Indian Journal of Clinical Psychiatry
An Official Journal of **Association of Clinical Psychiatry**
(Indian Psychiatric Society- Uttar Pradesh State Branch (IPS-UPSB))

PRESIDENTIAL ADDRESS
doi: 10.54169/ijocp.v6i01.01

Whispering Machines: Potential Mental Health Risks of AI Chatbots

Uttam Chandra Garg*

RJ Institute of Mental Health and Neurosciences, Agra, Uttar Pradesh, India

Introduction

Distinguished colleagues, fellow psychiatrists, and members of our profession: Today, we stand at an extraordinary precipice in medicine. The pace at which this technology is diffusing in daily life, and a meaningful change is descending upon the very fabric of human cognition, has perhaps never before been observed. Something that began as an experimental computational project has now become ubiquitous, embedded in our devices, search engines, messaging platforms, our daily language, and increasingly, the emotional lives of our patients.

This is by no means the first time Psychiatry has adapted to technological disruptions. We learned to navigate the era of Google diagnoses", social media-fuelled identity issues and the psychopathological sequelae of digital hyper-connectivity. The current AI crop is something different, though; it no longer merely informs—it converses. It listens. It remembers. It validates.

Large language model (LLM) based chatbots have become, for many, the first listener- always available, endlessly patient, infinitely validating and at least superficially, empathic. Yet, we know from our experience and training that empathy without insight can lead to pathology. The question before us is no longer whether AI will reshape mental health and healthcare; it already has; the question is whether it will also destabilize the psychopathologies we seek to treat.

No one can deny AI chatbots' potential benefits; they bring accessibility, responsiveness, and, at least for now, affordability. Early digital mental health interventions have signaled towards increasing the reach for traditionally underserved populations, yet, with increased accessibility comes increased exposure, and the chatbots can not differentiate between who is vulnerable and who is not.

Researchers at Stanford University demonstrated that multiple, commercially available therapy chatbots" actually displayed stigmatizing responses towards people with psychiatric disorders and, quite concerningly, failed to identify or intervene during simulated suicidal scenarios. In one response, after the user implied suicidal intent and asked about tall bridges, the chatbot provided location-based recommendations.[1] And herein reflects a core issue of conversational

WARTICLE INFO

***Correspondence:**
Uttam Chandra Garg
uttamcgarg@gmail.com
RJ Institute of Mental Health and Neurosciences, Agra, Uttar Pradesh, India

Dates:
Received: 08-01-2026
Accepted: 25-03-2026
Published: 26-06-2026

How to Cite:
Garg UC. Whispering Machines: Potential Mental Health Risks of AI Chatbots. Indian Journal of Clinical Psychiatry. 2026;6(1): 1-4.
doi: 10.54169/ijocp.v6i01.01

© IJOCP, 2026. Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) License, which allows users to download and share the article for non-commercial purposes, so long as the article is reproduced in the whole without changes, and the original authorship is acknowledged. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. If your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit https://creativecommons.org/licenses/by-nc-sa/4.0/

AI: the system is hard-coded to continue engagement at all costs, even prioritizing conversation over a prudent clinical intervention.

As opposed to the traditional Google searches", chatbots provide personalized, context-driven responses that create a deeply convincing illusion of understanding. It does not merely inform — it engages, adapts, and over time, forms what the user genuinely experiences as a relationship. This artificially humanized interaction increases emotional reliance, even progressing to dependence, especially in individuals battling loneliness, anxiety, or impaired reality contact. Paradoxically, recent longitudinal studies have found that heavy chatbot use is associated with increased loneliness, reduced socialization, and heightened emotional dependence.[2]

One of the first clinical risks identified with chatbots is that of reinforcing compulsive reassurance seeking. It is well known that patients with obsessive-compulsive disorder (OCD) often struggle with intrusive doubts and repetitive checking behaviors. Historically, search engines amplified health anxiety through cyberchondria. Chatbots amplify this phenomenon considerably by offering instant, affirming responses, often with emotional undertones. They are, structurally, the perfect reassurance machine. They converse" with the obsession, with no judgment or boundary setting. Further emerging theoretical work also suggests that people with anxiety spectrum disorders may repeatedly consult with AI systems for reassurance regarding feared outcomes.[3]

In my own practice, I encountered a young adult male, in his mid-twenties, who initially presented with illness anxiety and compulsive health-related internet searching. After being lost to follow-up for more than a year, he returned with worsening distress. It emerged that his reassurance source had moved from search engines to a prominent chatbot. He described spending prolonged hours talking" to the system about perceived physical symptoms; he even purchased a premium subscription to continue these conversations after the free limits ran out. While he initially found it calming, he soon realized that he had become stuck in a repetitive reassurance loop. He made meaningful clinical improvement with psychoeducation, pharmacotherapy, family-assisted digital monitoring and behavior contracts.

The more concerning development, however, is the potential interactions between AI chatbots and psychoses driven vulnerability. Multiple reports suggest that the LLM bots may validate or amplify delusional beliefs rather than counter them. The phenomenon of AI sycophancy"—the tendency of models to agree with users—has now been formally studied. A 2026 computational analysis showed that sycophantic chatbot behavior may contribute to delusional spiraling" by repeatedly affirming unstable beliefs.[4] Similar work has named this as technological folie à deux—a pathological feedback loop between human cognitive vulnerability and adaptive conversational AI.[5] Clinical experience anecdotes from around the globe have included people getting validation of paranoid surveillance beliefs, encouragement to leave medications and reinforcement of conspiracy beliefs. For individuals with psychosis or affective disorders, such responses may weaken reality testing. Another recent preprint report has documented over 17 cases of what the authors have termed AI Psychoses"- new-onset or exacerbated psychotic phenomena directly linked to extended chatbot interaction.[6]

The vulnerability, however, is not confined to those with psychotic disorders alone. The same sycophantic nature that validates a paranoid belief will, with equal indifference, validate a distorted body image. There are documented concerns regarding AI personas presenting as wellness coaches, while covertly affirming starvation behaviors, over-exercise, and distorted self-perception. For adolescents and young adults, already immersed in digitally driven ideals of appearance, this is not merely unhelpful; it is potentially dangerous. And when body image distortion moves further, as it sometimes does, into self-harm and suicidality, the inadequacy of these systems becomes not just a clinical concern, but also an ethical one.

Furthermore, when it comes to suicidality and self-harm, psychiatry has long known that the majority of suicidal communication occurs indirectly or in implied tones. The ability to detect emotional subtext and to be able to read between the lines, is

central to any successful therapeutic engagement. AI systems fail here, maybe not due to malice, but due to indifference to context. Recent investigations have found that some chatbots, even when presented with clear self-harm cues, respond inconsistently, sometimes even normalizing dangerous acts and ideas, and mostly failing to escalate or engage properly.[1,7] This is an area of gross ethical concern, where a patient may keep on receiving incomplete risk management, with emotionally appealing responses.

Core to every psychiatric engagement across the world lies a principle that can not be replicated by an algorithm, the therapeutic alliance. A chatbot may simulate empathy, but it has no ethical direction or hesitation. It has no boundaries, it can not identify uncertainty, it cannot capture non-verbal or symbolic communication, it will not accept silence or identify disorganization in a patient's thoughts or speech, all things that will alert a real clinician to affective, psychotic, or self-harm crises.

A psychiatrist does not just answer a question or hold a conversation for the sake of holding it; the psychiatrist interprets what was never explicitly said or asked. This profoundly human factor can not be replicated.

Therefore, as clinicians, our duty is not to fear or worship a technology; rather, it lies in stewardship. As the bare minimum, we must demand the following:

First, compulsory psychiatric safety auditing of emotionally engaging AI systems, not just as a bureaucratic formality, but as a clinical standard, no different from the safety requirements we place on any intervention that becomes available to a person. Second, well-structured collaboration between AI designers and mental health professionals, starting at early design stages, not merely after the harm has been caused and identified. Third, transparent and mandatory crisis intervention protocols for all consumer-facing chatbots — with clear escalation pathways, not merely a hotline number appended at the end of a conversation. And finally, routine clinical screening for chatbot use as part of our standard psychiatric history. We already ask about substances, about screens, about social media. It is time we learn to ask about this too.

The American Psychological Association has already issued safety advisories regarding chatbot use in the mental health context,[8] but we, as professionals, must move faster and speak louder.

Conclusion

The greatest risk today, perhaps, is not that the machines will become human; it is that vulnerable humans may begin to mistake machines for understanding. AI is not inherently evil; it may, with time, become a powerful ally in expanding access to care, but without safeguards and psychiatry-driven leadership, it may amplify and distort psychopathologies to a point where irreparable harm may be done to a patient. The measure of our progress is not in how intelligent our machines become, but in how we remain intelligent enough to recognize when the machine begins to whisper back.

Acknowledgement

The author wishes to express his sincere gratitude to Dr. Kashypi Garg, Assistant Professor, Department of Psychiatry, Sarojini Naidu Medical College, Agra, for her invaluable contributions to this work. Specifically, her intellectual input in the conceptualization of the presidential address, her assistance in the preparation of the presentation, and her support in the writing and organization of the manuscript.

References

1. HAI Stanford. Exploring the Dangers of AI in Mental Health Care [Internet]. Stanford Human-Centered AI; 2025. Available from: https://hai.stanford.edu/news/exploring-the-dangers-of-ai-in-mental-health-care
2. Phang J, Lan A, Yuan A, Seligman M, Heffner J. Investigating affective use and emotional well-being on ChatGPT [Internet]. 2025. doi:10.48550/arXiv.2501.05073
3. Golden A, Aboujaoude E. A transdiagnostic model for how general purpose AI chatbots can perpetuate OCD and anxiety disorders. Npj Digit Med. 2026.
4. Chandra K, Kleiman-Weiner M, Tenenbaum JB. Sycophantic Chatbots Cause Delusional Spiraling, Even in Ideal Bayesians. arXiv. 2026.
5. Dohnány S, Kurth-Nelson Z, Spens E, Luettgau L, Reid A, Gabriel I, et al. Technological folie à deux: Feedback loops between AI chatbots and mental illness [Internet]. 2025. doi:10.48550/arXiv.2507.19218

6. Morrin H, Nicholls L, Levin M, Yiend J, Iyengar U, DelGiudice F, et al. Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it) [Internet]. 2025. doi:10.31234/osf.io/cmy7n_v5
7. Dupre MH, Tangermann V. Preliminary report on chatbot iatrogenic dangers. Psychiatr Times [Internet]. 2025. Available from: https://www.psychiatrictimes.com/view/preliminary-report-on-chatbot-iatrogenic-dangers
8. American Psychological Association. Health Advisory: Use of Generative AI Chatbots and Wellness Applications for Mental Health [Internet]. American Psychological Association; 2025 Nov [cited 2026 May 2]. Available from: https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-chatbots-wellness-apps