JUST TECH

When Help Isn’t Fully Human: The Problem of Generative AI in Crisis Support

By: Stephen J. Neville

PLATFORMS & INFRASTRUCTURE

Neville, Stephen J. "When Help Isn’t Fully Human: The Problem of Generative AI in Crisis Support." Just Tech. Social Science Research Council. August 6, 2025. DOI: https://doi.org/10.35650/JT.3086.d.2025.

Across industries—including customer service, education, journalism, public health, and beyond—generative AI is being pitched as a solution to increase efficiency, cut costs, and boost productivity. Chatbots that summarize meetings, write code, or simulate human communication are marketed as transformative and essential for virtually every sector. As public and private organizations race to adopt these tools, there is an urgent need for measured adoption and greater public debate in key sectors where human care needs to be prioritized over efficiency and innovation. Nowhere is this more pressing than in crisis counseling, where vulnerable people reach out for genuine and trustworthy human connection and support.

AI at the Front Lines of Human Crisis

In the last decade, crisis helplines and peer-support platforms have increasingly turned to technology partnerships to manage growing demand and streamline their operations. Organizations such as The Trevor Project and RAINN have collaborated with companies like Google and Amazon to integrate AI systems into their crisis hotlines. In 2016, Twilio developed the messaging infrastructure for Crisis Text Line, which operates exclusively via text. Automated tools are built into these platforms, such as risk assessments that triage callers by scanning for keywords or phrases that might indicate high risk, so that the most urgent cases can be referred quickly to a human counselor.

The motivations for this kind of AI adoption are understandable. Many helplines are underfunded, understaffed, and overwhelmed. AI is marketed as a silver bullet to help with scale and speed. And in

some cases, there are unmistakable benefits: Machine learning tools can assist with quality assurance or be used to create training simulators so that new and inexperienced counselors can safely practice crisis scenarios without posing risks to real help-seekers.

Although many of these integrations proceed with minimal public scrutiny, AI is rapidly transforming how help is delivered, who gets it first, and how counselors and peer supporters communicate with vulnerable people.

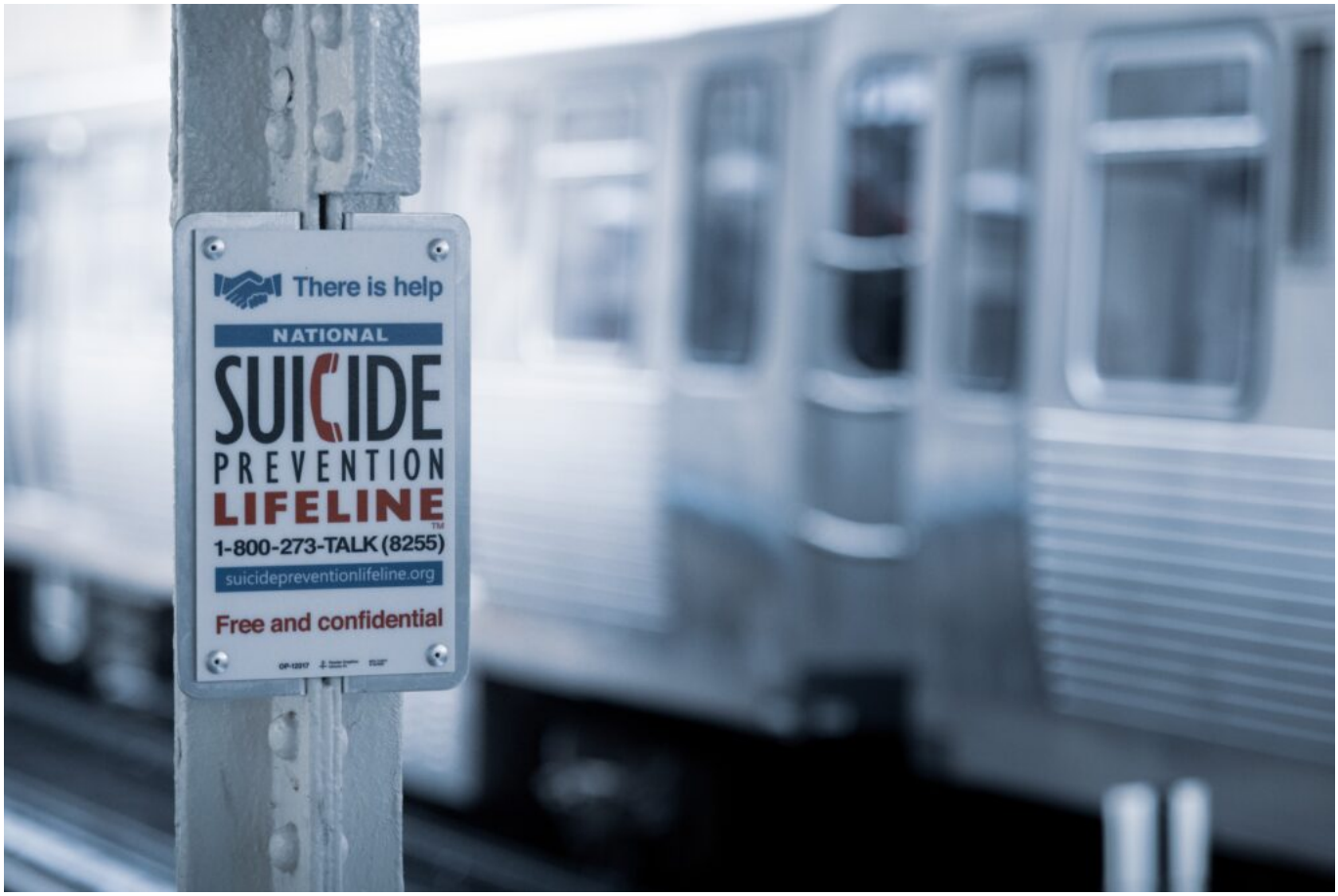
The Quiet Spread of AI Experiments

Some mental health apps and peer-support platforms have begun testing generative AI to "boost empathy." On the peer-support app, TalkLife, researchers used AI to revise messages written by human supporters.[1] A similar experiment[2] caused public outcry when the peer-to-peer support platform Koko, used GPT-3 to rewrite messages sent to help-seekers, without their informed consent.

The goal of these and similar experiments is to make peer support feel more empathetic, especially because it can be challenging to express empathy over text.[3] However, this can communicate the very opposite of empathy by making help-seekers feel disrespected and violated. In crisis contexts, where vulnerable people often reach out in desperation, consent and transparency matter deeply.

Unlike automating emails or chat support, generative AI in mental health settings carries far greater risks. Help-seekers are often in acute distress, seeking trustworthy and confidential support. When AI mediates these sensitive conversations—by editing responses or shaping how counselors respond—it can erode that trust and potentially threaten future help-seeking behaviors. Risks are compounded by tangible harms: Generative AI systems are known to produce nonsensical outputs or "hallucinations." In one high-profile case, Google's chatbot Gemini reportedly told a user to commit suicide amid a conversation about the challenges of aging. These kinds of mistakes are unacceptable in any setting, but in crisis support, they can be damaging and even deadly.

Deeper reflection on the role of AI is needed in this context. AI chatbots like Replika have been tested in loneliness interventions and suicide prevention,[4] and some help-seekers report feeling heard by them and could possibly prefer them over human therapy.[5] However, most people still want human connection when they reach out in a state of crisis.



Reaching out for Human Connection

A recent survey in Australia found that both the general public and help-seeking users of Lifeline, a major crisis service, expressed an overwhelming preference for speaking with a real person over an automated system.[6] Approximately half of all respondents said they would be less likely to use the service if they knew it relied on automated technology.

A preference for human connection reflects more than idealist sentiment. It points to the deeper importance of trust and relational care in crisis situations. Help-seekers are not just looking for information or quick answers; they are looking to be heard, seen, and supported by someone capable of emotional comprehension. Although chatbots can simulate empathetic communication, this is irrelevant in considering help-seekers' expectations for human commiseration.

Even when AI is used only to "augment" human work, the issue of consent remains critical. People in crisis may not be in a position to read or fully comprehend a platform's terms of service and data privacy policies. Even if AI is not autonomously interacting with help-seekers, it might be operating behind the scenes to analyze words, suggest replies or resources, and conduct experiments without obtaining meaningful consent from research participants.[7]

This relates to what researchers call a "transparency dilemma": Disclosing AI use can reduce trust but not disclosing it at all risks greater harm if users later find out.[8] This is highly pertinent in regard to

mental health support[9]: If people feel they can't trust a platform to keep their information private or worry they're being misled or manipulated somehow, they may not reach out at all.

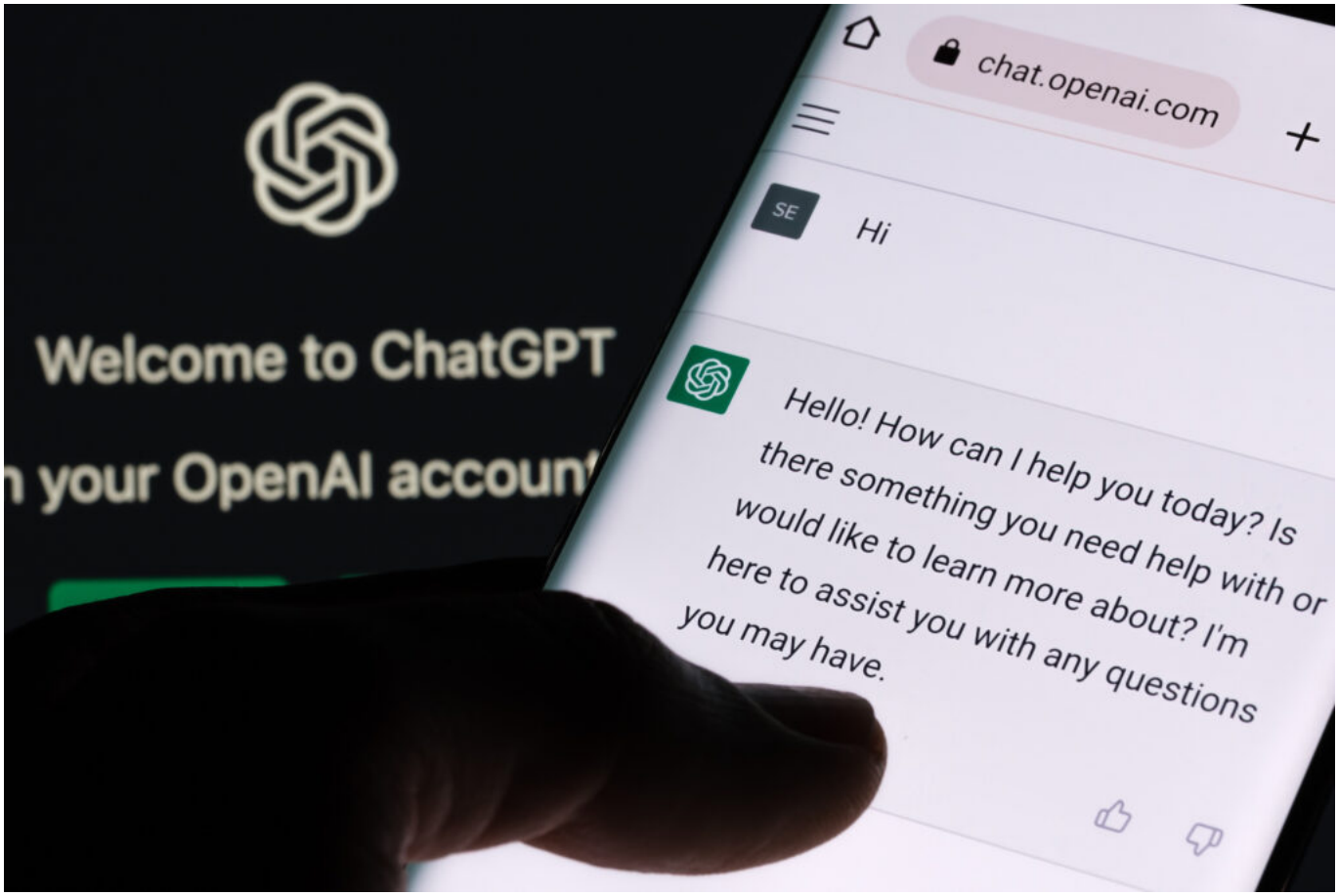
When Empathy Becomes a Business Model

In crisis counseling, consent becomes even more fraught when sensitive data is repurposed to create commercial products as this blurs the boundary between care and "surveillance capitalism"—an economic order entrenched by digital technologies that extracts data from human experience to create revenue streams.[10]

In 2022, investigative reporting revealed that Crisis Text Line shared anonymized conversation data with its for-profit spin-off, Loris.ai. Although Crisis Text Line did not sell data to Loris, they purposefully shared anonymized data collected from crisis conversations. This arrangement was intended to create an economically sustainable funding model insofar as Loris would develop a proprietary machine learning system for customer service applications and Crisis Text Line would share in the profits to carry out its life-saving nonprofit activities. Although no data was sold, highly sensitive data collected from an exceptionally vulnerable user base was used for secondary commercial activities.

The backlash was swift. The watchdog organization Electronic Privacy Information Center (EPIC) condemned the action for violating public trust. In response, Crisis Text Line quickly announced a formal end to the data sharing partnership. Nonetheless, the episode raises important questions about how crisis data are handled and how organizations might be striking a questionable balance between priorities of care and commercial automation.

The broader AI industry operates on a logic of extraction in treating everything—text, images, sounds, videos—as raw materials for algorithmic training. As scholar Kate Crawford writes, this reflects a kind of "ruthless pragmatism, with minimal context, caution, or consent-driven data practices…"[11] When this logic is applied in mental health care, the social harms and risks are magnified.



Behind the Scenes: Self-adoption of AI by Crisis Counselors

Through interviews with volunteer counselors at Crisis Text Line, I've learned that AI is entering crisis support not only through in-house partnerships and integrations but also informally. Some counselors are using generative AI tools like ChatGPT to help with proofreading tasks—acting as a kind of Grammarly for crisis counselors. Others use it to look up mental health topics or search for ways to respond to a particular type of crisis.

While these uses are well-meaning, they raise serious concerns. The Crisis Text Line platform provides counselors with a menu of mental health resources as well as answers to frequently asked questions that have been vetted by qualified experts. ChatGPT, by contrast, pulls from a vast array of online sources that may not align with evidenced-based research or professional guidelines. One volunteer described their approach this way:

> I use it to gain information that I can text, for example, "how to solve mental crisis on that issue," "how to solve mental health on depression" . . . I would go to ChatGPT and be like, "how to solve mental issues concerning loved ones," "concerning work," "concerning society," because ChatGPT will give me a rundown of everything I want to solve. [Then] I'll pick one or two points, summarize it, add my own idea, and give it [to the help-seeker]. That would be the perfect solution for the person.

This begins to raise serious problems of consent, as some help-seekers might be averse to engaging in dialogue and receiving recommendations shaped by AI. Additionally, help-seekers may assume they're

receiving support directly from a trained person using approved materials. In practice, the advice may have passed through a generative AI system with unknown sources and unknown biases.

Adding to this, one counselor I spoke with explained how they even copy and paste excerpts of real conversations into ChatGPT to get suggestions on how to respond. This informal use introduces serious privacy and data security concerns. When a conversation transcript is shared with ChatGPT or similar tools, sensitive data is collected and may be used to train the models that power ChatGPT or sold off in a future business transaction, based on OpenAI's privacy policy. Compounding these problems, personal data could be exposed in a future data breach or misused in various ways.

Increased digital literacy around these problems and meaningful public debate about how to responsibly integrate AI into crisis counseling are needed to avoid exposing vulnerable people to preventable harms in the fragile moments they reach out for help.

Toward Transparency and Accountability

There is an urgent need for crisis support organizations to update their policies in response to AI developments, not only to address applications at the institutional level but also ad-hoc use by volunteers and staff. This is especially true as the number of active weekly users of ChatGPT continues to balloon. Policies about automation should be clear, public, and ensure accountability.

Some organizations, like Trans Lifeline, have taken a strong stance against AI integration, explicitly stating that their support is always human-led and not shaped by machine learning systems. Others, including Crisis Text Line, have embraced AI for triage and training while maintaining that human counselors handle all conversations. When organizations are clear and upfront about how they use AI, it allows help-seekers and community advocates to make more informed decisions about where to turn for support or which platforms to recommend.

But even the clearest organizational policies can be undermined if individuals are left to navigate AI use on their own beyond the official tech stack. Help-seekers who choose a particular helpline because of its commitment to privacy or human care deserve assurance that this commitment extends to every aspect of a conversation.

As the world races to adopt AI, crisis counseling is one place where the pace of adoption should be tempered. When we move fast, things get broken—as we have learned time and time again from tech industry blunders and misdeeds. Help-seekers and the general public are not ready for AI in this high-risk setting, and responsible governance needs to be ensured amid the current AI hype.[12] Even as AI reshapes many aspects of our world, it's plausible that the expectation for human connection in times of crisis will endure. These expectations are not obstacles to innovation; they are ethical imperatives that must be respected as a condition of technological change.

Footnotes

1 Ashish Sharma et al., "Human–AI Collaboration Enables More Empathic Conversations in Text-Based Peer-to-Peer Mental Health Support," *Nature Machine Intelligence* 5, no. 1 (2023): 46–57, https://doi.org/10.1038/s42256-022-00593-2.

2 Katherine Cohen et al., "Improving Uptake of Mental Health Crisis Resources: Randomized Test of a Single-Session Intervention Embedded in Social Media," *Journal of Behavioral and Cognitive Therapy* 33, no. 1 (2023): 24–34, https://doi.org/10.1016/j.jbct.2022.12.001.

3 Carrie A. Moylan et al., "'It's Hard to Show Empathy in a Text': Developing a Web-Based Sexual Assault Hotline in a College Setting," *Journal of Interpersonal Violence* 37, no. 17–18 (2022): NP16037–59, https://doi.org/10.1177/08862605211025036.

4 Bethanie Maples et al., " Loneliness and Suicide Mitigation for Students Using GPT3-Enabled Chatbots," *NPJ Mental Health Research* 3 (2024): 1–6, https://doi.org/10.1038/s44184-023-00047-6.

5 Per Carlbring et al., "A New Era in Internet Interventions: The Advent of Chat-GPT and AI-Assisted Therapist Guidance," *Internet Interventions* 32 (April 2023):100621, https://doi.org/10.1016/j.invent.2023.100621.

6 Jennifer S. Ma et al., "Consumer Perspectives on the Use of Artificial Intelligence Technology and Automation in Crisis Support Services: Mixed Methods Study," *JMIR Human Factors* 9, no. 3 (2022): e34514, https://doi.org/10.2196/34514.

7 Timothy D. Reierson, "Commentary on 'Protecting User Privacy and Rights in Academic Data-Sharing Partnerships: Principles From a Pilot Program at Crisis Text Line,'" *Journal of Medical Internet Research* 26, no. 1 (2024): e42144, https://doi.org/10.2196/42144.

8 Oliver Schilke and Martin Reimann, "The Transparency Dilemma: How AI Disclosure Erodes Trust," *Organizational Behavior and Human Decision Processes* 188 (May 2025):104405, https://doi.org/10.1016/j.obhdp.2025.104405.

9 Patrick Brown et al., "Trust in Mental Health Services: A Neglected Concept," *Journal of Mental Health* 18, no. 5 (2009): 449–58, https://doi.org/10.3109/09638230903111122.

10 Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs, 2020).

11 Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale University Press, 2022), 95.

12 Pei Boon Ooi and Graeme Wilkinson, "Enhancing Ethical Codes with Artificial Intelligence Governance—a Growing Necessity for the Adoption of Generative AI in Counselling," *British Journal of Guidance & Counselling* 53, no. 1 (2024): 66–80, https://doi.org/10.1080/03069885.2024.2373180.