

WHEN PEOPLE BECOME DATA POINTS: THE POTENTIAL IMPACT OF AI IN MENTAL HEALTHCARE

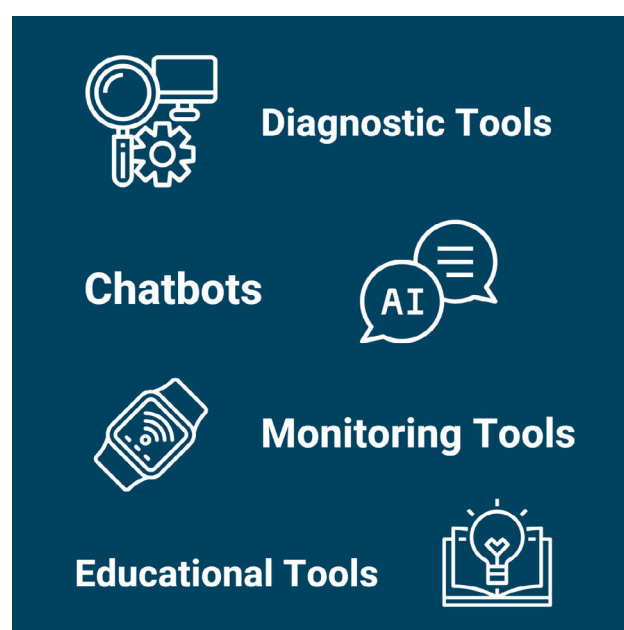


Image: © Pop Nukoonrat

Artificial intelligence (AI) holds great promise for transforming mental healthcare. From personalised treatment plans to early detection of mental health conditions, AI could make mental health services more accessible and effective. AI systems are already being developed for various applications, including diagnostic assessment, therapeutic support (such as chatbots), monitoring of mental health, and even educational tools aimed at promoting mental health literacy. These applications span both clinical and non-clinical settings, addressing a broad spectrum of conditions from depressive disorders and anxiety to non-medical issues, such as loneliness. And yet, despite all its potential, AI introduces a set of new risks that extend beyond individual patients to broader societal concerns, raising questions about equity, safety, and ethics.

To understand these complexities, it is useful to categorise AI applications in mental healthcare according to their primary purposes: screening, monitoring, diagnosis, treatment, and education. Screening tools often rely on natural language processing (NLP) or machine learning to detect signs of mental health conditions from social media activity or smartphone usage, or health records.¹ Monitoring tools, such as wearable devices, track biometric or behavioural data to identify patterns indicative of mental health changes.²

Diagnostic tools use neuroimaging, voice pattern analysis, and other forms of data to assist in identifying mental health conditions.³ Treatment applications include AI-driven cognitive behavioural therapy (CBT) chatbots or emotion recognition software for exposure therapy.⁴ Finally, educational tools provide psychoeducation and training for clinicians and patients using conversational AI or virtual learning environments.⁵ This typology not only underscores the versatility of AI in mental healthcare but also points towards the potential risks that arise with its adoption.



Potential risks can be identified at three levels. At the individual level, concerns include misdiagnosis, inappropriate treatment recommendations, and privacy breaches. At the collective level, issues such as biased datasets, accessibility barriers, and the marginalisation of vulnerable groups come to the forefront. At the societal level, challenges including over-surveillance, erosion of trust in healthcare, and the commodification of mental health services emerge, revealing broader implications for equity and justice. Addressing these risks requires a comprehensive and inclusive approach to AI development and governance. This policy paper applies a multi-level risk framework, based on Smuha's (2021) typology of individual, collective, and societal harms, to analyse these challenges and propose actionable solutions.⁶ By exploring these risks at multiple levels, this paper aims to provide pathways for responsibly integrating AI into mental healthcare where it provides health benefits for patients, while safeguarding individual rights, collective interests, and societal values.

INDIVIDUAL RISKS

- **Health risks:** Mental health tools pose significant health risks due to misdiagnoses, misleading feedback, lack of contextual understanding, and over-reliance on tool-generated advice, which can delay treatment, provide inappropriate responses in critical situations, and exacerbate users' distress, as illustrated by cases like *Woebot's*⁷ mishandling of abuse disclosures and *Eliza's*⁸ role in a suicide incident.
- **Privacy and autonomy:** AI mental health tools risk exposing sensitive user data to breaches and unauthorised sharing, undermining trust and autonomy through opaque recommendation processes and inadequate privacy safeguards, as demonstrated by *BetterHelp's*⁹ 2021 data-sharing controversy.

COLLECTIVE RISKS


- **Exclusion of marginalised population groups:** AI mental health tools often exclude marginalised populations due to biased training data, cultural and linguistic limitations, inaccessibility for low-income or disabled users, and a lack of personalised approaches, leading to disparities in mental health support and effectiveness.¹⁰
- **Exploitation of vulnerabilities:** Companies may exploit vulnerabilities of individuals living with mental health conditions by marketing unproven or overpriced AI tools or in-app purchases, risking health harms, undermining trust in mental healthcare, and leaving patients without effective support.
- **Medicalisation and reductionist disease conceptions:** The dominance of biological data in health records and the limitations of AI in capturing complex social and emotional aspects of mental health can lead to an overemphasis on biological determinants, medicalisation of everyday life, and an unjust focus on individual responsibility for mental health.¹¹
- **Additional risks for institutionalised people:** AI in institutional settings, for example in residential care facilities, prisons, or inpatient psychiatric care, risks amplifying the loss of autonomy and dignity for people living with mental health conditions by automating decision-making and standardising practices that are difficult to opt out of. Also, the field's history of paternalistic and derogatory 'best interest' motives for deciding for instead of with people, and exclusion and stigmatisation of people with psychosocial disabilities urges extra caution when using AI.¹²

SOCIETAL RISKS

- **Over-surveillance and dehumanisation:** AI-driven mental health tools risk eroding privacy through excessive data collection, fostering over-surveillance that deepens power imbalances, and may diminish human empathy in therapeutic contexts, leading to dehumanised care.
- **Erosion of trust in healthcare:** Errors, opacity, and unreliability in AI decision-making can undermine trust in mental healthcare systems, discouraging patients from seeking care and harming provider-patient relationships.
- **Increasing health disparities:** Biases in AI systems and systemic barriers exacerbate health disparities, perpetuating stereotypes, misdiagnoses, and inequitable access to care for marginalised populations, deepening societal inequalities.¹³

- **Efficiency over quality:** The prioritisation of cost-saving and time-saving AI technologies, driven by financial incentives, risks overshadowing considerations of care quality and patient well-being in mental healthcare.
- **Mental health AI as a profitable business:** The commodification of mental healthcare prioritises profit over patient needs, exploiting sensitive data for commercial gain while concentrating power in private entities, undermining equitable care and public trust.¹⁴



Addressing Individual Risks of AI in Mental Healthcare 	
1. Mandatory Evidence-Based Validation	<ul style="list-style-type: none"> • Introduce national certification requirements that mandate clinical trials for AI systems in mental healthcare before deployment, similar to medicine approvals. • Develop standardised protocols for testing the safety, efficacy, and reliability of AI tools, with independent oversight to ensure unbiased results.
2. Emergency Protocols for High-Risk Scenarios	<ul style="list-style-type: none"> • Implement legal mandates requiring AI systems to have “human-in-the-loop” mechanisms for handling emergency situations. • Develop a national registry of qualified mental health professionals available for immediate escalation when AI flags emergency situations.

3. Enhanced Privacy Protection	<ul style="list-style-type: none"> • Enact mental health-specific data protection laws that require explicit consent for data use and implement encryption standards for sensitive data storage and transmission. • Impose significant financial penalties for data breaches involving mental health information to deter misuse and negligence.
4. Transparency Certification Programmes	<ul style="list-style-type: none"> • Establish government-backed certification systems to verify AI tools' compliance with ethical and safety standards, including scientific evidence base, requiring clear labelling on approved tools. • Require AI developers to provide simplified, user-friendly disclosures on how their systems function, tailored for diverse literacy levels.

Addressing Collective Risks of AI in Mental Healthcare 

1. Diverse Dataset Mandates	<ul style="list-style-type: none"> • Enforce legal requirements for the use of datasets representing diverse demographics, including race, gender, socio-economic status, and geography. • Fund repositories of diverse mental health datasets, using strong privacy preserving techniques, which can be accessed under stringent conditions, in order to reduce barriers for smaller developers and ensure equity in AI system training.
2. Digital Inclusion Initiatives	<ul style="list-style-type: none"> • Subsidise access to validated mental health AI tools for underprivileged communities through public healthcare programmes. • Ensure public health clinics are equipped with AI tools designed for underserved populations, including those with disabilities or who do not speak the standard language.
3. Unethical Marketing	<ul style="list-style-type: none"> • Prohibit manipulative advertising that targets individuals, with regulatory bodies conducting regular audits of AI marketing practices and commercial modelling.
4. Accessibility Standards	<ul style="list-style-type: none"> • Develop and enforce accessibility guidelines for AI tools to ensure usability for individuals with disabilities, language barriers, or low digital literacy. • Introduce pricing caps for AI mental health tools and encourage open-source development through public grants and non-commercial incentives.

Addressing Societal Risks of AI in Mental Healthcare



1. Ethical Impact Assessments	<ul style="list-style-type: none"> Require developers to submit ethical impact assessments as part of the regulatory approval process, including evaluations of equity, fairness, and societal trust. Create standardised templates for these assessments, guided by mental health professionals, ethicists, and civil society groups.
2. Algorithmic Accountability Laws	<ul style="list-style-type: none"> Close loopholes that allow AI providers to circumvent regulation by using product disclaimers stating that the intended purpose is non-medical, while the functionalities of the product indicate a medical use case. Mandate routine audits of AI systems to ensure compliance with ethical and safety standards, with results disclosed in publicly accessible reports. Establish an independent regulatory body for monitoring and addressing complaints related to AI misuse in mental healthcare.
3. Community Advisory Boards	<ul style="list-style-type: none"> Form multi-stakeholder advisory boards to oversee the deployment of AI tools, ensuring alignment with societal values and the needs of marginalised communities.
4. Public Awareness Campaigns	<ul style="list-style-type: none"> Launch national education initiatives highlighting the appropriate use, limitations, and risks of mental health AI tools. Partner with schools, workplaces, and healthcare providers to distribute accessible educational materials
5. Preventing Commodification	<ul style="list-style-type: none"> Incentivise non-commercial AI development through public funding for tools addressing neglected areas of mental healthcare. Require AI developers to adhere to a “patient-first” ethos, prioritising quality of care over profit and responding to a demonstrable unmet need, together with strong public return on public investment principles, as conditions for public funding.

Suggested Methods

1. Stakeholder Engagement	<ul style="list-style-type: none"> Organise regular stakeholder forums involving developers, patients, clinicians, policymakers, and ethicists to review progress and update guidelines. Use participatory approaches, such as patient advisory groups, to ensure real-world applicability of AI tools in mental healthcare.
2. Bridging Research and Practice	<ul style="list-style-type: none"> Create a national fund to support the implementation of clinically effective applications, with proven efficacy in controlled settings, into healthcare settings and pathways. Develop a “sandbox” regulatory model allowing experimental deployments in supervised environments before full market release.
3. Monitoring and Oversight	<ul style="list-style-type: none"> Establish an independent committee to assess the long-term societal impact of mental health AI, publishing regular reports to inform adaptive regulations. Introduce metrics to evaluate the success of AI tools, including effects on healthcare disparities, patient outcomes, and ethical compliance.

REFERENCES

- Kevin W Jin et al, 'Artificial Intelligence in Mental Healthcare: An Overview and Future Perspectives' (2023). British Journal of Radiology 20230213. and: Abayomi Arowosegbe and Tope Oyelade, 'Application of Natural Language Processing (NLP) in Detecting and Preventing Suicide Ideation: A Systematic Review' (2023) 20 International Journal of Environmental Research and Public Health 1514.
- Alaa Abd-Alrazaq et al, 'Wearable Artificial Intelligence for Detecting Anxiety: Systematic Review and Meta-Analysis' (2023). Journal of Medical Internet Research e48754.
- Chang Su and others, 'Deep Learning in Mental Health Outcome Research: A Scoping Review' (2020) Translational Psychiatry.
- Alaa A Abd-alrazaq et al, 'An Overview of the Features of Chatbots in Mental Health: A Scoping Review' (2019). International Journal of Medical Informatics 103978.
- Fahad Alanezi, 'Assessing the Effectiveness of ChatGPT in Delivering Mental Health Support: A Qualitative Study' (2024) Journal of Multidisciplinary Healthcare.
- Nathalie A Smuha, 'Beyond the Individual: Governing AI's Societal Harm' (2021). Internet Policy Review.
- Zoha Khawaja and Jean-Christophe Bélisle-Pipon, 'Your Robot Therapist Is Not Your Therapist: Understanding the Role of AI-Powered Mental Health Chatbots' (2023). Frontiers in Digital Health 1278186.
- Imane El Atillah, '[AI Chatbot Blamed for "encouraging" Young Father to Take His Own Life](#)'. Euronews (31 March 2023).
- Leonardo Horn Iwaya and others, 'On the Privacy of Mental Health Apps' (2022). Empirical Software Engineering.
- Nii Tawiah and Judith P Monestime, 'Promoting Equity in AI-Driven Mental Health Care for Marginalized Populations' (2024) Proceedings of the AAAI Symposium Series.
- Jonah Bossewitch et al, 'Digital Futures in Mind: Reflecting on Technological Experiments in Mental Health & Crisis Support' (1 September 2022).
- '[Civil Commitment and the Mental Health Care Continuum: Historical Trends and Principles for Law and Practice](#)' (2 May 2021). Accessed 1 December 2024.
- Nicole Gross, 'A Powerful Potion for a Potent Problem: Transformative Justice for Generative AI in Healthcare' (2024). AI and Ethics.
- Zara Abrams, 'Monetizing Mental Health Technology' (2024) American Psychological Association, Monitor on Psychology.

Authors:

Hannah van Kolfschooten

Janneke van Oirschot

For more information:

Alice Beck: alice@haiweb.org

www.haiweb.org



Funded by
the European Union

This publication was funded by the European Union. Its contents are the sole responsibility of Health Action International and do not necessarily reflect the views of the European Union.

European
Artificial Intelligence
& Society Fund

This project has been supported by the European Artificial Intelligence & Society Fund (EAISF), a collaborative initiative of the Network of European Foundations (NEF). The sole responsibility for the project lies with the organiser(s) and the content may not necessarily reflect the positions of EAISF, NEF or European AI Fund's Partner Foundations.