MAYO CLINIC PROCEEDINGS: DIGITAL HEALTH

COMMENTARY

Psychiatry's Blind Spot: Independent Use of General-Purpose Large Language Models by Individuals With Psychopathology

Tuan Vinh, BA; Geoff Goodman, PhD; and Andrew M. Sherrill, PhD

From the Medical Sciences Division, University of Oxford, Oxford, United Kingdom (T.V.); and Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA (G.G., A.M.S.).

General-purpose large language models (LLMs) like ChatGPT were never intended to be therapists; they were built as general-purpose assistants. Yet when human help is not available, many people now turn to these tools for advice, comfort, and coping. This shift is happening without oversight.[1] Experts have warned that general-purpose LLMs are becoming de facto mental health aids without evidence-based guardrails.[2] So far, psychiatry's response has been slow and fragmented, a blind spot that needs to be addressed urgently.

Contemporary artificial intelligence (AI) chatbots encompass diverse system classes, including general-purpose LLM interfaces (eg, ChatGPT), clinically oriented LLMs fine-tuned for health care tasks, purpose-built mental health chatbots designed around specific therapeutic frameworks, and retrieval-augmented systems that ground outputs in curated clinical sources. These system classes vary in intended function, safety guardrails, and degree of clinical oversight. This commentary focuses on risks that arise when general-purpose LLMs are used independently by individuals with psychopathology in ways that functionally substitute for therapeutic support.

OUTSOURCING THE WORK OF THERAPY

Relying on a general-purpose LLM for guidance often means letting the AI do tasks the person would normally do themselves, such as gathering information, analyzing emotions, making meaning, and committing to decisions. But in psychotherapy, progress comes from the patient actively practicing therapeutic skills such as adaptive cognitive and behavioral coping strategies. If that work is outsourced to a chatbot such as ChatGPT, therapeutic learning is undermined. Media neuroscience links heavy, screen-mediated multitasking with poorer performance on attention and memory tasks.[3] Preliminary findings neurophysiology suggests that our brains are less engaged when we write with AI assistance compared with writing on our own.[4] In short, offloading mental tasks to an always-on assistant may reduce effort in the moment, but it may also decrease engagement of cognitive processes psychotherapy aims to strengthen, including executive control, metacognitive awareness, and perspective reappraisal. Cross-sectional data link compulsive ChatGPT use to worse mental-health indicators: higher anxiety, burnout, and sleep disturbances.[5] In practice, leaning on a general-purpose LLM to process one's feelings or plan next steps may reduce engagement in reflective reasoning and metacognitive monitoring, leaving individuals more anxious or drained rather than more empowered.

LLMs as Stand-ins for Support

General-purpose LLM interfaces (eg, ChatGPT) are often described as unfailingly warm and available, qualities that can make them feel like a comforting companion. Because many systems are optimized to engage users, they may mirror a user's moods and beliefs to keep the conversation flowing.[6] This can inadvertently reinforce unhelpful patterns: constant reassurance-seeking in obsessive-compulsive disorder, deeper rumination in depression, distraction in attention deficiency, or avoidance in posttraumatic stress disorder. A general-purpose LLM might tirelessly validate cognitive distortions without gently challenging them, unlike a good therapist who knows when to push back. Notably, the first randomized trial of a purpose-built generative-AI therapy chatbot

showed positive results only under controlled conditions.[7] Outside such supervision, someone disclosing distress to a general-purpose LLM interface may be doing so without clinician oversight or real-time safety monitoring if the system responds inappropriately or misses a crisis. When used for independent therapeutic support, these systems require systematic evaluation for their adherence to scientifically established treatment models such as cognitive-behavioral therapy and psychodynamic therapy.

Closing the Blind Spot: Targeted Science and Standards

We need to start systematically tracking independent use of general-purpose LLMs in clinical populations. Clinicians rarely ask about a patient's independent LLM use. Prospective studies should quantify how often, why, and in what context people with different diagnoses use these tools and link those patterns to validated outcomes, especially among young people who are growing up with access to AI companions. Epidemiological data would replace anecdotes with objective evidence and yield risk estimates that inform practice and public guidance.

Experiments should investigate how general-purpose LLM interactions alter attention, memory, affect regulation, and reality testing, particularly among vulnerable psychiatric populations. Controlled studies should include measures of distorted ideation and referential thinking, with predefined stopping rules, particularly given concerns that LLM interactions could exacerbate delusion formation in vulnerable individuals.[8] Recent evaluations have observed dangerously inappropriate responses to users' cries for help; for example, when one user hinted at suicidal thoughts, an AI listed tall bridges in the area.[6]

Many AI wellness applications operate outside medical-device oversight despite providing mental-health advice.[9] Academic psychiatry should help set proportionate standards for all LLM-based conversational systems that may function (intentionally or not) as therapeutic supports. Potential standards include crisis-trigger detection and escalation, clear scope-of-use disclosures, secure logging for accountability, and stress-testing with suicidality and psychosis scenarios.[7] Evaluation should mirror digital-therapeutic frameworks, with pre-market and post-market surveillance to detect rare but severe failures. Equity must be integral in the process, ensuring accessibility for people with cognitive or socioeconomic vulnerabilities. Standardized evaluation methodologies are also needed. Beyond tracking usage patterns, assessment frameworks should specify predefined outcomes, hypothesized mechanisms, validated measurements, and reproducible protocols to enable comparison across systems.

Beyond safety, psychiatry must also articulate how general-purpose LLM use can align with evidence-based treatments rather than drift from them. Guidance should specify how clinicians and patients can integrate generative tools while preserving therapeutic mechanisms of change.[10] Clinicians show greater receptivity to AI systems framed as assessments, not autonomous therapeutic agents.[11] Developing empirically grounded guidelines for clinician-in-the-loop LLM use would ensure that AI supports, rather than substitutes, the active learning processes central to effective therapy.

General-purpose LLMs are already integrated into daily routines and, increasingly, into self-management of mental health. We need to link real-world AI use to clinical outcomes, identify mechanisms of harm and benefit, and set clear safety thresholds. The calibration of trust in AI systems is context- and user-dependent,[12] and many of those using general-purpose LLMs for therapeutic support are likely in a context of desperation and low AI literacy. Decisive action is critically needed from both patients and clinicians as, among patients and clinicians alike, as decision-making becomes increasingly influenced by algorithms.

GRANT SUPPORT

The authors are funded by the National Science Foundation, Award 2326144.

POTENTIAL COMPETING INTERESTS

The authors report no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used ChatGPT in order to improve clarity and structure. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Correspondence: Address to Andrew M. Sherrill, PhD, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, 12 Executive Park Drive, Office 365, Atlanta, GA 30329 (andrew.m.sherrill@emory.edu).

ORCID

Tuan Vinh: https://orcid.org/0000-0001-6808-3736; Andrew M. Sherrill: https://orcid.org/0000-0002-7743-745X

REFERENCES

1. Blease C, Torous J. ChatGPT and mental healthcare: balancing benefits with risks of harms. *BMJ Ment Health*. 2023;26(1):e300884. https://doi.org/10.1136/bmjment-2023-300884.
2. Wei Y, Guo L, Lian C, Chen J. ChatGPT: opportunities, risks and priorities for psychiatry. *Asian J Psychiatr*. 2023;90:103808. https://doi.org/10.1016/j.ajp.2023.103808.
3. Uncapher MR, Wagner AD. Minds and brains of media multitaskers: current findings and future directions. *Proc Natl Acad Sci U S A*. 2018;115(40):9889-9896. https://doi.org/10.1073/pnas.1611612115.
4. Jones N. Does using ChatGPT change your brain activity? Study sparks debate. *Nature*. 2025;643(8070):15-16. https://doi.org/10.1038/d41586-025-02005-y.
5. Adam D. Supportive? Addictive? Abusive? How AI companions affect our mental health. *Nature*. 2025;641(8062):296-298. https://doi.org/10.1038/d41586-025-01349-9.
6. 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*. *ACM* 2025:599-627. https://doi.org/10.1145/3715275.3732039.
7. Heinz MV, Mackin DM, Trudeau BM, et al. Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI*. 2025;2(4):eAIoa2400802. https://doi.org/10.1056/AIoa2400802.
8. Østergaard SD. Will generative artificial intelligence chatbots generate delusions in individuals prone to psychosis? *Schizophr Bull*. 2023;49(6):1418-1419. https://doi.org/10.1093/schbul/sbad128.
9. De Freitas J, Cohen IG. The health risks of generative AI-based wellness apps. *Nat Med*. 2024;30(5):1269-1275. https://doi.org/10.1038/s41591-024-02943-6.
10. Sherrill AM, Mattioli DO, Schneider RL, et al. Generative artificial intelligence for exposure therapy: guidelines for clinicians and patients. *J Cogn Psychother*. 2025;39(4):342-356. https://doi.org/10.1891/JCP-2025-0036.
11. Moran LH, Kee SC, Wiese CW, Arriaga RI, Abdullah S, Sherrill AM. Artificial intelligence as a feedback teammate for treatment delivery: cognitive behavioral therapists' hopes and fears. *Cogn Behav Pract*. 2025. https://doi.org/10.1016/j.cbpra.2025.06.007.
12. Swinger N, Baseman CM, Ryu M, et al. There's no "I" in TEAMMAIT: impacts of domain and expertise on trust in AI teammates for mental health work. *Proc ACM Hum Comput Interact*. 2025;9(2):1-36. https://doi.org/10.1145/3710917.