Women in AI Ethics™

RESOURCE GUIDE
AI & MENTAL HEALTH

January 2025

Contributors: Leigh Goulbourne | Jeanette Schofield | Danita Delce
Editor: Mia Shah-Dand

AI & MENTAL HEALTH

January 31, 2025

Overview: The integration of artificial intelligence into healthcare systems - particularly for mental health diagnosis and support - raises profound ethical concerns that warrant careful examination. While AI technologies offer potential benefits, the current lack of robust accountability mechanisms and significant risk of harm have cast a shadow over their implementation in sensitive settings like mental health need greater scrutiny. Researchers caution against AI deployment without appropriate safeguards and there are a growing number of policy proposals to address patient privacy, algorithmic bias, and self-harm. More protective measures are essential for responsible innovation with AI in this critical field.

Our January 2025 Reading Circle explored this complex landscape of promising opportunities, ethical concerns, and accountability for AI mental health interventions. The curated research papers and articles included in this report offer a balanced perspective on the benefits and risks of AI usage in mental health.

For an informal discussion on other relevant topics in AI Ethics, join us for Women in AI Ethics™ (WAIE) reading circles on last Fridays at 11a ET.

Contributors: Leigh Goulbourne | Jeanette Schofield | Danita Delce

Editor: Mia Shah-Dand

TABLE OF CONTENTS

BACKGROUND: AI & MENTAL HEALTH

National Library of Science – Cureus: [Increased Screen Time as a Cause of Declining Physical, Psychological Health, and Sleep Patterns: A Literary Review](#) (2022)

Frontiers in Digital Health Journal: [Artificial intelligence in positive mental health](#)

Columbia Business School: [Macroeconomics of Mental Health](#) (2025)

THE GROWING USE OF AI IN MENTAL HEALTH

Increased screen time & stress

National Center for Health Statistics: [Daily Screen Time Among Teenagers: United States, July 2021–December 2023](#) (2024)

U.S. Surgeon General Advisory: [Parents under Pressure](#) (2024)

Disappearance of 'third places' & lack of community

The Brookings Institution: ["Third places" as community builders](#) (2016)

National Library of Science[: Closure of 'Third Places'? Exploring Potential Consequences for Collective Health and Wellbeing](#) (2019)

Lack of access to affordable mental healthcare services

AAMC Research and Action Institute: [Exploring Barriers to Mental Health Care in the U.S.](#) (2022)

Bankrate: [Bankrate's 2025 Annual Emergency Savings Report](#) (2025)

Presumption of privacy and comfort with digital technologies

Pew Research: [Part 2: Information Sharing, Friending, and Privacy Settings on Social Media](#) (2013)

Deloitte Insights: [To share or not to share](#) (2017)

Stanford Center for Human-Centered AI: [Rethinking Privacy in the AI Era: Policy Provocations for a Data-Centric World](#) (2024)

THE RISKS ASSOCIATED WITH AI USAGE IN MENTAL HEALTH

Manipulation and emotional dependence

Sage Journals - New Media & Society: [Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika](#) (2022)

University of Antwerp: [Manipulation and the AI Act: Large Language Model Chatbots and the Danger of Mirrors](#) (2025)

Lack of guardrails against misuse of private data

Research Gate: [Privacy Issues of AI](#) (2021)

SAGE Journal: [To chat or bot to chat: Ethical issues with chatbots in mental health](#) (2023)

Harvard Business School: [Chatbots and Mental Health: Insights into the Safety of Generative AI](#) (2023)

Cornell University: "[The Dark Side of AI Companionship: A Taxonomy of Harmful Algorithmic Behaviors in Human-AI Relationships](#)" (2025)

Model hallucinations

Cureus Journal of Medical Science "[Artificial Hallucinations in ChatGPT: Implications in Scientific Writing](#)" (2023)

Birmingham City University: [Cyber Risks of Machine Translation Critical Errors: Arabic Mental Health Tweets as a Case Study](#) (2024)

PREVENTION AND MITIGATION OF RISKS

Utrecht University: [Bias Discovery in Machine Learning Models for Mental Health](#) (2022)

National Library of Science - Cureus: [A Call to Address AI "Hallucinations" and How Healthcare Professionals Can Mitigate Their Risks](#) (2023)

Science Publishing Group: [Securing Well-Being: Exploring Security Protocols and Mitigating Risks in AI-Driven Mental Health Chatbots for Employees](#) (2024)

Taylor & Francis Learning, Media, & Technology Journal: ['No, Alexa, no!': designing child-safe AI and protecting children from the risks of the 'empathy gap' in large language models](#) (2024)

Computer Law & Security Review (Vol 54): [Better alone than in bad company: Addressing the risks of companion chatbots through data protection by design](#) (2024)

National Library of Science – JMIR: [Regulating AI in Mental Health: Ethics of Care Perspective](#) (2024)

Nature: [Detecting hallucinations in large language models using semantic entropy](#) (2024)

Technische Universitat Darmstadt: [Towards Privacy-aware Mental Health AI Models: Advances, Challenges, and Opportunities](#) (2025)

RELEVANT ARTICLES

NPR: [Eating disorder helpline takes down chatbot after it gave weight loss advice](#) (2023)

VICE: ['He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says](#) (2023)

Fast Company: [Gen AI & Gen Alpha The impacts of growing up in an innovation cycle](#) (2024)

News.com: ['I need to go outside': Young people 'extremely addicted' as Character.AI explodes](#) (2024)

People: [Teen, 14, Dies by Suicide After Falling in 'Love' with AI Chatbot. Now His Mom Is Suing](#) (2024)

The McGill International Review: [Where Have All the Great, Good Places Gone?: The Decline of the "Third Place"](#)

Urban: [How Third Places Contribute to Thriving Communities](#) (2024)

Bloomberg [Parents Have a Worse Relationship to Tech Than Their Kids](#) (2025)

CBS News: [Most Americans can't afford a $1000 emergency expense, report finds](#) (2025)

CONCLUSION

The use of AI in mental health is on the rise due to lack of affordable mental health resources, disappearing community spaces, and users' comfort with digital technologies among other reasons. There are a wide range of ethical issues related to use of AI in sensitive situations such as manipulation of vulnerable users and misuse of private user data. In response to growing awareness about the risks of AI usage, there are emerging mitigation strategies including privacy-aware protocols, hallucination-detecting algorithms, and child-safe design principles.

This resource list provides a snapshot of the AI & Mental Health landscape at a given point in time. We will continue to publish updated resource lists so that our members can stay updated on the latest developments and make informed decisions about development, deployment, and use of AI.

Support our work by signing up for paid subscription to our [newsletter on Substack](#).

You can find more information on Women in AI Ethics™ on [our website](#).

Join us on last Fridays at 11a ET for [AI Ethics Reading Circles](#).