Mental Health Europe

Equal rights.
Better mental health.
For all.

December 2024

# Artificial Intelligence in mental healthcare

This Mental Health Europe study has been co-authored by Piers Gooding*, Hannah van Kolfschooten** and Francesca Centola***.

*Associate Professor, La Trobe Law School, La Trobe University.

** Lecturer-researcher, Amsterdam Law School, University of Amsterdam, and AI consultant, Health Action International

*** Policy and Knowledge Officer, Mental Health Europe.

For any questions please reach out to: f.centola@mentalhealtheurope.org

# Contents

# Glossary

**'AI system'** is defined in Article 3 section 1 of the European Union (EU) Artificial Intelligence (AI) Act as:

"a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments."

This definition will be used in this report.

**Co-creation:** A collaborative approach involving all actors in mental health working together on an equal basis to develop and implement policies, services, programmes, research and communication that foster positive mental health according to a psychosocial model and human rights-based approach. (Mental Health Europe)

**Digital phenotyping** refers to the use of data from digital devices (such as smartphones and wearables) to identify behavioural patterns associated with particular states of mental health.

**Generative AI** refers to a subset of artificial intelligence focused on creating new content, ranging from text and images to audio, video, 3D models, and synthetic data. This field primarily uses machine learning algorithms, especially deep learning models, to discern patterns in training data and generate new outputs.

**People with lived experience:** people who experience or have experienced mental health issues. (Mental Health Europe)

# Summary

This study explores the opportunities, risks, and ethical considerations surrounding the use of artificial intelligence (AI) systems in mental healthcare and provides recommendations for their responsible implementation and regulation.

Healthcare forms one of the most popular sectors for AI deployment in the EU.[1] In mental healthcare, AI systems are used in diverse ways, from administrative tasks to communication platforms, professional decision support, digital therapies like chatbots and personal sensing technologies, and patient monitoring. Beyond healthcare settings, AI applications that use mental health-related data also extend to criminal justice, consumer products, education, and employment.

AI systems offer significant potential benefits, including improved accessibility to mental health support, particularly for underserved populations, and reducing administrative burdens in healthcare systems. Clinically, proponents advance AI for personalising treatments, improving diagnostic accuracy, and supporting timely interventions. Additionally, AI can accelerate mental health research and facilitate the development of novel therapeutic approaches, such as virtual reality therapies. Socially-

oriented opportunities include enhancing peer support and practitioner training, promoting networked collaboration, and fostering greater community connections.

However, AI systems in mental health also pose serious risks, which can be identified at three levels: individual, collective and societal[2]. At the individual level, concerns include safety risks, privacy violations and inadequate informed consent. Broader challenges include strengthening inequalities or creating new ones, oversurveillance, reinforcing individualistic views of mental health, depersonalisation of care, and diverting limited resources.

The report highlights key principles and possible courses of action for policymakers to address these risks, by balancing innovation with ethical responsibility and human-centered care in AI-driven mental health systems. It is argued that AI tools need to be developed with ethics, inclusivity, accuracy, safety and the genuine needs of end users in mind. Possible solutions mainly include robust regulation and oversight, transparency and explainability, as well as human rights-centric and co-creation approaches.

The active participation and leadership of the most affected communities - people with lived experience – should be regarded as a fundamental ethical and political principle guiding all discussions and decisions on digitalisation and mental health. Ensuring that they have an equal voice in shaping policies and innovation is essential to achieving fair and balanced outcomes and preventing the deepening of health inequalities in society.

## About this study

This study targets policymakers, service providers and any other stakeholder interested in the applications and impact of AI in mental healthcare.

Following an introductory background that sets Mental Health Europe's vision, the study provides an overview of the applications of AI in mental healthcare. It then assesses the opportunities and risks associated with these technologies, offering recommendations for policymakers on how to address the risks. The study highlights the role of lived experience in AI development and provides specific recommendations for civil society.

# Background

In 2022, Mental Health Europe published a Report Mental health in the digital age: Applying a human-rights based, psychosocial approach as compass. In that Report, Mental Health Europe put forward its vision for mental health in a digital world: digitalisation should not be considered as an end in itself, but as a means to a bigger end. A means towards a mentally healthier society, increased autonomy and agency over our own mental health, and better fitting mental healthcare.

We stressed the importance for new technologies to be applied not just because they are profitable or feasible, but because they respond to a real need. This is particularly important, considering the commercial interests at play. We advocated for a co-creation approach and a human rights framework to serve as the compass for any developments in digital mental health. Only with this collaborative approach – in every step of the process, from design to evaluation – can digital technologies align with real needs and work towards realising a vision of a society where everybody can fully enjoy their human rights and thrive.

Since 2022, many developments have taken place at the EU level, including the adoption of the AI Act, the first comprehensive regulation of AI by a major regulator anywhere. Entered into force in August 2024, this regulation aims to ensure that AI systems are safe and respect fundamental rights and values, to promote trust in AI technologies, to support innovation, and to enhance EU competitiveness in AI. The AI Act applies horizontally to all sectors in which AI is used, including healthcare.

The ambition of this Study is to focus on the impact of AI applications in mental healthcare, highlighting the risks specific to this sector and proposing measures to address them. By doing so, it lays the foundation for critical assessment of whether the AI Act is fit for purpose in the case of mental healthcare and how potential gaps can be addressed by policymakers. The study underscores the importance of collaboration between policymakers, people with lived experience, civil society organizations, and other relevant stakeholders in ensuring that these gaps are effectively tackled during the implementation phase of the regulation.

# Overview of AI applications in mental healthcare

In mental healthcare, AI systems are used in diverse ways. The tables below provide examples of AI applications, differentiating whether they target service users or mental health professionals.

## *AI applications for service users*

| Purpose | AI application | Examples |
|---|---|---|
| **Digital Therapies** | **Digital phenotyping**: AI analyses data from digital devices (e.g., smartphones, wearables) to identify behavioral patterns linked to mental health states (e.g., depression, anxiety). | MindLAMP2 (in which LAMP stands for Learn, Assess, Manage, Prevent) is an open source and freely available app which generates multiple, customisable data streams from a user (e.g. device motion, location activity) and offers activities that encourage mindfulness and self-reflection. MindLAMP2 has been used in a variety of mental health clinical and research use cases. |
| | **Chatbots**: Interactive AI tools acting as virtual counselors, mimicking therapeutic encounters, and providing advice. They may also guide users to access social services or support systems. | Woebot and Wysa are AI chatbots designed to mimic therapeutic interactions, offering mental health advice and coping strategies. |
| **Personalised Medicine** | AI leverages genetic, environmental, and behavioral data to create tailored treatment plans for individual service users. | IBM Watson Health processes patient data to recommend personalized treatment options for depression based on a combination of genetic and behavioral factors. |
| **Patient Monitoring and Control** | • AI can track individual health or crisis data (e.g., identifying those at risk of suicide or psychosis).<br>• Monitors medication adherence by flagging when patients stop taking prescribed treatments. | AI-Based Suicide Alerts at Facebook/Meta, i.e. pattern recognition software to detect users expressing suicidal intent. |

| | | |
|---|---|---|
| **Health Informatics for Service Users** | AI can support service users s in navigating healthcare systems, including:<br>• Managing personal health records.<br>• Assisting with regulatory reporting, form completion, and applying for social services or benefits using tools like ChatGPT. | MyChart helps service users manage their personal health records and communicate with providers. |
| **Education and Information Sharing** | AI can enhance mental health literacy by disseminating accurate information and combating stigma through platforms like apps or chatbots. | AI platforms like Ada Health provide mental health education and stigma-reducing content tailored to people. |

## *Uses for mental health professionals*

| Purpose | AI application | Examples |
|---|---|---|
| **Digital Therapies** | Some AI-integrated tools (e.g., digital phenotyping, chatbots) are recommended or prescribed by practitioners as part of treatment plans. | MindLAMP2, Woebot and Wysa. |
| **Professional Decision Support** | AI presents data to mental health practitioners to support decision-making or triggers actions without requiring input. | Tempus processes genomic and clinical data to assist mental health practitioners in creating personalized treatment plans. |
| **Patient and Population Monitoring** | • AI systems analyse data to identify at-risk individuals or population cohorts (e.g., suicide prevention).<br>• Track community health trends through administrative data or wearable device insights. | Crisis Text Line uses AI to prioritise high-risk cases based on language patterns in text messages.<br>AI systems like Babylon Health flag population health trends, such as increasing rates of depression in certain geographic areas. |
| **Support in Legal and Administrative Roles** | AI assists professionals in regulatory compliance, reporting, and analysing system-wide health outcomes. | DocuSign and similar AI tools assist professionals with regulatory documentation and compliance tasks. |

| | | |
|---|---|---|
| **Information Sharing and Administration** | • Managing electronic health records.<br>• Supporting clinical coding and appointment scheduling. | Epic Systems provides AI-driven clinical decision support integrated into electronic health record (EHR) systems to streamline record management and appointment scheduling; Zocdoc uses AI to match people with appropriate practitioners based on specialization, insurance, and availability. |
| **Communication Tools** | • Facilitating service-related communication, such as matching users with practitioners based on expertise, location, and other criteria.<br>• Providing plain language explanations of complex medical information for healthcare professionals. | MOST (Moderated Online Social Therapy) is a digital mental health platform for young people (12–25) that combines interactive therapeutic tools, a moderated peer network, and real-time support from clinicians and peer workers; Galaxy. AI's Medical Language Simplifier converts complex medical terminology into plain, understandable language. |

# Opportunities

Evidence supporting the benefits of AI systems in mental health context is limited (discussed later), as the field is largely at an experimental stage. Most commentary and research therefore consider what *opportunities* AI systems can offer. What is considered an opportunity will differ among individuals and groups—what one person or group see as an opportunity, may be seen by others as a misguided or undesirable aim.

It does not seem possible to neatly distinguish opportunities as primarily benefiting service users and mental health professionals, as they appear to be entwined. The same may be said of opportunities for society more generally. The primary opportunities for service users and mental health professionals noted in research on AI systems in mental healthcare include:

- **Improved accessibility of mental health support**, where chatbots and online platforms may help overcome geographical barriers and provide support to individuals in remote areas, people seeking support outside of typical working hours, or those who find it difficult to access traditional face-to-face services. Increased accessibility could particularly help underserved populations.[3] One example is the multilingual chatbot ChatPal, which was specifically developed to promote mental well-being among individuals living in sparsely populated areas, where traditional services are limited due to geographical distance.[4]
- **Potential to reduce administrative costs and address workforce shortages**: AI may assist mental health services, for example with administrative tasks such as scheduling appointments, managing service user flow, and generating reports, freeing up time to focus on providing direct support.[5]

From a more clinical perspective, opportunities discussed in research include:

- **Personalisation of treatment**: AI systems could analyse large volumes of data to identify patterns and predict how a person might respond to treatment, helping to create personalised responses tailored to individual needs.[6]
- **Timely support**: AI systems could help pre-empt when a person is going to experience crises, potentially facilitating timely support and preventing escalation. This aim would seem to require large amounts of personal information generated by wearables, electronic records, social media activity, and so on.[7]
- **Improved diagnostic accuracy**: AI systems have been used to improve diagnostic accuracy in some medical fields, such as mammography screening metrics.[8] Some proponents suggest this may be possible in the mental health context by analysing complex datasets and identifying subtle patterns to assist with diagnosis that humans might miss.[9]
- **Development of novel treatment approaches**: AI may facilitate the development of new therapeutic tools and interventions, such as virtual reality therapies. These techniques could offer alternative support options to complement, improve, or replace traditional support.[10]
- **Support for research and knowledge generation**: AI may accelerate mental health research by analysing large datasets and identifying patterns, such as risk factors, treatment targets, patterns for service use for particular socio-economic groups.[11]

Some have taken a more social approach to crisis support in seeking opportunities for AI systems, and suggest that a narrow focus on diagnosis and detection may be misguided. This social approach might characterise one key opportunity as **networked collaboration**, in which AI systems are used to improve social connection, peer support, or in the training and supervision of mental health practitioners.[12] Examples include developing better tools to help service providers enhance staff skills and empathetic understanding; networked, interactive media to engage support workers in their supervision and help to collectively improve; and technologies that facilitate service users to connect with peers and local or online communities. This social orientation often draws on work of people with lived experience of mental health issues and psychosocial disability.

These opportunities are presented with a degree of cautious optimism by most commentators and there may be some disagreement about precisely where the opportunities lie, which is to be expected during a period of experimentation.

# Risks

We will now present the risks noted in the literature. They entail harm at individual, as well as collective and societal level[13]. They jeopardise core values of healthcare, such as autonomy, dignity and trust[14].

## Safety risks

Health-related risks can occur because of errors or disinformation generated by or exacerbated by AI systems, as well as because of the inability of AI systems to understand various contextual factors. One significant concern is **bias in training data**. If this data is incomplete, unrepresentative, or reflects existing societal prejudices, the AI may perpetuate these biases, leading to inaccurate or discriminatory diagnoses.[15] For instance, if data used to train the AI underrepresents certain ethnic groups or socioeconomic backgrounds, the AI might misinterpret symptoms or behaviours common within those groups.

Another issue is **the potential for false positives and negatives.** AI can produce false positives, suggesting a particular response, such as flagging a suicide risk when it is not present, or false negatives, such as failing to identify a significant risk or crisis.[16] These errors have significant implications, potentially leading to unnecessary or inadequate service responses.

The proliferation of publicly available AI tools, like chatbots, introduces additional risks. These tools may provide **harmful advice or misinformation**, particularly when used for mental health support.[17] They might misinterpret user input, offer inaccurate information, or provide inappropriate responses, potentially creating a false sense of security for individuals seeking help.

Finally, **the inherent complexity of mental health poses a significant challenge for AI systems.** Mental health issues often manifest uniquely in individuals, shaped by subjective experiences, cultural contexts, environmental factors, and personal histories.[18] These nuances are difficult, if not impossible, to reduce to computational models, making it challenging for AI to accurately interpret individual needs or provide reliable support. This limitation is especially problematic in critical incidents, where accurate understanding and appropriate responses are crucial.

## Privacy

**Privacy risks** associated with using AI in mental healthcare stem from the **sensitive nature of mental health related-data** and the potential for harm if this information is mishandled or misused.

The increasing reliance on digital platforms and interconnected systems in health and social care has heightened the risk of **data breaches and unauthorised access**,[19] such as the largescale hacking of psychotherapeutic records affecting up to 30,000 people in Finland[20]. This is particularly concerning for AI systems that collect and analyse vast amounts of sensitive personal data, including medical records, therapy notes, and even social media activity.

Another significant issue is the **sharing of data with third parties.** There are growing concerns about AI systems passing on personal information to entities such as insurance companies, employers, or marketing firms without explicit consent. Such practices could result in discrimination, erode trust in mental healthcare services, and, in cases where informed consent has not been properly obtained, amount to misleading or deceptive conduct.

Additionally, the **lack of transparency and control** over how AI systems handle personal data presents a serious challenge. Individuals are often unaware of, or unable to control, how their information is collected, used, or shared. This lack of agency can heighten privacy concerns, discourage people from seeking support, and deter them from disclosing sensitive information necessary for effective support.

## Lack of informed consent

Informed consent is a fundamental principle of the protection of human rights in the healthcare sector.[21] AI systems pose a risk of **inadequate or compromised informed consent** when used in mental healthcare (and healthcare generally).

It can be hard for people, including service users and mental health practitioners, to **fully understand how AI systems work**, the kind of data they collect, and what could happen to that data. This lack of understanding makes it difficult to give truly informed consent, as people may not realise what they're agreeing to. Some ethicists have suggested that generative AI could in theory facilitate informed consent, if it can provide information that is at least more accurate, accessible, and trustworthy that

that offered by mental health practitioners.[22] Consent can be compromised where data collected by AI systems might also be used for **purposes other than those originally consented to,** including secondary use by government departments, or commercial data brokerage.[23]

A lack of transparent reporting on AI models undermines their replicability and hinders the identification of potential biases or errors.[24]

## New or amplified inequities

Concerns about the potential for **bias and discrimination** when using AI in mental healthcare include the use of **biased datasets.** AI algorithms learn from the data they are trained on. If these datasets contain biases or discriminatory content, the AI systems will likely inherit and reproduce them in their predictions and recommendations. While this may be true of existing mental health services, where known bias and inequities exist such as certain groups being treated less favourably than others, AI systems may worsen existing disparities and perpetuate them at a larger scale. The transparency issues noted earlier can also make it difficult to identify and address biases in AI systems[25].

AI systems may also create novel forms of discriminatory content, such as AI-generated images that reinforce harmful stereotypes. For example, one study found that using the word schizophrenia as a prompt to generate images, produced images depicting 'grotesque, unnatural facial features… blood, and expressions of horror'.[26] Similarly, a research group at Google demonstrated that social attitudes toward people with mental health conditions, describing them as bad and even violent, were encoded in AI systems designed to detect hate speech in written text.[27]

**Datasets used to train AI models often lack diversity** and do not adequately represent certain populations. This can lead to AI systems that are less accurate or effective for groups living with mental health issues, misinterpreting individual or community experiences, potentially leading to misdiagnosis, and inappropriate or inadequate care. This possibility is also exacerbated by the general failure to include people with lived experience in the creation, design, development and governance of technologies purportedly designed to benefit them.

In some cases, even when datasets are representative and high-quality, this will be inadequate to resolve **discrimination perpetuated by *human systems and institutions* that use AI systems**, such as insurance-based discrimination, where insurance premiums are raised if a person is predicted to have a mental health issue.

In addition to bias, AI tools can strengthen inequities if they are **not accessible** to certain groups (e.g., people with disabilities or low digital literacy).

## Depersonalisation of care

AI systems like chatbots, lack empathy, an element generally considered crucial for building trust and therapeutic relationships, even as they may be able to mimic it. Robots, chatbots, and animations can simulate emotions such as sadness, empathy, and curiosity, but these are merely programmed responses, and can mislead people into attributing human-like empathy to machines.

**AI systems face significant limitations in responding to human emotions.** AI systems are tools. While proficient at processing data that may help individuals or service providers illuminate aspects of a person's experience, they cannot grasp the nuanced complexity of human emotions. This can lead to

insensitive or inappropriate responses, particularly when addressing expressions of distress or vulnerability,[28] and may alienate users.

There is also a risk that the use of AI in mental healthcare could lead to the **depersonalisation of care.** By reducing human interaction and connection, AI may prioritise convenience and efficiency at the expense of the personalised, empathetic support that many individuals in crisis need.

Further, over-reliance on AI could **potentially erode mental health practitioners' empathy**. If AI systems are used to monitor service users or replace direct interaction, mental health practitioners may have fewer opportunities to practice and cultivate communication skill, and become less attuned to what is needed in care encounters.[29]

## Surveillance

Some commentators have raised concerns about the potential misuse of AI for surveillance and control, particularly when mental health data is involved. One area of concern is the use of AI systems **to predict suicide and self-harm**. While such technologies are often promoted as tools for prevention and intervention, they come with significant risks, including privacy violations, inaccuracies in predictions, and the possibility of unwarranted interventions. These interventions could include involuntary treatment or unnecessary involvement of law enforcement.[30]

**Biometric monitoring technologies also present surveillance risks.** By collecting and analysing physiological data such as facial expressions, voice patterns, gait, and eye movements, these systems may attempt to infer mental states and behaviours. Such uses blur the line between monitoring and invasive surveillance.

Another issue is the **sharing of sensitive mental health data with law enforcement and government agencies**, such as non-criminal data about suicide attempts being used by law enforcement or border authorities.[31]

These practices risk creating "a market for surveillance in the mental health context that perpetuates and even extends the worst power imbalances, inequities and harms of current mental health practices"[32].

## Reinforcing individualistic views of mental health

Most AI driven technology in the mental health context appears to be directed at detection and diagnosis, which draws the focus to the individual identified as requiring expert intervention[33]. This can lead to overemphasis on biological determinants, and an unjust focus on individual responsibility for mental health[34]. This framing reinforces individualistic views of mental health and makes invisible the broader socio-economic, relationship, and environmental factors, which – in interaction with personal health issues - shape people's mental health.

## Diverting limited resources

AI is often promoted by a highly active market of firms that try to sell tools to governments and service providers. These companies can be highly strategic, rhetorically sophisticated, and well organised in influencing policymaking, including procurement.[35]

AI marketing has included outlandish claims about using AI to solve the longstanding and complex issues of mental health, such as claims that 'AI-brain-chips… could "solve" autism and schizophrenia'.[36] Such AI hype, particularly where evidence is over-stated, can alter how funding is directed and draw resources away from where they are needed most.

Claims that AI and other digital technologies will necessarily increase efficiency in mental healthcare are often unsupported by evidence, even when technologies are presented as a lower cost of service relative to face-to-face support.[37] Advocacy organisation *Privacy International* has likewise argued that there remains little evidence that AI will necessarily lead to more efficient healthcare systems, despite a widespread assumption – boosted by technology vendors – that this will be the case.[38]

Even if AI systems were found to increase efficiency, efficiency should not be the principal or only goal, as other goals may be equally or even more important, such as creating caring systems.

# Conclusions and recommendations

The sources examined for this report acknowledge the need for robust research to support claims about AI systems, as well as addressing potential risks and challenges. They emphasise the need for ethical considerations, careful implementation, and forms of governance and oversight that involve people with lived experience of mental health problems, to maximise benefits while minimising harm.

When regulating AI uses in mental health care, it is crucial to reflect on human relationships. Indeed, "One common aim of AI is to break down tasks into individual components which can be repetitively undertaken. Yet, care is not just tasks, it is also emotion; it is a fundamental part of human relationships and it is a highly complex social interaction"[39].

It has been argued that the current EU regulatory framework scarcely addresses the unique impact of AI on human interactions and emotions, which are an integral part of mental health care[40]. This oversight would reinforce the limited accountability and responsibility of AI-developing companies in mental health[41]. It is also crucial to address the conflict of interests of companies, putting good quality mental health care at the centre, over profit considerations.

## What role for lived experience in the development of AI technologies?

Involving people with lived experience in developing AI systems for mental healthcare is crucial for creating technologies that are relevant, effective, and ethical. This perspective can help identify priorities that resonate with service users, such as privacy, equity, and human connection.[42] Including

these voices also helps address bias and discrimination, power imbalances, and accessibility barriers, improving the likelihood that AI systems are effective, inclusive and trustworthy.

Some researchers advocate for engaging individuals with lived experience throughout the research and development process, from defining problems to knowledge dissemination, and ongoing governance over the systems.[43] Recommendations for integrating lived experience perspectives include co-design of AI systems, reflective practices, transparent decision-making, prioritising their leadership, ensuring diversity, and fairly compensating contributions. The active involvement of those most impacted by algorithmic and data-driven technologies should not be seen merely as a required step of 'stakeholder engagement', but rather as an ethical necessity. When developers, public procurement authorities, mental health services and others, embed lived experience expertise in these ways, AI systems are more likely to reflect real-world needs and foster not just acceptance but genuine trust and utility in AI systems.

## Recommendations for policymakers

To address the **safety risk**

- ✓ Ensure **services are of high quality** by integrating AI with high standard of scientific integrity, including robust validation before widespread use.
- ✓ Establish an independent regulatory body, for monitoring and addressing complaints related to AI misuse in mental healthcare (**accountability**).
- ✓ **Implement mandatory reporting of adverse effects** and require AI developers and healthcare providers to report and track errors, biases, and unintended harms caused by AI systems.

To address the **privacy risk**

- ✓ Enforce robust **data protection** measures, including clear consent protocols and regulatory frameworks that prioritise individual and collective privacy rights.
- ✓ **Prohibit unauthorised data sharing** of mental health data and prevent AI systems from transferring mental health data to third parties (e.g., insurers, employers, advertisers) without explicit, informed consent.

To address the **lack of appropriate informed consent**

- ✓ Ensure **transparency and explainability** of AI, allowing people to access clear information about how these systems work, what data they use, how they make decisions, and who that data might be shared with.

To address the risk of **creating new inequalities or strengthening them**

- ✓ Mandate the **inclusion of diverse populations in AI training datasets** to reduce biases and improve accuracy for all communities.
- ✓ Ensure the **active involvement of people from communities in vulnerable and/or marginalised situations and people with lived experiences** in the design, development and testing of AI technologies, in order to reduce bias.

- ✓ Subsidise access to validated mental health AI tools for underprivileged communities, in order to ensure **digital inclusion**.
- ✓ Develop and enforce **accessibility guidelines**, for AI tools to ensure usability for individuals with disabilities, language barriers, or low digital literacy.

To address the **depersonalisation of care** risk

- ✓ Require **human oversight** in AI-driven mental healthcare, to ensure that empathy remains central to care.

To address the **surveillance risk**

- ✓ Require AI developers to notify individuals about their interaction with AI and to ensure an **opt-in approach**, or – where such an approach is not feasible- to at least ensure the possibility to opt-out of AI systems.
- ✓ **Regulate law enforcement use of mental health data** and restrict the sharing of mental health-related AI insights with police, immigration authorities, and other government agencies.

To address the **risk of reinforcing individualistic views of mental health**

- ✓ **Resist the temptation of 'techno-solutionism'**, the belief that all problems can be solved with an easy technological fix. Resources and efforts should be invested in addressing the broader determinants that shape mental health, shifting the focus from the individual to the broader context around him.

To address the **risk of diverting resources**

- ✓ Require AI developers to prioritise **quality of care over profit** and to prove that AI is addressing a real need and not creating any harm. Public funding can be used as a leverage in this direction. Co-creation can serve as a tool to ensure that AI developments respond to real needs.

## Recommendations for civil society

- ✓ **Monitor AI in Mental Healthcare**: Track the deployment of AI tools in mental health settings, document cases of bias, privacy violations, or harm, and advocate for policy interventions.
- ✓ **Centre Lived Experience in AI Development**: Facilitate structured discussions and consultations with individuals with lived experience of mental health issues to determine which AI applications are beneficial, which are harmful, and how AI can be designed to support, rather than undermine, mental healthcare.
- ✓ **Raise Awareness of AI Risks and Rights**: Launch public campaigns in collaboration with mental health and digital rights organizations to highlight the risks of AI in mental healthcare, emphasizing the need for human rights-centered regulation.
- ✓ **Engage in Policy and Legislative Processes**: Actively participate in consultations, expert panels, and policymaking efforts to ensure AI regulations prioritize mental health rights, accessibility, and ethical standards, preventing AI from reinforcing stigma or restricting access to care.

# Bibliography

[1] Hannah van Kolfschooten and Janneke van Oirschot, 'The EU Artificial Intelligence Act (2024): Implications for Healthcare' (2024) 149 *Health Policy* 105152 ('The EU Artificial Intelligence Act (2024)').

[2] Hannah van Kolfschooten and Janneke van Oirschot, *When People Become Data Points: The Potential Impact of AI in Mental Healthcare* (Health Action International, December 2024) <https://haiweb.org/wp-content/uploads/2024/12/AI-in-Mental-Healthcare.pdf>.

[3] *Shane Cross et al, 'Use of AI in Mental Health Care: Community and Mental Health Professionals Survey*' (2024) 11(1) *JMIR Mental Health* e60589.

[4] Courtney Potts et al, 'A Multilingual Digital Mental Health and Well-Being Chatbot (ChatPal): Pre-Post Multicenter Intervention Study' (2023) 25(1) *Journal of Medical Internet Research* e43051.

[5] World Health Organization, 'Artificial Intelligence in Mental Health Research: New WHO Study on Applications and Challenges' (6 February 2023) <https://www.who.int/europe/news-room/06-02-2023-artificial-intelligence-in-mental-health-research--new-who-study-on-applications-and-challenges>.

[6] Ibid.

[7] Anoushka Thakkar, Ankita Gupta and Avinash De Sousa, 'Artificial Intelligence in Positive Mental Health: A Narrative Review' (2024) 6 *Frontiers in Digital Health* 1280235.

[8] See eg, Nora Eisemann et al, 'Nationwide Real-World Implementation of AI for Cancer Detection in Population-Based Mammography Screening' [2025] *Nature Medicine* 1.

[9] Charlotte Blease and Adam Rodman, 'Generative Artificial Intelligence in Mental Healthcare: An Ethical Evaluation' (2024) 5(12) *Current Treatment Options in Psychiatry* <https://link.springer.com/article/10.1007/s40501-024-00340-x>.

[10] Thakkar, Gupta and De Sousa (n 7).

[11] Roberto Tornero-Costa et al, 'Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review' (2023) 10(1) *JMIR Mental Health* e42045.

[12] Jonah Bossewitch, 'Brave New Apps: The Arrival of Surveillance Psychiatry', *Mad In America* (9 August 2019) <https://www.madinamerica.com/2019/08/brave-new-apps-the-arrival-of-surveillance-psychiatry/>.

[13] van Kolfschooten & van Oirschot (n 2).

[14] Hannah van Kolfschooten, *'EU regulation of artificial intelligence: Challenges for patients' rights'*, (2022), 59, Common Market Law Review, Issue 1, pp. 81-112.

[15] Rhonda L Wilson et al, 'Artificial Intelligence: An Eye Cast towards the Mental Health Nursing Horizon' n/a(n/a) *International Journal of Mental Health Nursing* <https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13121>.

[16] Mason Marks, 'Artificial Intelligence-Based Suicide Prediction' (2019) 21(98) *Yale J.L. & Tech.* 24.

[17] Blease and Rodman (n 9).

[18] Tornero-Costa et al. (n 11).

[19] Thakkar, Gupta and De Sousa (n 7).

[20] Compensation uncertain for Vastaamo victims. (2021, June 20). *Yle Uutiset*. https://yle.fi/uutiset/osasto/news/compensation_uncertain_for_vastaamo_victims/11991155

[21] See eg, European Convention on Human Rights, arts 3, 8; Convention on the Rights of Persons with Disabilities, arts 3, 12, and 25.

[22] Blease and Rodman (n 9).

[23] See eg, Federal Trade Commission (U.S.), 'FTC to Ban BetterHelp from Revealing Consumers' Data, Including Sensitive Mental Health Information, to Facebook and Others for Targeted Advertising' (2 March 2023) <https://www.ftc.gov/news-events/news/press-releases/2023/03/ftc-ban-betterhelp-revealing-consumers-data-including-sensitive-mental-health-information-facebook>.

[24] World Health Organization (n 5).

[25] Hannah van Kolfschooten, 'The AI cycle of health inequity and digital ageism: mitigating biases through the EU regulatory framework on medical devices' (2023) *Journal of Law and the Biosciences* 10(2), July-December 2023, lsad031, https://doi.org/10.1093/jlb/lsad031.

[26] Morgan King, 'Harmful Biases in Artificial Intelligence' (2022) 9(11) *The Lancet Psychiatry* e48.

[27] Ben Hutchinson et al, 'Social Biases in NLP Models as Barriers for Persons with Disabilities' [2020] *arXiv:2005.00813 [cs]* <http://arxiv.org/abs/2005.00813>.

[28] Geoff White, 'Child Advice Chatbots Fail to Spot Sexual Abuse', *BBC News* (online, 11 December 2018) <https://www.bbc.com/news/technology-46507900>.

[29] See eg, Rhonda L Wilson et al, 'Artificial Intelligence: An Eye Cast towards the Mental Health Nursing Horizon' (2023) 32(3) *International Journal of Mental Health Nursing* <https://onlinelibrary.wiley.com/doi/abs/10.1111/inm.13121>; Blease and Rodman (n 9).

[30] Marks (n 16).

[31] Office of the Privacy Commissioner of Canada, 'Disclosure of Information about Complainant's Attempted Suicide to US Customs and Border Protection Not Authorized under

the Privacy Act' (21 September 2017) <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-federal-institutions/2016-17/pa_20170419_rcmp/>.

[32] Jonah Bossewitch et al, *Digital Futures in Mind: Reflecting on Technological Experiments in Mental Health and Crisis Support* (report, University of Melbourne, 1 September 2022) <https://automatingmentalhealth.cc/>.

[33] Ibid.

[34] Ibid.

[35] Piers Gooding, 'Market Forces in Automated Mental Health Services: New Claims in Algorithmic Care and Disability Justice' in Lieketseng Ned et al (eds), *The Routledge International Handbook of Disability and Global Health* (Routledge, 2024) 565.

[36] Isobel Asher Hamilton, 'Elon Musk Said His AI-Brain-Chips Company Could "solve" Autism and Schizophrenia', *Business Insider Australia* (14 November 2019) <https://www.businessinsider.com.au/elon-musk-said-neuralink-could-solve-autism-and-schizophrenia-2019-11>.

[37] Jacqueline Sin et al, 'Digital Interventions for Screening and Treating Common Mental Disorders or Symptoms of Common Mental Illness in Adults: Systematic Review and Meta-Analysis' (2020) 22(9) *Journal of Medical Internet Research* e20581.

[38] Privacy International, 'Our Analysis of the WHO Report on Ethics and Governance of Artificial Intelligence for Health', *Privacy International* (20 July 2021) <http://privacyinternational.org/news-analysis/4594/our-analysis-who-report-ethics-and-governance-artificial-intelligence-health>.

[39] Bossewitch et al (n 32).

[40] Tamar Tavory, 'Regulating AI in Mental Health: Ethics of Care Perspective' (2024) 11(1) *JMIR Mental Health* e58493.

[41] Ibid.

[42] Ian Muchamore, Panos Karanikolas and Piers Gooding, 'How Lived Experience Expertise Shapes Research and Development in Digital Mental Health' <https://apo.org.au/node/328006>.

[43] Kellie Gilbert et al, 'Digital Futures in Mind: Why Lived Experience Collaboration Must Guide Digital Mental Health Technologies' (2024) (online first) *Australian Journal of Social Issues* <https://onlinelibrary.wiley.com/doi/abs/10.1002/ajs4.355>.



Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission's CERV Programme. Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by
the European Union