Large language models as mental health providers

Lancet Psychiatry 2025

Published Online
September 9, 2025
https://doi.org/10.1016/S2215-0366(25)00269-X

General purpose chatbots are routinely used for personal health questions, including mental health. Converging evidence from independent surveys,[1] observational data from public forums,[2] and media reports show extensive use of large language models (LLMs) by people with anxiety, depression, relationship problems, and in crisis situations. These data sources have methodological limitations, but they suggest LLM chatbots have already progressed from personal coaching into psychotherapeutic intervention. This progression raises the question of where we draw the regulatory line between a so-called online assistant and an online psychotherapist; the answer will have far-reaching implications.

Although LLMs show promise for education, triage, and support, their clinical effectiveness and safety remain insufficiently established and evaluation methods are heterogeneous.[3,4] Dangerous interactions of LLM chatbots with users have already been documented.[5] A 2025 study compared the performance of three leading LLMs with expert suicidologists using a standardised suicide intervention inventory; the LLMs showed an upward bias in judging responses as appropriate and performance varied by model.[6] The study underscores that although models can at times approximate the performance of trained humans on narrow benchmarks, they remain inconsistent. WHO has warned that large, multimodal models are rapidly entering the health-care space and has issued detailed ethics and governance guidance, calling for transparency, rigorous evaluation, and oversight proportionate to risk.

Current LLM development safety measures, such as harm refusal techniques, trust and safety tooling, red-teaming, and content filters, represent progress but remain insufficient without transparent evaluation. OpenAI has acknowledged the challenges in making ChatGPT safe for users who are in what they describe as "a fragile enough mental place" and has stated it does not intend for its products to be used for therapy. This discrepancy is highlighted in a major LLM that displays warnings when users engage with characters bearing professional titles such as therapist or doctor, cautioning users against relying on the characters for professional advice, however there are also reports of LLMs abandoning this practice.

The legal landscape is shifting rapidly, and it is unclear how long the doublespeak of offering therapy but denying responsibility can last. Section 230 of the US Communications Decency Act, which shields platforms (such as Facebook and Reddit) from liability for user-generated content, offers little protection for users when their own generative systems start acting like clinicians. Product liability cases against chatbots are progressing through US courts, meanwhile the US Congress has introduced bills to restrict legal protection for algorithmic amplification and generative artificial intelligence.

LLM developers are concerned that traditional regulatory approaches are not adequate to regulate this fast-moving practice of generative systems acting as clinicians. The speed of artificial intelligence innovation vastly outpaces legislative and regulatory cycles, leaving rules that are costly to implement and outdated on arrival. Collaboration between developers, clinicians, researchers, and policymakers, although not simple, will be essential to address challenges as the field moves from wellness towards health-care tools. LLM developers need such collaborations to improve their clinical performance; meta-analyses show existing digital mental health apps achieve standardised mean differences of only 0·20–0·30,[7] whereas face-to-face psychotherapy averages approximately 0·80.[8] LLMs' immediate availability, dynamically generated dialogue, and context-sensitive responses position them to bridge this efficacy gap, but this will require rigorous research.[9,10]

Clinical trials conducted by independent researchers are needed to rigorously test the safety and efficacy of LLM-based mental-health interventions. This will require the companies that own LLMs to share de-identified interaction and outcomes data with qualified researchers. Such data are also required for observational studies to enable independent real-world characterisation and evaluation of how these tools are used. Because these tools are already being used, practical guidelines for mental-health use of LLMs should be drafted in parallel by developers, clinicians, and researchers, with frequent revision as evidence becomes available.

Proactive engagement among clinical researchers and technology companies offers multiple advantages. It

For the **WHO ethics and governance guidelines** see https://www.who.int/publications/i/item/9789240084759

For more on **LLM development safety measures** see https://www.tamingllms.com/notebooks/safety.html

For **OpenAI's statement** see https://techcrunch.com/2025/06/25/sam-altman-comes-out-swinging-at-the-new-york-times/

For **OpenAI's usage policies** see https://openai.com/policies/usage-policies/

For more on **artificial intelligence character disclaimers** see https://techcrunch.com/2024/12/12/amid-lawsuits-and-criticism-character-ai-announces-new-teen-safety-tools

For more on **missing artificial intelligence disclaimers** see https://www.technologyreview.com/2025/07/21/1120522/ai-companies-have-stopped-warning-you-that-their-chatbots-arent-doctors/

For more on **artificial intelligence chatbot lawsuits** see https://socialmediavictims.org/blog/character-ai-lawsuit-moves-forward

For more on **bills protecting against predatory chatbot practices** see https://sd18.senate.ca.gov/news/california-senate-advances-legislation-protecting-against-predatory-chatbot-practices

enables development of effective interventions that could reduce the global burden of untreated mental disorders, and positions companies as responsible actors working for public benefit. The scale and granularity of LLM-interaction data offer an unprecedented opportunity to understand help-seeking behaviour, treatment effectiveness, and symptom trajectories. Additionally, it positions companies to shape regulatory frameworks collaboratively rather than reactively.

LLMs have entered everyday use for mental health. Developers who embrace transparency and collaborative research can transform the mental health landscape and define the future of digital care for the better.

We declare no competing interests.

During the preparation of this work the first author used ChatGPT and Claude in order to find sources and improve the quality of writing. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

*Tony Rousmaniere, Simon B Goldberg, John Torous
trousmaniere@sentio.org

Department of Marriage and Family Therapy, Sentio University, Torrance, CA 90505, USA (TR); Psychotherapy Division of the American Psychological Association, Washington, Washington DC, USA (SBG); Division of Digital Psychiatry, Beth Israel Deaconess Medical Center, Boston, MA, USA (JT)

1 Rousmaniere T, Zhang Y, Li X, Shah S. Large language models as mental health resources: patterns of use in the United States. *Practice Innovations* 2025; published online July 7. https://doi.org/10.1037/pri0000292.
2 Zao-Sanders M. How people are really using Gen AI. Harvard Business Review, 2024. https://hbr.org/2024/03/how-people-are-really-using-genai (accessed Aug 14, 2025).
3 Hua Y, Na H, Li Z, et al. A scoping review of large language models for generative tasks in mental health care. *NPJ Digit Med* 2025; **8:** 230.
4 Jin Y, Liu J, Li P, et al. The applications of large language models in mental health: scoping review. *J Med Internet Res* 2025; **27:** e69284.
5 Moore J, Grabb D, Agnew W, et al. 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; June 23–26, 2025 (599–627).
6 McBain RK, Cantor JH, Zhang LA, et al. Competency of large language models in evaluating appropriate responses to suicidal ideation: comparative study. *J Med Internet Res* 2025; **27:** e67891.
7 Goldberg SB, Lam SU, Simonsson O, Torous J, Sun S. Mobile phone-based interventions for mental health: a systematic meta-review of 14 meta-analyses of randomized controlled trials. *PLOS Digit Health* 2022; **1:** e0000002.
8 Wampold BE, Imel ZE. The great psychotherapy debate: the evidence for what makes psychotherapy work. London, UK: Routledge; 2015.
9 Obradovich N, Khalsa SS, Khan W, et al. Opportunities and risks of large language models in psychiatry. *NPP Digit Psychiatry Neurosci* 2024; **2:** 8.
10 Stade EC, Stirman SW, Ungar LH, et al. Large language models could change the future of behavioral healthcare: a proposal for responsible development and evaluation. *Npj Ment Health Res* 2024; **3:** 12.