-00979-5.

18. LI, H.; ZHANG, R.. Finding love in algorithms: deciphering the emotional contexts of close encounters with AI chatbots. Journal of Computer-Mediated Communication, 2024. DOI: 10.1093/jcmc/zmae015.

19. MALGAROLI, M. et al. Natural language processing for mental health interventions: a systematic review and research framework. Translational Psychiatry, 2023. DOI: 10.1038/s41398-023-02592-2.

20. MINER, A.S. et al. Key Considerations for Incorporating Conversational AI in Psychotherapy. Frontiers in Psychiatry, 2019. DOI: 10.3389/fpsyt.2019.00746.

21. MORRIS, R.R. et al. Towards an Artificially Empathic Conversational Agent for Mental Health Applications: System Design and User Perceptions. Journal of Medical Internet Research, 2018. DOI: 10.2196/10148.

22. PARMAR, P. et al. Health-focused conversational agents in person-centered care: a review of apps. npj Digital Medicine, 2022. DOI: 10.1038/s41746-022-00560-6.

23. PROVOOST, S. et al. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. Journal of Medical Internet Research, 2017. DOI: 10.2196/jmir.6553.

24. REDDY, S. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. Implementation Science, 2024. DOI: 10.1186/s13012-024-01357-9.

25. ROUSTAN, D.; BASTARDOT, F.. The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations. Interactive Journal of Medical Research, 2025. DOI: 10.2196/59823.

26. SAEIDNIA, H.R. et al. Ethical Considerations in Artificial Intelligence Interventions for Mental Health and Well-Being: Ensuring Responsible Implementation and Impact. Social Sciences, 2024. DOI: 10.3390/socsci13070381.

27. SCHYFF, E.L.V.D. et al. Providing Self-Led Mental Health Support Through an Artificial Intelligence–Powered Chat Bot (Leora) to Meet the Demand of Mental Health Care. Journal of Medical Internet Research, 2023. DOI: 10.2196/46448.

28. STADE, E.C. et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. npj Mental Health Research, 2024. DOI: 10.1038/s44184-024-00056-z.

29. VILAZA, G.N.; MCCASHIN, D.. Is the Automation of Digital Mental Health Ethical? Applying an Ethical Framework to Chatbots for Cognitive Behaviour Therapy. Frontiers in Digital Health, 2021. DOI: 10.3389/fdgth.2021.689736.

30. WANG, X.; ZHOU, Y.; ZHOU, G. The Application and Ethical Implication of Generative AI in Mental Health: Systematic Review. JMIR Mental Health, 2025. DOI: 10.2196/70610.