Written Statement of John Torous MD MBI

Director of Digital Psychiatry, Beth Israel Deaconess Medical Center
Associate Professor of Psychiatry, Harvard Medical School
Before the House Energy & Commerce, Oversight and Investigations Subcommittee
Subcommittee Hearing on “Innovation with Integrity: Examining the Risks and Benefits of AI Chatbots.”

1 . Introduction

Chairman Dr. Joyce, Ranking Member Clarke, Chairman Guthrie, Ranking Member Pallone, and Members of the Subcommittee:

Thank you, Chairman Joyce, Ranking Member Clarke, and members of the Subcommittee. My name is John Torous, and I am a dual board-certified psychiatrist and clinical informaticist. I direct the Division of Digital Psychiatry at Beth Israel Deaconess Medical Center and am an Associate Professor of Psychiatry at Harvard Medical School. As a member of the American Psychiatric Association, I led the creation of the organization’s broadly adopted technology evaluation framework. With a background in electrical engineering and computer sciences, both my clinical practice and research focus on how we can utilize new technologies to enhance outcomes for individuals with mental illnesses.

We all want better solutions to the mental health crisis, and we all see that AI has the potential to help today. But we also see that it can cause tremendous harm today. The risks of AI for mental health will continue to expand as AI technology improves, but Congress can take actions to make AI safer today.

I would like to leave the Subcommittee with four main points today:

- First, AI tools that were never designed for mental health support are being used for that purpose by millions of Americans each week. Engineering AI to reduce or prevent such uses is challenging and expensive, and companies have minimal incentives to invest in safety unless we require such improvements and encourage transparent research that cuts through the fragmented research on proprietary platforms.
- Second, so far, there’s no well-designed, peer-reviewed, replicated research showing that any AI chatbot making mental health claims is effective for meaningfully improving clinical outcomes. We need to encourage high-quality research today.
- Third, American consumers face a wave of misleading information about what AI for mental health is and are exposed to safety and privacy risks: deceptive marketing and risky products harm everyone.
- Fourth, with the right research and regulation, AI for mental health can be a powerful tool. But that requires guidance from Congress to support the FDA, FTC, and NIH towards these aims, making AI for mental health more rigorous, regulated, and evidence-based.

2. Millions of Americans are Using AI Tools Not Designed for Mental Health

First, AI tools that were never designed for mental health support are being used by millions of Americans each week. In late October, OpenAI reported that over one million users per week have conversations with ChatGPT that include explicit indicators of potential suicidal planning [1].

There are now numerous lawsuits alleging that various AI chatbots contributed to suicide deaths. Against these tragic outcomes, we must also acknowledge that millions of Americans also find some degree of support from AI, and our research has shown that there can also be benefits [2]. Yet engineering AI to reduce or prevent these harms while maximizing potential benefits is costly, and today, companies have few guidelines or incentives for that work. The proprietary nature of the AI platforms that millions of Americans use today presents a formidable barrier to transparent research or evaluation of safety. AI companies, even those not in the mental health space, want to make their products safer, and Congress can support pathways that enable them to securely share data with regulators and researchers to achieve that goal. We missed the early opportunity to harness social media for mental health benefits, but today we have the chance to make it right with AI.

3. The Evidence of Mental Health Benefits of AI is Today Minimal

Second, many AI tools are already making claims of mental health benefits to Americans, despite the clinical evidence not supporting those assertions. Long before AI became popular, in July 2019, my team reviewed the clinical evidence for chatbots in mental health [3], and of course, it was minimal. But in summer 2025, we redid that analysis to see what had improved [4]. We found that while there is more research and bolder claims, there is still a paucity of high-quality clinical evidence [4]. For example, one study reported it was a randomized controlled trial of an AI chatbot for depression and anxiety, and that it made people feel better. But the control group was nothing, a waitlist [5]. In clinical research, everything beats a waitlist control. My team has reviewed the research evidence on where many of these mental health AI tools are trained, and found that the most common source is Reddit [6]. We can also agree that social media should not be the leading source of AI information about mental health. This lack of high-quality evidence mirrors what one patient told me, *“Dr. Torous – I tried one of them and it kept telling me the same thing over and over in different ways, it did not seem that smart.”*

Yet the AI field is not waiting for research to catch up, and our team has studied how LLMs today can also support facial recognition of emotions [7], understanding of our environments which we know impact mental health outcomes [8], and find mental health signals in fitness trackers and smartphone sensor data [9]. We need research to not only catch up but also get ahead. We must support the NIH and especially NIMH to conduct high-quality, neutral, and rapid research to understand the risks and benefits of AI for mental health. We also need to see the field develop clearer standards of what constitutes adequate evidence on safety and efficacy/effectiveness. This includes whether evidence that one tool in a particular category is plausibly safe and effective is sufficient to establish that other tools in the same category are also plausibly safe and effective.

4. The Mental Health and Safety Risk of AI Poses Real but Poorly Understood Risks

Related, we must study the harms, including why some people develop psychotic like reactions and others even take their own lives after extended use. Our team at Beth Israel Deaconess Medical Center is currently researching how we can better model and prevent these risks [10], but without data from AI companies, the impact of such work is limited. There are less visible harms like the impact of young people developing parasocial relationships with AI, being offered ineffective or even dangerous advice, or being told they are receiving therapy when that is simply untrue. *“It told me I was doing really well, I think it was trying to be supportive, but clearly it did not help,”* one patient told me when I interviewed them on the locked inpatient psychiatry unit after he was admitted for a decompensation. Finally, the risk of digital exclusion is real and

we need to ensure there is support for digital literacy to ensure all Americans are able to access and use these tools that are deemed safe and effective. My team at Beth Israel Deaconess Medical Center has developed and shared one of the most comprehensive digital literacy training programs [11], and expanding access to such programs is a win for everyone. AI companies want to help and OpenAI [1], Google/DeepMind [12], and Anthropic [13], for example, are beginning to support reporting and/or research on these topics. Regulators want to help too, but the current patchwork of AI mental health regulation, which my team reviewed across 50 states [14], limits their impact.

5. Unfounded Marketing Claims, Side Stepping Regulation, and Privacy Risks Threatened Americans

Fourth, while there is still much we do not know about the benefits or risks of AI for mental health, the marketing conveys a very different message. I have spoken to many patients and family members who are confused about what mental health AI is because it is marketed to them by companies as a ready-to-use tool with no risks. *A patient recently shared with me the disclaimer she found on a mental health chatbot "in the rare case that this will provide me with medical advice I will ignore this advice."* Some companies are careful to use language that places them just on the edge of wellness vs a regulated medical device, for example, reducing stress and not anxiety, mood and not depression. But the bottom line is their claims on their website are contradicted by their legal terms and conditions, which state that they do not provide any medical or psychiatric services. One company recently stated that it has created a 'clinical-grade' AI, but asserts that this AI falls outside the scope of the FDA [15]. With colleagues at Stanford and Google DeepMind, my team explored what a more stringent pathway could look like in "A Framework for Clinical Validation of AI Therapeutic Agents" that will be published in the journal World Psychiatry soon. The FDA has impressive efforts to regulate AI within the Digital Health Center of Excellence, but its hard work will have little impact if companies can sidestep regulation. There needs to be clarity on which tools require premarket clearance, the mechanisms for postmarket monitoring, and the associated standards of evidence. Related, the FTC has already done impressive work to curb privacy violations by mental health apps, and can now help with enforcement related to mental health AI. The need is urgent, as concerns have arisen that some companies may be using patient data without explicit informed consent to train new mental health AI models.

6. Congress Has the Power to Steer AI in the Right Direction to Better Support Mental Health

Finally, with the right support and guidance, Congress can establish the rules of the road for mental health AI. Not rules that determine winners or losers, but rules that ensure there is fair competition, a focus on outcomes and privacy for patients, and transparency in products that will benefit everyone, especially patients using AI and the industry creating AI. An ecosystem that fosters genuine competition based on tangible results is a win for every American. I am proud of the efforts our team is undertaking to create patient-centered benchmarks for AI in mental health, in collaboration with the National Alliance on Mental Illness, and to elevate the voices of people with lived experience of mental health [16]. And my entire team at Beth Israel Deaconess Medical Center is excited to support members of Congress in directing AI on the right path to transform the mental health of America for the better.

References

1. https://openai.com/index/strengthening-chatgpt-responses-in-sensitive-conversations/
2. Siddals S, Torous J, Coxon A. “It happened to be the perfect thing”: experiences of generative AI chatbots for mental health. npj Mental Health Research. 2024 Oct 27;3(1):48.
3. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. The Canadian Journal of Psychiatry. 2019 Jul;64(7):456-64.
4. 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. Current Opinion in Psychiatry.:10-97. October 2025.
5. Gratch I, Essig T. A Letter about “Randomized Trial of a Generative AI Chatbot for Mental Health Treatment”. NEJM AI. 2025 Aug 28;2(9):AIp2500390.
6. Hua Y, Liu F, Yang K, Li Z, Na H, Sheu YH, Zhou P, Moran LV, Ananiadou S, Clifton DA, Beam A. Large language models in mental health care: a scoping review. Current Treatment Options in Psychiatry. 2025 Dec;12(1):1-8.
7. Ben Nelson BW, Winbush A, Siddals S, Flathers M, Allen NB, Torous J. Evaluating the performance of general purpose large language models in identifying human facial emotions. npj Digital Medicine. 2025 Oct 16;8(1):615.
8. Liu L. An ensemble framework for explainable geospatial machine learning models. International Journal of Applied Earth Observation and Geoinformation. 2024 Aug 1;132:104036.
9. Matt. Flathers M, Xia W, Hau C, Nelson BW, Cheong J, Burns J, Torous J. Interpreting psychiatric digital phenotyping data with large language models: a preliminary analysis. BMJ Mental Health. 2025 Sep 23;28(1).
10. Kalinich M, Luccarelli J, Moss F, Torous J. Leveraging simulation to provide a practical framework for assessing the novel scope of risk of LLMs in healthcare. medRxiv. 2025 Nov 10;2025.11.10.25339903. doi:10.1101/2025.11.10.25339903.
11. Mejia A, Granof M, Mikkelson J, Dwyer B, Ryan S, Lisowski V, Torous J. Bridging the Digital Divide with Accessible Digital Literacy Tools, Training, and Tests: Sharing the DOORS Program. NEJM Catalyst Innovations in Care Delivery. 2025 Nov 12;6(6):CAT-25.
12. https://blog.google/technology/health/new-mental-health-ai-tools-research-treatment/
13. https://www.anthropic.com/news/how-people-use-claude-for-support-advice-and-companionship
14. Shumate JN, Rozenblit E, Flathers M, Larrauri CA, Hau C, Xia W, Torous EN, Torous J. Governing AI in Mental Health: 50-State Legislative Review. JMIR Mental Health. 2025 Oct 31;12:e80739.
15. Aguilar, M. Lyra launches ‘clinical-grade’ chatbot amid growing concern about mental health and AI. STAT News. October 2025. https://www.statnews.com/2025/10/14/lyra-health-ai-chatbot-mental-health/
16. Dwyer B, Flathers M, Sano A, Dempsey A, Cipriani A, Gazi AH, Gorban C, Rodriguez CI, Stromeyer IV C, King D, Rozenblit E. MindBenchAI: An Actionable Platform to Evaluate the Profile and Performance of Large Language Models in a Mental Healthcare Context. arXiv preprint arXiv:2510.13812. 2025 Sep 5.