

News and Perspective

# “Feasible but Fragile”: An Inflection Point for Artificial Intelligence in Mental Health Care

Kayleigh-Ann Clegg, JMIR Correspondent

**Key Takeaways**

- The future of artificial intelligence (AI) tools in mental health care is at an inflection point; regulators are taking both the potential benefits and the risks of these tools seriously.
- Whether these tools succeed or fail to meet their potential for improving mental health care depends on the extent to which stakeholders are able to successfully seize the moment and collaborate on transparent, high-quality research; establish and incentivize safety and efficacy; adopt patient-centric benchmarks; and think beyond traditional therapeutic models.

On November 18, 2025, a congressional hearing was held in Washington, DC, by the US House Energy and Commerce’s Subcommittee on Oversight and Investigations, examining the risks and benefits of artificial intelligence (AI) chatbots [1,2].

Marlynn Wei, MD, JD; Jennifer King, PhD; and John Torous, MBI, MD (director of digital psychiatry at Beth Israel Deaconess Medical School and associate professor of psychiatry at Harvard Medical School) provided expert testimony at the congressional hearing [3-5]. I sat down with Torous to discuss his reflections on the future of AI in mental health.

## An Inflection Point

Following on the heels of several lawsuits [6] and mounting concerns about the safety of commercially available AI chatbots and their widespread “off-label” use as psychological support [7,8], November’s congressional hearing was somewhat anomalous—in a good way.

“I actually am optimistic,” says Torous, “because we never saw a congressional oversight committee form in the early days when social media came out or when apps came out or when VR [virtual reality] came out. It’s exciting to see a body like Congress taking the time and attention to try to understand what the issue is.”

He remarks that it’s a different trajectory than we’ve seen over the past 25 years of digital health innovation, one that simultaneously signals that “we’re seeing the end of AI exceptionalism in mental health.” It suggests that regulators are taking the risks seriously and that AI—whether purpose-built or used de facto—will not be exempt from the same scrutiny applied to other clinical tools.

It’s also a potential inflection point from the otherwise rapid and underregulated growth and proliferation of AI tools for mental health, including many chatbots whose safety and efficacy remain to be definitively established [9,10]. The shape of the trajectory now—whether these tools succeed or

fail to materialize their potential for improving mental health care—depends on what we do from here.

## Shift the Incentive Structures

To shift toward the development of sustainable, effective AI tools for mental health, Torous says, we need to shift the incentive structures—through clear, enforceable standards—for companies developing them. Instead of optimizing for engagement, we should be optimizing for privacy, safety, and efficacy.

“It’s better that we have competition based on what are safer and more effective products,” he says, for both the end users and the companies themselves. Otherwise, “it’s only going to be the people with the most marketing budget who win. It’s anti-innovation if we don’t have some standards.”

We can draw lessons from the trajectories of other digital health tools for mental health, like mental health apps, and avoid repeating past mistakes: “We tried winning on number of clicks and engagement; it did not make apps safer. It did not make them more effective, and it kind of destroyed the market for apps,” he says. “Competing on engagement is a zero-sum game, it’s a race to the bottom—and the bottom is pretty bleak.”

## Prioritize Collaboration, Transparency, and Research

Developing the standards necessary to shift those incentive structures first and foremost requires collaboration between all stakeholders.

Establishing the “rules of the road” for that collaboration can be tricky, Torous says, but necessary. While there need to be penalties, for example, when companies act in bad faith, the core aims of regulation and oversight should be encouraging transparency and data sharing, and treating improvements in AI chatbots and other tools as a mutual endeavor.

Funding rigorous, well-designed, replicable research should follow. “I think the first study would need to be a large, open, transparent study to set the bar and understand what we’re dealing with,” says Torous. Foundational research is needed to determine and confirm, among other things, patterns of use, causes and base rates of potential risks, and potential benefits associated with use of both commercially available and purpose-built AI tools.

## Adopt Patient-Centric Benchmarks

Ensuring real benefit and minimizing harm from AI tools for mental health will require centering the patient and patient needs. The human should be the focus of any research conducted or standards adopted.

This is something that Torous and his research team are already working on, and a value that has been at the core of their work for many years. Building on their experience developing MindApps.org [11], an assessment framework and free app library to help people make informed decisions about which mental health apps might best suit their needs, they have partnered with the National Alliance on Mental Illness (NAMI) on a project they’re calling “MindBench.ai” [12].

MindBench.ai will function similarly as an assessment framework and platform that people can use to make decisions about large language models (LLMs) and LLM-based mental health tools. But instead of using static evaluation criteria, the team will develop dynamic, rigorous safety and performance benchmarks based on patient input.

“We’re going to meet with patient groups around the country,” says Torous. “And we’re going to say: ‘What are you most excited about? What are you most worried about? Let’s build these benchmarks that will help us understand if these chatbots meet your needs as real people on the ground.’ We’re really excited that this is going to bring the voice of patients to the forefront.”

In co-designing and standardizing benchmarks, the project will not only protect and help patients but also help incentivize and give companies something to work toward. “No one wakes up and says, ‘I’m so excited my chatbot is causing harm,’” Torous says. “I really have full confidence that, with guidance, companies will love it because they want to make [their products] better.”

While MindBench.ai will initially be focused on chatbots designed for mental health, if useful, it could eventually be expanded to incorporate other kinds of tools or used as an example to guide the development of similar frameworks.

**Keywords:** artificial intelligence; chatbots; mobile apps; telemedicine; digital health; psychiatry; mental health; mental disorders; disease models; theoretical

## Conflicts of Interest

John Torous is the editor in chief of *JMIR Mental Health*, a JMIR Publications journal, at the time of this publication.

## Think Beyond Traditional Models

There’s a further vision that Torous is excited about, one which speaks to a sometimes underrecognized harm: missed opportunity. It involves the importance of thinking outside the box and beyond traditional therapeutic models.

“The next generation [of mental health tools] is going to be much more powerful, much more personalized,” he says. “When you’re bringing in that much new information about people’s environments, their physiology, their emotions, and their words, we could consider redefining [existing nosologies], making functional definitions of mental illnesses.”

We could also, he says, move beyond using chatbots to replicate already-established kinds of treatment. “I think what we’re seeing about to happen in 2026 is the first time we can not only collect such personalized raw data about mental health, but we also have the computing power to now turn that into personalized insights,” he says. “You could imagine there’s a world where we could actually build new therapy treatments. We have this whole new world of personal insights that can drive personalized treatments.”

This immense potential, however, hinges on regulation, safety, and trust. “Of course there’s tremendous risk when you have a lot of personal data. And I think that’s also why it makes sense to have more safety and regulation in place, because then there’s a trust to really think about rebuilding what mental health treatment is and to use this [opportunity] for what it is.”

## Feasible but Fragile

Torous comes across as level-headed, pragmatic, and principled—prone to neither nostalgic hand-wringing nor uncritical techno-optimism. His vision of the potential for AI in mental health is both measured and forward-looking. For me, as a cautious psychologist and occasional hand-wringer, it inspires optimism.

“There’s real potential here,” he says. With rigorous, collaborative research and regulation to shape it, the “AI bubble” need not burst or engulf us all; instead, it can solidify into something genuinely transformative for mental health. “It’s feasible, but fragile.”

“Feasible, but fragile”—an apt description and a call to action for researchers, regulators, patients and patient advocates, innovators, and clinicians alike.

---

**References**

1. Innovation with integrity: examining the risks and benefits of AI chatbots. Congress.gov. Nov 18, 2025. URL: <https://www.congress.gov/event/119th-congress/house-event/118669> [Accessed 2025-12-11]
  2. O&I hearing: innovation with integrity: examining the risks and benefits of AI chatbots. House Energy and Commerce Committee YouTube page. Nov 18, 2025. URL: <https://www.youtube.com/watch?v=6rGZxWs-TA0> [Accessed 2025-12-11]
  3. Wei M. Testimony of Marlynn Wei, M.D., J.D. Democrats, Energy and Commerce Committee. Nov 18, 2025. URL: [https://democrats-energycommerce.house.gov/sites/evo-subsites/democrats-energycommerce.house.gov/files/evo-media-document/11.18.2025\\_oi\\_hearing\\_wei\\_testimony.pdf](https://democrats-energycommerce.house.gov/sites/evo-subsites/democrats-energycommerce.house.gov/files/evo-media-document/11.18.2025_oi_hearing_wei_testimony.pdf) [Accessed 2025-12-11]
  4. King J. Safeguarding data privacy and well-being for AI chatbot users. US House of Representatives: Document Repository. Nov 18, 2025. URL: <https://docs.house.gov/meetings/IF/IF02/20251118/118669/HHRG-119-IF02-Wstate-KingPhDJ-20251118.pdf> [Accessed 2025-12-11]
  5. Torous J. Written statement of John Torous MD MBI. US House of Representatives: Document Repository. Nov 18, 2025. URL: <https://docs.house.gov/meetings/IF/IF02/20251118/118669/HHRG-119-IF02-Bio-TorousMDMBIJ-20251118.pdf> [Accessed 2025-12-11]
  6. Batt M, Dinçer M. August & September 2025 Tech Litigation Roundup. Tech Policy Press. Oct 10, 2025. URL: <https://www.techpolicy.press/august-september-2025-tech-litigation-roundup/> [Accessed 2025-12-11]
  7. Use of generative AI chatbots and wellness applications for mental health: an APA health advisory. American Psychological Association. Nov 2025. URL: <https://www.apa.org/topics/artificial-intelligence-machine-learning/health-advisory-chatbots-wellness-apps> [Accessed 2025-12-11]
  8. Iftikhar Z, Xiao A, Ransom S, Huang J, Suresh H. How LLM counselors violate ethical standards in mental health practice: a practitioner-informed framework. Proc AAAI/ACM Conference AI Ethics Soc. 2025;8(2):1311-1323. [doi: [10.1609/aies.v8i2.36632](https://doi.org/10.1609/aies.v8i2.36632)]
  9. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry. Jul 2019;64(7):456-464. [doi: [10.1177/0706743719828977](https://doi.org/10.1177/0706743719828977)] [Medline: [30897957](https://pubmed.ncbi.nlm.nih.gov/30897957/)]
  10. Bodner R, Lim K, Schneider R, Torous J. Efficacy and risks of artificial intelligence chatbots for anxiety and depression: a narrative review of recent clinical studies. Curr Opin Psychiatry. Jan 1, 2026;39(1):19-25. [doi: [10.1097/YCO.0000000000001048](https://doi.org/10.1097/YCO.0000000000001048)] [Medline: [41198140](https://pubmed.ncbi.nlm.nih.gov/41198140/)]
  11. M-Health Index & Navigation Database. URL: <https://mindapps.org/> [Accessed 2025-12-11]
  12. MindBench.ai. URL: <https://mindbench.ai/> [Accessed 2025-12-11]
- 

*Please cite as:*

Clegg KA

*"Feasible but Fragile": An Inflection Point for Artificial Intelligence in Mental Health Care*

*J Med Internet Res* 2025;27:e89202

URL: <https://www.jmir.org/2025/1/e89202>

doi: [10.2196/89202](https://doi.org/10.2196/89202)

© JMIR Publications. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 16.Dec.2025