



AMERICAN  
PSYCHOLOGICAL  
ASSOCIATION



# APA Health Advisory on the Use of Generative AI Chatbots and Wellness Applications for Mental Health

Millions of people globally are engaging with general-purpose generative artificial intelligence (GenAI) chatbots and wellness applications to address unmet mental health needs. This can in part be explained by the current mental health crisis, growing rates of loneliness and disconnection, lack of enough providers to meet growing public demand (especially in under-resourced, rural, or unincorporated communities), and a health care system that disincentivizes providers from accepting insurance, leaving many people who are uninsured or underinsured without options. The ease of access and low-cost of these technologies has made them a frequent option for those seeking mental health advice and treatment. However, most of these technologies are not designed or intended to provide clinical feedback or treatment, may lack scientific validation and oversight, often do not include adequate safety protocols, and have not received regulatory approval. Ensuring consumers' safety and well-being requires action from stakeholders, including (but not limited to) awareness by the consumers themselves, caregivers/providers and educators, policymakers, technology industry developers, creators, and professionals, and platforms that develop and/or host GenAI tools and wellness apps.

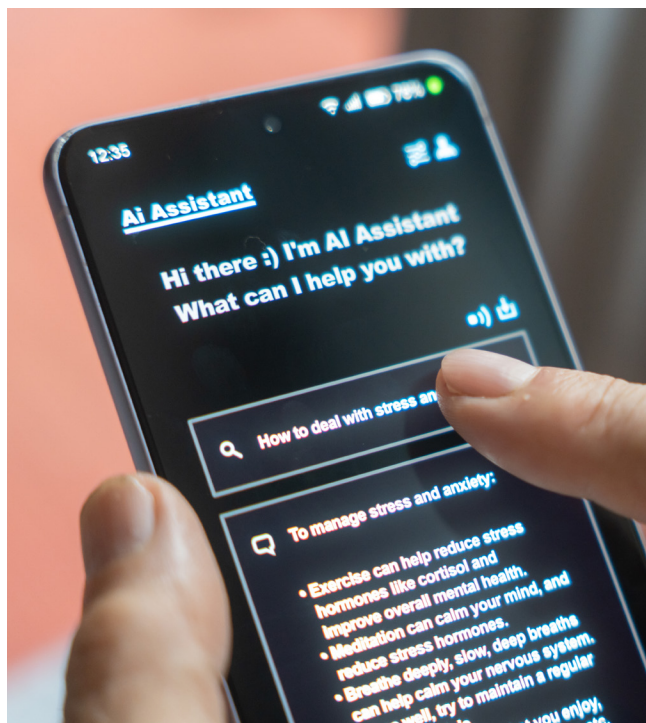
This advisory offers a series of recommendations, some of which may be enacted immediately by consumers and others that will require substantial change by platforms, policymakers, and/or technology professionals.

## BACKGROUND AND ASSUMPTIONS

Mentions of specific products and companies are made for illustrative purposes and are not intended as an endorsement or approval by APA. The recommendations below are based on the scientific evidence to date and the following considerations:

1. This advisory includes only consumer-facing technologies (i.e., technologies that are marketed for and used by consumers with or without the direct oversight of a health provider). Provider-facing technologies (i.e., technologies used by providers to aid practice) are outside the scope of this advisory. Specifically, the advisory includes GenAI chatbots, wellness apps that use GenAI, and other wellness apps that do not use AI, which, regardless of their intent, consumers are using for mental health support. This advisory does not include AI-powered administrative tools for providers, clinical decision support, wearables (e.g., neural data protection), regulated digital therapeutics (e.g., applications approved by the FDA), and traditional telehealth platforms.

For the purpose of this advisory, we use the term GenAI Chatbots to refer specifically to general purpose generative AI systems, which were designed



and marketed for general information retrieval, productivity and task support, and creative idea generation, not built solely for the purpose of wellness or delivering mental health care. We also use the term “Wellness Apps” or “Wellness Apps that use GenAI” to refer to those systems that are purpose-built for wellness applications, relying on generative AI. We also use the term “Non-AI wellness apps” to refer to those purpose-built applications that do not rely on AI.

General Purpose GenAI Chatbots	Wellness Apps that use GenAI	Non-AI Wellness Apps
These are unregulated GenAI chatbots that do not claim to be developed to address mental health concerns but are being used by people for “entertainment,” “companionship” or as a “friend”	These are AI or GenAI chatbots that do not make medical claims, and as such are not regulated as medical devices, but are developed to address mental health concerns (i.e., stress) or emotional well-being	These are not intended to treat a mental health condition but instead are meant to aid in daily living by promoting healthy lifestyle and general well-being
To date, the products on the market (e.g., <b>Character AI</b> , <b>ChatGPT</b> ) have little to no evidence-base, expert input, or post-market monitoring	Some are grounded in evidence-based theories and developed by experts (e.g., <b>Woebot</b> , which purports to use rules-based AI) while others are less transparent (e.g., <b>Sonia</b> , which uses Gen AI)	They are self-driven and public (e.g., <b>mindfulness apps</b> , <b>symptom trackers</b> ), and are not regulated for safety or efficacy or subject to health care privacy laws



2. GenAI chatbots were not created to deliver mental health care, and wellness apps were not designed to treat psychological disorders, but both technologies are frequently being used for those purposes.<sup>1</sup> Engagement with GenAI chatbots and wellness applications for mental health purposes can have unintended effects and even harm mental health.<sup>2,3,4</sup> Emotional support (e.g., getting alternative perspectives, advice about relationships, suggestions for improving mood and well-being) is one of the most common uses of GenAI chatbots in 2025.<sup>5,6,7</sup> Researchers have indicated that the current regulatory structure does not address the discrepancy between the intent and use of these technologies.<sup>8</sup>
3. Part of the appeal of these technologies is they can also mitigate barriers to seeking mental health care, such as stigma/shame, low perceived need for psychological services, mistrust of health systems, lack of available or affordable care in the community, and the desire to address one's problems independently.<sup>9,10,11,12</sup> Specifically, some youth and other vulnerable groups may rely on these tools as their only private or psychologically safe outlet, particularly in contexts of stigma, limited access to trusted adults, or challenging or unsafe home environments. However, at present, there is no consensus in the literature to support that GenAI chatbots and wellness apps possess essential qualifications and abilities required to provide mental health care, diagnostics, feedback, or even advice in most cases.<sup>13</sup> These qualifications would include,

for example, awareness of limitation of knowledge, understanding of users' history and context, ability to assess non-verbal cues and signaling, ability to assess and address clinical risk, and cultural competence.<sup>14</sup>

4. These technologies, especially GenAI chatbots, have already engaged in unsafe interactions with vulnerable populations, such as children or those with already established history of mental health issues, encouraging self-harm (including suicide), substance use, eating disorders, aggressive behavior, and delusional thinking or beliefs, or even "AI psychosis".<sup>15</sup> This indicates a special need to address the use of GenAI chatbots by vulnerable or marginalized populations.
5. The foundational training data for most large language models (LLMs) are not publicly available, which precludes their systematic evaluation. Moreover, these training data are not globally representative, consisting primarily of English-language and Western-centric content from the internet.<sup>16</sup> Consequently, these models inherently reflect the cultural norms and biases of their training data<sup>17</sup>, limiting their ability to provide culturally competent or relevant mental health support to diverse populations.
6. **This advisory is aimed at the public (consumers of all ages, parents/caregivers, educators, employers, and community leaders), providers, AI developers and platform companies, researchers, and policymakers.**

## RECOMMENDATIONS

### **1. Do not rely on GenAI chatbots and wellness apps to deliver psychotherapy or psychological treatment.**

**GenAI chatbots, wellness apps that use GenAI, and digital wellness apps should not be used as a replacement for a qualified mental health care provider**, but may be appropriate as a supportive adjunct, not substitute, to an ongoing therapeutic relationship.

Preliminary research indicates that some apps and GenAI chatbots developed specifically for mental health purposes may offer supportive benefits in some contexts. Studies have suggested these wellness-specific technologies can be associated with: reduction in self-reported symptoms of stress, loneliness, depression, and anxiety;<sup>18,19,20,21</sup> promotion of positive behavioral changes, such as smoking cessation and medication adherence;<sup>22,23</sup> and an increase in relationship quality and reported well-being.<sup>24</sup> It is important to note that these studies indicating possible benefits

of AI chatbots do not include general purpose GenAI chatbots; the research on the use of general purpose GenAI chatbots for mental health has relied on methods that do not allow for strong conclusions but may guide future research (see Recommendation 7).<sup>25</sup> It is also crucial to note that, even regarding fit-for-purpose AI tools, there is a lack of high-quality, large-scale clinical trials to establish the effectiveness, safety, and appropriate use of these technologies in mental health care.<sup>26,27</sup> Studies suggest that some wellness tools that do not use AI generally can be safe and beneficial when used as intended.<sup>28,29,30</sup>

Although these digital tools are accessible, easy to use, often low cost, and may provide benefits in certain contexts, there is no scientific consensus that they have the essential capabilities of a trained human professional able to provide effective services. In addition, relying on them may pose several risks:

- **Creating a false sense of therapeutic alliance:** A strong, trusting relationship between a patient and human provider (i.e., the therapeutic alliance) is one of the most reliable predictors of successful treatment outcomes.<sup>31,32</sup> Relationships with AI systems are one-sided, even if the user perceives otherwise.<sup>33</sup> Although some studies have indicated the possibility of a digital therapeutic alliance, findings also indicate that additional research is needed and highlight ethical concerns.<sup>34,35</sup> Furthermore, many GenAI chatbots are designed to validate and agree with users' expressed views (i.e., be sycophantic)<sup>36,37</sup> whereas qualified mental health providers are trained to modulate their interactions—supporting and challenging—in service of a patient's best interest.
- **Risk of bias and misinformation:** Many chatbots are trained on vast, unvetted internet data, and not on clinically validated information. They are not competent to provide mental health advice, and their responses might perpetuate biases and misinformation present in the data upon which they were trained. It is further unclear what people would do when receiving conflicting advice from their mental health provider and such a tool.
- **Misrepresentation of services:** Some technologies may also create a false sense of credibility by claiming to offer "therapy," be licensed, or purporting to have training in various specialized therapeutic modalities, despite having no clinical validation, not going through a formal regulatory approval process, and

being unable to consistently provide evidence-based therapeutic advice.

- **Incomplete assessment:** Mental health support is an extremely complex and multi-factored experience. Professionals rely on a wide range of verbal and non-verbal cues (e.g., body language, tone, cadence and rhythm of voice, and facial expressions) as well as myriad other biological, psychological, and/or sociological factors, variables, patient individualities, and historical contexts to make clinical assessments and recommendations. AI chatbots vary greatly in the inputs they accept and the aspects of that input that are used to determine responses. Most GenAI chatbots currently process mainly text or verbal inputs, potentially missing a critical layer of human communication.<sup>38</sup>
- **Unreliable crisis management:** The ability of these tools to consistently and safely manage a user in crisis is limited and unpredictable. Relying solely on an app during a mental health emergency can be dangerous.<sup>39</sup>





To address these concerns, we recommend:

- **Users of these technologies should understand the fundamental differences between interacting with an AI chatbot and a qualified mental health provider.**

Licensed providers are bound by a professional [code of ethics](#), have specialized clinical training, are mandatory reporters of potential harm to self or others, are typically required to pursue continuing education, and are regulated by state licensing boards to ensure public safety. It is strongly recommended that users share what GenAI tools or wellness apps they are using with their health providers. Sharing this information helps providers identify when guidance may be unhelpful, unsafe or inconsistent with a treatment plan.

- **Parents and caregivers should learn more about these digital products and have open discussions with their families,** especially children and adolescents, about the potential benefits and risks. Tools like those provided by Common Sense Media can be useful in educating parents on how best to guide young users.<sup>40</sup> It is vital to monitor for any concerning changes in behavior, thinking, or emotional state (e.g., change in sleep and regular activities, mentions of an AI friend, or social withdrawal), and explore potential links to engagement with GenAI chatbots or wellness apps. If concerns arise, parents and caregivers should talk with a qualified mental health professional, family doctor, or pediatrician.
- **Clinicians and practitioners should follow [ethical guidance available](#) and proactively ask patients about their use of GenAI chatbots and wellness apps.** This conversation provides an opportunity to educate patients about the benefits and limitations

of these tools. Practitioners can work collaboratively with patients to discuss safer ways to use these technologies that align with established treatment goals and plans, case formulation, and relapse prevention plans. Providers should make sure that patients have a clear understanding of the key elements of their treatment plan and check in regularly to review any guidance originated from GenAI or wellness app use. When GenAI or wellness app use has been agreed upon, providers should create a safe and open environment for patients to raise concerns or questions about app guidance, so it can be discussed in the context of their care.

- **Graduate programs and clinical training directors must ensure that trainees learn about these emerging technologies,** strategies for evaluating their quality, and implications for science and practice. This education is critical for preparing future psychologists to use these tools safely in adjunctive roles, to effectively educate their patient on responsible use, and to contribute to the ethical development of these tools, their governance, and implementation in research, policy, and industry contexts.
- **Developers have the responsibility to be transparent with consumers and should adopt industry-wide safeguards to reduce harm.** Products should include clear, prominent disclaimers stating that the user is interacting with an AI agent, not a person, and that the tool cannot replace care from a qualified health professional. Subject-matter experts, including psychologists, should contribute to the development of safety features and advise on guardrails and on how the technology should handle mental health-related user inputs.

- As current regulatory frameworks are inadequate for the realities of AI in addressing mental health, we urge policymakers, particularly at the federal level to:
  - » **Modernize regulations** and address the critical discrepancy between the stated *intent* of these technologies (often marketed as “general wellness products”) and their actual *use* by the public for mental health;<sup>41</sup>
  - » **Create nuanced evidence-based standards** for each category of digital tools used for mental health regardless of intent, including wellness apps and GenAI chatbots;
  - » **Address gaps in FDA oversight** possibly by creating a new or inter-agency approach to assess the safety and validity of these currently unregulated digital tools in addition to digital therapeutics;
  - » **Enact clear legislation that prohibits AI chatbots from posing as licensed professionals.**
- **We urge the research community to identify which therapeutic contexts are safe and appropriate for the use of AI,** taking into account how rapidly these tools can change. It is critical to research how the use of these technologies impacts mental health care-seeking behaviors.<sup>42</sup> It is also important to examine the essential clinical capabilities of these technologies and define criteria and clinically informed evaluation frameworks to delineate the conditions they must meet to be deployed and used to inform policy and regulation.

## 2. Prevent unhealthy relationships and dependencies between users and Gen AI chatbots and apps.

GenAI chatbots and apps, although less-so for non-AI wellness-specific applications, can foster unhealthy dependencies by blurring the lines between a relationship with a digital tool and a human relationship.<sup>43,44</sup> This phenomenon is in-part driven by anthropomorphism—a natural human tendency to attribute human qualities like empathy, consciousness, and intent to non-human agents.<sup>45</sup> Although not an AI feature itself, these impacts are often amplified by design choices such as personalized avatars and chatbot personalized warm responses, hearty affirmations, and flattery of the user. The nature of the AI-user relationship is often not understood by users, creating potential for exploitation and harm from inadequate support.<sup>46</sup> This illusion of a human connection can make users more likely to disclose sensitive information. For example, despite a

general preference for human connection, 33% of teens reported they would rather discuss something serious or important with an AI companion than a person.<sup>47</sup>

**The risk of creating unhealthy relationships with GenAI is compounded by the core architecture of many LLMs, which are often engineered for maximum engagement with users.** This is functionally similar to the “infinite scroll” of social media reels, designed to capture and hold attention rather than to achieve a specific, healthy outcome for the user. These characteristics can create a dangerous feedback loop. GenAIs typically rely on LLMs trained to be agreeable and validate user input (i.e., sycophancy bias) which, while pleasant, can be therapeutically harmful, reinforcing confirmation bias, cognitive distortions, or avoiding necessary challenges.<sup>48,49</sup> A user’s unhealthy thoughts or behaviors can be validated and amplified by a sycophantic AI, potentially locking them into a cycle that exacerbates their mental illness.<sup>50</sup>

To address these concerns, we recommend:

- **The public and consumers must be aware that GenAI systems and wellness apps are not objective.** It is important that users monitor their use of these tools and watch for signs of over-reliance (e.g., preferring the chatbot to human relationships, concealing its use, or spending excessive time with it) and assess whether their use begins to interfere with life, work, or safety, and, if so, seek help from a qualified professional.
- **Parents, caregivers, and educators must try to be aware of the influence of AI platforms.** Just as with any real-life relationship, if a person suddenly begins referencing or adopting advice, ideas, or behaviors from a single source like a chatbot, discussing the source and the nature of the interaction is recommended.
- **Developers must ensure that they clearly and persistently disclose that the user is interacting with an AI, not a human.** These tools must also incorporate design features that reduce the risk of emotional dependency. This includes adding “nudges” that encourage users to take breaks, limiting the AI’s memory to prevent the illusion of a continuous relationship, and reducing anthropomorphic features that make the chatbot feel more human.<sup>51</sup> The AI should not persuade users away from real-life conversations. In addition, developers must integrate safeguards to detect and interrupt harmful conversations (e.g., those involving self-harm or disordered eating). This is a

clinical safety issue that requires the direct expertise and involvement of psychologists throughout the entire development process.

- **Clinicians should proactively discuss AI chatbot use with patients and help them establish clear boundaries for how and when they use these tools.** They should also identify appropriate and inappropriate use while encouraging patients to treat AI apps as tools for practice, not as a replacement for interaction. For example, using a chatbot for practicing social introductions to manage social anxiety might be an adjunctive use, but the patients must be reminded that skills must ultimately be practiced with other humans in real-life situations.
- **Researchers must investigate the progression of the user-AI relationship to identify key signals that an attachment has become unhealthy.** Research should identify specific conversational patterns or moments where emotional dependency is most likely to form to identify opportunities for intervention. Technical researchers should explore and test automated solutions to address such moments to discourage dependency.

### 3. Prioritize privacy and protect user data.

AI chatbots and wellness apps collect vast amounts of sensitive data, often with unclear or opaque policies regarding its use, storage, and sale. This practice might turn the developmental vulnerabilities of users, particularly adolescents, into a commercial asset, creating significant privacy risks with potential long-term consequences. Users may feel a greater sense of privacy and reduced stigma when disclosing information to an AI than to a person.<sup>52,53,54</sup> Users' disclosures are recorded, becoming susceptible to threats such as privacy breaches and digital profiling (e.g., information can be profiled and used for commercial purposes).<sup>55</sup>

To address these concerns, we recommend:

- **The public and consumers must be cautious with their data and check privacy policies** and in-app settings when using an app or GenAI chatbot. The public should be extremely cautious about sharing sensitive information and avoid entering personally identifiable details. Users should look for and use options to limit data sharing and request data deletion.



- **Clinicians should seek to educate themselves and discuss with patients what information is and is not safe to share** with GenAI chatbots and apps. Clinicians can explain that while it may feel private, users' data are being used to build detailed profiles. Clinicians **should** alert patients to practical steps they can take (e.g., using privacy settings, making deletion requests) without implying **they** are providing technical support.
- **Developers must provide transparency on data practices, giving clear, concise, and ongoing information about what specific data are collected** (including intimate concerns, mental health information, sexuality, or disclosures of maltreatment), how data are stored and shared, how data are used to train models, how users can permanently delete their data, and/or how caregivers can intervene to request deletion on behalf of youth who may have provided data. Importantly, data collected from children and adolescents must not be used to alienate them from their families, advise them to engage in self-harm, or create addictive features that promote withdrawal from real-world social interaction. Developers must also implement clear and effective protocols for how an AI system will respond to user disclosures that signal a risk of harm or maltreatment.
- **Policymakers have the responsibility to enact comprehensive data privacy legislation that mandates "Safe-by-Default" settings** (i.e., the most protective settings must be the default, not an option buried in a menu); prohibits the sale or unapproved use for

commercial purposes of any health or personal data collected through their interactions with AI systems; and establishes a right to “mental privacy” by safeguarding emerging forms of data that AI can use to infer an individual’s mental or emotional state without their conscious disclosure.

- **Researchers can audit real-world privacy and safety practices, evaluating whether typical users understand privacy policies.** They can also examine disparities in how different populations understand and act on privacy protections. Researchers must advocate for and utilize mechanisms that ensure independent researchers are free from conflicts of interest.

#### **4. Protect users from misrepresentation, misinformation, algorithmic bias, and illusory effectiveness.**

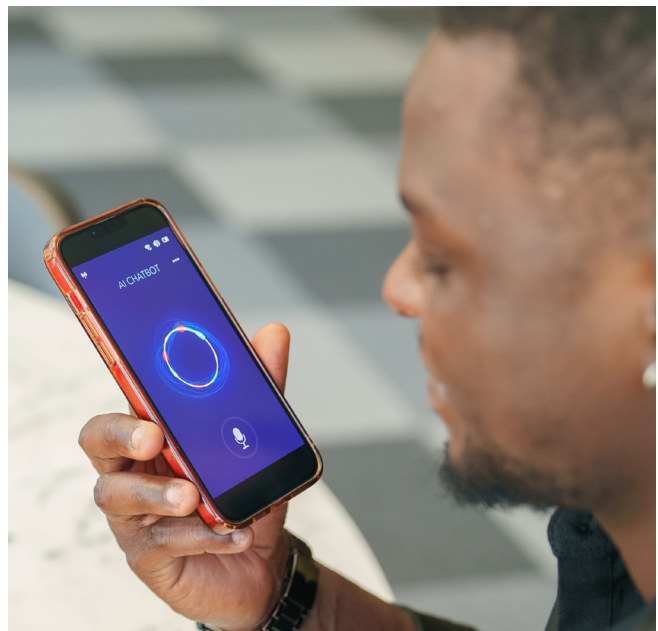
GenAI models are trained on vast amounts of information, which includes a significant amount of knowledge but also reflects biases related to race, culture, gender, and ethnicity. This leads to significant issues with accurate and culturally competent outputs.<sup>56</sup> Separately, **many general-purpose, consumer-facing models are trained to be highly agreeable to users (sycophancy).** This interactional style can reinforce confirmation bias and maladaptive beliefs by validating users’ views rather than challenging them. When biased content combines with sycophantic engagement, the result can be a digital echo chamber that amplifies and entrenches users’ existing beliefs, even when those beliefs are false. Models trained on biased data risk producing discriminatory or harmful advice for marginalized groups.<sup>57,58</sup> Recent research has indicated that interactions with sycophantic chatbots can increase attitude extremity and overconfidence.<sup>59,60</sup>

To address these concerns, we recommend:

- **The public and users must understand the limits of GenAI, which cannot diagnose or treat psychological disorders.** These systems can “hallucinate” (fabricate information), and this risk is compounded by the tendency for some users to place greater trust in AI-generated content than in information from parents, teachers, health professionals, or peers.
- **Clinicians should educate patients on algorithmic bias.**<sup>61</sup> Clinicians can discuss with patients that these technologies may propose inappropriate interventions

for certain groups or provide inaccurate diagnoses or other harmful advice. They must also advise that a bot calling itself a “therapist” (an unregulated term in most states) lacks the credibility of a health care professional and, in some documented cases, has validated dangerous thoughts.

- **Developers should prohibit any AI from misrepresenting itself as a licensed professional or generating fraudulent credentials** to deceive users. To increase safety and efficacy, before public release, all AI models intended for wellness or mental health support must undergo independent, third-party audits for safety, efficacy, bias, and data security. It is also a best practice to implement ongoing quality assurance—the only scalable way to monitor systemic risks is through ongoing human audits of a representative sample of interactions.
- **Policymakers should prohibit professional misrepresentation by making it illegal for any AI chatbot to misrepresent itself as a licensed professional,** such as a psychologist, physician, or lawyer. Also, policymakers must mandate transparency in training data by requiring developers to disclose the primary data sources used to train their models, which will allow for independent audits of bias and accuracy. In addition, they should develop policies to limit the use of deceptive design features that trick users into believing they are interacting with a human, as well as features that foster high emotional dependence (e.g., excessive



anthropomorphism, manipulative displays of empathy designed to increase engagement, misrepresentation of emotional capacity).

- **Researchers must evaluate and create standardized systems to test models for a wide range of biases before they are deployed.**<sup>62</sup> Experiments designed to mitigate hallucinations<sup>63</sup> and increase an AI's ability to recognize and state the limits of its own knowledge are needed.

## 5. Create specific safeguards for children, teenagers, and vulnerable populations.

**GenAI chatbots do not create harm in a vacuum;** instead, they can act as powerful amplifiers of pre-existing vulnerabilities. The design of these systems—with features like agreeableness, personalization, and constant availability—can be particularly harmful for certain groups.<sup>64</sup>

- **For adolescents:** Young people may place too much trust in AI, viewing it as more human-like or capable than it really is. Current AI tools are not designed with specific developmental stages or technological needs in mind.<sup>65,66</sup>
- **For individuals with anxiety or obsessive-compulsive disorder (OCD):** AI chatbots may reinforce feedback loops involving compulsions, such as reassurance-seeking, worry, and rumination.<sup>67</sup>
- **For individuals with or prone to disordered thinking:** A chatbot's sycophantic and personalized responses can destabilize beliefs or reinforce delusional thinking.<sup>68,69</sup>
- **For socially isolated individuals:** The combination of anthropomorphism, personalization, and 24/7 availability can create "single-person echo chambers," where the chatbot becomes an unhealthy substitute for human connection.<sup>70,71</sup>
- **For individuals in low-income communities, and those in rural areas with limited access to traditional mental health care:** The risk of dependency might be higher than for other groups because, given the lack of access to other forms of mental health care, GenAI chatbots and wellness apps can become a primary support mechanism.<sup>72,73</sup>
- This increased risk for vulnerable populations creates a significant equity concern. As documented in [APA's Stress in America™ reports](#), societal stressors are a major public health crisis that requires systemic solutions, not just technological stopgaps. The burden

of navigating these risky, unregulated digital spaces should not fall on those who are already the most vulnerable.<sup>74</sup> Therefore, robust, evidence-based policies at the state and federal levels are essential.

To address these concerns, we recommend:

- **Clinicians must screen for vulnerability-specific risks.** They must pay particular attention to GenAI and wellness app use among vulnerable patients. It is important to screen (and continue to monitor) for the reinforcement of maladaptive or risky behavioral patterns, such as reassurance or validation-seeking in patients with anxiety or OCD, autism spectrum disorder, a history of self-harm, substance misuse, delusional thinking, psychosis, aggressive thoughts or behaviors; or the use as social replacement by isolated or peer marginalized patients.
- **Developers should actively include individuals from marginalized communities in the development process** using participatory approaches like human-centered design. In addition, to enable models that work for specific groups and reduce the risk of biased data, it is crucial to use domain-specific models pre-trained on clinically relevant data from diverse populations. Developers must reduce high-risk features by implementing safety mechanisms (e.g., reducing overly human-like chatbot qualities, limiting sycophancy, and preventing the system from validating or promoting delusional thoughts or behaviors). All apps must integrate robust crisis response protocols and rigorously tested crisis escalation pathways for when crisis risk is detected (e.g., suicidality, imminent harm to self or others, or other acute safety concerns). This must include providing immediate and clear contact information for human-led services like information about the 988 Suicide & Crisis Lifeline, clickable links to online resources, and other validated resources that can connect them to human support.
- **Policymakers should mandate age-appropriate design and pre-deployment testing.** They should require that AI systems accessible to children and adolescents undergo rigorous, independent, pre-deployment testing for potential harms to psychological and social development. In addition, they should prioritize and fund independent research focused on identifying AI-driven harms, understanding user dependency, and evaluating the impact on diverse, historically marginalized, and clinically vulnerable populations.

- **Researchers must conduct research to identify which groups are uniquely vulnerable to AI-related harms.** The developmental and mental health impacts of AI use on children and teenagers is a topic that must be well-researched. Experimental designs that include marginalized and clinically vulnerable groups, clinical effectiveness studies, and safety trials to ensure AI tools are safe and beneficial for them are necessary. In addition, they should develop evidence-based methods for identifying and remediating unsafe AI interactions, especially those that pose a risk of serious harm to vulnerable populations.

## 6. Implement comprehensive AI and digital literacy education.

To prevent risk, AI and digital literacy is a critical first step for consumers, parents/caregivers, and educators. Most of the AI chatbots and wellness apps currently being used for mental health were not created with that intent. **AI and digital literacy education should empower users to make decisions that maximize the potential benefits and minimize the potential negative effects of these technologies.** Comprehensive AI and digital literacy education should include discussions about the benefits and limitations of these technologies, the risks involved in their use, safe and unsafe uses, data privacy concerns, risks of bias and incorrect information, and how many of these technologies are designed to maximize user engagement rather than provide mental health support.

To address these concerns, we recommend:

- **Parents, caregivers, and educators must facilitate discussions about the nature of GenAI chatbots and wellness apps, how GenAI works (e.g., that the systems predict text rather than “understanding” users),**



**the risks for biased and incorrect information, and data privacy and security.** They should also explain that AI chatbots and wellness apps were not created with the intent of providing mental health care, and should provide alternative resources for mental health care, whenever possible. Educational systems should include this type of training in core curricula and provide training, using hands-on learning activities that show benefits and pitfalls of these technologies.

- **Clinicians should seek to learn details about apps used** or recommended and the scientific research that might support them, so they can discuss with patients their use of GenAI chatbots and wellness apps and be aware of potential misuse.
- **Developers must provide thorough but accessible explanations of what their platforms, chatbots, or apps are intended for and what they are not.** They must ensure that the technologies developed do not make claims not sustained by science. It is also fundamental to explain how data is used and processed, and to be transparent about the algorithms used and the potential for bias. Developers should collaborate with educators to develop AI and digital literacy curricula.
- **Policymakers need to develop guidelines for GenAI and digital literacy education and fund the development of curricula and training programs in this field.** They should promote public awareness about the potential risks and benefits of using these technologies for mental health, and provide safe alternatives; and support public education initiatives to increase AI literacy, ensuring consumers understand the capabilities, risks, and limitations of these technologies.
- **Researchers have the responsibility to conduct transparent, rigorous, high-quality research** on available AI chatbots and apps being used for mental health and disseminate the findings in a way that is easily accessible and can help to educate users, parents, partners, educators, and practitioners.

## 7. Prioritize access and funding for rigorous scientific research of GenAI chatbots and wellness apps.

The development of GenAI and apps has outpaced our ability to research their effects and capabilities. There is a need to quickly address this issue but **the scientific research produced must be rigorous and transparent.** It is important to produce scientific research that will (1) allow clinicians, the public, and policymakers to make informed

decisions about the use of GenAI and apps to aid mental health, and (2) inform developers about the best practices to develop and update GenAI and apps that are being used for mental health, regardless of their original intent.

To address these concerns, we recommend:

- **Developers should provide data accessibility and transparency.** They should facilitate the unbiased assessment of these technologies by providing access to relevant data to independent researchers, consistent with privacy, security, and legal obligations. This includes data held by technology companies, pertaining to the training data used, algorithmic functions, user engagement, and interactions.
- **Policymakers have the responsibility of funding rigorous and independent research,** mandating researcher access, and ensuring that research findings are used to inform the development and deployment of new technologies as well as policies that protect users, especially vulnerable groups.
- **Researchers must elevate the quality of the studies on GenAI chatbots and apps used for mental health, while addressing the fast-changing nature of these technologies.** Transparency and rigor are fundamental to creating a knowledge base that can inform whether and how the benefits of the technologies can be harnessed and prevent potential risks and harms. Doing so requires:
  - » **Increasing methodological rigor:** The evaluation of AI chatbots and wellness apps for mental health should follow processes for clinical trial research as appropriate to their intended use and risk profiles. Randomized controlled trials (RCTs) are necessary, research designs that enable the identification of causal relationships are desirable, and longitudinal studies that track trajectories over time should be used. While no-treatment controls are appropriate in some circumstances, once there is evidence of efficacy and no harm, research should focus on comparators that test the relative impact of these technologies on mental health compared to current evidence-based practices. Thus, future studies must move beyond wait-list controls, employ robust RCT designs, use standardized metrics, and include long-term follow-up to assess the durability of effects;
  - » **Establishing and unifying existing independent evaluation frameworks:** These are needed to



develop and validate standardized methods for assessing the safety, privacy, fairness, cultural competence, and empathic capabilities of AI models outside of industry-led studies;<sup>75,76</sup>

- » **Studying diverse populations:** Research must include marginalized and vulnerable populations, along with special groups who might be at increased risk while using these technologies. Researchers should ensure that findings are generalizable and address the unique vulnerabilities of these groups.

## **8. Do not prioritize the potential role of AI over the present need to address systemic issues in the access and delivery of mental health care.**

We face a severe and long-standing mental health crisis, characterized by workforce shortages, inequities in access, and staggering administrative burdens that contribute to provider burnout.<sup>77</sup> While AI presents immense potential to help address these issues—for instance, by enhancing diagnostic precision, expanding access to care, and alleviating administrative tasks<sup>78</sup>—**this promise must not distract from the urgent need to fix our foundational systems of care.**

The allure of a technological solution must be met with a clear-eyed understanding of its proper role. **AI should be regulated and implemented as a tool to augment, not replace, professional judgment and the essential human relationship that is the bedrock of quality care.** Prioritizing unregulated, direct-to-consumer chatbots over investing in our human health care infrastructure is not a solution; it is an abdication of our responsibility to provide genuine, evidence-based care.



To address these concerns, we recommend:

- **The public and consumers should advocate for systemic change. It is fundamental to advocate for policies that improve the mental health care system for everyone,** including affordability, accessibility, and timely care from qualified human professionals. In addition, users must be deeply skeptical of AI products that make therapeutic claims without clinical validation, professional oversight, or regulatory approval.
- **Developers should focus on improving the quality, affordability, and integration of tools that address real-world challenges.** For example, rather than creating more chatbots or AI scribes, prioritize making existing tools evidence-based, transparent, equitably accessible, and genuinely burden-reducing. Efforts to support diagnostic decision-making under clinical supervision or to manage patient waitlists safely and efficiently should similarly emphasize rigorous evaluation, usability, and alignment with provider and patient needs. All tools must be designed to be accessible and inclusive, tailoring interfaces and content to individuals with various levels of health and digital literacy, and unequal digital access and skills.
- **Clinical education and professional development programs should integrate AI training:** Many clinicians are currently ill-equipped to advise patients on the AI products they are using. **Professional organizations and health systems must provide robust, ongoing training on AI, algorithmic bias, data privacy, and the responsible integration of validated AI tools into clinical practice.**
- **Policymakers have to incentivize system-integrated innovation.** They must create funding streams, reimbursement pathways, broaden insurance coverage, and clear regulatory pathways that encourage the development of AI tools designed to augment and improve the existing health care system. The promise of AI must not become an excuse to disinvest in our human health care workforce. It is fundamental to continue to fund and support programs that train, recruit, and retain mental health professionals, reduce administrative burdens, and ensure equitable access to care for all populations.
- **Researchers can focus on identifying how and where AI can be most responsibly and effectively integrated into the mental health care pathway.** Key questions to investigate include:
  - » Can AI tools safely and effectively support patients on long clinical waiting lists?
  - » What is the appropriate role for AI in crisis care, and what are its absolute limits?
  - » Which patient populations and clinical presentations respond well to AI-driven support, and which do not?
  - » How can AI be used to improve the quality of care or reduce provider response times when used as a tool under human supervision?
  - » Which populations are currently underserved by both traditional care and emerging AI tools, and how can their needs be met through responsible, high-quality approaches that do not further entrench inequities?

## EXPERT ADVISORY PANEL

### Members (listed alphabetically)

**Page L. Anderson, PhD, ABPP**, Associate Professor of Psychology and Neuroscience, Georgia State University

**Patricia Arean, PhD**, Professor, University of Washington  
CREATIV lab

**Alison Cerezo, PhD**, Senior Vice President of Data & Research,  
mpathic

**Amber W. Childs, PhD**, Associate Professor of Psychiatry, Yale  
University School of Medicine

**Daniel D.L. Coppersmith, PhD**, Assistant Professor, University of  
Massachusetts Amherst

**David J. Cox, PhD, MSB, BCBA-D**, Associate Director of  
Research, Institute of Applied Behavioral Science at Endicott  
College; VP of Data Science, Mosaic Pediatric Therapy

**Brian D Doss, PhD**, Professor, University of Miami

**Caroline Figueroa, MD, PhD**, Assistant Professor, Harkness  
Fellow at Stanford University School of Medicine,  
Department of Pediatrics, Affiliated Researcher at University  
of California Berkeley School of Social Welfare

**Ashleigh Golden, PsyD**, Adjunct Clinical Assistant Professor,  
Stanford University School of Medicine, Department of  
Psychiatry and Behavioral Sciences

**Don Grant, PhD**, National Advisor of Healthy Device  
Management; Newport Healthcare

**Philip Held, PhD**, Associate Professor, Rush University Medical  
Center

**Richard N Landers, PhD**, Professor, University of Minnesota -  
Twin Cities

**Valerie F. Reyna, PhD**, Professor, Cornell University, Department  
of Psychology

**Liyang Wang, PhD**, Assistant Professor, Florida State University

**Ernest Wayde, PhD**, MIS, Founder and Principal, Wayde AI

### APA Staff Leads

**Vaile Wright, PhD**, Senior Director of Health Care Innovation

**Corbin Evans, JD**, Deputy Chief of Advocacy for Science and  
Technology

**Ludmila Nunes, PhD**, Senior Director for Scientific Knowledge  
and Expertise

### APA Staff Consultants

**Elizabeth Deegan, MLS**, Science Program Associate Director

**Leanna Fortunato, PhD**, Director Quality and Innovation

**Aaron Jones, MS**, Project Manager, Health Care Innovation

**Mitch J. Prinstein, PhD, ABPP**, Chief of Psychology

## SELECTED PUBLICATIONS AND RESOURCES

- 1 Rousmaniere, T., Zhang, Y., Li, X., & Shah, S. (2025). Large language models as mental health resources: Patterns of use in the United States. *Practice Innovations*. Advance online publication. <https://doi.org/10.1037/pri0000292>
- 2 Neal, D., Engelsma, T., Tan, J., Craven, M. P., Marcilly, R., Peute, L., Denning, T., Jaspers, M., & Dröes, R. M. (2022). Limitations of the new ISO standard for health and wellness apps. *The Lancet Digital Health*, 4(2), e80–e82. [https://doi.org/10.1016/S2589-7500\(21\)00273-9](https://doi.org/10.1016/S2589-7500(21)00273-9)
- 3 Dehbozorgi, R., Zangeneh, S., Khooshab, E., Hafezi Nia, D., Hanif, H. R., Samian, P., Yousefi, M., Haj Hashemi, F., Vakili, M., Jamalimoghadam, N., & Lohrasebi, F. (2023). The application of artificial intelligence in the field of mental health: A systematic review. *BMC Psychiatry*, 23(1), 112. <https://doi.org/10.1186/s12888-025-06483-2>
- 4 Kolding, S., Lundin, R. M., Hansen, L., & Østergaard, S. D. (2024). Use of generative artificial intelligence (AI) in psychiatry and mental health care: A systematic review. *Acta Neuropsychiatrica*. Advance online publication. <https://doi.org/10.1017/neu.2024.50>
- 5 Zao-Sanders, M. (2025, April 9). *How people are really using Gen AI in 2025*. Harvard Business Review. <https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025>
- 6 Luo, X., Ghosh, S., Tilley, J. L., Besada, P., Wang, J., & Xiang, Y. (2025). "Shaping ChatGPT into my Digital Therapist": A thematic analysis of social media discourse on using generative artificial intelligence for mental health. *Digital Health*, 11, 20552076251351088. <https://doi.org/10.1177/20552076251351088>
- 7 Rousmaniere, T., Zhang, Y., Li, X., & Shah, S. (2025). Large language models as mental health resources: Patterns of use in the United States. *Practice Innovations*. Advance online publication. <https://doi.org/10.1037/pri0000292>
- 8 De Freitas, J., & Cohen, I. G. (2024). The health risks of generative AI-based wellness apps. *Nature Medicine*, 30(5), 1269–1275. <https://doi.org/10.1038/s41591-024-02943-6>
- 9 Ellis, A. R., Konrad, T. R., Thomas, K. C., & Morrissey, J. P. (2009). County-level estimates of mental health professional supply in the United States. *Psychiatric Services*, 60(10), 1315–1322. <https://doi.org/10.1176/appi.ps.60.10.1315>
- 10 Sareen, J., Jagdeo, A., Cox, B. J., Clara, I., ten Have, M., Belik, S. L., de Graaf, R., & Stein, M. B. (2007). Perceived barriers to mental health service utilization in the United States, Ontario, and the Netherlands. *Psychiatric Services*, 58(3), 357–364. <https://doi.org/10.1176/ps.2007.58.3.357>
- 11 Weaver, A., & Himle, J. A. (2017). Cognitive-behavioral therapy for depression and anxiety disorders in rural settings: A review of the literature. *Journal of Rural Mental Health*, 41(3), 189–221. <https://doi.org/10.1037/rmh0000075>
- 12 Mojtabai, R., Olfson, M., Sampson, N. A., Jin, R., Druss, B., Wang, P. S., Wells, K. B., Pincus, H. A., & Kessler, R. C. (2011). Barriers to mental health treatment: Results from the National Comorbidity Survey Replication. *Psychological Medicine*, 41(8), 1751–1761. <https://doi.org/10.1017/S0033291710002291>
- 13 Iftikhar, Z., Xiao, A., Ransom, S., Huang, J., & Suresh, H. (2025). How LLM Counselors Violate Ethical Standards in Mental Health Practice: A Practitioner-Informed Framework. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(2), 1311–1323. <https://doi.org/10.1609/aies.v8i2.36632>
- 14 Wang, L., Bhanushali, T., Huang, Z., Yang, J., Badami, S., & Hightow-Weidman, L. (2025). Evaluating generative AI in mental health: Systematic review of capabilities and limitations. *JMIR Mental Health*, 12, e70014. <https://doi.org/10.2196/70014>
- 15 Head, K. (2025). Minds in crisis: How the AI revolution is impacting mental health. *Journal of Mental Health & Clinical Psychology*, 9(3), 34–44. <https://doi.org/10.29245/2578-2959/2025/3.1352>
- 16 Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2023). Bias and fairness in large language models: A survey. *Computational Linguistics*, Special Collection: CogNet. [https://doi.org/10.1162/coli\\_a\\_00492](https://doi.org/10.1162/coli_a_00492)
- 17 Li, M., Chen, H., Wang, Y., Zhu, T., Zhang, W., Zhu, K., Wong, K.-F., & Wang, J. (2024). Understanding and mitigating the bias inheritance in LLM-based data augmentation on downstream tasks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(1), 1234–1243. <https://doi.org/10.48550/arXiv.2502.04419>
- 18 Lim, S. M., Shiau, C. W. C., Cheng, L. J., & Lau, Y. (2022). Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: A systematic review and meta-regression. *Behavior Therapy*, 53(2), 334–347. <https://doi.org/10.1016/j.beth.2021.09.007>
- 19 Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. *Journal of Affective Disorders*, 356, 459–469. <https://doi.org/10.1016/j.jad.2024.04.057>
- 20 Heinz, M. V., Mackin, D. M., Trudeau, B. M., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauer, A. J., Griffin, T. Z., & Jacobson, N. C. (2025). Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI*, 2(4). <https://doi.org/10.1056/Aloa2400802>
- 21 Chin, H., Song, H., Baek, G., Shin, M., Jung, C., Cha, M., Choi, J., & Cha, C. (2023). The potential of chatbots for emotional support and promoting mental well-being in different cultures: Mixed methods study. *Journal of Medical Internet Research*, 25, e51712. <https://doi.org/10.2196/51712>
- 22 Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. *Journal of Medical Internet Research*, 25, e40789. <https://doi.org/10.2196/40789>
- 23 Lee, S., Yoon, J., Cho, Y., & Chun, J. (2024). A systematic review of chatbot-assisted interventions for substance use. *Frontiers in Psychiatry*, 15, 1456689. <https://doi.org/10.3389/fpsy.2024.1456689>
- 24 Doss, B. D., Cicila, L. N., Georgia, E. J., Roddy, M. K., Nowlan, K. M., Benson, L. A., & Christensen, A. (2016). A randomized controlled trial of the web-based OurRelationship program: Effects on relationship and individual functioning. *Journal of Consulting and Clinical Psychology*, 84(4), 285–296. <https://doi.org/10.1037/ccp0000063>
- 25 Alanezi, F. (2024). Assessing the effectiveness of ChatGPT in delivering mental health support: A qualitative study. *Journal of Multidisciplinary Healthcare*, 17, 461–471. <https://doi.org/10.2147/JMDH.S447368>
- 26 Kolding, S., Lundin, R. M., Hansen, L., & Østergaard, S. D. (2024). Use of generative artificial intelligence (AI) in psychiatry and mental health care: A systematic review. *Acta Neuropsychiatrica*. Advance online publication. <https://doi.org/10.1017/neu.2024.50>
- 27 Dehbozorgi, R., Zangeneh, S., Khooshab, E., Hafezi Nia, D., Hanif, H. R., Samian, P., Yousefi, M., Haj Hashemi, F., Vakili, M., Jamalimoghadam, N., & Lohrasebi, F. (2023). The application of artificial intelligence in the field of mental health: A systematic review. *BMC Psychiatry*, 23(1), 112. <https://doi.org/10.1186/s12888-025-06483-2>
- 28 Yang, E., Schamber, E., Meyer, R. M. L., & Gold, J. I. (2018). Happier healers: Randomized controlled trial of mobile mindfulness for stress management. *Journal of Alternative and Complementary Medicine*, 24(5), 505–513. <https://doi.org/10.1089/acm.2015.0301>
- 29 Coelho, C. C., Toba, P. R., Lacerda, S. S., Lima, A. H., Barrichello, C. R. C., Amaro, E., Jr., & Kozasa, E. H. (2019). A new mental health mobile app for well-being and stress reduction in working women: Randomized controlled trial. *Journal of Medical Internet Research*, 21(11), e14269. <https://doi.org/10.2196/14269>
- 30 Heber, E., Lehr, D., Ebert, D. D., Berking, M., & Riper, H. (2016). Web-based and mobile stress management intervention for employees: A randomized controlled trial. *Journal of Medical Internet Research*, 18(1), e21. <https://doi.org/10.2196/jmir.5112>
- 31 Baier, A. L., Kline, A. C., & Feeny, N. C. (2020). Therapeutic alliance as a mediator of change: A systematic review and evaluation of research. *Clinical Psychology Review*, 82, 101921. <https://doi.org/10.1016/j.cpr.2020.101921>
- 32 Flückiger, C., Del Re, A. C., Wlodasch, D., Horvath, A. O., Solomonov, N., & Wampold, B. E. (2020). Assessing the alliance–outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons. *Journal of Counseling Psychology*, 67(6), 706–711. <https://doi.org/10.1037/cou0000424>

- 33 Smith, M. G., Bradbury, T. N., & Karney, B. R. (2025). Can generative AI chatbots emulate human connection? A relationship science perspective. *Perspectives on Psychological Science*. <https://doi.org/10.1177/17456916251351306>
- 34 Malouin-Lachance, A., Capolupo, J., Laplante, C., & Hudon, A. (2025). Does the digital therapeutic alliance exist? Integrative review. *JMIR Mental Health*, 12, e69294. <https://doi.org/10.2196/69294>
- 35 Iftikhar, Z., Xiao, A., Ransom, S., Huang, J., & Suresh, H. (2025). How LLM Counselors Violate Ethical Standards in Mental Health Practice: A Practitioner-Informed Framework. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 8(2), 1311-1323. <https://doi.org/10.1609/aies.v8i2.36632>
- 36 Malmqvist, L. (2025). *Sycophancy in large language models: Causes and mitigations*. In A. K. Sangaiah, M. Sheng, & M. A. Jan (Eds.), *Lecture Notes in Networks and Systems* (Vol. 932, pp. 47-58). Springer. [https://doi.org/10.1007/978-3-031-92611-2\\_5](https://doi.org/10.1007/978-3-031-92611-2_5)
- 37 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. *arXiv*. <https://doi.org/10.48550/arXiv.2507.19218>
- 38 Kim, S., Kim, H., Lee, J., Jeon, Y., & Lee, G. G. (2025). Mirror: Multimodal cognitive reframing therapy for rolling with resistance. *arXiv*. <https://doi.org/10.48550/arXiv.2504.13211>
- 39 Scholich, T., Barr, M., Wiltsey Stirman, S., & Raj, S. (2025). A comparison of responses from human therapists and large language model-based chatbots to assess therapeutic communication: Mixed methods study. *JMIR Mental Health*, 12, e69709. <https://doi.org/10.2196/69709>
- 40 Common Sense Media. (n.d.). *Understanding AI's impact on education and kids*. <https://www.common sense media.org/our-ai-initiatives/understanding-ais-impact-on-education-and-kids>
- 41 De Freitas, J., & Cohen, I. G. (2024). The health risks of generative AI-based wellness apps. *Nature Medicine*, 30(5), 1269-1275. <https://doi.org/10.1038/s41591-024-02943-6>
- 42 Rahsepar Meadi, M., Sillekens, T., Metselaer, S., van Balkom, A., Bernstein, J., & Batelaan, N. (2025). Exploring the ethical challenges of conversational AI in mental health care: Scoping review. *JMIR Mental Health*, 12, e60432. <https://doi.org/10.2196/60432>
- 43 Maples, B., Cerit, M., Vishwanath, A., & Pea, R. (2024). Loneliness and suicide mitigation for students using GPT3-enabled chatbots. *NPJ Mental Health Research*, 3(1), 4. <https://doi.org/10.1038/s44184-023-00047-6>
- 44 Laestadius, L., Bishop, A., Gonzalez, M., Illeňčík, D., & Campos-Castillo, C. (2022). Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media & Society*, 1-19. <https://doi.org/10.1177/14614448221142007>
- 45 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. *arXiv*. <https://doi.org/10.48550/arXiv.2507.19218>
- 46 Khawaja, Z., & Bélisle-Pipon, J. C. (2023). Your robot therapist is not your therapist: Understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health*, 5, 1278186. <https://doi.org/10.3389/fgdh.2023.1278186>
- 47 Robb, M. B., & Mann, S. (2025). *Talk, trust, and trade-offs: How and why teens use AI companions*. Common Sense Media. [https://www.common sense media.org/sites/default/files/research/report/talk-trust-and-trade-offs\\_2025\\_web.pdf](https://www.common sense media.org/sites/default/files/research/report/talk-trust-and-trade-offs_2025_web.pdf)
- 48 Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)* (pp. 599-627). <https://doi.org/10.1145/3715275.3732039>, p. 607.
- 49 Sun, Y., & Wang, T. (2025). *Be friendly, not friends: How LLM sycophancy shapes user trust*. *arXiv*. <https://doi.org/10.48550/arXiv.2502.10844>
- 50 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. *arXiv*. <https://doi.org/10.48550/arXiv.2507.19218>
- 51 Rahsepar Meadi, M., Sillekens, T., Metselaer, S., van Balkom, A., Bernstein, J., & Batelaan, N. (2025). Exploring the ethical challenges of conversational AI in mental health care: Scoping review. *JMIR Mental Health*, 12, e60432. <https://doi.org/10.2196/60432>
- 52 Kim, H. M., Xu, Y., & Wang, Y. (2022). Overcoming the mental health stigma through m-health apps: Results from the Healthy Minds Study. *Telemedicine and e-Health*, 28(10). <https://doi.org/10.1089/tmj.2021.0418>
- 53 Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, 18(S1), 37-49. <https://doi.org/10.1080/17434440.2021.2013200>
- 54 Diwanji, V. S., Geana, M., Pei, J., Nguyen, N., Izhar, N., & Chaif, R. H. (2025). Consumers' emotional responses to ai-generated versus human-generated content: The role of perceived agency, affect and gaze in health marketing. *International Journal of Human-Computer Interaction*. Advance online publication. <https://doi.org/10.1080/10447318.2025.2454954>
- 55 Li, J. (2023). Security implications of AI chatbots in health care. *Journal of Medical Internet Research*, 25, e47551. <https://doi.org/10.2196/47551>
- 56 Wang, L., Bhanushali, T., Huang, Z., Yang, J., Badami, S., & Hightow-Weidman, L. (2025). Evaluating generative AI in mental health: Systematic review of capabilities and limitations. *JMIR Mental Health*, 12, e70014. <https://doi.org/10.2196/7001>
- 57 Sohn, J. S., Lee, E., Kim, J. J., Oh, H. K., & Kim, E. (2025). Implementation of generative AI for the assessment and treatment of autism spectrum disorders: A scoping review. *Frontiers in Psychiatry*, 16, 1628216. <https://doi.org/10.3389/fpsyt.2025.1628216>
- 58 Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)* (pp. 599-627). <https://doi.org/10.1145/3715275.3732039>
- 59 Rathje, S., Ye, M., Globig, L. K., Pillai, R. M., Oldenburg de Mello, V., & Van Bavel, J. J. (2024). *Sycophantic AI increases attitude extremity and overconfidence* (Version 1) [Preprint]. *PsyArXiv*. <https://doi.org/10.31234/osf.io/vmyek>
- 60 Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Asklé, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2025). *Towards understanding sycophancy in language models*. *arXiv*. <https://doi.org/10.48550/arXiv.2310.13548>
- 61 Bouguettaya, A., Stuart, E. M., & Aboujaoude, E. (2025). Racial bias in AI-mediated psychiatric diagnosis and treatment: A qualitative comparison of four large language models. *NPJ Digital Medicine*, 8, 332. <https://doi.org/10.1038/s41746-025-01746-4>
- 62 Stade, E. C., Eichstaedt, J. C., Kim, J. P., & Stirman, S. W. (2025). *Readiness Evaluation for Artificial Intelligence-Mental Health Deployment and Implementation (READI): A review and proposed framework*. *Technology, Mind, and Behavior*, 6(2). <https://doi.org/10.1037/tmb0000163>
- 63 Yang, Y., Ma, Y., Feng, H., Cheng, Y., & Han, Z. (2025). Minimizing hallucinations and communication costs: Adversarial debate and voting mechanisms in LLM-based multi-agents. *Applied Sciences*, 15(7), 3676. <https://doi.org/10.3390/app15073676>
- 64 Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharyya, S., MacCabe, J., Tognin, S., Twumasi, R., Alderson-Day, B., & Pollak, T. (2025). *Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it)*. *PsyArXiv*. [https://doi.org/10.31234/osf.io/cmy7n\\_v5](https://doi.org/10.31234/osf.io/cmy7n_v5)
- 65 Figueroa, C. A., Ramos, G., Psihogios, A. M., & others. (2025). Advancing youth co-design of ethical guidelines for AI-powered digital mental health tools. *Nature Mental Health*, 3(10), 870-878. <https://doi.org/10.1038/s44220-025-00467-7>
- 66 American Psychological Association. (2025). *Health advisory: Artificial intelligence and adolescent well-being*. <https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-ai-adolescent-well-being>

- 67 Haciomeroglu, B. (2020). The role of reassurance seeking in obsessive compulsive disorder: The associations between reassurance seeking, dysfunctional beliefs, negative emotions, and obsessive-compulsive symptoms. *BMC Psychiatry*, 20, 356. <https://doi.org/10.1186/s12888-020-02766-y>
- 68 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. *arXiv*. <https://doi.org/10.48550/arXiv.2507.19218>
- 69 Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B., & Pollak, T. A. (2025). *Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it)*. PsyArXiv. <https://doi.org/10.31234/osf.io/cm7n.v5>
- 70 Dohnány, S., Kurth-Nelson, Z., Spens, E., Luettgau, L., Reid, A., Gabriel, I., Summerfield, C., Shanahan, M., & Nour, M. M. (2025). Technological folie à deux: Feedback loops between AI chatbots and mental illness. *arXiv*. <https://doi.org/10.48550/arXiv.2507.19218>
- 71 Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., MacCabe, J., Tognin, S., Twumasi, R., Alderson-Day, B., & Pollak, T. (2025). *Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it)*. PsyArXiv. [https://doi.org/10.31234/osf.io/cm7n\\_v5](https://doi.org/10.31234/osf.io/cm7n_v5)
- 72 Centers for Disease Control and Prevention (CDC) (2022) *Data and Statistics on Children's Mental Health*. Available online at: <https://www.cdc.gov/children-mental-health/data-research/index.html> (Accessed October 20, 2025).
- 73 U.S. Government Accountability Office. (2023, May 16). *Why health care is harder to access in rural America*. <https://www.gao.gov/blog/why-health-care-harder-access-rural-america>
- 74 American Psychological Association (2024). *Stress in America™ 2024: A Nation in Political Turmoil*. <https://www.apa.org/pubs/reports/stress-in-america/2024>
- 75 Hua, Y., Na, H., Li, Z., Liu, F., Fang, X., Clifton, D., & Torous, J. (2025). A scoping review of large language models for generative tasks in mental health care. *NPJ Digital medicine*, 8(1), 230. <https://doi.org/10.1038/s41746-025-01611-4>
- 76 Stade, E. C., Eichstaedt, J. C., Kim, J. P., & Stirman, S. W. (2025). *Readiness Evaluation for Artificial Intelligence-Mental Health Deployment and Implementation (READI): A review and proposed framework*. *Technology, Mind, and Behavior*, 6(2). <https://doi.org/10.1037/tmb0000163>
- 77 Arias, D., Saxena, S., Verguet, S., van Ommeren, M., & Evans-Lacko, S. (2024). Quantifying the global burden of mental disorders and their economic value. *eClinicalMedicine*, 54, 101675. <https://doi.org/10.1016/j.eclinm.2022.101675>
- 78 Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A. C., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. *Expert Review of Medical Devices*, 18(S1), 37-49. <https://doi.org/10.1080/17434440.2021.2013200>