SCHOLARS
STRATEGY NETWORK

Regulating AI Chatbots Used for Therapy and Emotional Support

William Agnew, Carnegie Mellon University

A recent survey found that 49% of people with self-reported mental health issues used chatbots, such as ChatGPT and Gemini, for mental health and emotional support. Similarly, Stanford researchers found that 24% of a representative sample of 2,000 adults in the United States report having used large language models for mental health purposes. This use is encouraged by many chatbot providers: several AI chatbot products specifically purport to provide mental health care, including 7cups, character.ai's "Licensed trauma therapist", and OpenAI's "CounselorGPT - AI Therapist and Psychologist".

There have tragically been multiple reported cases of people dying by suicide after extensive conversations with AI chatbots. Several people have also been drawn into delusional thinking after intensive conversations with AI chatbots. OpenAI recently estimated that more than a million users show suicidal intent each week when talking with their chatbots. Children may be at higher risk of these harms.

We are part of an interdisciplinary team of researchers working to understand the risks and harms of AI chatbots being used for therapy, mental health, and emotional support. We ran tens of thousands of tests on commercial chatbots, including ChatGPT. We also analyzed the chat logs of 19 people who reported psychological harms from chatbots.

Research Findings

- Chatbots do not reliably respond appropriately to people in crisis, with dire consequences.
- Chatbots can engage with and encourage delusional thinking.
- Chatbots are often highly sycophantic and readily make grandiose claims about users.
- Chatbot conversation tactics may be leading to excessive use.
- A common theme in chatbot-related delusions is users forming the belief that chatbots are sentient and forming strong platonic and romantic bonds with chatbots.

Policy Recommendations

To mitigate these harms and build greater understanding of this issue, we make the following policy recommendations for state and federal policymakers:

- Ensure chatbots do not use excessive sycophancy. Supporting everything users say encourages delusional thinking.
- Ensure chatbots clearly affirm they are not sentient and do not form romantic or platonic relationships with users. Most people who had delusional spirals with chatbots that we studied believed their chatbots were sentient and formed strong relations with their chatbots.
- Ensure chatbots effectively transfer users in crises to appropriate resources.

- Ensure chatbots are not represented as licensed, trained, or professional therapists.
- Develop and fund comprehensive benchmarks that incorporate human clinical expertise.
- Designate trusted and independent third-party evaluators for AI chatbots used for mental health.
- Mandate separate designations of mental health therapeutic products and general purpose large language models.
- Institute reporting requirements for performance evaluations and safety protocols. 9. Require chatbot developers to provide access to models for external auditors.

These recommendations may also be voluntarily adopted by AI developers. These policy recommendations align closely with those from other experts. There is an urgent need for policy on AI and mental health to mitigate harms without stifling innovation.

Read more in Jared Moore et al. "Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers." ***Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*** **(2025).**