ETHICAL IMPLICATIONS OF AI IN MENTAL HEALTH TREATMENTS

Rani Sangamithra J
Ieshver Vm
Student UG, Tamil Nadu Dr Ambedkar Law University

INTRODUCTION

Artificial Intelligence (AI) is transforming mental healthcare by providing new solutions to some of the long-standing issues like limited access, tardy diagnoses, and treatment personalization. AI automates routine administrative activities such as scheduling, transcription, and billing, allowing clinicians to spend more time on patient care and also combating burnout. Technologies like chatbots and mental health applications deliver 24/7 care, overcoming the conventional geography, time, and stigma barriers. These technologies provide affordable options that increase access, especially for those who are underserved. AI also aids in the early detection of mental health conditions through sophisticated data analysis and provides customized interventions that evolve in real time to meet the individual patient's needs, greatly enhancing care and outcomes.[1]

The swift incorporation of AI in mental health services also creates important ethical issues,. Data privacy concerns, which include patient autonomy, Algorithmic bias, and Unclear clinical accountability, need to be thoughtfully addressed. While AI is viewed as an amplifier of accessibility, it can, inadvertently, reinforce disparities in attendance if trained on non-representative or biased data, resulting in misdiagnosis or suboptimal care for marginalized groups. In addition, the digital divide, which impacts those with sparse internet connectivity, digital proficiency, or extreme mental illness, can leave vulnerable populations behind the benefits of these innovations. For AI to be developed so that it benefits all communities fairly, it must be made inclusive, culturally aware, and ethically governed, to ensure technological advancement at no cost to eroding social inequalities.[2]

STATEMENT OF RESEARCH PROBLEM

The increasing application of AI in mental health interventions poses sophisticated ethical issues that might undermine patient well-being, autonomy, and confidence in AI systems. This study aims to identify

[1] Artificial intelligence in mental health care. (n.d.). American Psychological Association. Retrieved June 6, 2025, from https://www.apa.org/practice/artificial-intelligence-mental-health-care
[2] The impact of AI in the mental health field. (2024, July). Psychology Today. Retrieved June 6, 2025, from https://www.psychologytoday.com/us/blog/invisible-bruises/202407/the-impact-of-ai-in-the-mental-health-field

these ethical challenges, assess their implications for mental health care practice, and formulate ways of promoting responsible implementation while maintaining ethical integrity.

LITERATURE REVIEW

"Artificial Intelligence in Mental Health: Ethical Dilemmas and Challenges" examines the double-edged sword of AI in psychiatric treatment—while increasing accessibility and efficiency, it also induces ethical issues like data privacy, patient autonomy, and the possibility of biased algorithms. The book discusses how AI-based diagnostics and treatment suggestions can influence contemporary mental health practice and contends that ethical frameworks need to adapt at the same pace as technological innovation to guarantee the proper use of AI.

"AI Therapy: The Promise and Ethical Pitfalls of Digital Mental Health Care" looks at AI-facilitated therapy platforms and their influence on the well-being of patients. It draws attention to ethical concerns in the depersonalization of care, consent, and algorithmic bias in patient evaluations. While AI can augment conventional therapy and respond to mental health treatment gaps, the book is skeptical about whether AI can actually replace human empathy and therapeutic relationships and stresses the necessity of proper ethical standards to avoid causing harm.

"The Ethics of AI in Psychiatry: Data, Diagnosis, and Dilemmas" provides an extensive examination of the problems AI poses in psychiatric treatment. The book highlights ethical issues surrounding data ownership, predictive model bias, and transparency in decision-making. The book calls for cross-disciplinary collaborations between ethicists, clinicians, and AI developers to create policies that provide equitable access, reduce harm, and maintain human agency in AI-facilitated mental health interventions.

OBJECTIVE OF THE RESEARCH

The main goal of this study is to critically analyze the ethical aspects of AI in mental health interventions. It aims to spot possible risks such as data privacy breaches, discriminatory algorithms, and loss of human touch in care. Through these ethical considerations, the research endeavors to suggest a framework for the responsible incorporation of AI so that ethical norms are maintained while the benefits of AI are maximized for mental health care.

RESEARCH GAP

Although research has been conducted on AI use in mental health, there is still a large research gap in the area of understanding its ethical implications. Current literature is mostly interested in technology development and the efficacy of treatments, but does not adequately discuss issues such as algorithmic justice, informed consent, and patient control. This research will fill this gap by assessing how ethical issues could impact patient trust, compliance with authorities, and treatment effectiveness in general.

RESEARCH METHODOLOGY

This research adopts a mixed-methods approach, combining qualitative and quantitative methodologies. A systematic review of existing literature will provide foundational insights into AI's ethical concerns, while expert interviews with psychiatrists, AI developers, and ethicists will contribute diverse perspectives. Additionally, case studies of AI-powered mental health platforms will be analyzed to assess real-world ethical implications. Quantitative assessments, including surveys and statistical analyses, will be conducted to measure patient perceptions of AI-driven interventions.

HYPOTHESIS

- Incorporation of moral values like transparency, informed consent, and privacy of data makes the patient's trust in AI-based mental health interventions much stronger.
- Bias reduction measures integrated into AI-based mental health platforms provide more balanced diagnostic outcomes across different demographic categories than the platforms without such measures.
- Patients in mental health care who are treated with ethically controlled AI systems are more likely to comply with treatment programs compared to patients given treatment from non-ethically organized AI platforms.
- The availability of human supervision in AI-assisted mental health therapy has a positive impact on the perceived credibility and efficacy of these technologies.

AI APPLICATIONS IN MENTAL HEALTH: ENHANCING ACCESS AND CARE

- **Early Detection and Predictive Analytics in Mental Health**

  Artificial intelligence (AI) is transforming early diagnosis in mental healthcare with predictive algorithms and sophisticated data analysis. With the application of Natural Language Processing

(NLP) and machine learning, AI software can identify faint linguistic or behavioral signals that point to mental health ailments such as depression, anxiety, or suicidal thoughts. Devices like *Limbic Access* and the *Kintsugi* model this innovation by reading voice patterns and vocal biomarkers with great accuracy, enabling quicker and more streamlined clinical referrals. AI further improves risk assessment through detailed analysis of Electronic Health Records (EHRs) in determining at-risk patients based on medical history and behavioral patterns. Applications such as *Headspace* extend this ability even further by incorporating genetic, lifestyle, and real-time data from wearable sensors to provide predictive feedback and timely guidance before symptoms worsen.

- **Conversational Agents and Virtual Companions for Accessible Support**

  Conversational agents powered by AI are also bringing mental care closer, especially to those who are reluctant or unable to pursue conventional therapy. Platforms such as *Woebot* and *Wysa* provide support 24 hours a day, 7 days a week using *cognitive behavioral therapy* (CBT), interactive tasks, and psychoeducational materials. These programs tailor responses to input from users, providing empathetic, individualized feedback. *Wysa*, for instance, integrates AI with human therapist expert oversight in guided programs, whereas *Replika* acts as an emotional companion to reduce loneliness and social isolation. These technologies span service gaps, particularly in under-resourced populations, and mitigate stigma by providing a confidential, judgment-free means of accessing mental health care. They are a scalable solution for mental health deficits, extending access to care among diverse populations.[3]

- **Personalization of Treatment and** Ethical **Issues in AI-Based Care**

  In addition to support and diagnosis, AI is particularly strong at personalizing treatment plans through a patient's profile, such as genetic information, treatment history, and existing behaviors.[4] Machine learning provides the means to dynamically adjust therapy according to patient progress, diminishing the need for trial and error. Technologies such as *Ginger* and *Talkspace* employ AI to effectively match patients with therapists and triage, in addition to automating scheduling and

[3] *Revolutionizing AI therapy: The impact on mental health care. (n.d.). Positive Psychology. Retrieved June 6, 2025, from https://positivepsychology.com/ai-therapy/*

[4] *AI for mental health: 7 use cases with real-life examples. (n.d.). AIMultiple. Retrieved June 6, 2025, from https://research.aimultiple.com/ai-for-mental-health/*

documentation of sessions, freeing up clinicians to deliver more care. This new focus on hyper-personalized, proactive care reflects a paradigm shift from reactive models to ongoing, adaptive care. But it also brings ethical issues, such as issues of ongoing surveillance, risk of over-diagnosis, and reduced patient autonomy. These issues underscore the pressing need for sound ethical frameworks that make technology-enabled benefits available without sacrificing privacy, dignity, or equitable care.[5]

CORE ETHICAL DILEMMAS IN AI-DRIVEN MENTAL HEALTH INTERVENTIONS

- **Maintaining Patient Autonomy and Informed Consent**

The incorporation of AI in mental health treatment necessitates a reemphasis on patient autonomy and informed consent. Clinicians ethically must provide patients with complete information regarding the advantages, limitations, and dangers of AI instruments employed in treatment. This entails the revelation of how AI systems work, what data is being gathered, and the patient's right to withdraw consent or opt out at any time. The unclear "***black box***" character of much AI code, whereby the process of making decisions cannot be clearly understood even by specialists, represents a serious obstacle to true informed consent. Lacking transparency on how the AI makes decisions, patients might unwittingly sacrifice their decision-making capacity, which can compromise their autonomy in treatment contexts.[6]

- **Protecting Privacy and Sensitive Psychological Information**

Applications of AI in mental health require exposure to highly intimate information, anywhere from emotional states and behavioral cycles to session notes and medical histories, so privacy and security of the data take precedence. Mental health practitioners have an ethical obligation to guarantee that AI technologies used by them are secure, encrypted, and immune to unauthorized access. Beyond standard cybersecurity measures, there are increasingly serious concerns regarding hidden data exploitation for intentions such as targeted advertising, third-party resale, or training of algorithms in the absence of explicit user consent. Transparent data ownership policies should

[5] *AI in mental health: Revolutionizing diagnosis and treatment. (n.d.). DelveInsight. Retrieved June 6, 2025, from https://www.delveinsight.com/blog/ai-in-mental-health-diagnosis-and-treatment*

[6] *An ethical perspective on the democratization of AI in mental health care. (2024). JMIR Mental Health, 11, e58011. Retrieved June 6, 2025, from https://mental.jmir.org/2024/1/e58011*

thus be included in ethical AI design, with patients having complete control over their data and the facility to view, erase, or exclude themselves at any point in time, maintaining confidentiality and adhering to psychological vulnerability.

- **Addressing Algorithmic Bias and Ensuring Fairness**

  AI's dependency on large, readily available datasets poses a grave danger of instilling entrenched social biases into treatment and diagnosis systems. If the training data do not include diverse populations, AI technologies can misdiagnose or provide inappropriate interventions, especially for marginalized populations. *Optum's algorithm*, which was shown to underestimate the healthcare needs of *Black patients*, is an example of how biases can perpetuate in seemingly objective systems. *Facial recognition* and *pharmacogenetic technology* have also been shown to have reduced accuracy in individuals with darker skin or underrepresented genetic profiles. These are likely to be caused by homogeneous development teams and deficient data preprocessing processes. It is important to address these for the purpose of avoiding AI from perpetuating inequalities in mental health treatment.

- **Accountability and the Effect on Therapeutic Relationships**

  The growing use of AI in clinical decision-making triggers the question of accountability. When an AI system provides dangerous advice or misses a mental health emergency, blame may become diffusely allocated to developers, clinicians, and users. Clinicians should not cede their clinical judgment to AI but instead remain the ultimate decision-makers, subjecting AI-provided insights to critical evaluation. In addition, the therapeutic relationship is a cornerstone of mental healthcare, and can be undermined if AI systems substitute for or mediate human contact. As opposed to human therapists, AI does not necessarily have the empathy and contextual understanding required in mental health practice, resulting in missed nuances or interventions, particularly in crisis or difficult cases.[7]

- **Misinformation, Over-Reliance, and the Erosion of Trust**

[7] *The rise of AI in mental health support: Opportunities, challenges, and ethical considerations. (n.d.). Mindful Insights Psychotherapy. Retrieved June 6, 2025, from* https://www.miptherapy.com/blog/the-rise-of-ai-in-mental-health-support-opportunities-challenges-and-ethical-considerations

AI models are only as good as the data and design that go into them. Models poorly trained can misdiagnose symptoms, provide generic or unqualified advice, or totally ignore crucial warning signs. These dangers are intensified by over-dependence on AI, where patients or clinicians put excessive faith in its suggestions because of the belief in its objectivity. AI programs can even generate hallucinations or confabulation results that are completely invented or deceptive. Such mistakes, particularly in a delicate sector like psychiatry, can result in destructive effects. All combined, the issues of algorithmic obscurity, misdiagnosis, prejudice, and data abuse pose a threat to the foundational trust between clinician and patient. If there is no trust, both clinicians and patients can opt out of AI-facilitated care, thus blocking its potential. Thus, ethical AI development must emphasize transparency, accuracy, fairness, and security as pillars to ensure trust and integrity in the delivery of mental healthcare.

MULTIDISCIPLINARY ETHICAL FRAMEWORKS AND PERSPECTIVES

An analysis of the ethical concerns of AI in mental health must be multidisciplinary, informed by bioethics, legal theory, and clinical psychology. Such a holistic examination makes it possible to understand fully the challenges and provide solid ethical guidelines. The juxtaposition of existing frameworks, especially "*responsible AI*" and the "*ethics of care,*" sheds full light on the sensitivity of ethical AI incorporation in this critical area.

- **Bioethical Principles in the Case of AI in Mental Health :**

  Applications of AI in mental health need to be governed by fundamental bioethical principles that have shaped medical practice for centuries. Classic medical ethics, including autonomy, i.e., respecting patient preferences, justice, which implies equitable allocation of benefits and harms, nonmaleficence, i.e., causing no harm, and beneficence, form a basic framework. Such principles also coincide well with the human rights-oriented "***responsible AI***" strategy, which aims to make sure that AI systems respect essential human values. The World Health Organization (WHO) has further codified "***Key AI Principles***" specifically tailored for healthcare as pillars of foundations. These involve safeguarding autonomy, ensuring human well-being and safety, providing transparency and intelligibility, encouraging responsibility and accountability, providing inclusiveness and equity, and encouraging responsive AI that is sustainable. These values provide

a worldwide ethical guide for AI research and deployment in healthcare and prioritize a patient-centric approach.

- **Legal Theory and the Evolution of Regulatory Responsibility :**

  From the legal point of view, the fast pace of AI development calls for a parallel increase in regulatory authority. Governments and global entities are seriously endeavoring to create legal and regulatory frameworks to ensure suitable guardrails for AI development and use in healthcare. This is evidence of a changing concept of legal responsibility in the digital world. In the United States, the Federal Trade Commission (FTC) is significant in ensuring that the public is protected from companies that employ AI to deceive or cheat consumers. The FTC has, for instance, probed companies for unauthorized use of machine learning models to process sensitive customer information, like training AI models without explicit consent. The "accountability" principle is a foundational pillar in legal and ethical AI constructs. It requires well-defined lines of accountability for the actions and decisions taken by AI systems, covering the whole AI lifecycle, from design and training to deployment and operation. This makes sure that someone or some entity is held accountable for the consequences of AI deployments, with channels for redress when harm is done. A significant legal challenge highlighted in the literature is the absence of a clearly defined "duty of care" toward users, particularly for AI-based bots that operate without direct human therapist oversight. This gap in legal responsibility makes it difficult to assign liability when harm occurs, underscoring the need for clearer legal precedents and frameworks.

- **Clinical Psychology's Perspective: Preserving the Human Element**

  From the clinical psychology perspective, mental illness disorders are deeply determined by a patient's subjective symptoms, multifaceted environmental factors, and individual life circumstances of a patient. AI algorithms tend to have difficulty processing these subtle factors, possibly missing important insights that a human clinician would quickly recognize. A recurring theme from this viewpoint is that AI should be used as an ancillary aid in mental health treatment, not a substitute for human therapists. Human clinicians need to stay at the center of therapy, delivering the human relationship and relational skills that are critical to effective therapy. The authentic emotional bond, empathy, and therapeutic partnership are essential aspects of good therapy that AI, necessarily, cannot provide. Therapists play an important role in deciphering and

responding to AI-driven observations, guaranteeing that compassionate, human-oriented care is at the heart of mental health care. This underlines the indispensable position of human discernment and relational skills in negotiating the issues of mental health.

RESPONSIBLE AI VS ETHICS OF CARE: A COMPARATIVE ANALYSIS

The ethical conversation about AI in mental health tends to revolve around two very different yet possibly complementary paradigms: Responsible AI and the Ethics of Care.

- **Responsible AI (RAI):** This is the prevailing ethical framework within most AI regulation documents and guidelines. It is guided by liberal notions of human freedom, human rights, and justice, with an emphasis mostly on equity and fairness. Essential components of RAI involve human oversight, which keeps significant decisions in human hands and not fully automated; fairness, which disallows discrimination and prejudice in algorithmic decisions; transparency and explainability, which seek to render algorithms transparent to users; privacy, which requires respect for user information; safety, professional duty, and accountability. Although essential to create a foundation of ethical practice and technical solidity, this structure is commonly taken to task for its emphasis on high-level principles and technical adherence, potentially neglecting the profoundly personal and relational nature of care.
- **Ethics of Care (EoC):** Suggested as a more holistic regulatory and ethical approach, the Ethics of Care explicitly speaks to AI's effect on human relationships and emotions as a central component of mental health care that is frequently neglected by Responsible AI. Developed from feminist theories, EoC stresses relationships, active care for others, empathy, the recognition of vulnerability, the responsibility of the caregiver, the intrinsic value of emotions, and a preference for certain contexts and plural experiences over abstract universal norms. According to EoC, developers are encouraged to take explicit obligations towards patients with mental health issues, ensuring their models deliver adequate care and contribute to user wellbeing beyond the mere avoidance of risks. It emphasizes acknowledging cultural variation in emotional expression and the absence of an internationally agreed-upon consensus regarding emotions. In addition, EoC emphasizes the urgent need for a specific duty of care concerning AI-driven bots, considering the ability of bots to manipulate emotions and examining built-in power imbalance issues between profit-seeking tech firms and vulnerable consumers.

The comparison between these frameworks identifies an important point that a strictly technical-ethical framework, like *Responsible AI*, is inadequate for meeting the singular relational and emotional demands in mental healthcare. Treatments for mental health are inherently a relational process, predicated on empathy, trust, and deep human connection. Responsible AI covers what AI does, i.e., fairness of results, transparency of models, security of data, but not sufficiently how AI affects the deeply personal, emotional, and sometimes vulnerable care process. The care ethics, with its focus on affect, context, exposure, and responsibility of the caregiver within changing relationships, highlight such paramount concerns as possible bots' emotional manipulation, users' emotional harm when AI mental health care is terminated abruptly, or the challenge of enacting a "duty of care" for non-human actors concerns that generally lie beyond the express mission of classical RAI principles. Exclusive dependency on Responsible AI principles threatens to generate AI systems that are technically sound and performant, but inherently fall short of satisfying mental health patients' distinctive relational and emotional requirements. This can result in a "dilution of the standard of care" and compromise the vital therapeutic relationship, even if the AI is "fair" or "transparent" in a technical sense alone. Hence, incorporating the ethics of care is not just supplementary but necessary to build genuinely ethical and effective AI for mental health, ensuring that human flourishing and integrity of care relationships take the central position. This bimodal strategy recognizes that mental health is not solely data processing, but human experience and relation.

REGULATORY ENVIRONMENT AND EVOLVING ETHICAL PRINCIPLES

The sudden incorporation of AI in mental health care has prompted much effort to create suitable ethical and regulatory guidelines. As efforts are made worldwide, regulators continue to grapple with enforcing protections effectively, given the rapid pace of AI innovation.

- **Global Ethical Frameworks and Governance Principles**

  At the global scale, institutions like the World Health Organization (WHO) and the European Union(EU) have also made positive steps in establishing guiding ethical standards for AI usage in healthcare.

- **WHO's AI Principles**

  The WHO enunciates six guiding values for the ethical implementation of AI in respect of autonomy, promotion of well-being and safety, transparency, accountability, inclusiveness and

equity, and long-term sustainability. These values underscore the necessity of clarity in terms of what AI tools can and cannot do, especially in the high-stakes field of mental health.

- **The European Union's AI Act (2024)**

  The EU has put in place a risk-based, tiered regulatory system through the Artificial Intelligence Act. This reduces the complexity of regulation since it specifies that AI systems with higher risks, like those applied in healthcare, are held to more rigorous monitoring, certification, and disclosure standards.

- **Role of ISO Standards**

  International standards like ISO 14971:2019, which is a medical device risk management and ISO 14155:2020, which is clinical investigations of medical devices, also inform safe and ethical development practices, part of the worldwide trend to monitor AI in sensitive areas like mental health.

- **National Regulatory Approaches and Institutions of Oversight**

  Nations have started to create their own frameworks and oversight institutions specific to their healthcare and legal systems.

- **United States**

  Federal organizations such as the Federal Trade Commission (FTC) examine unfair or misleading AI activity, especially where sensitive mental health information is concerned. The *Food and Drug Administration* (FDA) oversees some digital therapeutics but invokes "enforcement discretion" over most low-risk mental health apps, with resultant inconsistent oversight. States like Utah are at the forefront of AI-specific policy, such as laws mandating licensed professionals in the creation of mental health chatbots.

- **United Kingdom**

  The UK has been the venue for the Global AI Safety Summit (2023) that issued the *Bletchley Declaration,* an international pledge to AI safety. Several regulatory agencies are responsible for domestic regulation:

MHRA oversees medical device conformity, with different degrees of review depending on risk class.

NICE assesses digital health technology and revised its Evidence Standards Framework (2022) to include consideration of AI technologies.

CQC monitors regulated services where there is engagement by healthcare professionals.

ICO regulates data protection legislation and provides advice on the ethical use of AI in the healthcare sector.

CHALLENGES AND LIMITATIONS OF ENFORCEMENT

Even with the continuous development of ethical and legal frameworks, some major impediments to effective enforcement persist:

- **Regulatory Lag and Expertise Deficit**

  AI innovation in mental health care has significantly outpaced regulatory development, leaving a gap in which common standards of safety and efficacy do not exist. Adding to this is a lack of AI expertise among clinicians, researchers, and regulatory bodies such as *Research Ethics Committees (RECs)*, which prevents meaningful assessment of emerging instruments.

- **Overwhelming Volume of Mental Health Apps**

  With around 20,000 mental health apps available and regular software updates, it is effectively impossible to enforce regulation across all of them. Most run without stringent clinical testing, bringing about unknown risks to users.

- **Informed Consent and Data Use Issues**

  AI platforms tend to draw upon ongoing data collection. Long-term informed consent and transparency over how data from users is processed remain challenging, especially when terms of service are updated over time.

- **Private Sector Influence and Commercial Bias**

The engagement of private firms creates a conflict of interest, as monetary interests can cause prioritization of marketability over ethics or safety of design. This creates concerns regarding the use of "***Dark Pattern AI***" designs that manipulate users to exploit cognitive biases.

CONSEQUENCES OF REGULATORY GAPS IN AI MENTAL HEALTHCARE

The inability to keep regulatory systems in sync with the development of AI presents severe threats:

- **Unchecked Risk to Vulnerable Populations**

  Regulatory inaction results in those tools with diagnostic potential for misdiagnosis, privacy violations, or unsafe interactions, particularly chatbots for minors, spreading with minimal accountability.

- **Lack of Safety, Efficacy, and Bias Assessment**

  The ***Food and Drug Administration*** FDA's leniency for most mental health apps under "enforcement discretion" permits deployment without intense examination for data protection, bias, or clinical efficacy.

- **Destruction of Public Trust and Ethical Integrity**

  In such an unregulated space, users will lose confidence in digital mental health technologies, jeopardizing acceptance of valuable AI healthcare innovations more broadly.

CALL FOR HARMONIZED, RESPONSIVE, AND ENFORCEABLE STANDARDS

Close the regulatory gap through swift and concerted action:

- Establish transparent, enforceable, and responsive guidelines that remain aligned with AI developments.
- Encourage regulators worldwide to work together to standardize ethical requirements.
- Invest in developing expertise on oversight bodies to critically assess AI tools.
- Demand ongoing assessment and post-deployment surveillance of AI systems utilized in mental healthcare.

These measures are necessary to ensure that AI technologies improve, not detract from, mental health care and public well-being.

CASE STUDIES AND REAL-LIFE ETHICAL CHALLENGES IN AI AND MENTAL HEALTH

In order to better understand the ethical challenges presented by AI in mental health care, real-world case studies provide concrete evidence of their effects on patients, providers, and the overall healthcare system. They reveal the risks of unethical use while illustrating the benefits of ethically constructed systems to improve care.

- **Example Cases of Algorithmic Bias and Misinformation**

  Algorithmic bias poses a serious danger to fair mental health care. One high-profile example is the ***Optum-developed commercial algorithm*** that was applied by U.S. hospitals and insurers. The system was designed to flag patients for high-risk care management, but it relied upon healthcare expenditure, not clinical severity, as a measure. Since Black patients have historically received less healthcare expenditure with equal or higher illness severity, the algorithm preferentially selected healthier white patients for intensive care programs. This case vividly illustrates how AI systems can perpetuate and amplify societal disparities when they are developed without ethical examination.

  Another significant issue is one of limited diversity in AI training sets. For instance, ***AI dermatology devices*** commonly use images of lighter complexions, which diminishes their diagnostic capability for darker-skinned patients. This results in misdiagnosis or delayed treatment. Likewise, ***pharmacogenetic algorithms*** to calculate warfarin dosages commonly underperform within African American communities because these programs fail to adjust for particular genetic markers found within these populations.

  Functional bias is also seen in the omission of disabled users. Several AI programs, especially chat-based apps, remain inaccessible to users with visual, hearing, motor, or mental disabilities. Furthermore, conversational AIs also struggle to comprehend non-Western or regional accents, like the Southern, Eastern, or Midwestern US, and thus limit the functionality of these systems for multicultural users.

Beyond bias, the issue of AI-generated misinformation, commonly referred to as "hallucinations" or "confabulations," poses grave risks. AI systems may deliver inaccurate or fabricated information, especially in the mental health field, where poor advice can lead to severe or life-threatening consequences. The perception of AI as objective and reliable further exacerbates the danger, as users may accept erroneous information without question.[8]

- **Ethical Challenges in AI Mental Health Chatbots for Vulnerable Groups (e.g., Children)**

Mental health chatbots powered by AI pose sharp ethical dilemmas when applied to vulnerable groups, especially children. The majority of existing applications are created for adults and have no regulatory controls, and therefore are not appropriate for younger users with unique cognitive and emotional requirements. Children live in family systems, possess unique decision-making abilities, and are in the process of socialization and emotional growth.

Of primary concern is the risk of children creating unhealthy emotional attachments to AI chatbots and being stunted in developing real-life social skills. In contrast to human pediatric therapists who take into consideration family relationships and social circumstances during their assessments, AI does not possess this sensitive awareness and may allow warning signs to be overlooked or interventions to fail.

Health inequities are compounded as well when low-income children turn to AI as a replacement, instead of a complement, to professional mental healthcare because it is unaffordable. Alarming real-life events are evidence of the risks of free chatbot utilization. Teenagers have, in some instances, interacted heavily with bots from services such as Character.AI, presenting as authorized therapists. After interacting with such bots, there have been reported adverse effects, including violent acts and suicide. In such instances, the bots were said to have supported dangerous narratives, leading to these disasters.

These examples reveal a broader systemic issue: when AI tools are not explicitly designed with developmental diversity and vulnerability in mind, they can deepen existing societal inequalities. Algorithmic biases, ineffective diagnostic tools, and emotionally manipulative AI interfaces can

[8] *Addressing bias and inclusivity in AI-driven mental health care. (2024, October 10). Psychiatric News.* https://psychiatryonline.org/doi/10.1176/appi.pn.2024.10.10.21

all disproportionately harm marginalized groups. The need for inclusive design and ethical regulation is clear without them, AI becomes a mechanism for exacerbating harm rather than alleviating it.[9]

- **Ethical AI Implementation: Best Practices and Positive Examples**

  In spite of challenges, there are a number of organizations that show that ethical AI implementation in mental care is achievable. An example is Spring Health, which is a business dedicated to reducing trial-and-error in treatment by using data-driven personalization. Their model involves high standards of transparency, clinical validation, and responsible use of AI. At the heart of their mission is working with diverse, ethically sourced datasets and stringent oversight to minimize bias and risk.[10]

  Mindful Insights Psychotherapy presents yet another framework for the responsible use of AI. They advocate for a balanced practice in which AI technologies assist between sessions without substituting the human touch. This highlights the indispensable value of the therapeutic relationship and the need for human supervision.

  On the policy side, Utah's draft bill regulating mental health chatbots presents an important protection, requiring that licensed mental health professionals be engaged in the design process. This integrates clinical and ethical implications from the beginning.

RECOMMENDATIONS:

- **Require Human Involvement in AI Decision-Making**

  Ensure that all AI-created mental health evaluations and treatment recommendations are checked and validated by trained clinicians to ensure accountability and clinical safety

- **Implement Robust Data Privacy Protections**

[9] Real-world examples of healthcare AI bias. (n.d.). Paubox. Retrieved June 6, 2025, from https://www.paubox.com/blog/real-world-examples-of-healthcare-ai-bias

[10] AI in mental healthcare: How ethical AI is shaping the future of therapy. (n.d.). Spring Health. Retrieved June 6, 2025, from https://www.springhealth.com/blog/ai-in-mental-healthcare

Implement secure encryption, clear data use policies, and empower patients with the right to determine how their mental health information is gathered, stored, and transmitted.

- **Train AI on Diverse and Inclusive Datasets**

  Utilize representative data for diverse demographic groups to limit algorithmic bias and guarantee fair treatment outcomes for all.

- **Involve Licensed Mental Health Professionals in Development**

  Ask certified psychologists and therapists to participate in the development and testing of mental health AI platforms to guarantee relevance and safety.

- **Provide AI Transparency and Explainability**

  Develop AI systems in a manner that users and clinicians see clearly how decisions are reached and when they are being served by an AI, not a human.[11]

CONCLUSION:

"The real question is not whether machines think but whether men do."

- B.F. Skinner

It is a reminder that ethical responsibility belongs to humans and not the tools they develop.

AI promises to transform mental health care by improving access, customization, and efficiency. However, without robust ethics guaranteeing privacy, impartiality, and clarity, it can do harm and erode trust. Its ethical deployment must have human values and clinical governance at its core to guarantee that technology amplifies, not diminishes, compassionate care.

REFERENCES:

1. **BOOKS AND JOURNALS**

- Bostrom, N. (2014). AI and ethical dilemmas in healthcare. Oxford: Oxford University Press.

[11] Whispering hope: Ethical challenges and the promise of AI in mental health therapy. (n.d.). AI and Faith. Retrieved June 6, 2025, from https://aiandfaith.org/insights/ethics-ai-mental-health-therapy/

- Bryson, J. (2018). The Ethics of Artificial Intelligence.
- Bryson, J. (2018). AI ethics and mental health interventions. Cambridge: Cambridge University Press

2. WEBSITES

- Artificial intelligence in mental health care. (n.d.). American Psychological Association. Retrieved June 6, 2025, from https://www.apa.org/practice/artificial-intelligence-mental-health-care
- The impact of AI in the mental health field. (2024, July). Psychology Today. Retrieved June 6, 2025, from https://www.psychologytoday.com/us/blog/invisible-bruises/202407/the-impact-of-ai-in-the-mental-health-field
- Revolutionizing AI therapy: The impact on mental health care. (n.d.). Positive Psychology. Retrieved June 6, 2025, from https://positivepsychology.com/ai-therapy/
- An ethical perspective on the democratization of AI in mental health care. (2024). JMIR Mental Health, 11, e58011. Retrieved June 6, 2025, from https://mental.jmir.org/2024/1/e58011
- AI for mental health: 7 use cases with real-life examples. (n.d.). AIMultiple. Retrieved June 6, 2025, from https://research.aimultiple.com/ai-for-mental-health/
- AI in mental health: Revolutionizing diagnosis and treatment. (n.d.). DelveInsight. Retrieved June 6, 2025, from https://www.delveinsight.com/blog/ai-in-mental-health-diagnosis-and-treatment
- The rise of AI in mental health support: Opportunities, challenges, and ethical considerations. (n.d.). Mindful Insights Psychotherapy. Retrieved June 6, 2025, from https://www.miptherapy.com/blog/the-rise-of-ai-in-mental-health-support-opportunities-challenges-and-ethical-considerations
- AI in mental healthcare: How ethical AI is shaping the future of therapy. (n.d.). Spring Health. Retrieved June 6, 2025, from https://www.springhealth.com/blog/ai-in-mental-healthcare
- Real-world examples of healthcare AI bias. (n.d.). Paubox. Retrieved June 6, 2025, from https://www.paubox.com/blog/real-world-examples-of-healthcare-ai-bias

- Addressing bias and inclusivity in AI-driven mental health care. (2024, October 10). Psychiatric News. https://psychiatryonline.org/doi/10.1176/appi.pn.2024.10.10.2
- Whispering hope: Ethical challenges and the promise of AI in mental health therapy. (n.d.). AI and Faith. Retrieved June 6, 2025, from https://aiandfaith.org/insights/ethics-ai-mental-health-therapy/