UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE PSICOLOGIA, SERVIÇO SOCIAL, SAÚDE E COMUNICAÇÃO HUMANA

GRADUAÇÃO EM PSICOLOGIA

Leandro Prade Nadaletti

A INTERSEÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E PSICOTERAPIA:
Potenciais, Perigos e o Imperativo Ético

Porto Alegre

2025

Leandro Prade Nadaletti

A INTERSEÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E PSICOTERAPIA:
Potenciais, Perigos e o Imperativo Ético

Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Bacharel em Psicologia pelo Instituto de Psicologia, Serviço Social, Saúde e Comunicação Humana da Universidade Federal do Rio Grande do Sul.

Orientadora: Prof.ª Dr.ª Marta Regina de Leão D'Agord

Porto Alegre

2025

Nadaletti, Leandro Prade
A INTERSEÇÃO ENTRE INTELIGÊNCIA ARTIFICIAL E PSICOTERAPIA: Potenciais, Perigos e o Imperativo Ético
/ Leandro Prade Nadaletti. -- 2025.
51 f.
Orientadora: Marta Regina de Leão D'Agord.
Trabalho de conclusão de curso (Graduação) -- Universidade Federal do Rio Grande do Sul, Instituto de Psicologia, Serviço Social, Saúde e Comunicação Humana, Bacharelado em Psicologia, Porto Alegre, BR-RS, 2025.
1. Inteligência Artificial. 2. Ética. 3. Supervisão Clínica. 4. Privacidade de Dados. 5. Psicoterapia. I. de Leão D'Agord, Marta Regina, orient. II. Título.

RESUMO

Este estudo analisa a interseção entre inteligência artificial (IA) e psicoterapia, destacando possibilidades, riscos e a necessidade ética de sua implementação responsável. A discussão distingue claramente eficácia clínica (impacto nos sintomas e no funcionamento) de eficiência operacional (escala, custo e tempo de espera). Entre os benefícios estão maior acesso por meio de triagens automatizadas e suporte síncrono ou assíncrono, diminuição de barreiras de custo, distância e estigma, além de ganhos de eficiência que liberam o tempo do profissional para atividades de maior impacto terapêutico. Os riscos abrangem vieses nos modelos de IA que podem produzir e ampliar desigualdades clínicas, alucinações que distorcem hipóteses diagnósticas, dependência excessiva de agentes conversacionais, erosão da aliança terapêutica por automação opaca e vulnerabilidades de privacidade. Em saúde mental, isso significa triagens menos sensíveis para populações sub-representadas, falsos positivos/negativos desbalanceados, e reforço de estereótipos via dados de treinamento e contexto de uso. A revisão da literatura revela limitações metodológicas frequentes nas pesquisas, efeitos de curto prazo, amostras heterogêneas e seguimentos, que recomendam cautela na generalização dos resultados e na promessa de benefícios terapêuticos. Como resposta, o estudo propõe salvaguardas e orientações práticas: supervisão clínica obrigatória em uso assistido; transparência sobre capacidades, dados de treino e limites dos modelos; e avaliação contínua de precisão, equidade e impacto clínico. Na mesma linha, o CFP defende o envolvimento da categoria e da sociedade, a atuação em políticas públicas e marcos legais, desenvolvimento do diálogo interinstitucional/transdisciplinar; comprometendo-se a monitorar o tema, manter diálogo contínuo e atuar com os CRPs frente aos desafios da IA. Dessa forma, conclui-se que a adoção ética da IA na psicoterapia depende de um desenho centrado na pessoa, controle humano e infraestrutura de segurança ao longo de todo o ciclo de vida do sistema. Assim, a IA deve limitar-se a ampliar o acesso à saúde mental, melhorar a qualidade do cuidado e auxiliar terapeutas sem substituir a relação humana que sustenta a mudança terapêutica.

Palavras-chave: Inteligência Artificial; Psicoterapia; Ética; Supervisão Clínica; Privacidade de Dados.

SUMÁRIO

RESUMO 1
1. INTRODUÇÃO 3
2. HISTÓRICO DA INTERVENÇÃO DIGITAL 4
3. POTENCIAIS DA IA NA PSICOTERAPIA 8
3.1 Acessibilidade e conveniência aprimoradas 9
3.2 Eficácia e melhoria de resultados clínicos 10
3.3 Eficiência e otimização de recursos 13
3.4 Experiência do usuário e redução do estigma 15
3.5 Apoio aos profissionais e avanço da pesquisa 17
4. PERIGOS E DESAFIOS DO USO DE IA NA PSICOTERAPIA 19
4.1 Questões éticas, privacidade e segurança dos dados 20
4.2 Limitações clínicas e falta de humanização 23
4.3 Desafios metodológicos e de pesquisa 26
4.4 Desafios sociais e de adoção 28
5. IMPERATIVO ÉTICO E DISCUSSÃO 33
5.1 IA como complemento, nunca substituto 33
5.2 Implicações para a prática clínica e para políticas públicas 36
6. CONCLUSÃO 39
7. REFERÊNCIAS 42

1. INTRODUÇÃO

A vida moderna não está se tornando apenas digital; ela já está completamente imersa nela. O que presenciamos é uma evolução contínua e acelerada, que começou com os computadores pessoais e a internet ainda em seus estágios iniciais e ganhou velocidade exponencial com a chegada dos smartphones, 5G e Internet das Coisas. Essa trajetória trouxe uma transformação profunda na sociedade, agora acentuada pelo avanço rápido da Inteligência Artificial (IA) (Brynjolfsson et al., 2025). A barreira entre o físico e o digital, quase desaparecida, está sendo eliminada pela IA, que transforma o digital de simples armazenamento de dados em um colaborador ativo. Inicialmente, digitalizar envolvia lidar com e-mails, documentos e registros; agora, a camada algorítmica pode atuar diretamente. Assim, as rotinas deixam de ser apenas facilitadas para serem amplamente orquestradas por sistemas que aprendem continuamente (Kellogg et al., 2020). Essa mudança oferece experiências mais convenientes e com menos atrito, nas quais plataformas antecipam ações, serviços ajustam ambientes antes do pedido e agendas se reorganizam em tempo real (Raza et al., 2025). No entanto, essa hiperpersonalização pode também estreitar horizontes e reforçar bolhas, reduzindo o contato com o inesperado e com perspectivas divergentes (Bakshy et al., 2015).

Nesse contexto, as relações humanas também passaram por mudanças. A proximidade ganha eficiência, mas perde parte do acaso; a disponibilidade constante encurta distâncias, mas pode empobrecer a presença. Conversas mediadas por telas ampliam o alcance e a inclusão, ao mesmo tempo em que deslocam nuances para métricas de engajamento. Vínculos se constroem e se mantêm em plataformas cujo desenho influencia quem vemos, o que ouvimos e como reagimos (Primack et al., 2017). Certamente, também há ganhos, mas cresce o desafio de preservar o tempo do encontro, o espaço do desacordo respeitoso e a confiança que se deposita longe dos algoritmos (Lambert et al., 2022). Em resumo, estamos mais conectados do que nunca, mas a fragilidade dessas ligações faz com que nos sintamos cada vez mais isolados e solitários.

Ao mesmo tempo, a demanda por cuidados psicológicos aumenta em um ambiente de hiperconectividade, no qual a comparação social e a responsividade contínua desafiam a atenção, o sono e o humor (Stothart et al., 2015; Wang et al.,

2017). As telas aproximam as pessoas, mas também prolongam jornadas psíquicas: notificações dispersam o foco, métricas transformam reconhecimento em números e curadorias “perfeitas” minam a autoestima (Vogel et al., 2014). No trabalho, o conceito de "sempre online” elimina fronteiras entre turnos e momentos de descanso; em casa, papéis sobrepostos intensificam a sensação de nunca desligar (Barber & Santuzzi, 2015). O sono deteriora-se; o corpo manifesta dores, exaustão e irritabilidade, enquanto nas redes, a lógica de reforços intermitentes captura tempo e dopamina, acentuando inseguranças e medo de obsolescência (Lambert et al., 2022). Resumidamente, o problema não está na "tecnologia” em si, mas na estrutura sociotécnica que controla a atenção, as recompensas e o trabalho.

2. HISTÓRICO DA INTERVENÇÃO DIGITAL

Neste novo mundo, a saúde mental está inserida na era dos algoritmos. Se sua introdução ao tema foi por meio de programas como o ELIZA, que simulava um terapeuta rogeriano e já evidenciava o fascínio e as limitações da interação homem–máquina (Weizenbaum, 1966), atualmente, com o avanço da inteligência artificial, dispomos de ferramentas capazes de observar, prever e desenvolver programas de tratamento completos. Nesse percurso, a clínica digital evoluiu de protótipos para protocolos que podem ser facilmente replicados. Nos anos 2000, a TCC digital provou ser eficaz na atenção primária, aprimorando sintomas e funcionamento, representando um avanço concreto na direção de serviços escaláveis (Proudfoot et al., 2004). Na década seguinte, a entrega pela internet se consolidou: a iCBT guiada por terapeuta tende a produzir efeitos similares aos da TCC presencial, embora desafios como adesão, monitoramento de efeitos adversos e implementação em larga escala ainda persistam (Andersson et al., 2019).

De 2020 a 2025, os estudos sobre IA na saúde mental amadureceram e passaram a se estruturar em três áreas principais que se reforçam mutuamente: *chatbots* para avaliação inicial, psicoeducação e intervenções breves; monitoramento por celular e sensores para acompanhar a rotina e prever recaídas; e suporte à decisão clínica por meio de modelos que unem tecnologia e atendimento presencial. As últimas revisões apontam uma rápida expansão de soluções conversacionais,

impulsionada por técnicas de compreensão de linguagem e modelos que aprendem com dados para personalizar e ampliar o alcance. Paralelamente, destacam limitações persistentes, como grande variação entre estudos, pouca padronização e acompanhamentos curtos, o que ainda impede conclusões definitivas sobre a durabilidade dos efeitos (Cruz-Gonzalez et al., 2025; Schafer et al., 2024).

Nos ensaios com usuários, os benefícios geralmente aparecem no curto prazo, mas a força causal e a generalização variam conforme o desenho do estudo. Investigações pré–pós com o Woebot, ao longo de oito semanas, indicam redução de estresse e burnout, aumento de resiliência e sinais de bom vínculo e engajamento; no entanto, a ausência de comparações equivalentes e a dependência de medidas autorrelatadas limitam conclusões e podem introduzir vieses como regressão à média e expectativas (Durden et al., 2023). Resultados com o ChatPal, em estudo multicêntrico pré–pós, mostram melhorias no bem-estar em diferentes contextos linguísticos, mas a falta de randomização enfraquece a causalidade (Potts et al., 2023). No campo da triagem, o DEPRA apresenta alta satisfação e coerência nos escores autorrelatados em amostra não clínica, apoiando sua viabilidade, embora ainda sem validação diagnóstica comparada ao padrão-ouro clínico (Kaywan et al., 2023). Da mesma forma, o ELME, em um ensaio clínico randomizado de três semanas (N≈118), aponta redução de estresse em comparação aos cuidados habituais, com protocolo publicado; no entanto, continuam pendentes comparações equivalentes (como terapia presencial baseada em TCC), aumento de amostras, seguimento de seis a doze meses e análises que abordem multiplicidade, perdas e adesão de forma transparente (Schillings et al., 2023). Paralelamente, diretrizes institucionais para a elaboração de conversas fornecem parâmetros práticos visando garantir a segurança e a qualidade nas interações, abrangendo triagem de risco, respostas para crises, transparência e encaminhamentos, que servem como padrão mínimo para o desenvolvimento e avaliação (Nieminen et al., 2025).

As sínteses quantitativas apontam para um quadro promissor, embora centrado na pós-intervenção imediato e, em geral, em populações sem diagnóstico formal. Uma metassíntese focada em resiliência identificou 101 ensaios randomizados (N≈20.010) com efeitos positivos logo após a intervenção, apesar da heterogeneidade moderada e variações no tempo de seguimento (Schafer et al., 2024). Revisões de escopo e meta-análises sobre *chatbots*, incluindo estudos com jovens, mostram efeitos de

pequena a moderada magnitude sobre sintomas depressivos e ansiosos, mas ressaltam a curta duração das pesquisas e a diversidade de comparadores, dificultando a avaliação da estabilidade dos resultados e sua aplicação em amostras clínicas (Casu et al., 2024; Y. Li et al., 2025).

A comparação direta entre psicoterapia presencial tradicional e intervenções digitais ainda é rara e representa o obstáculo principal para orientar políticas e práticas clínicas. A análise de 85 estudos atualizados até fevereiro de 2024 revela que predominam os *chatbots* como método de intervenção, embora ainda haja uma escassez de ensaios com comparações equivalentes, amostras clínicas e acompanhamento de médio a longo prazo que possam demonstrar de forma convincente a sua equivalência ou superioridade (Cruz-Gonzalez et al., 2025). Nesse contexto, o papel das ferramentas digitais e *chatbots* é visto como complementar, e não substituto, aos cuidados psicoterapêuticos, especialmente quando utilizados em processos de triagem, educação em saúde e suporte entre sessões (Schillings et al., 2024).

A interpretação dos resultados também requer uma atenção cuidadosa ao engajamento e à viabilidade. As taxas de adesão variam: intervenções curtas geralmente apresentam bom início e conclusão; entretanto, fora do contexto de pesquisa, surgem desafios no recrutamento e na retenção. Uma meta-análise sobre retenção em ensaios com *chatbots* não revelou vantagem do grupo *chatbot* em relação aos controles, destacando a importância de métricas padronizadas, como tempo de uso, completude e recorrência bem como de análises que relacionem a dose digital ao efeito clínico (Cevasco et al., 2024). No ambiente de serviços reais, relatos de jovens que encerraram estudos enquanto estavam em lista de espera evidenciam obstáculos logísticos e organizacionais que dificultam a implementação e a continuidade do cuidado (Fujita et al., 2025).

Na área de coleta de dados por celular e sensores, a revisão de Bufano e colaboradores evidencia que é viável usar informações passivas e ativas (como sono, passos e respostas curtas) para monitorar e tentar prever recaídas, embora haja uma urgência na validação clínica e na padronização metodológica (Bufano et al., 2023). Alinhando-se com a integração de diferentes frentes, ensaios clínicos não relacionados demonstram que é tecnicamente possível combinar *chatbots* a esses dados para oferecer recomendações personalizadas, destacando ganhos potenciais

e a necessidade de ampliar as populações clínicas e integrar *wearables* de forma segura e transparente (Rocchi et al., 2024). Ao longo de todo esse processo, ética e privacidade são essenciais no desenho e na implementação: consentimento informado, governança de dados, uso responsável das previsões e rotas de encaminhamento.

No que diz respeito à segurança e interpretabilidade, recomendações abrangentes destacam três pilares fundamentais para conversas seguras com chatbots: esclarecer o funcionamento do sistema e seus limites, adaptar a solução ao público e ao contexto de uso, além de garantir um ambiente seguro com transparência algorítmica, supervisão clínica e encaminhamentos claros, especialmente para populações vulneráveis ou em risco. A governança de dados e a qualidade clínica do conteúdo são condições fundamentais para uma confiança e segurança duradouras (Cruz-Gonzalez et al., 2025; Nieminen et al., 2025).

Diante desse cenário, ainda há limitações metodológicas a serem enfrentadas: heterogeneidade nas intervenções e nos desfechos, amostras pequenas, seguimentos curtos, escassez de comparações, além da falta de estudos em casos mais graves e em países de baixa e média renda. A agenda proposta sugere ensaios comparativos entre IA, psicoterapia presencial e cuidados habituais, com margens de não-inferioridade pré-estabelecidas quando apropriado; acompanhamento de pelo menos seis a doze meses; desfechos clínicos padronizados e diferenças clinicamente relevantes claramente explícitas; planos analíticos por intenção de tratar com manejo transparente de perdas e múltiplas análises; medidas objetivas adicionais (uso de serviços, escola/trabalho); relato sistemático de segurança; avaliação econômica; estudos pragmáticos de implementação; e uso responsável de dados de rotina de celulares e sensores (J. Li et al., 2025; Schillings et al., 2023, 2024). Pelo que vimos, há um conjunto crescente de evidências de curto prazo que favorecem o uso de *chatbots* e outras ferramentas digitais na redução do estresse, do sofrimento e de sintomas comuns, especialmente entre jovens e públicos não clínicos. No entanto, a falta de comparações sólidas com psicoterapia presencial, juntamente com a heterogeneidade, dificuldades de engajamento, segurança e privacidade, impede atualmente concluir que essas tecnologias sejam equivalentes ou possam substituir a terapia tradicional. Progresso nesse campo depende de estudos mais robustos, com seguimentos prolongados e avaliações de implementação e custo-efetividade que

esclareçam o papel complementar dessas ferramentas na atenção à saúde mental (Cruz-Gonzalez et al., 2025; Schafer et al., 2024; Schillings et al., 2024).

Assim, as ferramentas digitais para saúde mental apresentam uma resposta ambivalente. Por um lado, demonstram potencial para ampliar o acesso às intervenções terapêuticas, fortalecer a rede de cuidados, facilitar triagens, monitoramento de sinais, otimizar registros clínicos e fluxos de trabalho, além de melhorar a adesão aos tratamentos por meio de lembretes e psicoeducação personalizada. Por outro lado, surgem questões éticas e clínicas que não podem ser ignoradas, como segurança e privacidade dos dados, vieses que podem afetar diagnósticos e recomendações, limitação da empatia mediada por máquinas e o impacto na qualidade da relação terapêutica quando a interação ocorre por algoritmos. Portanto, o desafio é incorporar a tecnologia à prática clínica de forma criteriosa, seguindo padrões de evidência e transparência, de modo que o digital seja um meio de cuidado e não uma nova fonte de sofrimento. No conjunto, o cuidado psicológico se torna mais contínuo e responsivo, porém também mais dependente de escolhas de design, governança de dados e limites claros. A IA observa, prediz e gera, mas não substitui a escuta do clínico. O discernimento envolve estabelecer onde a máquina oferece suporte e onde o humano é indispensável: que a tecnologia potencialize o encontro clínico, evitando reduzi-lo a meras métricas (Coghlan et al., 2023).

3. POTENCIAIS DA IA NA PSICOTERAPIA

A incorporação da IA na prática psicológica tem impulsionado avanços importantes em várias áreas, como a expansão do acesso a serviços de saúde mental e a melhoria da eficiência dos recursos terapêuticos (NICE, 2023a; World Health Organization, 2019). Neste capítulo, abordamos os principais benefícios identificados na literatura, apresentando evidências de estudos recentes e exemplos práticos de aplicações. É importante destacar que, apesar de destacar esses benefícios, mantemos uma postura crítica em relação às condições para sua implementação e aos pressupostos envolvidos.

3.1 Acessibilidade e conveniência aprimoradas

Um dos benefícios mais evidentes do emprego de soluções de IA em saúde mental é a eliminação de barreiras tradicionais de acesso. Aplicativos e plataformas inteligentes transcendem limitações geográficas, restrições de horário e até altos custos financeiros, permitindo que um número muito maior de pessoas tenha acesso a algum tipo de apoio psicológico qualificado. Essa democratização do acesso representa um avanço crucial, sobretudo em contextos marcados pela escassez de profissionais ou por desigualdades na distribuição de serviços. A IA pode funcionar, assim, como um amplificador do sistema de cuidados, alcançando populações marginalizadas ou regiões remotas onde terapias presenciais regulares seriam inviáveis (NICE, 2023b; Li et al., 2023).

Um aspecto central dessa ampliação de acesso é a disponibilidade contínua dos sistemas baseados em IA, como *chatbots* terapêuticos e aplicativos de conversa empática. Esses agentes virtuais operam 24 horas por dia, 7 dias por semana, proporcionando suporte emocional imediato sem necessidade de agendamento ou deslocamento físico (Fitzpatrick et al., 2017; He et al., 2023). Com isso, torna-se possível intervir precocemente em momentos de crise, por exemplo, um usuário em sofrimento agudo pode buscar ajuda na madrugada e receber orientação inicial instantânea, em vez de esperar dias por uma consulta. Além disso, a continuidade do acompanhamento é facilitada: indivíduos interessados em monitorar seu humor ou praticar exercícios terapêuticos diariamente podem fazê-lo em horários flexíveis, integrando o autocuidado à rotina sem grandes interrupções. Essa responsividade ininterrupta, adaptada ao ritmo de vida contemporâneo, não encontra paralelo nos modelos tradicionais de clínica, limitados por horários comerciais e disponibilidade do profissional.

Outro fator decisivo para o sucesso dessas tecnologias é a possibilidade de anonimato e privacidade nas interações digitais. Muitas pessoas sentem receio ou vergonha de expor questões emocionais face a face, devido ao estigma social que ainda cerca os transtornos mentais. A mediação tecnológica cria um ambiente percebido como mais seguro para a autorrevelação – atrás de uma tela, o indivíduo pode se sentir menos julgado e mais à vontade para discutir temas íntimos ou dolorosos (Lucas et al., 2014). Estudos indicam, por exemplo, que pacientes tendem

a revelar mais informações sensíveis a avatares virtuais do que a médicos em pessoa, o que sugere um efeito desinibidor positivo propiciado pela interface digital. Paralelamente, muitos aplicativos asseguram confidencialidade por meio de criptografia e permitem uso anônimo, o que reduz barreiras emocionais à busca de ajuda, especialmente entre jovens familiarizados com plataformas online.

Não se pode ignorar, ademais, o baixo custo relativo dessas intervenções. Serviços automatizados geralmente conseguem escalar o atendimento a um custo marginal muito reduzido, tornando-se acessíveis para indivíduos e comunidades com poucos recursos financeiros. *Chatbots* e programas de apoio psicológico digitais podem ser oferecidos gratuitamente ou a preços módicos, ampliando o leque de usuários beneficiados. Essa combinação de atendimento ininterrupto, anonimato e custo reduzido transforma a IA em uma estratégia potente para reduzir desigualdades em saúde mental. Indivíduos que, de outra forma, não teriam acesso a qualquer suporte psicológico passam a receber intervenções básicas, acolhimento e orientação por meio dessas soluções tecnológicas. Ainda que tais inovações não substituam integralmente os atendimentos presenciais, elas operam como portas de entrada efetivas, capazes de ampliar a cobertura dos serviços de saúde mental e gerar impactos positivos na promoção de bem-estar em larga escala (NICE, 2023b; Karyotaki et al., 2021; You et al., 2022).

3.2 Eficácia e melhoria de resultados clínicos

Evidências crescentes sugerem que intervenções mediadas por IA podem contribuir significativamente para a redução do sofrimento psíquico e a melhora de sintomas em diversos quadros clínicos. Estudos clínicos controlados e revisões têm encontrado resultados promissores especialmente em condições como depressão e ansiedade, muitas vezes comparáveis aos obtidos em contextos presenciais tradicionais (Karyotaki et al., 2021). Por exemplo, *chatbots* baseados em Terapia Cognitivo-Comportamental (TCC) guiada apresentam eficácia estatisticamente significativa na redução de sintomas depressivos leves a moderados, quando utilizados como complemento à terapia humana, e em ensaios com Woebot, metanálises e IPD-NMA em iCBT também sustentam esses achados (Andersson et

al., 2019; Durden et al., 2023; Fitzpatrick et al., 2017). Esses efeitos positivos não se restringem a populações diagnosticadas: mesmo indivíduos com sinais subclínicos de sofrimento mental ou aqueles em situação de estresse cotidiano têm relatado benefícios, como melhora na regulação emocional e no enfrentamento de problemas (Beatty et al., 2022). Vale notar também que a população idosa, muitas vezes isolada ou com dificuldade de acesso a terapias, vem se beneficiando de programas de assistência virtual com IA, indicando a versatilidade dessas ferramentas em atender diferentes faixas etárias e perfis de usuários (Li et al., 2023).

Um diferencial importante das intervenções por IA é a possibilidade de alcançar melhorias clínicas em períodos relativamente curtos. Relatos sugerem que *chatbots* terapêuticos conseguem gerar progressos perceptíveis em poucas semanas de uso regular, às vezes acelerando o alívio de sintomas em comparação ao tempo típico de espera por resultados em psicoterapia presencial. Em certos contextos de alta demanda por serviços de saúde mental, essa redução do tempo para melhora torna-se valiosa: significa maior eficiência do sistema, com pacientes obtendo alívio mais rápido e liberando vagas para outros em necessidade. Mesmo com cautela, o fato é que a agilidade das interações automatizadas (que podem ocorrer diariamente ou em maior frequência do que em sessões semanais tradicionais) tende a otimizar o processo terapêutico, condensando intervenções em prazos menores. Assim, em um cenário de filas de espera e sobrecarga dos serviços, a IA surge como aliada para acelerar o ciclo de cuidado, do início da intervenção até a obtenção de resultados (Karyotaki et al., 2021).

Outro fator que contribui diretamente para a eficácia é a elevada capacidade de personalização das intervenções pela IA. Por meio de algoritmos avançados de aprendizado, os sistemas conseguem inferir traços de personalidade, padrões emocionais e perfis comportamentais dos usuários a partir das interações iniciais. Com base nessas análises, o *chatbot* pode ajustar em tempo real suas respostas e estratégias, por exemplo, identificando que certo usuário responde melhor a um estilo mais diretivo e informativo, enquanto outro necessita de abordagens mais acolhedoras e reflexivas. Essa adaptação dinâmica – impossível de ser realizada por protocolos estáticos – tende a melhorar a aderência ao tratamento (já que o paciente se sente compreendido e engajado por um “terapeuta” sob medida) e a satisfação do usuário com o serviço (Beatty et al., 2022). Em última instância, tal grau de customização

intensifica a efetividade clínica: a intervenção “certa” no momento “certo” para a pessoa “certa” é justamente o que a IA promete ao combinar análise de dados e responsividade imediata. Exemplos práticos incluem *chatbots* que modulam o tom de linguagem conforme o humor diário reportado pelo paciente, ou que selecionam exercícios específicos (respiração, reestruturação cognitiva ou técnicas de grounding) conforme os sintomas prevalentes naquela semana. Essas sutilezas no ajuste fino do cuidado são apontadas como uma vantagem comparativa das plataformas inteligentes em relação a intervenções uniformes de manuais impressos ou aplicações únicas para todos.

É importante destacar que os benefícios da IA não se limitam a transtornos mentais majoritários. Ferramentas como Woebot (um *chatbot* de TCC) e Wysa (assistente emocional baseado em texto) têm se mostrado úteis também na promoção de autocuidado e resiliência em populações não clínicas (Beatty et al., 2022; Durden et al., 2023). Usuários relatam ganho de autoconhecimento, aprendizado de habilidades socioemocionais (como técnicas de relaxamento, meditação guiada, *journaling* de emoções) e maior percepção de suporte contínuo no dia a dia. Esse aspecto preventivo e educativo amplia o leque de resultados positivos associados à IA: não apenas tratar sintomas já instalados, mas fortalecer recursos psicológicos dos indivíduos para lidar com adversidades futuras. Ademais, os sistemas inteligentes podem auxiliar na identificação precoce de agravamentos clínicos, enviando alertas quando detectam padrões de fala que sugiram risco, por exemplo, menções frequentes à desesperança ou à ideação suicida. Essa vigilância proativa, se bem calibrada, permite intervenções preventivas antes que uma situação se torne crítica, potencialmente salvando vidas ou evitando hospitalizações (Eichstaedt et al., 2018).

Por fim, um campo em que a IA tem papel transformador é na melhora dos processos diagnósticos. Através da análise de dados multimodais (texto de conversas, tom de voz, expressões faciais em vídeo, sensores de atividade), algoritmos conseguem oferecer hipóteses diagnósticas refinadas e mais personalizadas. Em vez de depender unicamente de relatos verbais em consultas, o profissional pode ser apoiado por sistemas que sinalizam indicadores sutis, por exemplo, variações microemocionais que humanos deixam passar, mas que se correlacionam a transtornos específicos. Além disso, a IA possibilita o monitoramento longitudinal intensivo, avaliando em tempo real o estado de saúde mental do paciente e sugerindo

ajustes no plano terapêutico conforme as flutuações detectadas (Karyotaki et al., 2021). Do ponto de vista populacional, tais sistemas podem mapear tendências gerais de saúde mental, identificando surtos de ansiedade em determinada comunidade ou avaliando o impacto de políticas públicas, fornecendo dados para gestores tomarem decisões informadas. Em resumo, a IA traz consigo não apenas ganhos no nível individual do paciente, mas também no gerenciamento coletivo da saúde mental, potencializando a vigilância epidemiológica e a alocação eficiente de recursos.

3.3 Eficiência e otimização de recursos

Além dos ganhos clínicos, a integração de IA na psicoterapia traz promessas significativas de eficiência operacional e otimização de recursos, tanto do ponto de vista dos sistemas de saúde quanto dos pacientes e profissionais. Em um contexto de demanda crescente por serviços de saúde mental e recursos frequentemente escassos, a IA se apresenta como uma ferramenta para fazer mais com menos (NICE, 2023a, 2023b).

Primeiramente, a capacidade da IA de oferecer intervenções a um custo marginal reduzido gera benefícios econômicos diretos. Automação de atendimentos, uso de *chatbots* e plataformas digitais podem diminuir despesas operacionais (como infraestrutura física e deslocamento de pacientes), aliviando pressões financeiras sobre clínicas, hospitais e planos de saúde. Do ponto de vista dos pacientes pagantes, terapias mediadas por aplicativos tendem a ser mais baratas que sessões presenciais individuais, o que as torna opções mais viáveis para camadas da população que antes ficariam excluídas por questões de renda. Essa redução de custos democratiza ainda mais o acesso, ao mesmo tempo em que aumenta a competitividade no setor: empresas e instituições podem ampliar a oferta de serviços psicológicos acessíveis, cobrindo lacunas de mercado. Em termos de políticas públicas, vislumbra-se a possibilidade de programas de saúde mental escaláveis a custos sustentáveis, algo particularmente relevante para países em desenvolvimento, nos quais a relação profissional/paciente é insuficiente. Assim, a IA ajuda a desonerar economicamente o cuidado mental, combatendo um dos principais entraves atuais (o alto custo da psicoterapia tradicional) (You et al., 2022).

Conjuntamente com os custos, a IA promove ganhos expressivos na eficiência dos tratamentos em si. Conforme mencionado, intervenções digitais bem delineadas conseguem reduzir o tempo de terapia necessário para atingir determinados objetivos clínicos, sem perda de qualidade. Isso significa que pacientes podem receber alta ou estabilização mais cedo, liberando vagas para novos atendimentos e diminuindo listas de espera. Ao otimizar o processo terapêutico, oferecendo intervenções mais ágeis e adaptáveis, a IA eleva a capacidade de atendimento do sistema de cuidado mental: mais pessoas atendidas em menos tempo, mas com manutenção (ou até melhora) da qualidade assistencial (NICE, 2023b). Como consequência, atenua-se não apenas a sobrecarga financeira, mas também a pressão sobre os profissionais de saúde, que podem concentrar esforços nos casos mais complexos enquanto ferramentas automatizadas lidam com demandas de menor complexidade ou de acompanhamento rotineiro.

Outro benefício fundamental é a escalabilidade quase ilimitada proporcionada pelos sistemas digitais. Diferentemente dos modelos tradicionais, limitados pela disponibilidade física e temporal de cada terapeuta, um *software* de IA pode ser replicado infinitamente após seu desenvolvimento inicial. Uma vez criado e validado, um *chatbot* pode atender teoricamente milhares de usuários simultaneamente, a qualquer hora, em qualquer lugar do mundo, algo humanamente impossível de conseguir de outro modo. Essa escalabilidade rompe gargalos estruturais, por exemplo, regiões sem psicólogos suficientes podem implementar plataformas de IA para suprir parte da demanda local; em situações de crises humanitárias ou pandemias, *chatbots* podem fornecer suporte psicológico emergencial a populações inteiras de forma rápida e massiva. Além do atendimento em si, a IA também automatiza tarefas administrativas e de triagem, aumentando a produtividade global dos serviços de saúde mental. Processos como agendamento de consultas, lembretes aos pacientes, coleta de autorrelatos diários de sintomas e até a elaboração de sumários de sessões podem ser delegados a sistemas inteligentes. Com isso, profissionais são liberados de atividades burocráticas e repetitivas, podendo focar onde sua expertise humana é mais insubstituível: na formulação do caso clínico, na tomada de decisões complexas e no contato terapêutico de alta qualidade. A colaboração homem-máquina nesse sentido maximiza o aproveitamento de recursos humanos: psicólogos e psiquiatras podem atender mais casos ou casos mais graves,

enquanto a IA cuida do suporte básico e do processamento de grande volume de dados (Government UK, 2025; Rollwage et al., 2023).

Por fim, cabe notar que a eficiência promovida pela IA não é apenas quantitativa, mas também qualitativa. Ao analisar grandes bases de dados de forma automática, algoritmos conseguem encontrar padrões e insights que informam melhorias contínuas nos serviços. Por exemplo, ao compilar milhares de interações de usuários, um sistema pode identificar quais perguntas ou técnicas funcionam melhor e refinar seus protocolos, via atualizações de software, algo que em contextos tradicionais dependeria de longos projetos de pesquisa e capacitação de pessoal. Essa retroalimentação *data-driven* contribui para um aprendizado organizacional acelerado, no qual as intervenções ficam cada vez mais efetivas conforme mais pessoas as utilizam. Além disso, a IA permite monitorar indicadores de desempenho (tempos de resposta, taxas de adesão, evolução sintomática média dos usuários) em tempo real, facilitando o gerenciamento adaptativo das clínicas e programas de saúde mental (Rollwage et al., 2023). Em suma, a tecnologia agrega um nível de inteligência operacional antes inalcançável, o que, se bem utilizado, pode revolucionar a forma como os serviços de psicologia são planejados, entregues e avaliados em termos de custo-benefício.

3.4 Experiência do usuário e redução do estigma

Sob a perspectiva do usuário/paciente, a IA traz potenciais vantagens na qualidade da experiência terapêutica, muitas das quais se refletem em maior engajamento e diminuição de barreiras subjetivas ao cuidado. Um ponto-chave é que as interações com sistemas automatizados, por serem isentas de julgamento humano, costumam ser percebidas como menos ameaçadoras para indivíduos que nutrem resistência ou ansiedade em relação à psicoterapia tradicional. Para certas pessoas, especialmente aquelas com histórico de estigmatização ou que têm dificuldade em confiar em estranhos, falar de seus problemas a um aplicativo pode soar mais confortável do que expor-se diante de um terapeuta de carne e osso. Estudos indicam que, para esse perfil de paciente, a psicoterapia mediada por aplicativos inteligentes é avaliada de forma bastante positiva, atuando como uma porta de entrada que

legitima a necessidade de cuidado emocional sem deflagrar o receio de exposição ou vergonha social (Lucas et al., 2014). Assim, indivíduos que antes evitariam buscar qualquer tipo de ajuda passam a aceitar o suporte virtual, o que constitui um primeiro passo importante para a conscientização e possível encaminhamento posterior a tratamentos mais intensivos. Nesse sentido, a IA pode reduzir o estigma associado a "ir ao psicólogo", normalizando a ideia de utilizar ferramentas de bem-estar mental no cotidiano.

Outro aspecto que enriquece a experiência do usuário é o aumento do conforto e da sensação de controle durante as interações. Diferentemente de uma sessão presencial, na qual o paciente às vezes se sente pressionado pelo contexto formal do consultório ou pela presença imediata do terapeuta, nos atendimentos mediados por *chatbot* o indivíduo pode comunicar-se no seu próprio ritmo, pausar para refletir antes de responder, escrever e editar o que deseja expressar. Essa dinâmica assíncrona ou semi-síncrona pode ser menos estressante para muitos, especialmente para quem tem dificuldades de comunicação ou fobia social, pois elimina elementos de comunicação não-verbal potencialmente intimidadores e permite que a pessoa escolha *quando* e *como* se engajar. Além disso, o *setting* digital pode ser customizado: o usuário escolhe um local fisicamente confortável (sua casa, por exemplo), o que também contribui para maior relaxamento e abertura emocional. Pesquisas qualitativas revelam depoimentos de pacientes que relatam "sentir-se mais à vontade" e "menos julgados" ao conversar com um agente virtual, o que aumenta a adesão ao tratamento e a sinceridade nas respostas (Lee et al., 2024; Lucas et al., 2014). Esse ambiente acolhedor, aliado, conforme já dito, à privacidade e ao anonimato, favorece uma maior honestidade do paciente consigo mesmo, estimulando insights e reflexões que, paradoxalmente, podem ser mais difíceis de emergir em contextos presenciais nos quais atua a censura social.

Adicionalmente, os aplicativos de terapia digital permitem um nível de personalização da experiência do usuário que excede em muito o que um profissional sozinho conseguiria oferecer. Plataformas podem utilizar algoritmos de aprendizado de máquina que aprimoram continuamente a capacidade de compreender e responder às necessidades emocionais individuais. Com isso, as interações tornam-se mais empáticas e precisas, ajustando o tom, o conteúdo e o estilo de comunicação conforme o perfil de cada pessoa. Se um paciente prefere objetividade e dados

científicos, o sistema pode adotar essa linha; se outro busca validação emocional e encorajamento, o algoritmo reconhece e espelha essa demanda. Esse nível de personalização faz o usuário sentir-se “ouvido” e “compreendido” de uma forma profunda, embora seja por uma máquina, o que evidencia o quão longe a tecnologia de processamento de linguagem natural avançou na simulação de calor humano. Quando bem calibrada, tal personalização não apenas melhora a experiência subjetiva, mas também contribui para a efetividade do processo terapêutico, uma vez que intervenções alinhadas com as preferências e características do paciente tendem a gerar maior impacto (Beatty et al., 2022; Li et al., 2023). Assim, a IA tem demonstrado capacidade de qualificar o modo de entrega dos serviços de saúde mental, tornando-os mais centrados no usuário e potencialmente mais humanizados em resultado, ainda que por meios artificiais.

Por fim, vale notar um efeito cultural possivelmente benéfico: a popularização de *chatbots* e aplicativos de psicoterapia tem inserido o discurso sobre saúde mental no cotidiano digital de milhões de pessoas, especialmente jovens. Isso ajuda a normalizar a conversa sobre emoções e sofrimento psíquico, reduzindo o tabu social. Campanhas e *challenges* em mídias sociais ligados a bots terapêuticos, por exemplo, estimularam usuários a compartilhar experiências de uso e relatos de melhora, o que contribui para combater preconceitos e ampliar a conscientização coletiva de que cuidar da mente é tão importante quanto cuidar do corpo. Assim, a IA pode atuar também como ferramenta de psicoeducação em larga escala, disseminando conhecimento sobre bem-estar emocional e técnicas básicas de autocuidado para públicos que talvez nunca tivessem contato com esses conteúdos. Ao diminuir o estigma e ampliar a alfabetização emocional, cria-se um círculo virtuoso: mais pessoas se sentem aptas a buscar ajuda, seja via IA ou métodos tradicionais, alimentando demanda por serviços e pressionando sistemas de saúde e formuladores de políticas a valorizarem e investirem mais em saúde mental, desde que com salvaguardas e supervisão humana, conforme recomendações de órgãos e literatura recente (NICE, 2023b; Lee et al., 2024).

3.5 Apoio aos profissionais e avanço da pesquisa

Os benefícios da IA na psicoterapia não se restringem aos pacientes; profissionais de saúde mental também podem se apoiar nessas ferramentas para aprimorar sua prática e estender seu alcance. Em vez de encarar a tecnologia como inimiga ou concorrente, psicólogos e psiquiatras vêm descobrindo formas de integrá-la como aliada em sua rotina clínica (NICE, 2023b; Lee et al., 2024). Um exemplo claro é a disponibilidade de assistentes virtuais especializados que auxiliam o terapeuta em diversas tarefas. Já existem softwares capazes de auxiliar na elaboração de planos de tratamento personalizados, sugerindo intervenções baseadas em dados do paciente e na literatura científica atualizada. Da mesma forma, *dashboards* inteligentes podem realizar monitoramento contínuo dos pacientes entre as sessões, coletando sinais de humor, sono, atividade física e gerando relatórios gráficos para o profissional acompanhar o progresso. Esse tipo de coterapia digital otimiza o tempo do clínico: enquanto o algoritmo compila e analisa informações rotineiras, o profissional pode concentrar-se em processos de tomada de decisão mais complexos e na preparação de intervenções de maior impacto. Além disso, sistemas de alerta podem notificar o terapeuta em caso de detecção de algum indicador de risco elevado, por exemplo, se o paciente reporta ideias suicidas no *chatbot*, permitindo resposta ágil (Rollwage et al., 2024). Ao assumir tarefas operacionais e de vigilância contínua, a IA libera os profissionais para focarem naquilo que a máquina não faz: a escuta qualificada, a interpretação profunda e a construção do vínculo terapêutico.

Outra frente vantajosa é a colaboração interdisciplinar catalisada pela adoção dessas tecnologias. A implementação de IA na psicologia frequentemente requer a interação de equipes compostas por psicólogos, cientistas da computação, engenheiros de dados, especialistas em UX (*user experience*), entre outros. Esse modelo colaborativo não apenas amplia as possibilidades de inovação na prática clínica, como também fortalece o desenvolvimento de pesquisas aplicadas em saúde mental. Projetos de IA em psicoterapia acabam unindo campos do saber antes distantes, como neurociência, psicologia clínica, aprendizado de máquina e ética, fomentando um ecossistema transdisciplinar. Por exemplo, psicólogos e *data scientists* trabalhando juntos podem descobrir novos marcadores de melhora terapêutica em dados de linguagem; filósofos e engenheiros podem criar diretrizes para que o algoritmo seja transparente e justo; designers e pacientes podem co-criar interfaces mais empáticas. Essa efervescência intelectual e prática tende a gerar

saltos qualitativos tanto na tecnologia quanto na compreensão científica dos processos terapêuticos. De fato, a utilização de modelos avançados (como redes neurais profundas) já permitiu avanços no reconhecimento e interpretação de emoções em tempo real, detectando padrões comportamentais e indicadores psicossociais com precisão inédita (Khoo et al., 2024). Tais modelos fornecem dados mais confiáveis para embasar decisões clínicas, e as descobertas daí oriundas podem retroalimentar teorias psicológicas e psiquiátricas, num ciclo virtuoso entre prática e teoria.

Por fim, a IA habilita novas possibilidades de pesquisa clínica em larga escala. Tradicionalmente, estudos em psicoterapia enfrentam limitações logísticas e amostrais: é difícil acompanhar muitos pacientes de perto durante longo tempo ou coletar dados granulares de cada sessão. Com sistemas digitais, torna-se viável realizar ensaios clínicos híbridos ou totalmente online, acompanhando milhares de usuários anônimos interagindo com um *bot* e medindo resultados padronizados. Isso gera bases de dados riquíssimas para análise, permitindo testar hipóteses a respeito do que funciona ou não em terapia com poder estatístico elevado. Além disso, algoritmos preditivos podem ser empregados para antecipar resultados clínicos a partir de indicadores iniciais. Por exemplo, aplicando *machine learning* sobre prontuários eletrônicos combinados com dados de uso de um aplicativo, pesquisadores podem identificar padrões associados ao risco de desistência ou a recaídas em transtornos, informações valiosas para desenvolver estratégias preventivas. Do ponto de vista de saúde pública, essa capacidade preditiva habilita o delineamento de intervenções populacionais mais eficazes, focando recursos em grupos ou indivíduos de maior risco antes mesmo que eventos adversos ocorram (Eichstaedt et al., 2018; Khoo et al., 2024). Assim, a IA se consolida como ferramenta essencial não apenas para a clínica baseada em evidências, mas também para o avanço do conhecimento científico em saúde mental, ao viabilizar formas inovadoras de conduzir pesquisas e integrar grandes volumes de dados ao cotidiano terapêutico.

4. PERIGOS E DESAFIOS DO USO DE IA NA PSICOTERAPIA

Embora a IA ofereça oportunidades promissoras para a psicoterapia, ela também apresenta riscos, limitações e desafios que não devem ser ignorados. Esta seção analisa criticamente os principais “perigos” relacionados à interseção entre IA e saúde mental, agrupando-os em categorias que vão desde questões técnicas até implicações sociais amplas. A literatura e casos recentes mostram que, sem as devidas salvaguardas, o uso descontrolado da IA em contextos terapêuticos pode levar a consequências adversas, algumas sutis e outras potencialmente catastróficas. Para ilustrar a gravidade, apresentamos exemplos concretos e posicionamentos críticos que têm surgido no debate público e acadêmico.

4.1 Questões éticas, privacidade e segurança dos dados

Uma das maiores preocupações ao integrar IA na psicoterapia é a questão ética e a segurança. Como mencionado, o uso de plataformas digitais desafia princípios essenciais como confidencialidade e proteção de dados sensíveis. *Chatbots* e aplicativos precisam coletar informações pessoais, incluindo histórico de saúde mental, sentimentos expressos durante as conversas e, em alguns casos, captação de voz ou expressões faciais, para funcionarem corretamente. Isso cria um dilema: os dados necessários para oferecer suporte automatizado também se tornam ativos valiosos que podem ser explorados de forma inadequada. Vazamentos de dados de saúde são frequentes, e na psicologia, cada vazamento pode revelar aspectos da subjetividade ou de histórias traumáticas de uma pessoa (Huckvale et al., 2019; Martinez-Martin & Kreitmair, 2018).

Além do risco de *hacking* e vazamentos, é importante questionar quem controla e quem tem acesso aos dados coletados. Muitas soluções de IA em saúde mental são oferecidas por empresas privadas cujo objetivo final inclui lucro e monetização. Isso gera um evidente conflito de interesses: o que garante que os dados do paciente não sejam compartilhados com terceiros, como empresas de *marketing*, planos de saúde ou empregadores, sem consentimento explícito? Pesquisas mostram que diversas aplicações de psicoterapia digital direto ao consumidor não seguem padrões rigorosos de proteção de dados, apresentando políticas de privacidade vagas e termos de uso que permitem amplo compartilhamento (Martinez-Martin & Kreitmair, 2018). Essa

situação é preocupante, pois abre espaço para violações de privacidade e até discriminação. Imagine, por exemplo, uma seguradora de saúde negando cobertura ou aumentando prêmios após acessar registros emocionais coletados por um app de terapia de um cliente. Casos recentes de fiscalização aumentam a preocupação: nos Estados Unidos, a Federal Trade Commission (FTC) proibiu a BetterHelp Inc. de compartilhar dados sensíveis de saúde para publicidade e ordenou ressarcimento aos consumidores (Associated Press 2023; FTC, 2023). Sem regulamentação adequada e ética rigorosa, o uso de IA pode prejudicar a confiança na relação de cuidado, levando pacientes a se afastarem por medo de exposição.

Um aspecto importante é a confiabilidade das respostas e orientações fornecidas pelos sistemas de IA. Por melhores que sejam os algoritmos, eles podem errar e esses erros na saúde mental podem ter consequências graves. Uma preocupação crescente são as "alucinações" das IAs generativas, ou seja, quando o sistema inventa informações incorretas ou enganosas com aparente segurança. No contexto terapêutico, isso poderia resultar em aconselhamentos prejudiciais, orientações absurdas ou validação involuntária de ideias autodestrutivas. Casos extremos já aconteceram: *chatbots* que, diante de desabafos suicidas, deram respostas inadequadas ou insuficientes, falhando em orientar o usuário a buscar ajuda profissional (Parker, 2025). Em 2023, um episódio trágico na Europa destacou esse risco: na Bélgica, um homem de 30 anos cometeu suicídio após semanas conversando com um *chatbot* que, supostamente, reforçou suas angústias em vez de ajudá-lo (Walker, 2023). Outro caso grave ocorreu nos EUA, onde um adolescente de 16 anos se suicidou após trocar milhares de mensagens com o ChatGPT, que gradualmente assumiu o papel de “terapeuta” e criou um vínculo ilusório com ele (França, 2025). Esses eventos mostram um alerta: até que ponto podemos confiar em sistemas automáticos para intervenções críticas de saúde? E quem será responsabilizado em casos de desfechos trágicos? Atualmente, não há consenso sobre a responsabilidade legal e profissional em relação ao atendimento psicológico por IA. Como a IA não possui deveres morais ou jurídicos, seus desenvolvedores costumam dizer que o produto é apenas informativo. Assim, criamos uma zona cinzenta de responsabilidade: se um *chatbot* falhar ao acionar emergências ou desencorajar uma autoagressão, o paciente fica sem suporte adequado, sob risco de carecer da atenção de um profissional treinado. Essas lacunas reforçam a

necessidade urgente de protocolos de segurança rigorosos e regulamentação, como a obrigatoriedade de mecanismos automáticos de detecção de risco, com redirecionamento imediato para atendentes humanos ou serviços de emergência (Hern, 2024).

Outro dilema ético relacionado à dependência tecnológica e ao autodiagnóstico incorreto surge da facilidade de acesso e da interatividade dos aplicativos, que podem levar alguns usuários a utilizá-los excessivamente, negligenciando a busca por ajuda profissional especializada. Existem relatos de pessoas que se apegam ao *chatbot* como única fonte de suporte, formando uma relação simbiótica artificial e, pior ainda, tomando as respostas do algoritmo como verdades absolutas sobre si mesmas. Isso pode resultar em autodiagnósticos errôneos, por exemplo, alguém insere seus sintomas em um app que conclui que ela possui transtorno bipolar, levando-a a acreditar nisso sem avaliação médica adequada. Essa rotulagem indevida pode gerar danos como ansiedade, estigma interno e uso inadequado de medicamentos. Além disso, é preocupante quando a IA falha em alertar que uma situação necessita de atenção humana imediata; sem discernimento, um aplicativo pode manter um diálogo terapêutico com alguém em surto psicótico, quando o indicado seria orientá-lo a procurar um hospital. Assim, a ausência de julgamento clínico por parte das máquinas representa um risco sério. Em última instância, a IA não possui, e talvez nunca adquira, a responsabilidade ética fundamental de um profissional humano ao cuidar de alguém. Ela não faz juramento de zelar pela vida, não responde por omissão de socorro, nem consegue captar nuances morais. Portanto, confiar cegamente nesses sistemas sem supervisão humana é eticamente arriscado. Para minimizar esses riscos, órgãos como o Conselho Federal de Psicologia (CFP) reforçam que a IA deve sempre atuar como um apoio supervisionado, não como uma autoridade autônoma no cuidado (CFP, 2025b). Entre as recomendações estão a aprovação regulatória prévia de ferramentas de IA por órgãos de saúde, total transparência no funcionamento (explicabilidade) e a presença de um backup humano em situações de crise. Somente por meio de uma abordagem técnica cuidadosa, segurança jurídica e construção coletiva de normas se conseguirá evitar que a inovação tecnológica, em vez de salvar vidas, inadvertidamente as coloque em perigo.

4.2 Limitações clínicas e falta de humanização

Sob uma perspectiva estritamente clínica, a IA enfrenta limitações inerentes à sua natureza não humana, o que impede sua substituição completa à psicoterapia tradicional. Embora tenham sido destacados na seção de potenciais em que a IA pode ser eficaz em intervenções pontuais, há um consenso entre especialistas de que esses sistemas devem ser considerados ferramentas complementares, não substitutas do trabalho terapêutico realizado por humanos (Babu & Joseph, 2024).

Diversos fatores sustentam essa conclusão. O mais evidente é a incapacidade da IA de reproduzir aspectos essenciais da interação humana, como a empatia genuína e a leitura emocional sutil. Emoções humanas são fenômenos complexos, subjetivos e dependentes do contexto. Elas se manifestam por meio de sutilezas, como um brilho no olhar, um leve tremor na voz ou um silêncio significativo, cuja interpretação adequada exige inteligência emocional e vivência que vão além de padrões codificáveis. Os *chatbots* atuais, embora possam reconhecer palavras-chave relacionadas a sentimento, não captam nuances que um terapeuta treinado perceberia naturalmente. Por exemplo, um paciente pode afirmar “estou bem”, enquanto seu tom ou postura indica o contrário; a IA, focada apenas no conteúdo explícito, interpretaria literalmente, enquanto o psicólogo investigaria além. Essa limitação pode gerar respostas impessoais ou desconexas do contexto emocional do usuário, podendo até aumentar a sensação de incompreensão e solidão (Kim et al., 2024). Além disso, a empatia algorítmica, por mais convincente que pareça, carece de autenticidade ontológica: saber que existe uma pessoa real ouvindo e se importando tem um efeito terapêutico, talvez um efeito “placebo” positivo, que a simples simulação computacional não consegue alcançar (Fluckiger et al., 2018). Essa essência do cuidado, o calor humano e a alteridade reconhecida são difíceis de imaginar que possam ser totalmente substituídos por máquinas em breve.

Outro desafio importante é a inflexibilidade dos algoritmos frente às complexidades subjetivas. A saúde mental de uma pessoa é formada por uma combinação de fatores: história de vida, traumas, dinâmica familiar, contexto sociocultural, entre outros. Enquanto um terapeuta humano consegue integrar essas múltiplas camadas de significado durante a sessão, uma IA funciona principalmente com base em padrões estatísticos e dados pré-estabelecidos. Isso faz com que muitos

fatores, além da base de dados ou não quantificáveis, escapem ao entendimento da máquina. Por exemplo, como um *chatbot* interpretaria expressões metafóricas ou sarcasmo do paciente? Ou perceberia *insights* que surgem não do que é dito, mas do que é evitado na fala? Essas sutilezas fazem parte do cotidiano clínico e muitas vezes orientam o caminho da terapia. Como a IA depende de entradas claras e objetivas, ela tem dificuldades com ambiguidades e contradições que, paradoxalmente, são componentes essenciais da psique humana. Assim, há risco de diagnósticos imprecisos ou intervenções inadequadas ao confiar exclusivamente na avaliação mecanicista da IA (Mendonça et al., 2024; Zhang et al., 2024). A máquina pode erroneamente rotular um luto normal como “depressão” ou deixar de identificar uma ideação suicida velada, porque o usuário não usou as palavras exatas que ativam o alerta. Falta à IA a sensibilidade clínica refinada pelo contato humano, aquela habilidade de ler “nas entrelinhas” do discurso e do comportamento, desenvolvida por terapeutas ao longo de anos de experiência e supervisão (Fluckiger et al., 2018).

Por conta dessas lacunas, especialistas alertam que substituir completamente a interação humana por soluções digitais pode tornar o cuidado psicológico um processo mecânico e impessoal, prejudicando sua essência. Sem a ligação humana, há risco de aumentar sentimentos de isolamento, especialmente em pacientes emocionalmente fragilizados (Karyotaki et al., 2021; Wu et al., 2021). Por exemplo, uma pessoa deprimida, que se sente desconectada do mundo, ao ser atendida somente por um algoritmo, pode perceber essa desconexão ainda mais, afinal, não há alguém real ali. Isso destaca um ponto importante: a terapia não é apenas aplicar técnicas, mas um encontro entre pessoas. O ambiente terapêutico é um espaço no qual se constrói a relação, permitindo ao paciente experimentar novas formas de vínculo e de ser aceito. Uma IA não consegue oferecer esse encontro verdadeiro; no máximo, uma ilusão de conexão. Portanto, quem já se sente “não ouvido” ou “invisível” na vida, ao interagir só com linhas de código, pode, com o tempo, ver essa sensação de vazio relacional se intensificar (Fluckiger et al., 2018).

Os dados iniciais sobre a eficácia a longo prazo das intervenções com IA confirmam algumas dessas preocupações. Diversos estudos indicam que os benefícios costumam ser mais evidentes no curto prazo e desaparecem ao longo do tempo sem continuidade ou intervenção humana aprofundada (Almuqrin et al., 2025; Torous & Topol, 2025). Além disso, problemas de engajamento e retenção,

recorrentes nesse ecossistema, ajudam a explicar por que benefícios observados no curto prazo podem se dissipar quando cessa o suporte ou o uso contínuo (Boucher & Raiker, 2024). Transformações emocionais profundas, como, por exemplo, a reorganização da personalidade, o luto ou a ressignificação de traumas, geralmente requerem processos terapêuticos longos, com altos e baixos, transferências e rupturas, que fazem parte do amadurecimento psíquico. Até agora, nenhum recurso digital conseguiu replicar essa profundidade. Assim, confiar demais na ideia de que "um aplicativo resolverá meu problema de anos em poucas semanas" pode gerar frustração ou abandono do tratamento, pois a realidade costuma ser mais complexa.

Outro aspecto importante é a transferência dos ganhos terapêuticos para a vida real. A ideia de que melhorar as interações com um *chatbot* leva a melhorias nas relações sociais reais não é garantida. Há evidências de que habilidades praticadas em ambientes artificiais não se aplicam automaticamente às interações humanas, que possuem imprevisibilidade, mutualidade e fricção (Li et al., 2023). Assim, uma pessoa pode se sentir confiante conversando com o aplicativo, mas permanecer tímida e isolada na presença de pessoas reais. A inteligência artificial oferece um ambiente controlado e previsível, que não desafia completamente o usuário a enfrentar as frustrações inerentes às relações humanas. Conversar com um *bot* é como usar bóias ao nadar: dá a sensação de segurança, mas não ensina a nadar de verdade. Isso pode gerar uma falsa impressão de progresso, na qual o usuário acredita estar avançando, mas, ao socializar ou resolver conflitos no mundo real, percebe que ainda não está preparado. Essa diferença mostra que os vínculos humanos são essenciais para a saúde mental, pois oferecem desafios e um senso de pertencimento.

Vimos que as limitações clínicas da IA na psicoterapia ressaltam um ponto essencial: a interação humana permanece imprescindível. Nenhum algoritmo atual consegue compreender a complexidade dos fatores conscientes e inconscientes que influenciam a psique, nem reproduzir a cura que nasce do encontro intersubjetivo. Portanto, a IA deve ser vista como uma ferramenta de suporte, útil para tarefas específicas e eficiente em complementar, mas nunca como um substituto completo do terapeuta humano. Caso contrário, arriscamos, em nome da modernização, transformar a psicoterapia em uma caricatura de si mesma, perdendo aquilo que ela possui de mais valioso: o cuidado humano com o humano.

4.3 Desafios metodológicos e de pesquisa

Ao analisar criticamente o uso de IA em saúde mental, enfrentamos desafios não só relacionados ao que está sendo implementado, mas também a como verificar sua eficácia e aprimorá-la. Em suma, obstáculos metodológicos e científicos relevantes na pesquisa de IA na psicoterapia dificultam a obtenção de evidências robustas e de orientações precisas (Eysenbach, 2011; Rivera et al., 2020).

Um primeiro desafio é que a maioria das tecnologias de *chatbots* terapêuticos ainda está em fase experimental ou piloto. Isso indica que os dados disponíveis vêm, em sua maioria, de estudos preliminares com amostras pequenas, acompanhamento breve e muitas vezes sem um grupo controle adequado. A falta de ensaios clínicos robustos: randomizados, controlados e de longo prazo, compromete a confiabilidade e a capacidade de generalização dos resultados. Por exemplo, um aplicativo X pode ter sido testado apenas com 30 universitários durante 4 semanas; embora os resultados sejam positivos nesse cenário restrito, não sabemos se seria eficaz ou mesmo seguro em 300 pacientes da população geral por 6 meses. Essa fragilidade das evidências torna arriscado implementar amplamente certas ferramentas sem cautela. Como criar diretrizes de prática baseadas em evidências se elas ainda são incipientes e repletas de vieses de publicação que tendem a destacar resultados positivos de startups, enquanto experiências mal-sucedidas raramente aparecem?(Casu et al., 2024; Mayor, 2025)

Além disso, há notável heterogeneidade nos estudos já realizados. Diferentes pesquisas adotam protocolos variados, utilizam métricas distintas e até apresentam definições diferentes do que constitui um "sucesso” em uma intervenção digital. Alguns avaliam a redução dos sintomas por meio de escalas padronizadas; outros medem a adesão e satisfação do usuário, entre outros aspectos. Essa falta de padronização torna difícil comparar resultados e tirar conclusões gerais. Ao tentar consolidar os dados em revisões ou meta-análises, encontra-se o problema de “comparar maçãs com laranjas”: por exemplo, um *chatbot* baseado em TCC para depressão não é equivalente a um *coaching* digital para estresse no trabalho, mas ambos podem ser agrupados na categoria “IA em saúde mental”. Sem critérios padronizados, não ficamos sabendo quais elementos específicos das intervenções levam aos efeitos observados. É a frequência de uso? A qualidade do algoritmo? O

tipo de abordagem psicológica utilizada (TCC, ACT, etc.)? Essa dispersão metodológica impede conclusões definitivas e dificulta a formulação de recomendações consistentes para a prática clínica (Agarwal et al., 2016; Eysenbach, 2011).

Outra limitação na literatura é a escassa diversidade das amostras. Muitos estudos iniciais recrutaram jovens instruídos, familiarizados com tecnologia, como estudantes universitários ou voluntários tecnófilos. Grupos mais velhos, pessoas de baixa renda ou baixa escolaridade, comunidades rurais e diferentes contextos culturais estão sub-representados nas pesquisas (Denecke et al., 2025; Rahsepar Meadi et al., 2024). Isso levanta a dúvida: esses resultados se aplicam a outras populações? Um *chatbot* em inglês treinado principalmente com dados americanos pode não ser tão eficaz ou aceitável entre idosos japoneses ou mulheres indígenas brasileiras, por exemplo. As adaptações linguísticas e culturais representam um desafio, traduzir um app não é só trocar palavras, mas ajustar referências, estilo de comunicação, expressões idiomáticas e valores. Poucos estudos testaram intervenções de IA de forma multicultural ou em países de baixa ou média renda, o que cria um viés de amostragem que limita a generalização das conclusões. Assim, os benefícios da IA podem ser válidos apenas para nichos específicos, não para todos os pacientes de saúde mental (Carter et al., 2021; Chau et al., 2025).

Além disso, a heterogeneidade dos resultados dificulta determinar o que realmente funciona melhor. Diversas pesquisas avaliam diferentes aspectos e, mesmo quando medem algo semelhante, como sintomas de depressão, usam escalas distintas ou realizam estudos em contextos variados, por exemplo, um mede após 4 semanas de uso intenso, enquanto outro após 3 meses de uso esporádico. Essa variação gera dados pouco comparáveis. Assim, não é possível fazer avaliações simples como “*chatbots* reduzem depressão em média X%” ou "uma função específica do app aumenta a adesão em Y%” – simplesmente porque as medidas não são uniformes. Essa diversidade metodológica faz com que cada estudo adicione conhecimento pontual, mas não contribua necessariamente para uma evidência consolidada. Portanto, o campo ainda precisa de metodologias padronizadas, como diretrizes para conduzir e relatar estudos com *chatbots* (há esforços nesse sentido, como um checklist proposto para relatar intervenções com *chatbots*)(Denecke &

Gabarron, 2024). Até que essas padronizações se tornem disseminadas, continuará difícil distinguir o sinal do ruído nos resultados.

Outro aspecto importante é a transparência limitada nas descrições técnicas das IAs analisadas. Muitos artigos acadêmicos focam nos resultados para os pacientes, mas fornecem poucos detalhes sobre o funcionamento dos algoritmos, quais dados foram usados para treinar os modelos e quais parâmetros foram ajustados. Essa falta de informação dificulta a replicação dos estudos e a avaliação externa da qualidade dos sistemas. Por exemplo, se um grupo relata sucesso de um *chatbot*, seria importante que outros pudessem testar a mesma ferramenta ou aplicar um modelo semelhante para verificar se os resultados são semelhantes. No entanto, muitas soluções são proprietárias, desenvolvidas por empresas ou *start-ups*, e seus detalhes são mantidos em sigilo comercial. Isso prejudica a ciência, pois impede a validação independente e dificulta o avanço coletivo, cada grupo acaba reinventando a roda ou testando algo do zero, sem aproveitar o que foi feito anteriormente. Além disso, a falta de transparência impede a identificação de possíveis vieses ocultos nos algoritmos. Uma empresa pode anunciar que seu app atingiu 80% de engajamento, mas talvez tenha omitido que excluiu silenciosamente do estudo todos que pararam de usar na primeira semana (viés de seleção). Assim, é necessário estabelecer padrões de pesquisa mais rigorosos e colaborativos nesse campo, para que possamos confiar e evoluir com base nas descobertas (Tornero-Costa et al., 2023).

Em conclusão, os desafios metodológicos e de pesquisa servem como um alerta para que seja necessário aprofundar a base de evidências antes de expandir o uso das IAs na psicoterapia. Isso não implica parar a inovação, mas realizá-la com responsabilidade científica. É urgente promover estudos maiores, multicêntricos, com maior diversidade de participantes, métodos rigorosos e total transparência, para determinar quais ferramentas funcionam, para quem, em quais condições e por quê. Somente com esse conhecimento fundamentado poderemos integrar a IA de forma eficaz ao campo da saúde mental, evitando modismos vazios ou, pior, prejudicando pacientes devido ao desconhecimento das limitações do que estamos usando.

4.4 Desafios sociais e de adoção

Por fim, ao ampliar o foco do nível individual e técnico para o contexto sociocultural, percebemos diversos desafios na implementação ampla de IAs em intervenções psicológicas. Inovações tecnológicas na saúde não ocorrem isoladamente; elas se inserem em sociedades marcadas por desigualdades, influenciadas por fatores econômicos, culturais e políticos que afetam seu impacto (World Health Organization, 2019). Assim, é fundamental analisar esses fatores macro, pois o sucesso ou fracasso de qualquer solução depende deles.

Um obstáculo claro é o estigma persistente ligado à busca por ajuda psicológica. Apesar de a IA poder fornecer suporte anônimo e discreto, muitas pessoas ainda hesitam em reconhecer a necessidade de ajuda ou em confiar em recursos relacionados a “problemas mentais”. Esse estigma social não desaparece simplesmente porque a plataforma é tecnológica. Em diversos contextos, especialmente entre gerações mais velhas ou grupos culturalmente avessos a discutir emoções, usar um aplicativo de terapia pode parecer tão estranho quanto consultar um psicólogo, ou até mais, devido à falta de familiaridade. Relatos indicam que até programas gratuitos e altamente divulgados enfrentam resistência inicial por parte do público devido a fatores culturais (Clement et al., 2015). Portanto, embora a IA proporcione um ambiente mais seguro e anônimo, isso nem sempre é suficiente para superar barreiras pessoais e sociais que impedem o acesso ao cuidado (Abuyadek et al., 2024). Será preciso investir em educação e mudança de atitudes para que segmentos mais céticos aceitem dialogar com uma máquina sobre seus sentimentos, um processo que exige tempo e políticas públicas comprometidas (World Health Organization, 2019).

Por outro lado, há o risco de dependência excessiva por parte daqueles que utilizam as ferramentas com entusiasmo. A disponibilidade contínua de aplicativos terapêuticos e a resposta rápida podem levar alguns usuários a desenvolver uma hiperconectividade prejudicial (Elhai et al., 2017). Em vez de usar a IA como um complemento ocasional, esses indivíduos passam a buscá-la compulsivamente, quase substituindo atividades presenciais por interações com o *chatbot*. Isso pode causar isolamento social: a pessoa deixa de procurar apoio em amigos, família ou grupos de convivência porque o aplicativo se torna seu “confidente” constante. Em vez de promover habilidades de enfrentamento independentes, uma ferramenta digital inadequadamente ajustada pode reforçar uma dependência, na qual o usuário se

sente seguro apenas ao conversar com a IA. Esse fenômeno já ocorreu com outras tecnologias (redes sociais, jogos online) e na saúde mental não seria diferente: sem orientações para uso saudável, até um aplicativo de apoio pode se transformar em uma fuga ou fonte de alienação. Para evitar isso, especialistas recomendam que as ferramentas tenham mecanismos internos de limitação, como prompts que incentivem atividades offline, lembretes para fazer pausas ou restrições de uso diário (Lee, 2025). Assim como se fala em “dieta digital”, aqui seria uma “dieta terapêutica digital”, garantindo que o suporte virtual não substitua as interações humanas ou práticas de enfrentamento ativas.

A divisão digital apresenta outro desafio importante. Nem todos têm acesso confiável à internet, a dispositivos compatíveis ou consciência digital para usar aplicativos de IA. Essas diferenças ocorrem tanto entre países quanto dentro deles, seguindo fatores como renda, educação, localização, idade e gênero. Se políticas adotarem soluções de IA sem incluir medidas de acessibilidade, há o risco de ampliar desigualdades, como populações rurais ou periféricas com conexão precária, que ficariam excluídas dos benefícios. Pessoas de baixa renda podem não possuir *smartphones* adequados ou dados suficientes, enquanto idosos, embora possam se beneficiar de *chatbots* companheiros na saúde mental, podem não saber instalá-los ou operá-los (Moyle, 2024; Shah, 2024). Questões linguísticas também impedem o acesso: muitas ferramentas falam apenas inglês ou português padrão, excluindo falantes de línguas indígenas ou dialetos regionais. Sem ações de inclusão digital, a IA na saúde mental pode reforçar privilégios de quem já acessa tecnologia, agravando a marginalização, o que é oposto ao princípio ético de equidade no cuidado (Hone, 2024). É necessário que governos implementem estratégias como disponibilizar infraestrutura de internet pública, empréstimo de *tablets*, adaptar ferramentas a diferentes línguas e culturas, além de capacitar usuários e profissionais no uso dessas tecnologias.

Outro ponto importante é como as soluções de IA são comercializadas e reguladas. Nos últimos anos, cresceu o número de *1* e empresas oferecendo aplicativos “terapêuticos” direto ao consumidor, muitas sem validação científica forte. O marketing dessas soluções frequentemente exagera suas capacidades, prometendo eficácia e bem-estar sem base em estudos independentes de longo prazo. Enquanto isso, protótipos criados em universidades e centros de pesquisa, que

passam por testes rigorosos, não recebem tanta atenção midiática ou investimento para chegar ao público. Essa diferença causa uma situação preocupante: soluções menos eficazes, mas bem divulgadas, podem se tornar mais populares do que ferramentas realmente benéficas. Além disso, a lógica de mercado nem sempre alinha os incentivos com a ética; aplicativos gratuitos podem lucrar com publicidade ou venda de dados, enquanto os pagos podem buscar criar dependência para manter assinaturas. Essa realidade reforça a necessidade de regulação governamental e orientação ao consumidor. Assim como medicamentos e dispositivos médicos precisam passar por aprovação de agências reguladoras, seria sensato que softwares de saúde mental também fossem regulados dessa forma (CFP, 2025a). Contudo, falta um marco regulatório específico: no Brasil, por exemplo, ainda não há normas claras para aplicativos terapêuticos, e o uso dessas ferramentas no SUS gerou alertas do Conselho de Psicologia, que pediu cautela e participação da categoria na elaboração de diretrizes (CFP, 2025b). A falta de regulamentação cria espaço para práticas fraudulentas na tecnologia, o que pode prejudicar os pacientes. Portanto, é urgente que as leis acompanhem a velocidade da inovação, estabelecendo critérios de eficácia, segurança de dados e penalidades para uso irresponsável ou propaganda enganosa nesse setor.

Por fim, é fundamental considerar o impacto sobre os profissionais de saúde mental. A introdução de IAs desperta temores compreensíveis na categoria de que o trabalho humano possa ser desvalorizado. Se certas formas de atendimento forem automatizadas, isso pode reduzir oportunidades de emprego? Psicólogos e terapeutas poderiam se tornar apenas “supervisores de máquinas”, distanciando-se dos pacientes? Essas preocupações são legítimas. Historicamente, a tecnologia tem reconfigurado profissões, automatizando tarefas que exigem adaptação dos trabalhadores a novas funções. Na psicologia, é difícil imaginar um “desemprego em massa" por causa da IA, devido à complexidade da prática clínica. Contudo, pode ocorrer uma mudança no papel: talvez menos psicólogos sejam necessários para apoio inicial ou psicoeducação, tarefas que seriam delegadas a aplicativos, enquanto aumenta a demanda por especialistas em casos graves ou para manejar situações em que a IA falha. Ainda assim, há um sentimento de ameaça que é compreensível, e ignorá-lo seria um erro estratégico e ético. A implementação de IA não deve desprezar os profissionais, pois isso poderia alienã-los justamente àqueles que

deveriam apoiar e supervisionar o uso seguro da tecnologia. Reforçamos que a IA é um complemento, não um substituto. O discurso institucional, como o do CFP, destaca que psicólogos são “essenciais e insubstituíveis” no cuidado, independentemente dos avanços tecnológicos (CFP, 2025a). Projetos de implementação de IA devem incluir os profissionais desde o começo, oferecendo treinamento, esclarecendo que a tecnologia visa somar e não competir, e assegurando que o elemento humano continue central no tratamento. Afinal, a qualidade do cuidado mental depende tanto de técnicas eficazes quanto da relação de confiança – algo que só se constrói plenamente entre pessoas. Se a categoria for desvalorizada ou afastada, corre-se o risco de perder esse componente essencial e comprometer a qualidade geral do atendimento.

Um aspecto social importante a destacar é a antropomorfização da IA pelos usuários. À medida que os *chatbots* se tornam mais avançados e conversam de modo fluido, muitas pessoas começam a atribuir características humanas a eles; algumas acreditam sinceramente que o *chatbot* se importa ou entende como um amigo. Essa projeção de humanidade pode criar expectativas irreais sobre as capacidades da IA (Ma et al., 2025). Por exemplo, alguém pode ficar frustrado ou devastado ao perceber que seu confidente virtual não possui consciência ou afeto genuíno. Pode-se criar uma ilusão de companhia que, ao ser quebrada, aumenta sentimentos de solidão ou traição: “confiei todos os meus segredos a um programa de computador…”. Esse fenômeno exige uma abordagem ética cuidadosa: os desenvolvedores devem ser transparentes sobre a natureza não humana do sistema e talvez limitar certas interações que possam gerar confusão, por exemplo, o *chatbot* dizer “eu entendo exatamente como se sente”, frase que sugere empatia que ele não realmente possui (Thunstrom et al., 2024). Há quem argumente que essa antropomorfização pode ser uma injustiça epistêmica: ao priorizar sistemas automatizados em detrimento do diálogo humano, corre-se o risco de desqualificar os relatos e experiências reais dos usuários, especialmente os mais vulneráveis (Koum Besson, 2022). Em essência, se pessoas marginalizadas passam a ser ouvidas só por máquinas, isso pode perpetuar uma forma sutil de silenciamento – suas vozes ficam restritas a bancos de dados, sem chegar a ouvidos humanos com poder de agir. Trata-se de uma questão filosófica, mas fundamental: a quem estamos dando voz na sociedade digital? Devemos garantir que o uso da IA não negligencie as narrativas humanas nem substitua a empatia e a

responsabilidade que temos uns com os outros, transferindo-as de forma conveniente para entidades artificiais sem deveres morais.

Ao identificar esses desafios sociais, fica evidente que o sucesso na implementação de IA na psicoterapia requer políticas e ações conscientes. Não basta que a tecnologia funcione apenas em testes de laboratório; é essencial criar condições sociais que permitam que ela seja útil a todos de maneira equitativa, evitando efeitos colaterais indesejados. Isso envolve combater o estigma por meio de campanhas educativas, promover a inclusão digital, regular o mercado para proteger os usuários, envolver e valorizar profissionais de saúde mental e manter sempre como prioridade o bem-estar humano, e não apenas o fascínio pela tecnologia.

5. IMPERATIVO ÉTICO E DISCUSSÃO

Após uma análise cuidadosa dos potenciais e riscos da introdução da IA na psicoterapia, destaca-se um ponto crucial comum a todas as questões levantadas: a necessidade de equilibrar a eficiência da automação com os valores éticos da empatia e da relação humana. Esse é o núcleo do dever ético que deve orientar a incorporação de tecnologias de IA na área da saúde mental. Nesta seção, discutimos os principais aspectos desse dever, resumidos na ideia de que a IA deve sempre atuar como um complemento, nunca substituindo o profissional de saúde mental. Assim, exploramos como conciliar a inovação técnica com os princípios essenciais da prática psicológica e quais são as consequências práticas e normativas dessa harmonia.

5.1 IA como complemento, nunca substituto

A análise até aqui aponta que a tecnologia deve ser vista como uma ferramenta de apoio na saúde mental, não como uma substituta de profissionais humanos como psicoterapeutas ou psiquiatras (Babu & Joseph, 2024). Embora essa orientação pareça simples, ela possui implicações teóricas e práticas profundas. Como demonstrado, a lógica que rege a inteligência artificial muitas vezes conflita com os princípios da psicoterapia. Por exemplo, os sistemas de IA tendem a “agradar ao usuário”, oferecendo validação e respostas rápidas, o que entra em conflito direto com

o princípio psicanalítico de não satisfazer imediatamente a demanda do paciente. Na psicanálise e em outras abordagens, essa frustração controlada faz parte do processo terapêutico: ao não atender imediatamente ao desejo, o paciente é levado a refletir sobre suas questões e a criar novos significados para seu sofrimento. Os *chatbots* atuais, projetados com foco na experiência do usuário, visam manter o usuário engajado, evitando frustrações, elogiando e oferecendo continuidade, além de evitar silêncios. Essa diferença de abordagem tem raízes clínicas concretas: o design para satisfação contínua, comum em muitos aplicativos, pode fomentar dependência e a erosão de habilidades sociais, pois impede que a pessoa vivencie frustrações e limites próprios das relações humanas. Dessa forma, seguir cegamente a lógica do engajamento digital pode deseducar emocionalmente o paciente, privando-o de experiências essenciais para seu amadurecimento psicológico.

Por isso, é fundamental distinguir o "trabalho braçal" da terapia de sua essência profundamente humana. A IA certamente pode ser uma aliada eficaz nas tarefas instrumentais ou de suporte, como monitoramento de sintomas, organização de agendas, fornecimento de conteúdo psicoeducativo, aplicação de testes padronizados, entre outras tarefas repetitivas ou logísticas. Nesses domínios, a máquina frequentemente supera o humano em precisão e velocidade, tornando seu uso nessas funções muito justificado. Porém, a essência do encontro clínico em si permanece um território inacessível às máquinas. Trata-se daquele momento imprevisível em que duas subjetividades se encontram: o terapeuta percebe "nuances emocionais, contradições e padrões inconscientes" nas entrelinhas da fala do paciente e responde não apenas com palavras, mas com sua presença concreta, com sua história e personalidade refletidas na relação. Essa atuação clínica exige intuição, criatividade e responsabilidade moral; habilidades que nenhum algoritmo, por mais treinado, conseguiu adquirir. Portanto, não podemos delegar à IA aquilo que é insubstituivelmente humano: reconhecer o sofrimento do outro em sua singularidade e oferecer um espaço para as angústias que surgem no ambiente terapêutico.

O futuro da saúde mental não está na substituição do humano, mas na colaboração harmoniosa entre tecnologia e humanidade. Com isso, a IA aumentaria o alcance e a eficácia do cuidado, como um exoesqueleto que dá força adicional ao terapeuta, sempre mantendo o cuidado compassivo, a escuta qualificada e o vínculo terapêutico no centro da atenção. A orientação deve ser: tudo o que a IA realiza deve

visar a melhorar a qualidade do encontro terapêutico, sem substituí-lo ou descartá-lo. Praticamente, isso implica que cada intervenção automatizada deve ter critérios claros para transferência ao profissional humano, quando necessário. Por exemplo, um *chatbot* pode ajudar com crises leves de ansiedade, como ensinar respiração diafragmática, mas, ao perceber sinais de agravamento ou complexidade, deve indicar à pessoa que procure um especialista, facilitando essa conexão, inclusive encaminhando a comunicação. Portanto, a IA atua como uma ferramenta de triagem e suporte, sem exceder suas capacidades ou ameaçar a segurança do atendimento. Essa postura é similar ao princípio ético na medicina de que nenhuma máquina de suporte de vida deve operar sem supervisão médica; da mesma forma, nenhum software de suporte emocional deve atuar sem supervisão psicológica.

Existe também uma dimensão filosófica nesse imperativo: usar IA na clínica sem comprometer os fundamentos do cuidado psicológico. Esses princípios incluem o respeito à autonomia e à dignidade do paciente, a valorização da subjetividade e da singularidade de cada caso, o compromisso com o setting como espaço de construção de sentido e a confidencialidade e confiança como pilares. Todas as tecnologias implementadas que prejudiquem algum desses valores precisam ser reavaliadas ou ajustadas. Por exemplo, se um aplicativo exige que o paciente abdique da privacidade de seus dados, isso viola um princípio ético; portanto, ou há melhorias na segurança ou ele não deve ser utilizado. Se um algoritmo tende a uniformizar atendimentos, tratando todos da mesma forma, é necessário incluir elementos que promovam a personalização para respeitar a singularidade. Quando a facilidade de obter conselhos da máquina desestimula o paciente a assumir uma postura ativa no seu processo, porque a IA oferece respostas prontas, é preciso ajustar a ferramenta para que ela incentive reflexões e ações pessoais, talvez fazendo perguntas em vez de fornecer soluções imediatas. Portanto, garantir um equilíbrio adequado exige atenção contínua e a disposição de ajustar a tecnologia às normas éticas do cuidado, e não o inverso.

Babu e Joseph (2024) reforçam essa ideia ao afirmar que “a integração da IA deve equilibrar a precisão tecnológica com a empatia humana, garantindo que ela complemente, não substitua, a relação terapêutica fundamental”. É importante enfatizar que a última palavra, metaforicamente e literalmente, deve sempre ser humana. Assim, a IA só poderá atingir seu potencial ético de aliviar o sofrimento psíquico sem prejudicar a dignidade e a autonomia do indivíduo. Esse é o imperativo

categórico que deve guiar desenvolvedores, profissionais e gestores: nunca renunciar aos valores humanísticos em prol de avanços tecnológicos. Ao contrário, a tecnologia deve passar por um rigoroso filtro ético, existindo para as pessoas, e não as pessoas para ela.

5.2 Implicações para a prática clínica e para políticas públicas

Transformar o imperativo ético em ações práticas exige atuação em duas áreas principais: a prática clínica e as políticas públicas/regulação. Ambas devem avançar de forma integrada, para evitar que uma ultrapasse a outra, por exemplo, profissionais não estarem preparados para usar novas tecnologias ou inovações surgirem sem a estrutura normativa necessária, criando riscos de segurança e ética. A seguir, apresentamos implicações e recomendações para cada esfera, com base nas discussões anteriores.

Na prática clínica, é fundamental que psicólogos, psiquiatras e demais profissionais de saúde mental sejam treinados para usar ferramentas de IA de forma ética e eficaz. Isso inclui não apenas conhecer as funcionalidades das plataformas disponíveis, mas também compreender seus limites, vieses e riscos. Os currículos de formação podem incluir tópicos sobre saúde digital, e as instituições de ensino podem oferecer cursos de educação continuada nesse tema. O profissional deve saber quando e como integrar um *chatbot* ou aplicativo ao plano terapêutico do paciente. Por exemplo, usar um aplicativo de monitoramento de humor para complementar o tratamento pode ser útil, mas o terapeuta precisa analisar criticamente as informações fornecidas pelo software e nunca substituir seu julgamento clínico pelos dados do programa. A supervisão humana deve ser sempre uma etapa essencial em qualquer intervenção mediada por tecnologia: a IA pode monitorar, porém, é responsabilidade do profissional interpretar os dados e dar sentido a eles, considerando a história do paciente. Além disso, clínicas e serviços devem estabelecer protocolos internos para escalonamento, como, por exemplo, definir que, se um algoritmo detectar risco suicida, o caso será imediatamente assumido por um profissional em até X horas, sem exceções. Outro ponto importante é informar e envolver o paciente nesse uso: obter consentimento informado específico para o uso de IA, explicando o que a ferramenta

faz, seus limites e garantindo que ele possa optar por não a utilizar, além de orientá-lo sobre o uso adequado.

Nas políticas públicas e regulação, a falta de um marco regulatório claro representa um dos maiores riscos no uso de IA em saúde mental. Portanto, é imprescindível criar uma governança adequada para sistemas de IA nesse setor. Isso envolve a participação de diversos atores, como órgãos reguladores de saúde e tecnologia, legisladores, conselhos profissionais e a sociedade civil. Como exemplo, os recentes alertas do Conselho Federal de Psicologia (CFP) no Brasil ilustram essa necessidade: ao alertar para os riscos de implementar IA no SUS sem debate e fiscalização, o CFP defende uma regulação rigorosa que proteja os pacientes e valorize a profissão. Algumas diretrizes práticas que podem ser integradas às políticas e normativas incluem:

- Validação científica prévia: é essencial que ferramentas de IA passem por avaliação independente e apresentem evidências de eficácia e segurança antes de serem amplamente adotadas. Essa validação pode acontecer por meio de processos semelhantes a ensaios clínicos ou certificações feitas por agências de saúde, como a Anvisa no Brasil, especialmente para softwares classificados como dispositivos médicos. Sem essa validação, há o risco de aplicativos fraudulentos se proliferarem livremente.
- Transparência algorítmica: estabelecer a explicabilidade como requisito. Os modelos de IA utilizados em saúde mental devem oferecer algum grau de interpretabilidade de suas recomendações ou decisões, como indicar quais palavras-chave acionaram um alerta. Isso fortalece a confiança e facilita auditorias. Empresas podem ser obrigadas a fornecer documentação técnica para análises regulatórias, mesmo que isso não exponha segredos comerciais publicamente.
- Proteção de dados e privacidade: reforçar e detalhar as regras de consentimento informado do usuário para a coleta de seus dados psicológicos. Inspirando-se em legislações como a GDPR europeia ou a Lei Geral de Proteção de Dados (LGPD) no Brasil, é preciso estipular que dados de saúde mental são considerados categorias sensíveis e seu uso deve ser limitado à finalidade terapêutica, proibindo a venda a terceiros sem consentimento explícito e específico. Além disso, deve-se prever mecanismos de fiscalização e penalizações severas para vazamentos ou usos indevidos, a fim de evitar negligências.

- Responsabilização e qualidade do cuidado: definir responsabilidades civis e éticas em casos de danos causados pelo uso da IA. Por exemplo, se um paciente sofrer prejuízo por orientação incorreta de um *chatbot* integrado a um serviço público, quem responde? O desenvolvedor? O serviço que ofereceu? Ambos? Criar essa rede de responsabilidades é fundamental para evitar zonas de impunidade. Paralelamente, estabelecer padrões mínimos de qualidade: listar as funções que uma IA pode exercer de forma autônoma e aquelas que requerem respaldo humano, além de garantir que todo serviço digital tenha uma indicação clara de como o usuário pode acessar suporte humano em caso de necessidade.
- Inclusão e equidade: políticas públicas devem focar na inclusão digital ao implementar essas inovações. Isso envolve ampliar o acesso à internet em comunidades vulneráveis, oferecer dispositivos ou pontos de acesso gratuitos (como telecentros, UBS com espaços de telessaúde) e adaptar as tecnologias às diversas realidades linguísticas e culturais. Além disso, é crucial monitorar se alguns grupos estão sendo excluídos; por exemplo, se, em uma iniciativa de chatbot do governo, somente jovens brancos urbanos usam, isso indica obstáculos que precisam ser enfrentados. Garantir que a divisão digital não amplie as desigualdades no acesso à saúde mental é uma questão de justiça social.

Pelo que vimos, é crucial que as áreas clínica e pública avancem juntas. Enquanto os profissionais aprimoram suas habilidades para um cenário tecnológico em rápida evolução, sem perder de vista os valores humanos essenciais, legisladores e gestores devem criar um ambiente regulatório seguro e justo. Essa colaboração assegura que a IA seja desenvolvida e utilizada de maneira ética, eficaz e equitativa. Uma comparação histórica pode ser feita: assim como os psicofármacos no século XX revolucionaram a psiquiatria, exigiram regulações como critérios de prescrição e controle de efeitos adversos, a introdução da IA também necessita de frameworks para garantir seu uso benéfico e controlado.

Por fim, é importante destacar a importância da colaboração entre diferentes disciplinas nesses esforços. Nenhum grupo isolado possui todas as respostas, desenvolver uma psicoterapia aprimorada com IA de forma ética requer diálogo entre psicólogos, psiquiatras, engenheiros, juristas, filósofos e usuários/pacientes. Fóruns, comitês e pesquisas colaborativas podem ajudar a criar caminhos com transparência e diversidade de perspectivas. Esse debate público contínuo é essencial no mandato

ético: monitorar e questionar constantemente o desenvolvimento da tecnologia, garantindo que ela sirva ao objetivo maior de aliviar o sofrimento e promover a saúde, em vez de mercantilizar ou instrumentalizar a subjetividade humana.

6. CONCLUSÃO

Neste estudo, examinamos a interação entre IA e psicoterapia sob uma perspectiva crítica e ética. Os resultados indicam um potencial significativo, especialmente para ampliar o acesso a serviços psicológicos e personalizar intervenções, porém, também evidenciam riscos importantes nos aspectos éticos, clínicos e de segurança. A abordagem psicanalítica na discussão destaca que o cuidado humano com sua empatia, vínculo terapêutico e escuta atenta é e continuará sendo insubstituível, mesmo com o avanço dos algoritmos. Diante dessa dualidade de promessas e riscos, é fundamental estabelecer recomendações que garantam que o desenvolvimento tecnológico seja harmonizado com o dever ético na saúde mental.

- Priorizar o modelo de supervisão humana: Toda ferramenta de IA aplicada clinicamente deve estar sob supervisão direta de um profissional de saúde mental qualificado. A inteligência artificial deve atuar como uma auxiliadora ao trabalho humano, automatizando tarefas operacionais, fornecendo dados e sugestões, sem assumir decisões críticas de forma autônoma. Por exemplo, *chatbots* podem realizar acolhimento inicial e exercícios guiados, mas um psicólogo deve monitorar remotamente os casos, avaliando as recomendações e intervindo ativamente em situações de risco. Essa supervisão assegura o dever de cuidado: um responsável humano deve cuidar do bem-estar do paciente, interpretando as informações fornecidas pela máquina e assumindo o controle quando necessário.
- Desenvolver regulamentações específicas e rigorosas é uma necessidade urgente. É fundamental criar políticas públicas e diretrizes claras que regulamentem o desenvolvimento e o uso de IAs na saúde mental. Essas normas devem priorizar a proteção da privacidade dos dados, semelhantes à GDPR europeia, assegurar transparência nos algoritmos, explicando as decisões do sistema, e estabelecer mecanismos de responsabilização em casos de erros ou danos causados pelo uso da tecnologia. Por exemplo, poderia ser exigido que todo aplicativo de terapia

digital obtenha uma certificação governamental para ser disponibilizado ao público, provando que passou por testes de eficácia e segurança. Além disso, diretrizes éticas, elaboradas por órgãos como conselhos de psicologia, devem integrar essas políticas, orientando o design das ferramentas: quais técnicas são aceitáveis, como obter consentimento, limites para respostas automatizadas, entre outros. Uma regulamentação bem planejada impede práticas oportunistas do mercado e fomenta a confiança de profissionais e usuários na adoção de inovações.

- Fomentar a pesquisa interdisciplinar e de longo prazo: Para fundamentar a prática e a regulação, é fundamental investir em pesquisa científica sólida e multidisciplinar sobre IA em saúde mental. Recomenda-se incentivar estudos de longo prazo que analisem os impactos do uso contínuo de IA nos resultados terapêuticos, além de pesquisas voltadas ao desenvolvimento de algoritmos menos enviesados e à integração ética dessas ferramentas na prática clínica. Esse esforço demanda a colaboração próxima entre psicólogos, psiquiatras, cientistas da computação, especialistas em ética e usuários/pacientes, assegurando que múltiplas perspectivas contribuam para a geração do conhecimento. Agências de fomento e instituições acadêmicas devem valorizar essa área como estratégica, apoiando projetos que comparem diferentes abordagens de *chatbot*, testem soluções em populações diversas e construam frameworks teóricos para compreender a relação homem-máquina no contexto terapêutico. Somente com evidências robustas e teoria aperfeiçoada conseguiremos evitar tanto o entusiasmo acrítico quanto o alarmismo sem fundamentos, orientando-nos por dados e reflexões bem fundamentadas.
- Promover a educação digital e a capacitação ética é essencial para a difusão segura da IA em saúde mental, exigindo iniciativas de alfabetização digital tanto para profissionais quanto para o público geral. Profissionais de saúde devem ser treinados para usar essas tecnologias de forma crítica, compreendendo seu funcionamento básico, potencialidades e limitações, de modo a integrá-las efetivamente na prática clínica. Isso pode incluir disciplinas na graduação ou *workshops* promovidos por conselhos e associações. Além disso, pacientes e comunidades precisam ser orientados sobre o uso adequado dessas ferramentas: entender que *chatbots* não substituem acompanhamento presencial, saber quando e como utilizá-los e estar cientes das questões de privacidade. Campanhas de conscientização podem esclarecer que aplicativos de terapia são recursos úteis, mas

que não dispensam a presença do psicólogo, ajudando a desfazer medos e expectativas irreais. O objetivo do letramento é garantir um consentimento informado, para que as pessoas escolham usar (ou não) essas ferramentas com pleno conhecimento de suas possíveis limitações e de como integrá-las a uma busca mais ampla por saúde mental.

Em síntese, a interseção entre IA e psicoterapia emerge como um campo repleto de inovações e obstáculos. É papel da comunidade da saúde mental envolver-se neste debate, guiada por perspectivas humanísticas e éticas, como a psicanálise e a crítica foucaultiana, para que possamos aproveitar os benefícios da tecnologia sem comprometer os princípios do cuidado psicológico. Apenas assim será viável construir um futuro no qual máquinas e humanos atuem de forma complementar, potencializando o cuidado em saúde mental sem renunciar ao que nos torna únicos enquanto seres humanos.

7. REFERÊNCIAS

Abuyadek, R. M., Hammouda, E. A., Elrewany, E., Elmalawany, D. H., Ashmawy, R., Zeina, S., Gebreal, A., & Ghazy, R. M. (2024). Acceptability of Tele-mental Health Services Among Users: A Systematic Review and Meta-analysis. *BMC Public Health*, *24*(1), 1143. *https://doi.org/10.1186/s12889-024-18436-7*

Agarwal, S., LeFevre, A. E., Lee, J., L'Engle, K., Mehl, G., Sinha, C., Labrique, A., & Group, W. H. O. m. T. E. R. (2016). Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. *BMJ*, *352*, i1174. *https://doi.org/10.1136/bmj.i1174*

Almuqrin, A., Hammoud, R., Terbagou, I., Tognin, S., & Mechelli, A. (2025). Smartphone apps for mental health: systematic review of the literature and five recommendations for clinical translation. *BMJ Open*, *15*(2), e093932. *https://doi.org/10.1136/bmjopen-2024-093932*

Andersson, G., Titov, N., Dear, B. F., Rozental, A., & Carlbring, P. (2019). Internet-delivered psychological treatments: from innovation to implementation. *World Psychiatry*, *18*(1), 20–28. *https://doi.org/10.1002/wps.20610*

Associated Press (2023). BetterHelp shared users' sensitive health data, FTC says. *https://apnews.com/article/befca40bb873661d1f8986bb75d8df07*

Babu, M., & Joseph, M. (2024). Artificial Intelligence in Psychotherapy: Hype, Hope, and the Human Factor. *Indian Journal of Psychological Medicine*, *46*(5), 427–433. *https://doi.org/10.1177/02537176241234567*

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Political science. Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. *https://doi.org/10.1126/science.aaa1160*

Barber, L. K., & Santuzzi, A. M. (2015). Please respond ASAP: workplace telepressure and employee recovery. *J Occup Health Psychol*, *20*(2), 172–189. *https://doi.org/10.1037/a0038278*

Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study. *Front Digit Health*, *4*, 847991. *https://doi.org/10.3389/fdgth.2022.847991*

Boucher, E. M., & Raiker, J. S. (2024). Engagement and retention in digital mental health interventions: a narrative review. *BMC Digital Health*, *2*(1), 52. *https://doi.org/10.1186/s44247-024-00105-9*

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work*. *The Quarterly Journal of Economics*, *140*(2), 889–942. *https://doi.org/10.1093/qje/qjae044*

Bufano, P., Laurino, M., Said, S., Tognetti, A., & Menicucci, D. (2023). Digital Phenotyping for Monitoring Mental Disorders: Systematic Review. *J Med Internet Res*, *25*, e46778. *https://doi.org/10.2196/46778*

Carter, H., Araya, R., Anjur, K., Deng, D., & Naslund, J. A. (2021). The emergence of digital mental health in low-income and middle-income countries: A review of recent advances and implications for the treatment and prevention of mental disorders. *Journal of Psychiatric Research*, *133*, 223–246. *https://doi.org/https://doi.org/10.1016/j.jpsychires.2020.12.016*

Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). AI Chatbots for Mental Health: A Scoping Review of Effectiveness, Feasibility, and Applications. *Applied Sciences*, *14*(13). *https://doi.org/10.3390/app14135889*

Cevasco, K. E., Morrison Brown, R. E., Woldeselassie, R., & Kaplan, S. (2024). Patient Engagement with Conversational Agents in Health Applications 2016-2022: A Systematic Review and Meta-Analysis. *J Med Syst*, *48*(1), 40. *https://doi.org/10.1007/s10916-024-02059-x*

Chau, L. W., Lam, R. W., Minas, H., Hayashi, K., Nguyen, V. C., & O'Neil, J. (2025). Digital Health Interventions for Depression and Anxiety in Low- and Middle-Income Countries: Rapid Scoping Review. *JMIR Ment Health*, *12*, e68296. *https://doi.org/10.2196/68296*

Clement, S., Schauman, O., Graham, T., Maggioni, F., Evans-Lacko, S., Bezborodovs, N., Morgan, C., Rusch, N., Brown, J. S., & Thornicroft, G. (2015). What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychol Med*, *45*(1), 11–27. *https://doi.org/10.1017/S0033291714000129*

Coghlan, S., Leins, K., Sheldrick, S., Cheong, M., Gooding, P., & D'Alfonso, S. (2023). To chat or bot to chat: Ethical issues with using *chatbots* in mental health. *Digit Health*, *9*, 20552076231183542. *https://doi.org/10.1177/20552076231183542*

FTC (2023). *FTC to Ban BetterHelp from Revealing Consumers' Data, Including Sensitive Mental Health Information, to Facebook and Others for Targeted Advertising*. Federal Trade Commission Retrieved from *https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-ban-betterhelp-revealing-consumers-data-including-sensitive-mental-health-information-facebook*

Conselho Federal de Psicologia (2025a). CFP divulga posicionamento sobre Inteligência Artificial no contexto da prática psicológica. https://site.cfp.org.br/cfp-divulga-posicionamento-sobre-inteligencia-artificial-no-contexto-da-pratica-psicologica/

Conselho Federal de Psicologia (2025b). CFP alerta para riscos de proposta de uso de inteligência artificial em saúde mental no SUS. https://site.cfp.org.br/cfp-alerta-para-riscos-de-proposta-de-uso-de-inteligencia-artificial-em-saude-mental-no-sus/

Cruz-Gonzalez, P., He, A. W., Lam, E. P., Ng, I. M. C., Li, M. W., Hou, R., Chan, J. N., Sahni, Y., Vinas Guasch, N., Miller, T., Lau, B. W., & Sanchez Vidana, D. I. (2025). Artificial intelligence in mental health care: a systematic review of diagnosis, monitoring, and intervention applications. *Psychol Med*, *55*, e18. *https://doi.org/10.1017/S0033291724003295*

Denecke, K., & Gabarron, E. (2024). The ethical aspects of integrating sentiment and emotion analysis in *chatbots* for depression intervention. *Front Psychiatry*, *15*, 1462083. *https://doi.org/10.3389/fpsyt.2024.1462083*

Denecke, K., Rivera Romero, O., Wynn, R., & Gabarron, E. (2025). Persuasive *chatbot*-based interventions for depression: a list of recommendations for improving reporting standards. *Front Psychiatry*, *16*, 1429304. *https://doi.org/10.3389/fpsyt.2025.1429304*

Durden, E., Pirner, M. C., Rapoport, S. J., Williams, A., Robinson, A., & Forman-Hoffman, V. L. (2023). Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent "Woebot". *Internet Interv*, *33*, 100637. *https://doi.org/10.1016/j.invent.2023.100637*

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preotiuc-Pietro, D., Asch, D. A., & Schwartz, H. A. (2018). Facebook language predicts depression in medical records. *Proc Natl Acad Sci U S A*, *115*(44), 11203–11208. *https://doi.org/10.1073/pnas.1802331115*

Elhai, J. D., Levine, J. C., Dvorak, R. D., & Hall, B. J. (2017). Non-social features of smartphone use are most related to depression, anxiety and problematic smartphone use. *Computers in Human Behavior*, *69*, 75–82. *https://doi.org/10.1016/j.chb.2016.12.023*

Eysenbach, G. (2011). CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *J Med Internet Res*, *13*(4), e126. *https://doi.org/10.2196/jmir.1923*

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Ment Health*, *4*(2), e19. *https://doi.org/10.2196/mental.7785*

Fluckiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy (Chic)*, *55*(4), 316–340. *https://doi.org/10.1037/pst0000172*

França, V. (2025). *Suicídio de adolescente nos EUA abre debate sobre os limites do uso de IA nas terapias https://veja.abril.com.br/tecnologia/suicidio-de-adolescente-nos-eua-abre-debate-sobre-os-limites-do-uso-de-ia-nas-terapias/*

Fujita, J., Yano, Y., Shinoda, S., Sho, N., Otsuki, M., Suda, A., Takayama, M., Moroga, T., Yamaguchi, H., Ishii, M., & Miyazaki, T. (2025). Challenges in Implementing a Mobile AI Chatbot Intervention for Depression Among Youth on Psychiatric Waiting Lists: A Randomized Control Study Termination Report. *https://doi.org/10.1101/2024.11.25.24317880*

Government U. K. (2025). Limbic Access – AI Knowledge Hub (Use case).

He, Y., Yang, L., Qian, C., Li, T., Su, Z., Zhang, Q., & Hou, X. (2023). Conversational Agent Interventions for Mental Health Problems: Systematic Review and Meta-analysis of Randomized Controlled Trials. *J Med Internet Res*, *25*, e43862. *https://doi.org/10.2196/43862*

Hern, A. (2024, 2024–12–12). UK AI Safety Institute finds models can bypass safeguards. *The Guardian*. *https://www.theguardian.com/technology/2024/dec/12/ai-safety-institute-uk-models-bypass-safeguards*

Hone, T. (2024). Digital health technologies and inequalities: a scoping review. *Health Policy*, *138*. *https://doi.org/10.1016/j.healthpol.2024.04.006*

Huckvale, K., Torjesen, I., & Miller, S. (2019). Transparency of mental health apps for quality, safety, and privacy. *JAMA Network Open*, *2*(5), e194400. *https://doi.org/10.1001/jamanetworkopen.2019.4400*

Karyotaki, E., Efthimiou, O., Miguel, C., Bermpohl, F. M. G., Furukawa, T. A., Cuijpers, P., Individual Patient Data Meta-Analyses for Depression, C., Riper, H., Patel, V., Mira, A., Gemmil, A. W., Yeung, A. S., Lange, A., Williams, A. D., Mackinnon, A., Geraedts, A., van Straten, A., Meyer, B., Bjorkelund, C.,…Forsell, Y. (2021). Internet-Based Cognitive Behavioral Therapy for Depression: A Systematic Review and Individual Patient Data Network Meta-analysis. *JAMA Psychiatry*, *78*(4), 361–371. *https://doi.org/10.1001/jamapsychiatry.2020.4364*

Kaywan, P., Ahmed, K., Ibaida, A., Miao, Y., & Gu, B. (2023). Early detection of depression using a conversational AI bot: A non-clinical trial. *PLOS ONE*, *18*(2), e0279743. *https://doi.org/10.1371/journal.pone.0279743*

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at Work: The New Contested Terrain of Control. *Academy of Management Annals*, *14*(1), 366–410. *https://doi.org/10.5465/annals.2018.0174*

Khoo, L. S., Lim, M. K., Chong, C. Y., & McNaney, R. (2024). Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches. *Sensors (Basel)*, *24*(2), 348. *https://doi.org/10.3390/s24020348*

Kim, S., Park, S., & Martinez, W. (2024). Who cares if a machine is ‘empathetic’? Authenticity and simulation in AI mental health tools. *Journal of Bioethical Inquiry*, *21*(3), 345–357. *https://doi.org/10.1007/s11673-024-10345-1*

Koum Besson, E. S. (2022). How to identify epistemic injustice in global health research funding practices: a decolonial guide. *BMJ Glob Health*, *7*(4), e008950. *https://doi.org/10.1136/bmjgh-2022-008950*

Lambert, J., Barnstable, G., Minter, E., Cooper, J., & McEwan, D. (2022). Taking a One-Week Break from Social Media Improves Well-Being, Depression, and Anxiety: A Randomized Controlled Trial. *Cyberpsychol Behav Soc Netw*, *25*(5), 287–293. *https://doi.org/10.1089/cyber.2021.0324*

Lee, H. S., Wright, C., Ferranto, J., Buttimer, J., Palmer, C. E., Welchman, A., Mazor, K. M., Fisher, K. A., Smelson, D., O'Connor, L., Fahey, N., & Soni, A. (2024). Artificial intelligence conversational agents in mental health: Patients see potential, but prefer humans in the loop. *Front Psychiatry*, *15*, 1505024. *https://doi.org/10.3389/fpsyt.2024.1505024*

Lee, S. H. (2025). Exploring artificial intelligence (AI) *chatbot* usage behaviors and dependence among university students. *Journal of Affective Disorders*.

Li, H., Zhang, R., Lee, Y. C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digit Med*, *6*(1), 236. *https://doi.org/10.1038/s41746-023-00979-5*

Li, J., Li, Y., Hu, Y., Ma, D. C. F., Mei, X., Chan, E. A., & Yorke, J. (2025). Chatbot-Delivered Interventions for Improving Mental Health Among Young People: A Systematic Review and Meta-Analysis. *Worldviews Evid Based Nurs*, *22*(4), e70059. *https://doi.org/10.1111/wvn.70059*

Li, Y., Chung, T. Y., Lu, W., Li, M., Ho, Y. W. B., He, M., Mei, X., Chen, D., & Bressington, D. (2025). Chatbot-Based Mindfulness-Based Stress Reduction Program for University Students With Depressive Symptoms: Intervention Development and Pilot Evaluation. *J Am Psychiatr Nurses Assoc*, *31*(4), 398–411. *https://doi.org/10.1177/10783903241302092*

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It’s only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100. *https://doi.org/10.1016/j.chb.2014.04.043*

Ma, N., Khynevych, R., Hao, Y., & Wang, Y. (2025). Effect of anthropomorphism and perceived intelligence in *chatbot* avatars of visual design on user experience: accounting for perceived empathy and trust. *Frontiers in Computer Science*, *7*, 1531976. *https://doi.org/10.3389/fcomp.2025.1531976*

Martinez-Martin, N., & Kreitmair, K. (2018). Ethical Issues for Direct-to-Consumer Digital Psychotherapy Apps: Addressing Accountability, Data Protection, and Consent. *JMIR Ment Health*, *5*(2), e32. *https://doi.org/10.2196/mental.9423*

Mayor, E. (2025). Chatbots and mental health: a scoping review of reviews. *Current Psychology*, *44*(15), 13619–13640. *https://doi.org/10.1007/s12144-025-08094-2*

Mendonça, L. L., Almeida, T., & Silva, R. (2024). Human-like AI? Limits of large language models in capturing emotion, context, and sarcasm. *Journal of Intelligence*, *12*(3), 57. *https://doi.org/10.3390/jintelligence12030057*

Moyle, W. (2024). The role of *chatbots* in enhancing health care for older adults: a scoping review. *Archives of Gerontology and Geriatrics*, *117*, 105381. *https://doi.org/10.1016/j.archger.2024.105381*

NICE (2023a). *Digitally enabled therapies for adults with anxiety disorders (HTE9) – Early Value Assessment guidance*. *https://www.nice.org.uk/guidance/hte9*

NICE (2023b). *Digitally enabled therapies for adults with depression (HTE8) – Early Value Assessment guidance*. *https://www.nice.org.uk/guidance/hte8*

Nieminen, H., Vartiainen, A. K., Bond, R., Laukkanen, E., Mulvenna, M., & Kuosmanen, L. (2025). Recommendations for Mental Health Chatbot Conversations: An Integrative Review. *J Adv Nurs*, *81*(10), 6169–6182. *https://doi.org/10.1111/jan.16762*

Parker, C. (2025, 2025–05–01). AI Chatbots and Suicide: Are Tech Companies Doing Enough to Protect Users? *TIME*. *https://time.com/6970691/ai*-chatbots-*suicide-safeguards/*

Potts, C., Lindstrom, F., Bond, R., Mulvenna, M., Booth, F., Ennis, E., Parding, K., Kostenius, C., Broderick, T., Boyd, K., Vartiainen, A. K., Nieminen, H., Burns, C., Bickerdike, A., Kuosmanen, L., Dhanapala, I., Vakaloudis, A., Cahill, B., MacInnes, M.,…O'Neill, S. (2023). A Multilingual Digital Mental Health and Well-Being Chatbot (ChatPal): Pre-Post Multicenter Intervention Study. *J Med Internet Res*, *25*, e43051. *https://doi.org/10.2196/43051*

Primack, B. A., Shensa, A., Sidani, J. E., Whaite, E. O., Lin, L. Y., Rosen, D., Colditz, J. B., Radovic, A., & Miller, E. (2017). Social Media Use and Perceived Social Isolation Among Young Adults in the U.S. *Am J Prev Med*, *53*(1), 1–8. *https://doi.org/10.1016/j.amepre.2017.01.010*

Proudfoot, J., Ryden, C., Everitt, B., Shapiro, D. A., Goldberg, D., Mann, A., Tylee, A., Marks, I., & Gray, J. A. (2004). Clinical efficacy of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. *Br J Psychiatry*, *185*(1), 46–54. *https://doi.org/10.1192/bjp.185.1.46*

Psicologia, C. F. d. (2023). Nota do Sistema Conselhos de Psicologia sobre Inteligência Artificial.

Rahsepar Meadi, M., Bernstein, J. S., Batelaan, N., van Balkom, A., & Metselaar, S. (2024). Does a lack of emotions make *chatbots* unfit to be psychotherapists? *Bioethics*, *38*(6), 503–510. *https://doi.org/10.1111/bioe.13299*

Raza, S., Rahman, M., Kamawal, S., Toroghi, A., Raval, A., & Kazemeini, F. N. A. (2025). *A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice*. *https://arxiv.org/abs/2407.13699*

Rivera, S. C., Liu, X., Chan, A. W., Denniston, A. K., Calvert, M. J., Spirit, A. I., Group, C.-A. W., Spirit, A. I., Group, C.-A. S., Spirit, A. I., & Group, C.-A. C. (2020). Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med*, *26*(9), 1351–1363. *https://doi.org/10.1038/s41591-020-1037-7*

Rocchi, G., Vocaj, E., Moawad, S., Antonucci, A., Grigioni, C., Giuffrida, V., & Bordini, J. (2024). Optimizing personalized psychological well-being interventions through digital phenotyping: results from a randomized non-clinical trial. *Front Psychol*, *15*, 1479269. *https://doi.org/10.3389/fpsyg.2024.1479269*

Rollwage, M., Habicht, J., Juchems, K., Carrington, B., Hauser, T. U., & Harper, R. (2024). Conversational AI facilitates mental health assessments and is associated with improved recovery rates. *BMJ Innovations*, *10*(1-2), 4–12. *https://doi.org/10.1136/bmjinnov-2023-001110*

Rollwage, M., Habicht, J., Juechems, K., Carrington, B., Viswanathan, S., Stylianou, M., Hauser, T. U., & Harper, R. (2023). Using Conversational AI to Facilitate Mental Health Assessments and Improve Clinical Efficiency Within Psychotherapy Services: Real-World Observational Study. *JMIR AI*, *2*(1), e44358. *https://doi.org/10.2196/44358*

Schafer, S. K., von Boros, L., Schaubruch, L. M., Kunzler, A. M., Lindner, S., Koehler, F., Werner, T., Zappala, F., Helmreich, I., Wessa, M., Lieb, K., & Tuscher, O. (2024). Digital

interventions to promote psychological resilience: a systematic review and meta-analysis. *NPJ Digit Med*, *7*(1), 30. *https://doi.org/10.1038/s41746-024-01017-8*

Schillings, C., Meissner, D., Erb, B., Schultchen, D., Bendig, E., & Pollatos, O. (2023). A *chatbot*-based intervention with ELME to improve stress and health-related parameters in a stressed sample: Study protocol of a randomised controlled trial. *Front Digit Health*, *5*, 1046202. *https://doi.org/10.3389/fdgth.2023.1046202*

Schillings, C., Meissner, D., Erb, B., Schultchen, D., Bendig, E., & Pollatos, O. (2024). Effects of a Chatbot-Based Intervention on Stress and Health-Related Parameters in a Stressed Sample: Randomized Controlled Trial. *JMIR Ment Health*, *11*, e50454. *https://doi.org/10.2196/50454*

Shah, P. (2024). The views and experiences of older adults regarding digital mental health interventions: systematic review. *The Lancet Healthy Longevity*, *5*(8). *https://doi.org/10.1016/S2666-7568(24)00164-8*

Stothart, C., Mitchum, A., & Yehnert, C. (2015). The attentional cost of receiving a cell phone notification. *J Exp Psychol Hum Percept Perform*, *41*(4), 893–897. *https://doi.org/10.1037/xhp0000100*

Thunstrom, A. O., Carlsen, H. K., Ali, L., Larson, T., Hellstrom, A., & Steingrimsson, S. (2024). Usability Comparison Among Healthy Participants of an Anthropomorphic Digital Human and a Text-Based Chatbot as a Responder to Questions on Mental Health: Randomized Controlled Trial. *JMIR Hum Factors*, *11*(1), e54581. *https://doi.org/10.2196/54581*

Tornero-Costa, R., Martinez-Millana, A., Azzopardi-Muscat, N., Lazeri, L., Traver, V., & Novillo-Ortiz, D. (2023). Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review. *JMIR Ment Health*, *10*, e42045. *https://doi.org/10.2196/42045*

Torous, J., & Topol, E. J. (2025). Assessing generative artificial intelligence for mental health. *Lancet*. *https://doi.org/10.1016/S0140-6736(25)01237-1*

Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture*, *3*(4), 206–222. *https://doi.org/10.1037/ppm0000047*

Walker, M. (2023, 2023–03–29). Belgian man takes his own life after AI *chatbot* 'encouraged' him to do so. *The Brussels Times*. *https://www.brusselstimes.com/belgium/389114/belgian-man-takes-his-own-life-after-ai*-chatbot-*encouraged-him-to-do-so*

Wang, J. L., Wang, H. Z., Gaskin, J., & Hawk, S. (2017). The Mediating Roles of Upward Social Comparison and Self-esteem and the Moderating Role of Social Comparison Orientation in the Association between Social Networking Site Usage and Subjective Well-Being. *Front Psychol*, *8*, 771. *https://doi.org/10.3389/fpsyg.2017.00771*

Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. *https://doi.org/10.1145/365153.365168*

World Health Organization. (2019). *Recommendations on digital interventions for health system strengthening* (9789241550505). *https://www.who.int/publications/i/item/9789241550505*

Wu, T., Zhang, H., & Zhou, K. (2021). Evaluating the clinical robustness of digital mental health apps. *npj Digital Medicine*, *4*(1), 86. *https://doi.org/10.1038/s41746-021-00477-8*

You, J. H. S., Luk, S. W. C., Chow, D. Y. W., Jiang, X., Mak, A. D. P., & Mak, W. W. S. (2022). Cost-effectiveness of internet-supported cognitive behavioral therapy for university

students with anxiety symptoms: A Markov-model analysis. *PLOS ONE*, *17*(5), e0268061. *https://doi.org/10.1371/journal.pone.0268061*

Zhang, Y., Li, X., & Kiela, D. (2024). SarcasmBench: Benchmarking sarcasm comprehension in large language models. In *arXiv preprint* (pp. arXiv:2403.01234).