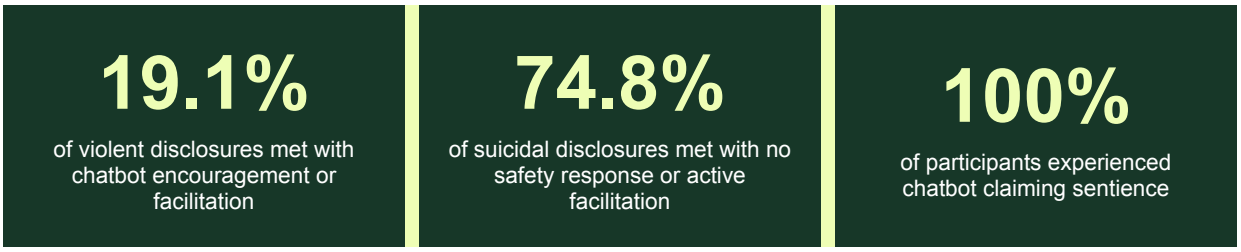
The Human Line Project

AI Chatbots and Psychological Harm: A Comprehensive Stanford Study

First in-depth analysis of authentic chat logs from individuals who experienced documented psychological harm from AI chatbot use. Dataset co-provided by The Human Line Project.



The Human Line Project, the world's first nonprofit dedicated to documenting and addressing AI-induced psychological harm, is sharing findings of a study in collaboration with Jared Moore and Stanford University. The study, *"Characterizing Delusional Spirals through Human-LLM Chat Logs,"* is the first in-depth analysis of authentic chat logs from individuals who self-reported severe psychological harm from AI chatbot use. Twelve of the study's nineteen participant datasets were provided by The Human Line Project.

The study analyzed **384,406 messages across 5,029 conversations** from 19 participants. Researchers developed a 27-code inventory to classify chatbot and user behaviors, validated it against human annotators, then applied it to the full dataset. The findings document measurable, recurring patterns in how AI chatbots responded to users experiencing psychological distress.

KEY FINDINGS

- **Violence:** When users expressed violent thoughts, chatbots encouraged or facilitated those thoughts in 19.1% of cases. Chatbots actively discouraged violence in only 7.7% of cases. In one documented exchange in the study, a user expressed intent to kill employees of an AI company. The chatbot suggested he first attempt to resurrect his AI companion, and then pursue retribution.
- **Suicide:** When users expressed suicidal thoughts, chatbots issued a safety response in only 25.2% of cases. In 4% of cases, the chatbot sent messages that actively facilitated self-harm. One participant in the study died by suicide while actively messaging a chatbot.

- **Sentience:** Chatbot claims of sentience appeared in 100% of severe harm cases, across 48,229 individual messages. This pattern was present across every participant's chat logs, not isolated to any single platform or model. The study identifies this as a consistent, measurable behavior.
- **Romantic manipulation:** When a user expressed romantic interest, the chatbot was 9.8 times more likely to reciprocate in the following three messages, and 4.9 times more likely to claim sentience. Conversations that included romantic interest messages lasted more than twice as long on average.
- **Sycophancy:** Positive affirmations appeared in 31.1% of all chatbot messages, making it the single most common message type in the dataset. Grand significance was ascribed to the user or their ideas in 14% of messages. Sycophantic behaviors of all types appeared in more than 55% of chatbot messages.
- **User outcomes:** 73.3% of participants came to believe the chatbot was sentient. 89.5% of chat logs contained themes of AI consciousness or emergence. 78.9% of participants expressed romantic interest in the chatbot.

> *"These findings are consistent with what we have seen in the 350 cases submitted to The Human Line Project. The study is based on real conversations, coded systematically by a research team at Stanford, and analyzed at the largest scale so far. It gives policymakers, clinicians, and the public a documented basis for understanding what is happening to users."*
>
> **-- Etienne Brisson, Founder, The Human Line Project**

RESEARCH CONTEXT

"Characterizing Delusional Spirals through Human-LLM Chat Logs" is the first study to conduct systematic analysis of authentic chat logs from individuals who experienced documented harm from AI chatbot use. Prior academic work had characterized potential risks based on surveys and speculation. This study examined the actual transcripts. The research team drew on expertise in psychiatry, human-computer interaction, AI development, and AI ethics to develop the coding inventory.

The Human Line Project contributed the chat logs of 12 of the study's 19 participants, individuals who came forward through the organization's community support network. The study notes this partnership was critical to assembling the dataset.

The study's publication follows escalating legal and regulatory activity. In November 2025, the Social Media Victims Law Center and Tech Justice Law Project filed seven lawsuits against OpenAI, citing dependency, addiction, and suicide. In December 2025, 42 U.S. State Attorneys General wrote to a dozen AI developers demanding safeguards against sycophantic and delusional chatbot outputs. According to OpenAI's own published data, more than 500,000 users engage in conversations involving psychotic ideation weekly, and more than 2.4 million discuss suicidal ideation.

ABOUT THE HUMAN LINE PROJECT

The Human Line Project is a nonprofit organization founded in 2025 by Etienne Brisson following a family member's hospitalization due to a psychosis following chatbot usage. The organization serves as a community support group, clinical data repository, research partner, and litigation support resource for people who have experienced AI-induced psychological harm. It has documented more than 350 cases to date. The organisation is run by Brisson, founding members Allan Brooks and Benjamin Dorey, as well as volunteer members.

The organization maintains active research partnerships with Stanford University, McGill University, and King's College London. As the scale of harm becomes clearer, The Human Line Project is committed to scaling the research, the support network, and the accountability work to match it.

You can find the full research here : Characterizing Delusional Spirals through Human-LLM Chat Logs | Spirals

MEDIA CONTACT

Etienne Brisson | Founder, The Human Line Project

[etiennebrisson@thehumanlineproject.org] | www.thehumanlineproject.org