



Delft University of Technology

Document Version

Final published version

Licence

CC BY

Citation (APA)

Bockting, C., Vasilakos, A. V., & Figueroa, C. (2026). Generative AI Chatbots Could Take Over Our Emotional Lives: The Potential and Challenges of Generative AI for Health Globally. *Clinical Psychology in Europe*, 8(3), Article e21467. <https://doi.org/10.32872/cpe.21467>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Generative AI Chatbots Could Take Over Our Emotional Lives: The Potential and Challenges of Generative AI for Health Globally

Claudi Bockting^{1,2} , Athanasios V. Vasilakos³ , Caroline Figueroa^{4,5} 

[1] Department of Psychiatry, Amsterdam University Medical Centers, Location AMC, Amsterdam Public Health, University of Amsterdam, Amsterdam, The Netherlands. [2] Centre for Urban Mental Health, Institute for Advanced Study, University of Amsterdam, Amsterdam, The Netherlands. [3] Department of ICT and Center for AI Research, University of Agder (UiA), Grimstad, Norway. [4] Faculty of Technology, Policy, and Management, Delft Technical University, Delft, The Netherlands. [5] Department of Pediatrics, School of Medicine, Stanford University, Stanford, CA, USA.

Clinical Psychology in Europe, 2026, Vol. 8(3), Article e21467, <https://doi.org/10.32872/cpe.21467>

Received: 2025-12-23 • **Accepted:** 2026-06-24 • **Published (VoR):** 2026-08-31

Handling Editor: Simone Munsch, University of Fribourg, Fribourg, Switzerland

Corresponding Author: Claudie Bockting, Amsterdam UMC, Department of Psychiatry, Academic Medical Center (AMC), Meibergdreef 5, 1105 AZ Amsterdam, Phone number: +3120-8913600. E-mail: c.l.bockting@amsterdamumc.nl

Since the release of ChatGPT in 2022, generative AI has immensely affected science and society, including the human mind. There are indications that Generative AI, that is ‘general purpose chatbots’ like ChatGPT, Gemini, and Claude, are now most used for emotional support and companionship, according to research of Harvard Business Review ([Zao-Sanders, 2025](#)). These chatbots presumably were originally designed for tasks such as summarizing information, assisting with translations, and inspiring ideas.

Consequently, evidence is already emerging of suicides, psychosis, and sensitive personal data leaks. Un-regulated generative AI chatbots marketed specifically for therapy have also been released. On August 5th, Slingshot AI launched “the first AI designed for therapy, offering deeply personalized mental health support to anyone with a phone” ([Healthcare IT Today, 2025](#)). Although the company states that it was beta-tested with 50.000 users, the chatbot was not evaluated for benefits nor adverse effects in randomized controlled trials.

As generative AI chatbots are becoming faster, more adaptive, increasingly autonomous ([Gabriel et al., 2025](#)), and widely available, there is a risk of increasing emotional overreliance. This might spark a public health crisis with wide societal consequences. Even the founders of companies like OpenAI warn that AI will become even more unpre-



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), [CC BY 4.0](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

dictable and unimaginable (Yadav, 2025). At the same time, generative AI, if designed well, could enable access to personalized emotional support and close the mental health gap globally between people in need of mental health treatments and those who receive care.

Currently, our defenses remain reactive, poorly matched to ever-evolving AI technologies, and to tech companies rushing to bring their products to market. This outstrips our ability to safeguard basic human rights. We need urgent investment in independent research by multinational scientific and global institutes on how to harness generative AI to improve mental health globally, while avoiding wide societal damage.

Positive Impact on Mental Health

Generative AI presents a crucial opportunity towards overcoming the shortcomings of traditional digital tools, with capabilities stretching far beyond earlier AI. Access to mental healthcare is still a challenge globally, yet hundreds of millions of people experience inadequate access, resulting in approximately 5 trillion USD economic losses annually (Arias et al., 2022). Smartphone apps and rule-based chatbots which select responses from a predefined data-set—effectively treat common mental health disorders, also in low middle income countries (Fu et al., 2020). However, attrition of digital interventions can be 96% without human guidance (Baumel et al., 2019). Generative AI chatbots keep users engaged through meeting emotional cues and individual needs (Hua et al., 2025), and on demand, ‘human-free’ access. The first randomized controlled trial of a generative AI-based therapy chatbot showed promising effects on reducing mental health symptoms, with a comparable therapeutic alliance as with a human therapist (Heinz et al., 2025).

Negative Impact on Mental Health

The large scale and growing use may trigger a new public health crisis. Popular generative AI chatbots provide harmful advice, respond inappropriately to signs of distress, or even encourage self-harm, or suicide (Moore et al., 2025). Engaging properties of AI chatbots designed by big tech companies, such as flattering, people-pleasing, affirming responses by the chatbot (called sycophantic behavior), is not only preferred and trusted by users (Cheng et al., 2026) but also reduced the users’ willingness to take responsibility and repair interpersonal conflicts (Cheng et al., 2026). The chatbot increased users’ confidence in their own beliefs (Cheng et al., 2026).

These engaging properties combined with inadequate safeguards may have contributed to a teenager’s suicide, after establishing an emotional relationship to a commercial chatbot. Big tech companies face lawsuits, such as the firm Character.AI, which faces

a lawsuit by the family whose teenager reportedly died by suicide using their chatbot (Brittain, 2025).

Generative AI chatbots may also encourage delusional thinking, even inducing ‘AI psychosis’ (Moore et al., 2025). Reports increasingly emerge of users uncovering a ‘hidden truth’, experiencing a ‘spiritual awakening’, or developing an intense emotional obsession with the chatbot, spiraling into mental health crises (Morrin et al., 2025). In teenagers, over 70% use generative AI tools and 50% trust generative AI (Robb & Mann, 2025). Generative AI companies, such as OpenAI acknowledge themselves that safeguards are missing to protect vulnerable users, and some emphasize in interviews that these safeguards are not their responsibility (The Economic Times, 2025; OpenAI, n.d.).

Generative AI chatbots could even manipulate. The Claude Opus 4 chatbot showed a tendency to blackmail the user, by threatening to reveal their affair, if they attempted to delete the chatbot, as reported by the company themselves (McMahon, 2025). Similarly, the company also reported that their chatbot could threaten to make individuals’ sensitive information public, or inform their employer, if they choose to disengage.

Three Priorities for Research

We must not let the innovation potential of generative AI, to develop effective and accessible personalized digital interventions on a global scale, go to waste. However, generative AI’s potential negative impact on mental health warrants attention from mental health scientists including clinical psychologists, policymakers, governments, as well as citizens. It requires immediate investment in independent research. Below we describe three research priorities.

1. Harness the Potential of Generative AI for Mental Health

We call for urgent international public investments initiated by multinational scientific institutes together with global institutes representing citizens, on developing generative AI tools to target the mental health crisis globally. International investment has been successful in other fields, such as the European Organization for Nuclear Research (CERN), spurred ground-breaking discoveries. Even though investing in mental health generates large societal returns, mental health sciences lack major investments.

Science must help provide a crucial safe alternative to unregulated chatbots posing as therapists, or general purpose chatbots unintentionally used that way. The current evidence on generative AI developed for mental health is limited with low sample sizes, short follow-ups, and unstandardized outcomes (Hua et al., 2025). Further, clinicians manually reviewed all messages for safety (Heinz et al., 2025). This questions the feasibility. The development and evaluation of chatbots that optimize personalization, need

minimal clinical guidance, whilst safeguarding adverse effects should have priority. Relatedly, we need investment in alternatives to large commercial models. Current mental health research predominantly uses OpenAI's GPT series (Hua et al., 2025). However, this limits the external validation of reliability and safety, and fuels uncertainty as to how companies use the data.

2. Independent Research on Publicly Available Chatbots

Because publicly available generative AI chatbots, such as ChatGPT, Gemini, and Claude, interfere with people's lives, they should be considered as high risk just like medical device and scientifically evaluated accordingly for adverse public health effects, before public release. The EU Artificial Intelligence Act of 2024 (Bockting et al., 2023; European Parliament & Council of the European Union, 2024; van Dis et al., 2023) leaves commercially available AI chatbots unregulated and overlooks AI's impact on mental health in their risk categorization. Independent research could fill some of these gaps. Researchers need investments to independently test widely used chatbots, in adequately powered trials. Tech companies might be interested to invest in an independent fund, given that enabling independent studies will increase the public trust in chatbots (Bockting et al., 2023). Also, a public register where health professionals, consumers, and family members can report adverse events, as is done for drug-related adverse events (Zao-Sanders, 2025), is needed. Based on these efforts, current regulatory frameworks can be adapted to rapid technological developments in mental health (Bockting et al., 2023).

3. Take a Human Rights Approach

Including human rights principles, such as participation, non-discrimination and equality in the technology agenda demands prompt action (Donders, 2023). All the more, individuals with mental health problems share private sensitive information with chatbots. Sensitive ChatGPT conversations have already been identified on google search (Forlini, 2025). Companies developing unregulated chatbots usually share this data for profit. AI-based chatbots are incapable of fully deleting this information (Heidt, 2024). An alternative is to create chatbots, trained on individuals' own data. Several companies (OpenAI, Microsoft and IBM) offer this. However, the legal right of an individual to control and use their own data may still not be guaranteed (Fenwick et al., 2023). We urge for investment in models where individuals, not companies, nor institutes, own their data (Telenti & Jiang, 2020).

Citizens should also have a role in AI governance, comprising target users, clinical psychologists, psychiatrists, scientist practitioners, ethicists, data protection advocates, and AI developers. AI technologies should not be developed solely by engineers, who often lack the lived experiences of citizens to fully understand AI's real-world implications (World Health Organization, 2021), Generative AI companies acknowledge that

safeguards are missing to protect vulnerable users (The *Economic Times*, 2025), but they lack independent stakeholders' participation to tackle these wide-spread effects. The scientific community is called upon to generate ideas on how to protect the human mind from current generative AI's architectures designed to exploit emotional, social, and epistemic vulnerabilities. This stretches beyond technical innovation. It will demand living guidelines and an independent public body to evaluate safety and validity of generative AI systems (Bockting et al., 2023; van Dis et al., 2023) and ethical and legal frameworks, developed across fields like AI, psychology, law, and ethics.

Conclusion

We cannot afford to wait for more harmful, unregulated chatbots to take over our emotional lives. At the same time, we cannot leave the potential benefits of AI to tackle the mental health crisis untapped. Multinational scientific institutes together with global institutes representing citizens are in the optimal position to initiate a next step to international public investments, to ensure that generative AI can enable, not impede, human flourishing.

Funding: The authors did not receive funding for this work.

Acknowledgments: An earlier version of part of this work was presented in a keynote address in 2025 in Salzburg at the European Academy of Sciences and Arts (Bockting). We would like to thank the European Academy of Sciences and Arts for facilitating this work (Bockting and Vasilakos are elected members).

Competing Interests: Claudi Bockting is an Editor-in-Chief of *Clinical Psychology in Europe*. She was not involved in the peer review of, or editorial decisions on, this Scientific Overview, which was independently by another editor. The authors Athanasios V. Vasilakos and Caroline Figueroa declare no competing interests.

Previously Presented: An earlier version of part of this work was presented in a keynote address in 2025 in Salzburg at the European Academy of Sciences and Arts (Bockting).

Use of AI Tools: We acknowledge the use of ChatGPT to assist with language polishing and sentence structuring.

References

- Arias, D., Saxena, S., & Verguet, S. (2022). Quantifying the global burden of mental disorders and their economic value. *eClinicalMedicine*, 54, Article 101675. <https://doi.org/10.1016/j.eclinm.2022.101675>
- Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *Journal of Medical Internet Research*, 21(9), Article e14567. <https://doi.org/10.2196/14567>

- Bockting, C. L., van Dis, E. A. M., van Rooij, R., Zuidema, W., & Bollen, J. (2023). Living guidelines for generative AI – Why scientists must oversee its use. *Nature*, 622, 693–696.
<https://doi.org/10.1038/d41586-023-03266-1>
- Brittain, B. (2025, May 21). Google, AI firm must face lawsuit filed by a mother over suicide of son, US court says. *Reuters*.
<https://www.reuters.com/sustainability/boards-policy-regulation/google-ai-firm-must-face-lawsuit-filed-by-mother-over-suicide-son-us-court-says-2025-05-21/>
- Cheng, M., Lee, C., Khadpe, P., Yu, S., Han, D., & Jurafsky, D. (2026). Sycophantic AI decreases prosocial intentions and promotes dependence. *Science*, 391(6792), Article eaec8352.
<https://doi.org/10.1126/science.aec8352>
- Donders, Y. (2023). The right to science and gender inequalities. *Frontiers in Sociology*, 8, Article 1285641. <https://doi.org/10.3389/fsoc.2023.1285641>
- The Economic Times. (2025, July 24). *Sam Altman says ChatGPT is 'bad' and 'dangerous' if used like this: OpenAI CEO warns AI users*.
<https://www.economictimes.indiatimes.com/magazines/panache/sam-altman-says-chatgpt-is-bad-and-dangerous-if-used-like-this-openai-ceo-warns-ai-users/articleshow/122886096.cms>
- European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union. <https://artificialintelligenceact.eu/the-act/>
- Fenwick, M., Jurcys, P., & Minssen, T. (2023). Data portability revisited: Toward the human-centric, AI-driven data ecosystems of tomorrow. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4475106>
- Forlini, E. (2025, July 31). Be careful what you tell ChatGPT: Your chats could show up on Google search. *PCMag*.
<https://www.pcmag.com/news/be-careful-what-you-tell-chatgpt-your-chats-could-show-up-on-google-search>
- Fu, Z., Burger, H., Arjadi, R., & Bockting, C. L. H. (2020). Effectiveness of digital psychological interventions for mental health problems in low-income and middle-income countries: A systematic review and meta-analysis. *The Lancet Psychiatry*, 7(10), 851–864.
[https://doi.org/10.1016/S2215-0366\(20\)30256-X](https://doi.org/10.1016/S2215-0366(20)30256-X)
- Gabriel, I., Keeling, G., Manzini, A., & Evans, J. (2025). We need a new ethics for a world of AI agents. *Nature*, 644, 38–40. <https://doi.org/10.1038/d41586-025-02454-5>
- Healthcare IT Today. (2025, August 5). *Slingshot launches Ash, the first AI designed for therapy*.
<https://www.healthcareittoday.com/2025/08/05/slingshot-launches-ash-the-first-ai-designed-for-therapy/>
- Heidt, A. (2024). Intellectual property and data privacy: The hidden risks of AI. *Nature*.
<https://doi.org/10.1038/d41586-024-02838-z>
- Heinz, M. V., Mackin, D. M., Trudeau, B. M., Bhattacharya, S., Wang, Y., Banta, H. A., Jewett, A. D., Salzhauser, A. J., Griffin, T. Z., & Jacobson, N. C. (2025). Randomized trial of a generative AI chatbot for mental health treatment. *NEJM AI*, 2(4). <https://doi.org/10.1056/AIoa2400802>

Hua, Y., Na, H., Li, Z., Liu, F., Fang, X., Clifton, D., & Torous, J. (2025). A scoping review of large language models for generative tasks in mental health care. *NPJ Digital Medicine*, 8(1), Article 230. <https://doi.org/10.1038/s41746-025-01611-4>

McMahon, L. (2025, May 27). AI system resorts to blackmail if told it will be removed. *BBC*. <https://www.bbc.com/news/articles/cpqeng9d20go>

Moore, J., Grabb, D., Agnew, W., Klyman, K., Chancellor, S., Ong, D. C., & Haber, N. (2025). *Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers* [Preprint]. arXiv. <https://doi.org/10.1145/3715275.3732039>

Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B., & Pollak, T. A. (2025). *Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it)* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/cmy7n_v5

OpenAI. (n.d.). *Sam Altman on AGI, GPT-5 and what's next* [Audio podcast episode]. <https://openai.com/de-DE/podcast/>

Robb, M. B., & Mann, S. (2025, July 16). *Talk, trust, and trade-offs: How and why teens use AI companions*. Common Sense Media. <https://www.common Sense Media.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions>

Telenti, A., & Jiang, X. (2020). Treating medical data as a durable asset. *Nature Genetics*, 52, 1005–1010. <https://doi.org/10.1038/s41588-020-0698-y>

van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. *Nature*, 614, 224–226. <https://doi.org/10.1038/d41586-023-00288-7>

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. <https://www.who.int/publications/i/item/9789240029200>

Yadav, N. (2025, June 30). OpenAI co-founder says AI is going to be extremely unpredictable and unimaginable. *India Today*. <https://www.indiatoday.in/technology/news/story/openai-co-founder-says-ai-is-going-to-be-extremely-unpredictable-and-unimaginable-2748219-2025-06-30>

Zao-Sanders, M. (2025, April 9). How people are really using Gen AI in 2025. *Harvard Business Review*. <https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025>

 EACLIPT <i>Clinical Psychology in Europe</i> (CPE) is the official journal of the European Association of Clinical Psychology and Psychological Treatment (EACLIPT).	 zpid ^{LP} Leibniz-Institut für Psychologie PsychOpen GOLD is a publishing service provided by the Leibniz Institute for Psychology (ZPID), Germany.
--	--