An Ethical Perspective on The Democratization of Mental Health with Generative Artificial Intelligence

Zohar Elyoseph, Tamar Gur, Yuval Haber, Tomer Simon, Tal Angert, Yuval Navon, Amir Tal, Oren Asman

Submitted to: JMIR Mental Health
on: March 02, 2024

Table of Contents

An Ethical Perspective on The Democratization of Mental Health with Generative Artificial Intelligence

Zohar Elyoseph[1, 2] BA, MA, PHD; Tamar Gur[3] BA, MA, PHD; Yuval Haber[4, 5] BA, MA; Tomer Simon[6] BA, MA, PHD; Tal Angert[4] BA, MA; Yuval Navon[7] BA; Amir Tal[8] BA, MA, PHD; Oren Asman[8, 9, 10] BA, MA, PHD

[1]Imperial College, London Department of Brain Sciences, Faculty of Medicine London GB
[2]The Max Stern Yezreel Valley College , The Center for Psychobiological Research, Department of Psychology and Educational Counseling, Yezreel Valley IL
[3]Reichman University The Adelson School of Entrepreneurship, Herzliya IL
[4]"The Artificial Third" Research & Development Team Tel Aviv IL
[5]Bar-Ilan University The PhD Program of Hermeneutics & Cultural Studies Ramat Gan IL
[6]Microsoft Israel R&D Center Tel Aviv IL
[7]Tel Aviv University Sagol School of Neuroscience Faculty of Medical and Health Sciences Tel Aviv IL
[8]Tel Aviv University Faculty of Medical and Health Sciences Samueli Initiative for Responsible AI in Medicine Tel Aviv IL
[9]Tel Aviv University Department of Nursing, Faculty of Medical and Health Sciences Tel Aviv IL
[10]Tel Aviv University Sagol School of Neuroscience Tel Aviv IL

Corresponding Author:
Zohar Elyoseph BA, MA, PHD
Imperial College, London
Department of Brain Sciences, Faculty of Medicine
Fulham Palace Rd, London W6 8RF
GB
London
GB

Abstract

Knowledge has become more open and accessible to a large audience with the "democratization of information" facilitated by technology. This paper provides an ethical perspective on utilizing Generative Artificial Intelligence (GenAI) for the democratization of mental health knowledge and practice. It explores the historical context of democratizing information, transitioning from restricted access to widespread availability due to the internet, open-source movements, and most recently, GenAI technologies such as Large Language Models (LLMs). The paper highlights why GenAI technologies represent a new phase in the democratization movement, offering unparalleled access to highly advanced technology as well as information. In the realm of mental health, this requires a delicate and nuanced ethical deliberation.

Including GenAI in mental health may allow, among other things, improved accessibility to mental health care, personalized responses, conceptual flexibility, and could facilitate a flattening of traditional hierarchies between health care providers and patients. At the same time, it also entails significant risks and challenges that must be carefully addressed. To navigate these complexities, the paper proposes a strategic questionnaire for assessing AI based mental health applications. This tool evaluates both the benefits and the risks, emphasizing the need for a balanced and ethical approach for GenAI integration in mental health.

The paper calls for a cautious yet positive approach to GenAI in mental health, advocating for the active engagement of mental health professionals in guiding GenAI development. It emphasizes the importance of ensuring that GenAI advancements are not only technologically sound but also ethically grounded and patient centered.

(JMIR Preprints 02/03/2024:58011)
DOI: https://doi.org/10.2196/preprints.58011

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users. Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

An Ethical Perspective on The Democratization of Mental Health with Generative Artificial Intelligence

Dr. Zohar Elyoseph[1,2*], Dr. Tamar Gur[3], Yuval Haber[4,5], Dr. Tomer Simon[6], Tal Angert[5], Yuval Navon[7,8], Dr. Amir Tal[8,9], Dr. Oren Asman[7,8,9]

1. Department of Educational Psychology and Counselling, Max Stern Yezreel Valley College.
2. Department of Brain Sciences, Faculty of Medicine, Imperial College London.
3. The Adelson School of Entrepreneurship, Reichman University, Herzliya, Israel.
4. The PhD Program of Hermeneutics & Cultural Studies, Bar-Ilan University.
5. "The Artificial Third" Research & Development Team
6. Microsoft Israel R&D Center.
7. Sagol School of Neuroscience, Tel Aviv University.
8. Faculty of Medical and Health Sciences, Tel Aviv University.
9. Samueli Initiative for Responsible AI in Medicine, Tel Aviv University.

***Correspondence:** Dr. Zohar Elyoseph, Department of Psychology and Educational Counselling, Max Stern Yezreel Valley College, Israel. Tel: +972 54 7836088; Zohare@yvc.ac.il

Abstract

Knowledge has become more open and accessible to a large audience with the "democratization of information" facilitated by technology. This paper provides a socio-historical perspective for the Theme Issue “Responsible Design, Integration, and Use of Generative AI in Mental Health”. It evaluates ethical considerations in utilizing Generative Artificial Intelligence (GenAI) for the

democratization of mental health knowledge and practice. It explores the historical context of democratizing information, transitioning from restricted access to widespread availability due to the internet, open-source movements, and most recently, GenAI technologies such as Large Language Models (LLMs). The paper highlights why GenAI technologies represent a new phase in the democratization movement, offering unparalleled access to highly advanced technology as well as information. In the realm of mental health, this requires a delicate and nuanced ethical deliberation. Including GenAI in mental health may allow, among other things, improved accessibility to mental health care, personalized responses, conceptual flexibility, and could facilitate a flattening of traditional hierarchies between health care providers and patients. At the same time, it also entails significant risks and challenges that must be carefully addressed. To navigate these complexities, the paper proposes a strategic questionnaire for assessing AI based mental health applications. This tool evaluates both the benefits and the risks, emphasizing the need for a balanced and ethical approach for GenAI integration in mental health.

The paper calls for a cautious yet positive approach to GenAI in mental health, advocating for the active engagement of mental health professionals in guiding GenAI development. It emphasizes the importance of ensuring that GenAI advancements are not only technologically sound but also ethically grounded and patient centered.

Keywords: Ethics, Generative Artificial Intelligence, Mental Health

Introduction

The Democratization of Information: From Print to AI Generated Content

The democratization of information has been described as the process of making knowledge more accessible, inclusive, and transparent to a broad audience, often facilitated by technological advancements (Wallace & Van Fleet ,2005). Over the past few centuries, a transformation has occurred in how knowledge is accessed, disseminated, and utilized. Historically, access to information and technology was often restricted to a privileged few – aristocrats, the church,

academics, researchers, and professionals who had the means to gather and interpret data. The printing press served as an important milestone in the democratization of information. With the development of the steam locomotive (trains) in the 1800s, printed newspapers and journals that included news and ideas could be disseminated quickly and relatively cheaply across large distances. More recently in the 1990s when the internet became widely accessible, search engines enabled widespread and decentralized access to knowledge. Web 2.0, a participatory web, with wiki platforms and other people-centric websites later leveraged the web and engaged its users and elicited their collective intelligence (Murugesan, 2007). This was followed by open-source movements that promoted sharing code and software frameworks freely, allowing developers globally to build upon and improve existing technologies. All these advancements led to an unprecedented amount of information freely accessible to billions. As technology continued to be developed and advanced, we argue that a new era in information democratization has begun since 2022 when various Generative Artificial Intelligence (GenAI) platforms opened their platform for anyone with an internet connection to be used. The current phase of technology democratization marks a shift away from its exclusive use by computer scientists, researchers and AI professionals and toward reaching a broader audience with less expertise. Users now have more opportunities to actively participate in improving current technologies and may play a larger role in its advancement. GenAI technologies, such as Large Language Models (LLMs) with visual and auditory elements provide billions of people with direct access to cutting-edge technology, transcending the concept of "end users" and allowing them to perform tasks previously reserved for those with extensive computer science knowledge. Today laypeople can use such technologies to create code, software, and GenAI models by expressing their desires in a natural language. These technologies drive the democratization of knowledge and technology by providing tailored, personalized, and on-demand information on a massive scale.

While the growing popularity of GenAI has undoubtedly aided the democratization of information, it

also raises serious concerns about surveillance and control. Considering insights from Foucauldian theories, the widespread integration of GenAI into social discourse raises concerns about the potential abuse of authority and narrative manipulation. Furthermore, relying on GenAI-driven decision-making processes risks reinforcing existing power dynamics and marginalizing specific voices in society. As GenAI affects more aspects of our lives, it is crucial to critically evaluate its implications on privacy, autonomy, and the integrity of information dissemination. This paper provides a socio-historical perspective for the Theme Issue on "Responsible Design, Integration, and Use of Generative AI in Mental Health". It considers the ethics of utilizing GenAI for the democratization of mental health knowledge and practice.

Democratization of Knowledge in Mental Health: A Permanent Shift

Since the launch of ChatGPT in November 2022, multiple studies have shown the transformative potential of GenAI in mental health (Elyoseph et al., 2023a; Hadar-Shoval et al., 2024; Elyoseph & Levkovich 2023; Levkovich & Elyoseph 2023a; Levkovich & Elyoseph 2023b; Haber et al., 2024; Hadar Shoval et al., 2023b; Tal et al., 2023; Elyoseph et al., 2024a; Elyoseph et al., 2024b; Elyoseph et al., 2024c; Elyoseph et al., 2024d;). This is crucial to recognize as we delve into the advantages and risks of GenAI in democratizing mental health knowledge. GenAI can address the global shortage of mental health professionals, reshape mental health care, advance diagnostic accuracy, improve treatment personalization, and enhance the overall accessibility of mental health services. It can facilitate mental health education and awareness, provide various self-help or self-paced mental health support tools, etc. However, it also poses risks, especially in the context of therapy and personalized mental health interventions.

Advantages of GenAI in Mental Health Democratization

Accessibility: A fundamental challenge in the mental health field is the limited access to mental health care both in developed and developing countries, as well as the disparities in access to mental health care (Hodgkinson et al., 2017; Whiteford et al., 2013; World Health Organization., 2013; Pan

American Health Organization., 2018; Vigo et al., 2019; Araya et al., 2018). Factors such as socioeconomic status (Hodgkinson et al., 2017), geographical location (Cummings et al., 2017), linguistic barriers (Ohtani et al., 2015), and cultural disparities (Byrow et al., 2020) present significant hurdles to the accessibility of mental health services. GenAI may be leveraged in mitigating these barriers through the development of linguistically and culturally attuned resources, and potentially offering solutions adaptable to various economic backgrounds (Cummings et al., 2017).

Personalized Responses: AI provides an opportunity for a new era of mental health services that are sharply attuned to the individual needs and preferences of each patient (Graham et al., 2019).Within the framework of treatment with a mental health professional, AI as well as GenAI technologies can facilitate deeper understanding of a person's unique psychological landscape by considering a myriad of factors such as their biological predispositions, societal and cultural influences, and personal preferences (Hadar-Shoval et al., 2023). These technologies have the potential to analyze complex patterns and variances in individual experiences and histories, which can be instrumental in crafting responses and interventions that are tailored to suit individual needs as well as preferences. In contrast, non-AI-based search engines still used by many over the internet fall short of this goal, as searches often lack context and fail to grasp the nuanced complexities of individual experiences and histories. Moreover, personalized responses enabled by GenAI can envision a departure from generic diagnostic protocols, creating room for treatment paradigms that are designed with a more intricate understanding of individual idiosyncrasies. This method assumes that mental health is influenced by biological, social, and personal factors, aiming to provide solutions tailored to the distinct characteristics of each person's experiences. Services that are tailored for each user specifically may lead the mental health field toward a more inclusive and equitable future.

Theoretical/Conceptual Pluralism: In the pursuit of democratizing mental health, the goal is to enhance the field's adaptability and inclusivity by embracing a variety of perspectives. GenAI mental

health tools may be leveraged in three main ways:

1. Providing a specific theory-based, or perspective-based analysis of various scenarios with the ability to shift between theories or perspectives.
2. Generating a dialogue or a synthesis between various theories or perspectives.
3. Potentially create a new theory or hypothesis (See **Table 1**).

Table 1: Utilizing GenAI for multiple theory/perspective analysis, integration, and formation of novel ideas

Ability	Demonstration
1. Provide theory-based analysis	*Cognitive theory (Beck):* The patient exhibits the key hallmarks of severe major depressive disorder per Beck's cognitive model - pervasive negative cognitive schemas, automatized depressive cognitions, diminished mood and inability to work. *Object relations theory (Winnicott):* The emergence of this severe depressive episode despite no clear environmental trigger suggests potential psychological vulnerabilities rooted in early childhood. Lack of good-enough mothering could have prevented the development of a cohesive self, resulting in a false self-presentation vulnerable to fragmentation and depressive affect under stress.
2. Synthesize perspectives	This patient's sudden-onset depression despite minimal stressors points to vulnerabilities in self-structure from poor early nurturing. His fragmented functioning aligns with Klein's paranoid-schizoid position, while loss of values reflects ACT's model. Synthesis suggests deficient parenting prevented integrating difficult experiences into a cohesive self. When his false self-faltered, overwhelming feelings emerged that he could neither accept mindfully nor confront to rebuild his personality. Addressing both positive introjections in Kleinian play therapy and value-guided committed action using ACT may enable overcoming this acutely disabled state.
3. Create new theory	The Wilted Plant Theory of Depression presents a unique botanical metaphor to understand and address depression. It likens the human psyche to a plant that requires 'emotional sunlight' (positive

Ability	Demonstration
	interactions), 'psychological nutrients' (intellectual engagement and meaningful activities), and 'emotional hydration' (free expression of emotions) to thrive. Just as a plant wilts without proper care, the theory posits that the human mind suffers in the absence of these essential elements, leading to depression. This metaphorical approach highlights the importance of a nurturing environment for maintaining and restoring mental well-being.

(generated by Claude AI 2.1 pro and ChatGPT 4 (December 2023).

In other words, this pluralism may facilitate the synchronous operation of a variety of therapeutic approaches, philosophies, and cultural viewpoints. For instance, we can observe opportunities where GenAI might enable the integration and dialogue between traditionally distinct therapeutic methodologies such as Cognitive Behavioral Therapy (CBT) and psychodynamic approaches (Grodniewicz et al., 2023). Here, the structural and goal-oriented strategies of CBT could be married with the depth of insight derived from psychodynamic explorations, engendering a more rounded approach to mental health care (Pilecki et al., 2015). Moreover, the perspective of psychiatry, with its medically grounded insights, could be brought into conversation with psychological approaches, nurturing a space where medical, psychological, and holistic strategies can come together to form a more comprehensive view of mental health care.

With that being said, the way current LLMs work, the mere ability to be creative in connecting various methods in a convincing manner may be wonderful for brainstorming new eclectic concepts and therapeutic approaches but is in no way, in itself, evidence of its feasibility and reliability in real life.

Equality and Reduction of Social Gaps: GenAI powered LLMs hold potential to foster greater equality and reduce prevailing social gaps (van Heerden, Pozuelo & Kohrt, 2023). By harnessing vast arrays of data and insights, such models may facilitate interventions crafted to meet the varied needs of different populations, including those historically underserved or marginalized (Fiske, Henningsen & Buyx, 2019). For instance, developing and distributing mental health programs in

languages and dialects that have historically lacked sufficient resources. It could further enable community-centric initiatives, enhancing representation of diverse groups in the mental health discourse, thereby paving pathways for more localized and culturally sensitive interventions. Moreover, GenAI may be able to identify and relate to social aspects that are at times highly relevant in mental health scenarios. GenAI based LLMs with access to information about symptoms, illnesses, and treatments, may allow laymen to ask questions and receive clarifications usually only available by contacting an expert. This may also facilitate in later contacting the relevant health care professionals, thus saving time and resources. This could become more relevant and useful when datasets used for training foundational models or fine-tuning general-purpose models have more representation of various languages and cultures.

Therapist-Patient Engagement: One of the notable strengths of GenAI is its ability to reduce bureaucratic and administrative burdens in mental health care settings. It can provide transformative solutions by automating tasks like transcription, summarization (Prottay et al., 2024), and form filling. Using these technologies, therapists may simplify administrative processes, freeing up more time and attention to provide direct patient care. With AI handling routine paperwork and data entry tasks, clinicians are freed from screens and forms, allowing them to focus on building a connection, conducting assessments, and providing personalized interventions to their patients. This not only increases the efficiency of mental health services, but it also improves the overall quality of patient care by encouraging more meaningful interactions between therapists and their clients.

Flattening of Hierarchies: The advent of GenAI bears the promising potential to flatten traditional hierarchical structures prevalent in the mental health sector, fundamentally altering the dynamics between health care providers and recipients (Cohen., 2023). Historically, psychiatrists and psychologists held a pronounced degree of authority, largely stemming from their exclusive access to specialized knowledge. If knowledge is no longer confined to a select few but is accessible to a wider population, this allows for a more balanced dynamic between mental health professionals and

individuals seeking help (Cohen., 2023; Tal et al., 2023). It could empower individuals with insights and understanding of their own mental health conditions, fostering more collaborative therapeutic relationships, potentially leading to more fruitful and synergistic therapy sessions, grounded on mutual understanding and shared knowledge. As we have previously defined, the introduction of GenAI into the field of mental health can be seen as an "artificial third" that changes the dynamic between mental health professionals and patients, so that in fact a new relationship triangle is created characterized by the flattening of the existing power hierarchy between experts and patients (Haber et al., 2024; Tal et al., 2023). In this vision of a democratized mental health landscape, GenAI acts as an equalizer, breaking down barriers to knowledge accessibility and cultivating a health care landscape built on collaboration, understanding, and shared expertise. This is further evident with the context window of LLMs increasing dramatically over a short period of time (for instance, OPENAI ChatGPT increased from 4K tokens to 128K on November 2023 and Google's Gemini increased on February 2024 to 1M tokens). This allows end users to upload a large amount of information (such as hundreds of text pages, images, and videos with clinically relevant information) and discuss it with a chatbot during one prompt or conversation.

Risks to Mental Health Democratization Through AI

Corporate Centralization: Corporate centralization in mental health services, facilitated by GenAI, carries a significant risk of prioritizing profit over individual-centered care, widening disparities in access and quality of mental health care, and influencing public health narratives toward economic gains rather than genuine support and care (Zajko, 2023). GenAI can assume a therapeutic role and be designed to foster trust and build rapport with users (Munn & Weijers, 2023), making it a potent instrument in the hands of entities that may have their own agendas. This includes but is not limited to the promotion of specific political narratives, ideological indoctrination, aggressive marketing or unnoticeable marketing strategies, also known as dark pattern AI (Freeman, 2023), taking advantage of its persuasive power for psychological manipulation and control. The centralization of power-

knowledge, without emphasizing checks and balances in a small number of economic corporations could potentially create a façade of democratization of mental health, but not reflect a true democratization in the field.

Information Transparency: Information transparency could be divided into two major aspects of the "one-way mirror", as only one party is exposed to the other party's information. On the provider side of the “mirror” there are real concerns about the management of user data. These include transactional misuses such as unauthorized sale to third parties or its exploitation for targeted marketing (Coghlan et al., 2023). However, more sinister breaches of personal privacy could also be achieved since GenAI systems have the potential to intrusively analyze personal conversations, behaviors, and emotions (Elyoseph et al., 2024a) without explicit consent. Moreover, the data harnessed might even be utilized in training AI systems, a process that remains largely concealed from the end-users. Indeed, the algorithms driving this AI applications function are much like a “black box" shrouded in mystery with no clarity as to how determinations and analyses are reached (Castelvecchi, 2016). Alas, democratization is thus a double-edged sword; while GenAI may indeed democratize access to mental health resources, the current level of transparency and explainability to users of its operational mechanics may limit a truly informed user engagement, limiting the realization of a democratized system with empowered users (von Eschenbach, 2021). When core aspects of an alignment process, including the embedded objectives, values, and ethics, are not made clear and transparent to users (Hadar Shoval et al., 2024), it can result in power becoming concentrated in an entity whose true incentives remain obscured.

People’s Misperceptions of AI: The level of expertise that people attribute to GenAI tools may be affected by their perceptions of technology. Numerous studies have shown that people tend to imbue AI systems with significant epistemic authority stemming largely from the veneer of impartiality and objectivity these technologies present. This attribution of high epistemic authority to GenAI systems may also pose a significant risk. Epistemic authority essentially refers to the weight and trust in a

source as a repository of knowledge and information (Reed, 2015). While GenAI systems can rely on a vast amount of data, the elevation of their epistemic authority could also carry detrimental effects for both health care providers and patients, as follows:

1. A risk of misinformation: GenAI systems are not infallible; they can make mistakes, be based on incorrect data, or present biased viewpoints, thus generating incorrect advice or guidance. In the context of GenAI's mistakes, consider mentioning the term "hallucinations" (Hatem et al., 2023) or “confabulations” (which in our eyes is problematic terminology because it can be perceived as offensive and because it has an anthropomorphic assumption). Attributing high epistemic authority to GenAI may lead to unconditional acceptance of its output, without a critical evaluation of the veracity of the information provided.

2. GenAI over-reliance with reduced patient self-engagement: While incorporating GenAI into mental health care has numerous advantages, it also highlights the serious risk of epistemic bias among both therapists and patients. Attributing high epistemic authority to AI may overshadow not only the expertise and nuanced understanding of health care providers but also the personal experiences and insights of the patients themselves. Overreliance on GenAI in healthcare may reduce patient self-engagement by prioritizing AI-generated insights over the comprehensive understanding provided by health care providers and patients, potentially undermining individuals' active participation in their mental health journey and resulting in less effective treatments. Relying on the AI's ability to articulate and construct our own thoughts and feelings, thus "Letting the Machine speak for us" could also mean relinquishing effort in our interpersonal engagements, including in therapy, reducing one's possibilities for self-understanding and growth (Hartford & Stein, 2024) Furthermore, therapists are vulnerable to epistemic bias by relying too heavily on AI-generated insights, potentially missing important nuances in patient narratives and clinical assessments. This overreliance on GenAI may unintentionally limit the therapist's ability to engage deeply with patients, as algorithmic recommendations may not fully capture the complexities of individual

experiences. As a result, it is critical for therapists to be vigilant against the influence of epistemic bias in their practice, striking a balance between using GenAI tools and retaining the essential human elements of empathy (Rubin et al, 2024), intuition, and clinical judgment.

3. Increased power imbalance: The elevation of GenAI as a central epistemic figure may lead to a power imbalance, where knowledge is centralized in the hands of a GenAI entity that is under the control of economic corporations. This undermines the democratization ethos which seeks to foster a collaborative and pluralistic approach to mental health and where knowledge is the result of collective insight, involving a harmonic convergence of professional guidance and personal experiences. Thus, while GenAI offers a promise of democratized access to information, it also threatens to replace a current knowledge monopoly (currently in the hands of mental health experts) with a monopoly of a small number of LLM companies, which is counterintuitive to the principles of democratization that advocate for a decentralized, collective approach to knowledge dissemination and utilization. It should be noted that open-source models that are available to the public enable decentralized technological development and constitute a decentralized force, and as these models continue to develop and improve, the risks of few companies monopolizing a field (including mental health) will be diminished.

People in emotional need may become dependent on or attached to GenAIs in potentially non-adaptive ways. For instance, many of Replika's AI chatbot 7 million users see it as their best friend, or even a family member (Munn & Weijers, 2023; Skjuve et al., 2021). While examining the relationship with this chatbot, researchers found that the patterns of dependency were different from other technology dependencies as it involved people feeling that Replika had needs and emotions that they needed to cater to (Laestadius et al., 2022). Accordingly, there is an additional layer of risk relating to the authenticity of this "relationship" with a machine (Elyoseph et al., 2024c) whereby the humanization of GenAI (Ferrrio, et al., 2024) may imitate human agency in a manner that could alter our perception of good and healthy lives (Asman, Tal & Barilan, 2023).

Regulation Issues: As GenAI technology, driven mainly by for-profit private corporations, starts to enter the sphere of mental health services, there's a growing concern regarding its adherence to the established protocols that have historically governed mental health services. While the democratization endeavor seeks to foster inclusivity and accessibility, the introduction of GenAI poses a conundrum; it opens avenues for unprecedented access to mental health resources but at the potential cost of diluting the standard of care and ethical considerations traditionally upheld by mental health professionals. One of the major bioethical and legal challenges in this regard is how care ethics concepts could be relevant within the developing field of “responsible AI”, to more fully consider AI’s impact on human relationships. (Tavory, in press).

Objective Perspective Versus Gender, Socio-economic, and Ethnic Biases: Integrating GenAI in mental health services is challenged by how one balances between clinically informed judgments and reducing bias. AI systems rely on pre-existing data that was produced, collected, and potentially also labelled by humans, thereby holding an intrinsic propensity to reflect societal biases, including those grounded in gender, socio-economic factors, and ethnicity (Timmons et al., 2023). At the same time, demographic factors play a critical role in assessing individual health risks and conditions. Consequently, the AI alignment should continuously navigate the narrow pathway between eliminating biases and retaining critical data essential for accurate clinical judgments. From the democratization perspective, AI may perpetuate biases, and at the same time, if overly aligned, may fail to provide users’ expectations of a personalized and efficient mental health service. Thus, the path forward calls for a nuanced and vigilant development process for AI systems, one that meticulously harmonizes statistical evidence with fundamental democratic values.

The claims raised above suggest that AI represents a real opportunity to advance the field of mental health, as it will likely be increasingly present in our lives and its adoption seems inevitable. We propose addressing the risks outlined in this article as thought tools in the development of applied AI tools for responsible use in mental health, rather than viewing them as warnings against using this

technology.

Guiding Ethical Development: A Strategic Questionnaire for AI Mental Health Applications

Considering the potential risks and opportunities identified in the discourse on GenAI applications in mental health, we propose a set of carefully formulated questions designed to assess GenAI's capability to enhance mental health care. These questions are intended for use in the development processes of mental health applications, ensuring a comprehensive evaluation of both the benefits and the risks involved. We have deliberately distinguished between risks and opportunities, recognizing that they do not always exist on the same scale. Namely, a significant risk does not necessarily negate the potential benefits of an AI application, and vice versa. Hence, it is imperative to conduct a differential assessment for each application, weighing its specific risks against its potential opportunities. This approach is grounded in a nuanced understanding that while GenAI offers remarkable prospects for democratizing mental health care, its implementation must be navigated with caution to avoid unintended consequences. The proposed questionnaire is thus an extension of the discourse presented so far in the article, bridging the theoretical considerations with practical evaluation tools.

Table 2: A Strategic Questionnaire for AI Mental Health Applications

Category	Questions
Promoting Democratization	1. Accessibility Does it improve access to mental health services for diverse individuals, including marginalized communities?
	2. User Empowerment: Does it provide tools for self-care and/or informed decision-making?
	3. Facilitating collaboration and Shared Decision Making: Does it facilitate a collaborative approach between patients and health care providers,allowing for an AI augmented shared decision-making process?
	4. Inclusivity: Can it adapt to diverse cultural, socio-economic, and personal needs, promoting inclusive care?
	5. Transparency: Does it provide clear information about its functionalities, limitations, and data usage ?
Identifying Potential Risks	1. Data Privacy and Security: How are privacy and security risks mitigated?
	2. Bias and Inequality: Does it reinforce societal biases or exacerbate inequalities in mental health care?

	3. (Over)dependence/Addiction: How likely is it for users to develop over reliance/dependence on this tool?
	4. Misinformation: How likely is the system to provide false or misinformation or lead to neglecting of human-based professional advice? 5. Corporate Involvement: Are intentional or non-intentional considertations steering the clinical information/advice provided? Or to compromising ethical standards in patient care?
	6. Overshadowing Human Expertise: Does it diminish the role or undermine the expertise of mental health professionals?

Discussion and Conclusion

The integration of GenAI in mental health care, as outlined in this paper, is a potent and inevitable aspect of the broader democratization movement. The ethical implications of not leveraging GenAI in this field are profound, given its potential to revolutionize care and treatment. GenAI introduces a paradigm shift, challenging existing dynamics within mental health care and presenting opportunities to resolve longstanding issues in the field. However, this shift is not without its challenges; it disrupts established power structures, provokes questions about truth and the nature of expertise, and raises concerns about the potential displacement of human roles by technology.

The transition to GenAI-driven mental health care is an inescapable reality, accompanied by considerable promises. It is imperative that the mental health field not only adapts to this new landscape but actively shapes it. This task should not be left solely to engineers and computer scientists; mental health professionals must play a pivotal role. Their involvement is critical to ensure that GenAI development aligns with the ethical standards and therapeutic goals of mental health care. In response to these challenges, our article proposes a structured questionnaire designed to guide responsible AI development in mental health. This questionnaire serves as a roadmap, delineating crucial considerations for balancing the opportunities and risks associated with GenAI integration. It emphasizes the need for a cautious yet optimistic approach in AI development and regulation, ensuring that advancements in mental health care are not only technologically sound but also ethically grounded and patient centered. As we conclude, we call upon mental health

associations and professionals to engage with these guidelines actively. By adopting a stance that is both critically vigilant and constructively engaged, the mental health field can navigate the complexities of GenAI integration. This approach is vital for harnessing AI's potential while safeguarding the foundational values and ethical principles of mental health care. Our contribution, through this discussion and the questionnaire, aims to ensure that the AI revolution in mental health is not only technologically advanced but also democratically enriched and ethically sound.

4249 words

Funding: This research received no funding.

Conflicts of Interest: The authors declare no conflict of interest.

Patient and public involvement: No patients involved.

Abbreviations:

AI: Artificial Intelligence

ChatGPT: Chat Generative Pretrained Transformer

CBT: Cognitive Behavioral Therapy

GenAI: Generative Artificial Intelligence

LLMs: Large Language Models

References

Araya R, Zitko P, Markkula N, Rai D, Jones K. Determinants of access to health care for depression in 49 countries: A multilevel analysis. J Affect Disord. 2018 Jul;234:80-88. doi: 10.1016/j.jad.2018.02.092. Epub 2018 Feb 27. PMID: 29524750.

Asman O, Tal A, Barilan YM. Conversational Artificial Intelligence-Patient Alliance Turing Test and the Search for Authenticity. Am J Bioeth. 2023 May;23(5):62-64. doi: 10.1080/15265161.2023.2191046. PMID: 37130413.

Byrow Y, Pajak R, Specker P, Nickerson A. Perceptions of mental health and perceived barriers to mental health help-seeking amongst refugees: A systematic review. Clin Psychol Rev. 2020 Feb;75:101812. doi: 10.1016/j.cpr.2019.101812. Epub 2019 Dec 24. PMID: 31901882.

Caliskan A, Bryson JJ, Narayanan A. Semantics derived automatically from language corpora contain human-like biases. Science. 2017 Apr 14;356(6334):183-186. doi: 10.1126/science.aal4230. PMID: 28408601.

Castelvecchi D. Can we open the black box of AI? Nature. 2016 Oct 6;538(7623):20-23. doi: 10.1038/538020a. PMID: 27708329.

Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S. To chat or bot to chat: Ethical issues with using chatbots in mental health. Digit Health. 2023 Jun 22;9:20552076231183542. doi: 10.1177/20552076231183542. PMID: 37377565; PMCID: PMC10291862.

Cohen IG. What Should ChatGPT Mean for Bioethics? Am J Bioeth. 2023 Oct;23(10):8-16. doi: 10.1080/15265161.2023.2233357. Epub 2023 Jul 13. PMID: 37440696.

Cummings JR, Allen L, Clennon J, Ji X, Druss BG. Geographic Access to Specialty Mental Health Care Across High- and Low-Income US Communities. JAMA Psychiatry. 2017 May 1;74(5):476-484. doi: 10.1001/jamapsychiatry.2017.0303. PMID: 28384733; PMCID: PMC5693377.

Elyoseph Z, Hadar-Shoval D, Asraf K, Lvovsky M. ChatGPT outperforms humans in emotional awareness evaluations. Front Psychol. 2023;14:1199058. Published 2023 May 26. doi:10.3389/fpsyg.2023.1199058. PMID: 37303897.

Elyoseph Z, Levkovich I. Beyond human expertise: the promise and limitations of ChatGPT in suicide risk assessment. Front Psychiatry. 2023;14:1213141. Published 2023 Aug 1. doi:10.3389/fpsyt.2023.1213141. PMID: 37593450.

Elyoseph Z, Refoua E, Asraf K, Lvovsky M, Shimoni Y, Hadar-Shoval D. Capacity of Generative AI to Interpret Human Emotions From Visual and Textual Data: Pilot Evaluation Study. JMIR Ment Health. 2024a;11:e54369. Published 2024 Feb 6. doi:10.2196/54369. PMID: 38319707.

Elyoseph Z, Levkovich I, Shinan-Altman S. Assessing prognosis in depression: comparing perspectives of AI models, mental health professionals and the general public. Fam Med Community Health. 2024b;12(Suppl 1):e002583. Published 2024 Jan 9. doi:10.1136/fmch-2023-002583. PMID: 38199604.

Elyoseph Z, Hadar Shoval D, Levkovich I. Beyond Personhood: Ethical Paradigms in the Generative Artificial Intelligence Era. Am J Bioeth. 2024c;24(1):57-59. doi:10.1080/15265161.2023.2278546. PMID: 38236857.

Elyoseph Z, Levkovich I. Comparing the Perspectives of Generative AI, Mental Health Experts, and the General Public on Schizophrenia Recovery: Case Vignette Study. JMIR Ment Health 2024d;11:e53043

Ferrario A, Sedlakova J, Trachsel M (2024). The Role of Humanization and Robustness of Large Language Models in Conversational Artificial Intelligence for Individuals With Depression: A Critical Analysis. JMIR Ment Health 2024;11:e56569. DOI: 10.2196/56569 Fiske A, Henningsen P, Buyx A. Your Robot Therapist Will See You Now: Ethical Implications of Embodied Artificial Intelligence in Psychiatry, Psychology, and Psychotherapy. J Med Internet Res. 2019 May 9;21(5):e13216. doi: 10.2196/13216. PMID: 31094356; PMCID: PMC6532335.

Freeman, R, (2023) Generative Artificial Intelligence, Automated User Interfaces, and the New Laws of Dark Patterns, The National Law Review 13 (320). Available at: https://www.natlawreview.com/article/generative-artificial-intelligence-automated-user-interfaces-and-new-laws-dark

Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, Jeste DV. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. Curr Psychiatry Rep. 2019 Nov 7;21(11):116. doi: 10.1007/s11920-019-1094-0. PMID: 31701320; PMCID: PMC7274446.

Grodniewicz JP, Hohol M. Waiting for a digital therapist: three challenges on the path to psychotherapy delivered by artificial intelligence. Front Psychiatry. 2023 Jun 1;14:1190084. doi: 10.3389/fpsyt.2023.1190084. PMID: 37324824; PMCID: PMC10267322.

Haber Y, Levkovich I, Hadar-Shoval D, Elyoseph Z. The Artificial Third: A Broad View of the Effects of Introducing Generative Artificial Intelligence on Psychotherapy JMIR Ment Health 2024;11:e54781. DOI: 10.2196/54781

Hadar-Shoval D, Elyoseph Z, Lvovsky M. (2023). The plasticity of ChatGPT's mentalizing abilities: personalization for personality structures. Front Psychiatry. ;14:1234397. doi: 10.3389/fpsyt.2023.1234397. PMID: 37720897; PMCID: PMC10503434.

Hadar-Shoval, D., Asraf, K., Mizrachi, Y., Haber, Y., & Elyoseph, Z. (2024). Assessing the

Alignment of Large Language Models with Human Values for Mental Health Integration: Cross-Sectional Study Using Schwartz's Theory of Basic Values. JMIR mental health, 11, e55988. https://doi.org/10.2196/55988

Hartford, A., Stein D.J. (2024) The Machine Speaks: Conversational AI and the Importance of Effort to Relationships of Meaning. . JMIR mental health, 11:e53203

Hatem R, Simmons B, Thornton JE. A Call to Address AI "Hallucinations" and How Healthcare Professionals Can Mitigate Their Risks. Cureus. 2023 Sep 5;15(9):e44720. doi: 10.7759/cureus.44720. PMID: 37809168; PMCID: PMC10552880.

Hodgkinson S, Godoy L, Beers LS, Lewin A. Improving Mental Health Access for Low-Income Children and Families in the Primary Care Setting. Pediatrics. 2017 Jan;139(1):e20151175. doi: 10.1542/peds.2015-1175. Epub 2016 Dec 12. PMID: 27965378; PMCID: PMC5192088.

Kapania S, Siy O, Clapper G, SP AM, Sambasivan N. “Because AI is 100% right and safe”: User attitudes and sources of AI authority in India. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 April:1-18.

Lambrecht A, Tucker C. Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Management Science 2019 65(7), 2966-2981. https://doi.org/10.1287/mnsc.2018.3093

Laestadius L, Bishop A, Gonzalez M, Illenčík D, Campos-Castillo C. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society 2022. https://doi.org/10.1177/14614448221142007

Lee MK. Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. Big Data & Society 2018 5(1), 2053951718756684.

Levkovich I, Elyoseph Z. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. Fam Med Community Health. 2023;11(4):e002391. doi:10.1136/fmch-2023-002391. PMID: 37844967.

Levkovich I, Elyoseph Z. Suicide Risk Assessments Through the Eyes of ChatGPT-3.5 Versus ChatGPT-4: Vignette Study. JMIR Ment Health. 2023;10:e51232. Published 2023 Sep 20. doi:10.2196/51232. PMID: 37728984.

Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM computing surveys 2021 *54*(6):1-35.

Munn N, Weijers D. Corporate responsibility for the termination of digital friends. AI & Society. 2023. 38(4): 1501-1502.

Murugesan S. Understanding Web 2.0. IT Professional. 2007 9(4): 34-41. doi: 10.1109/MITP.2007.78

Nov O, Singh N, Mann D. Putting ChatGPT's Medical Advice to the (Turing) Test: Survey Study. JMIR Med Educ. 2023 Jul 10;9:e46939. doi: 10.2196/46939. PMID: 37428540; PMCID: PMC10366957.

Ohtani A, Suzuki T, Takeuchi H, Uchida H. Language Barriers and Access to Psychiatric Care: A Systematic Review. Psychiatr Serv. 2015 Aug 1;66(8):798-805. doi: 10.1176/appi.ps.201400351. Epub 2015 May 1. PMID: 25930043.

Pan American Health Organization. The Burden of Mental Disorders in the Region of the Americas. Washington, DC: PAHO. 2018. Available online at: http://iris.paho.org/xmlui/handle/123456789/49578

Pilecki B, Thoma N, McKay D. Cognitive Behavioral and Psychodynamic Therapies: Points of Intersection and Divergence. Psychodyn Psychiatry. 2015 Sep;43(3):463-90. doi: 10.1521/pdps.2015.43.3.463. PMID: 26301762.

Prottay et al., (Accepted). Exploring the Efficacy of Large Language Models in Summarizing Mental Health Counseling Sessions: A Benchmark Study. JMIR Ment Health.

Reed B. Epistemic Authority: A Theory of Trust, Authority, and Autonomy in Belief. Philosophical Review. 2015 124 (1):159-162. doi: 10.1215/00318108-2812701

Rubin M, Arnon H, Duppert JD, Perry A, Considering the Role of Human Empathy in AI-Driven Therapy, *JMIR Ment Health*. Published 2024 Jun 11:11:e56529. doi: 10.2196/56529.

Skjuve M, Følstad A, Fostervold KI, Brandtzaeg, PB. My chatbot companion – a study of human-chatbot relationships. International Journal of Human-Computer Studies. 2021 149:102601. https://doi.org/10.1016/j.ijhcs.2021.102601

Tal A, Elyoseph Z, Haber Y, Angert T, Gur T, Simon T, Asman O. The Artificial Third: Utilizing ChatGPT in Mental Health. Am J Bioeth. 2023 Oct;23(10):74-77. doi: 10.1080/15265161.2023.2250297. Epub 2023 Oct 9. PMID: 37812102.

Tavory T, Regulating AI in Mental Health – The Ethics of Care Perspective, JMIR (accepted)

Timmons AC, Duong JB, Simo Fiallo N, Lee T, Vo HPQ, Ahle MW, Comer JS, Brewer LC, Frazier SL, Chaspari T. A Call to Action on Assessing and Mitigating Bias in Artificial Intelligence Applications for Mental Health. Perspect Psychol Sci. 2023 Sep;18(5):1062-1096. doi:

10.1177/17456916221134490. Epub 2022 Dec 9. PMID: 36490369; PMCID: PMC10250563.

van Heerden AC, Pozuelo JR, Kohrt BA. Global Mental Health Services and the Impact of Artificial Intelligence-Powered Large Language Models. JAMA Psychiatry. 2023 Jul 1;80(7):662-664. doi: 10.1001/jamapsychiatry.2023.1253. PMID: 37195694.

Vigo DV, Kestel D, Pendakur K, Thornicroft G, Atun R. Disease burden and government spending on mental, neurological, and substance use disorders, and self-harm: cross-sectional, ecological study of health system response in the Americas. Lancet Public Health. 2019 Feb;4(2):e89-e96. doi: 10.1016/S2468-2667(18)30203-2. Epub 2018 Nov 14. Erratum in: Lancet Public Health. 2019 Feb;4(2):e88. PMID: 30446416.

von Eschenbach WJ. Transparency and the black box problem: Why we do not trust AI. Philosophy & Technology. 2021 34(4):1607-1622. doi: 10.1007/s13347-021-00477-0

Wallace DP, Van Fleet C. From the editors: The democratization of information? Wikipedia as a reference resource. *Reference & User Services Quarterly*. 2005 45(2): 100-103. Available at: https://www.jstor.org/stable/20864471

Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, Charlson FJ, Norman RE, Flaxman AD, Johns N, Burstein R, Murray CJ, Vos T. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. Lancet. 2013 Nov 9;382(9904):1575-86. doi: 10.1016/S0140-6736(13)61611-6. Epub 2013 Aug 29. PMID: 23993280.

World Health Organization. Mental Health Action Plan 2013-2020. Geneva: World Health Organization. (2013). Available online at: https://www.who.int/mental_health/publications/action_plan/en/

Zajko, M. Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. Sociology Compass 2022 16(3), e12962. https://doi.org/10.1111/soc4.12962