



This article is available to subscribers. [Subscribe now.](#)











**SUBSCRIBE NOW**

ALREADY A SUBSCRIBER? [SIGN IN.](#)

[ORIGINAL ARTICLE](#)



# Randomized Trial of a Generative AI Chatbot for Mental Health Treatment

**Authors:** [Michael V. Heinz, M.D.](#) , [Daniel M. Mackin, Ph.D.](#) , [Brianna M. Trudeau, B.A.](#) , [Sukanya Bhattacharya, B.A.](#) , [Yinzhou Wang, M.S.](#) , [Haley A. Banta](#) , [Abi D. Jewett, B.A.](#) , [Abigail J. Salzhauer, B.A.](#) , [Tess Z. Griffin, Ph.D.](#) , and [Nicholas C. Jacobson, Ph.D.](#)  [Author Info & Affiliations](#)

Published March 27, 2025 NEJM AI 2025;2(4) DOI: 10.1056/AIoa2400802  
VOL. 2 NO. 4 Copyright © 2025



Evaluating clinical AI:  
Be empowered with the  
evidence you need.

**SUBSCRIBE NOW**



Have questions about this content? [Try AI Companion.](#)



# Abstract

## BACKGROUND

Generative artificial intelligence (Gen-AI) chatbots hold promise for building highly personalized, effective mental health treatments at scale, while also addressing user engagement and retention issues common among digital therapeutics. We present a randomized controlled trial (RCT) testing an expert-fine-tuned Gen-AI-powered chatbot, Therabot, for mental health treatment.

## METHODS

We conducted a national, randomized controlled trial of adults (N=210) with clinically significant symptoms of major depressive disorder (MDD), generalized anxiety disorder (GAD), or at clinically high risk for feeding and eating disorders (CHR-FED). Participants were randomly assigned to a 4-week Therabot intervention (N=106) or waitlist control (WLC; N=104). WLC participants received no app access during the study period but gained access after its conclusion (8 weeks). Participants were stratified into one of three groups based on mental health screening results: those with clinically significant symptoms of MDD, GAD, or CHR-FED. Primary outcomes were symptom changes from baseline to postintervention (4 weeks) and to follow-up (8 weeks). Secondary outcomes included user engagement, acceptability, and therapeutic alliance (i.e., the collaborative patient and therapist relationship). Cumulative-link mixed models examined differential changes. Cohen's d effect sizes were unbounded and calculated based on the log-odds ratio, representing differential change between groups.

## RESULTS

Therabot users showed significantly greater reductions in symptoms of MDD (mean changes: -6.13 [standard deviation {SD}=6.12] vs. -2.63 [6.03] at 4 weeks; -7.93 [5.97] vs. -4.22 [5.94] at 8 weeks;  $d=0.845-0.903$ ), GAD (mean changes: -2.32 [3.55] vs. -0.13 [4.00] at 4 weeks; -3.18 [3.59] vs. -1.11 [4.00] at 8 weeks;  $d=0.794-0.840$ ), and CHR-FED (mean changes: -9.83 [14.37] vs. -1.66 [14.29] at 4 weeks; -10.23 [14.70] vs. -3.70 [14.65] at 8 weeks;  $d=0.627-0.819$ ) relative to controls at postintervention and follow-up. Therabot was well utilized (average use >6 hours), and participants rated the therapeutic alliance as comparable to that of human therapists.

## CONCLUSIONS

This is the first RCT demonstration of a fully Gen-AI therapy chatbot for treating clinical-

Have questions about this content? Try AI Companion.



level mental health symptoms. The results were promising for MDD, GAD, and CHR-FED symptoms. Therabot was well utilized and received high user ratings. Fine-tuned Gen-AI chatbots offer a feasible approach to delivering personalized mental health interventions at scale, although **further research with** larger clinical samples is needed to confirm their effectiveness and generalizability. (Funded by Dartmouth College; ClinicalTrials.gov number, [NCT06013137](#).)

Already have an account? [Sign In](#)

**CONTINUE READING THIS ARTICLE. SELECT AN OPTION BELOW:**

## This Article is Available to Subscribers

**Subscribe Now for Full Access**

**Buy this article for \$55.**

Are you a member of an institution such as a university or hospital?

[Learn more about Institutional Access](#)

[Access through your institution](#)

### NOTES

A [data sharing statement](#) provided by the authors is available with the full text of this article.

All participants provided informed written consent prior to their participation. The Dartmouth Hitchcock Institutional Review Board approved the research protocol. The trial was preregistered through ClinicalTrials.gov as [NCT06013137](#).

Supported by Dartmouth College.

[Disclosure forms](#) provided by the authors are available with the full text of this article.

We are extraordinarily grateful for the efforts of the many dedicated people who made Therabot possible. We thank those who contributed in meaningful ways to the creation and curation of training data, including Victor A. Moreno, Chloe S. Park, Jimena Abejon Fuertes, Jonathan J.

Cartwright, Anna C. St. Jean, Erica L. Simon, Isabel R.

Hillman, Enoc A. Garza, Alexandr

Haddox, Mingyue Zha, Camilla 1 [Have questions about this content? Try AI Companion.](#)

Song, Cameron M. Hasund, Avijit Singh, Daniel W. Smith,



Rachel E. Quist, Kaitlyn I. Romanger, Chaehyun Lee, Anjali G. Dhar, Ivy N. Mayende, Eleanor M. Rodgers, Rachel Zhang, Jenny Song, Veronica E. Abreu, Russell T. Rapaport, Mary M. Basiliou, Sofia M. Yawand-Wossen, Nathan J. Kung, Jenny Y. Oh, Ashna J. Kumar, Eda Naz Gokdemir, Janelle E. Annor, Ganza, Belise Aloysie Isingizwe, Chloe N. Malave, Ezinne E. Anozie, Tara L. Karim, Nhi D. Nguyen, Krista E. Schemitsch, Helen M. Young, Mia G. Russo, Rachel E. Quist, Tonya I. Tolino, Mckenzi B. Popper, Daniel G. Amoateng, and Dr. Seo Ho (Michael) Song. We thank those who contributed in meaningful ways to the software development, including Jason Kim, John F. Keane, Dr. George D. Price, Dr. Matthew D. Nemesure, Ore E. James, Caroline C. Hall, Brendan W. Keane, Lisa Aeri Oh, Ly H. Nguyen, Dr. William R. Haslett, Vivian N. Tran, Alexander M. Ye, Atziri Enriquez, Sarah M. Chacko, Sofia Jayaswal, D.J. M. Matusz, Jose Hernandez Barbosa, Alyssia M. Salas, Ella J. Gates, and Tianwen Chen. We thank the team from Amazon Web Services (AWS), especially Stefan Mationg and Dr. Jianjun Xu, who provided valuable technical support for Therabot.

---

## SUPPLEMENTARY MATERIAL

---

### Supplementary Appendix (aioa2400802\_appendix.pdf)

[DOWNLOAD](#) 333.28 KB

### Disclosure Forms (aioa2400802\_disclosures.pdf)

[DOWNLOAD](#) 606.88 KB

### Data Sharing Statement (aioa2400802\_data-sharing.pdf)

[DOWNLOAD](#) 28.80 KB

#### BROWSE

Research  
Review  
Case Study  
Commentary  
Perspective  
Other  
Current Issue  
Issue Index

#### RESOURCES

Events  
Author Center  
Submit a Manuscript  
Subscribers  
Institutional Administrators  
Media  
Business Model  
Agents  
Permissions & Licensing  
Reprints  
NEJM CareerCenter

#### ABOUT US

About NEJM AI  
NEJM Group  
Editors & Publishers  
Editorial Policies  
FAQs  
Help  
Contact Us





#### SUBSCRIPTIONS

Subscribe  
Renew  
Activate Subscription  
Create Account  
Manage Account  
Pay Bill  
Institutional Sales  
Institution Administration Center

#### STAY CONNECTED

Contact Us  
Create Account  
Email Alerts  
Remote Access  
NEJM CareerCenter

#### FOLLOW US

 X  
 LinkedIn  
 YouTube  
 Bluesky

Have questions about this content? [Try AI Companion.](#)



**PUBLICATIONS**

[The New England Journal of Medicine](#)

[NEJM Clinician](#)

[NEJM Evidence](#)

[NEJM AI](#)

[NEJM Catalyst Innovations in Care Delivery](#)

NEJM AI is a product of NEJM Group, a division of the Massachusetts Medical Society.

Copyright © 2026 **Massachusetts Medical Society**. All rights reserved, including those for text and data mining, AI training, and similar technologies. ISSN 2836-9386.

[Copyright](#) | [Terms](#) | [Privacy Policy](#) | [Préférences de cookies](#)



Have questions about this content? [Try AI Companion.](#)

