

AI-Driven Psychological Profiling on Social Media: Mechanisms, Ethical Breaches, and Regulatory Challenges in Data Inference

Evan Bose^{1*}, Chaitanya Anil Kumar¹, Meenakshi N.²

Abstract

This literature review examines AI-driven psychological profiling on social media, analyzing 21 academic studies that focus on machine learning techniques such as supervised learning, deep neural networks, sentiment analysis, and natural language processing. These methodologies infer mental health indicators—such as depression, anxiety, and stress—from users' digital footprints, encompassing linguistic patterns, engagement metrics, and temporal behaviors. While these tools offer potential for early detection of psychological distress, they also raise significant ethical concerns. Key issues include the absence of explicit informed consent, lack of transparency due to algorithmic opacity, data minimization challenges, and the risk of bias amplification. The study critically assesses how biases in training data can disproportionately impact underprivileged populations, resulting in unfair access to interventions and distorted profiling results. The effects of AI-based profiling on autonomy and stigmatization of mental health are also examined. The assessment also emphasizes the absence of worldwide uniformity in regulatory methods and criticizes current legislative frameworks, such as the CCPA and GDPR, which provide scant protections for inferred psychometric data. It suggests multidisciplinary frameworks to improve data sovereignty, privacy, and accountability while promoting more robust enforcement and public monitoring. The review's conclusion highlights the necessity of long-term study and teamwork to fill up knowledge gaps in algorithmic accountability and ethical governance.

Keywords: AI ethics, psychological profiling, algorithmic opacity, mental health, legal protections.

INTRODUCTION

Background

Artificial intelligence (AI) has made significant strides in the field of mental health, indicating that technology has the potential to improve mental health. From early cognitive models to sophisticated interventions, AI is increasingly recognized for its role in addressing the global mental health crisis.

Tools such as chatbot-based therapy platforms—Woebot, Wysa, Talkspace, and BetterHelp—are proving effective in alleviating symptoms of depression and anxiety.

In today's information-rich environment, user-generated data collected online offers valuable insights into individuals' thoughts, emotions, and behaviours. By examining mobile phone sensor data and social media content, researchers seek to gain a deeper understanding of conditions like depression and anxiety [1]. The development of conversational agents, such as Woebot and Wysa, exemplifies how AI can facilitate the delivery of cognitive behavioural therapy (CBT) and provide emotional support.

*Author for Correspondence

Evan Bose
E-mail: evanbose14@gmail.com

¹Student, Department of Psychology, Bangalore University, Bangalore, Karnataka, India

²Assistant Professor, Department of Psychology, Bangalore University, Bangalore, Karnataka, India

Received Date: November 19, 2024

Accepted Date: December 17, 2024

Published Date: January 15, 2025

Citation: Evan Bose, Chaitanya Anil Kumar, Meenakshi N. AI-Driven Psychological Profiling on Social Media: Mechanisms, Ethical Breaches, and Regulatory Challenges in Data Inference. *Recent Trends in Social Studies*. 2025; 2(1): 1–7p.

The integration of AI and psychology presents opportunities for personalized interventions in mental health assessment and treatment [2]. According to Thieme et al. (2023), machine learning algorithms can analyze enormous volumes of data to find early risk factors for mental health illnesses [3].

Purpose of the Review

This analysis explores the transformative impact of AI on positive mental health outcomes. Integrating AI-driven solutions into mental health practices has significantly enhanced accessibility, personalization, and effectiveness of interventions. The emergence of digital mental health as a sub-discipline has seen AI-powered tools such as chatbots and virtual assistants providing immediate support and customized therapeutic approaches [4]. Furthermore, wearable technology that provides real-time monitoring helps people make well-informed decisions about their mental health.

In order to manage complicated medical datasets and enhance our comprehension of mental health diagnosis and treatment, deep learning techniques are essential. Machine learning technologies offer valuable insights into the implementation of AI in psychological therapies, highlighting advancements made in effectively treating mental disorders [5].

Ethical considerations surrounding AI-driven psychological evaluations are paramount. Addressing these concerns involves examining the efficacy and cost-effectiveness of chatbots delivering cognitive behavioural therapy, as well as analysing sentiment to forecast consumer behaviour. It is also essential to consider the impact of demographic characteristics on psychological applications to optimize AI-based treatment strategies [6].

MACHINE LEARNING TECHNIQUES IN PSYCHOLOGICAL PROFILING

Supervised Learning

Supervised learning is a key machine learning technique utilized in AI-driven psychological analysis, involving the training of algorithms on labelled data to predict outcomes based on identified patterns [7]. Various software tools, such as Weka, scikit-learn, and R, facilitate the development of these models, enabling researchers to tackle tasks like detecting bipolar disorder and recognizing anxiety.

In cybersecurity, supervised learning aids in profiling hackers by analysing personality traits defined by the OCEAN model (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism). Moreover, supervised learning techniques are instrumental in predicting mental health outcomes by studying linguistic patterns, engagement metrics, and temporal behaviours [8].

Deep Neural Networks

Deep neural networks (DNNs) are increasingly recognized as powerful tools in psychological profiling, particularly for analysing extensive and unstructured datasets. These machine learning models excel at identifying complex patterns and making predictions, which is particularly beneficial in mental health diagnostics. They have shown promise in conditions like dementia, utilizing large datasets such as those from the Alzheimer's Disease Neuroimaging Initiative [9].

DNN methods are especially well-suited for managing high-dimensional data, revealing new relationships between mental illnesses and various latent factors [10]. This holistic approach enhances the understanding of individual psychological profiles, which aids predictive modelling and reduces biases in the profiling process [11]. In summary, deep neural networks represent a promising direction for advancing AI-driven psychological profiling.

Sentiment Analysis

The analysis of emotions holds a significant place in AI-powered psychological profiling, particularly when it comes to understanding and predicting consumer behaviour. As highlighted sentiment analysis is utilized to dissect emotions within textual data, offering valuable insights into consumer behaviour.

By scrutinizing linguistic patterns and engagement metrics, AI can identify emotional signals and sentiments expressed by individuals. This data can be essential for customizing personalized treatment plans and therapy recommendations based on an individual's emotional state.

Additionally, Thakkar et al., 2024 emphasizes the importance of AI in fostering positive mental health results through sentiment analysis [5]. AI-driven solutions, like chatbots and virtual assistants, provide immediate support for mental health concerns by analysing sentiments to detect potential issues. Through machine learning algorithms, therapeutic interventions can be tailored based on real-time data from wearable devices, enhancing decision-making processes. It is evident from these references that sentiment analysis plays a crucial role in harnessing AI technologies for psychological profiling.

Natural Language Processing

One important area of artificial intelligence (AI) that greatly aids in psychological profiling is natural language processing (NLP) (Ripla, 2024). Machine translation, text mining, speech recognition, sentiment analysis, and other methods are all used in NLP [12]. Additionally, NLP can be utilized to evaluate written or spoken responses during jury selection processes. By analysing the language used by potential jurors, NLP algorithms provide critical information regarding their psychological profiles and decision-making tendencies.

AI-driven platforms, such as Talkspace, also leverage NLP algorithms to assess language patterns and emotional states during therapy sessions. This analysis helps generate personalized feedback and recommendations for therapists, enhancing the quality of care provided.

INFERENCE OF MENTAL HEALTH CONDITIONS

Linguistic Patterns

Linguistic patterns play a critical role in AI-powered psychological profiling, particularly in cybersecurity. Research has shown that analysing language can help identify potential threats through the usage of specific linguistic features. In a context, specific linguistic indicators have been identified from financial transaction records to predict personality traits [13]. The use of particular mobile payment applications and varying transaction amounts may be linked to levels of Neuroticism or Conscientiousness. Such insights derived from digital footprints provide valuable information about an individual's personality and behavior patterns.

Engagement Metrics

Engagement metrics play a crucial role in AI-driven psychological profiling, serving as vital indicators of how individuals interact with technology. This highlights the importance of engagement metrics in understanding user interactions with AI tools, providing insights into mental well-being.

By analysing engagement metrics such as user activity levels, response times, and content preferences, AI algorithms can gain deeper insights into an individual's psychological state. This information can be instrumental in deducing mental health conditions based on linguistic cues, temporal behaviours, and other relevant variables.

Temporal Behaviours

Temporal behaviours are crucial in AI-driven psychological profiling, particularly for analysing longitudinal data to enhance the accuracy of psychiatric diagnoses as symptoms evolve. AI techniques excel at interpreting patterns from temporal data, offering a deeper understanding of mental health conditions over time. This analysis helps identify shifts in linguistic patterns, engagement metrics, and behavioural trends that may indicate changes in an individual's mental well-being.

Furthermore, these temporal behaviours enable the development of personalized interventions and treatment plans tailored to individual trajectories. By analysing fluctuations in mood, behaviour, and

engagement levels over time, AI systems can create customized support strategies that address specific needs throughout a person's mental health journey.

ETHICAL ISSUES

Informed Consent

In the context of AI-driven psychological profiling, informed consent is a critical consideration. The General Data Protection Regulation (GDPR) emphasizes the lawful, fair, and transparent handling of personal data, requiring individuals to be informed about how their data is processed and to provide appropriate consent [14]. This regulation imposes strict conditions when dealing with special categories of personal data, such as health information used in AI profiling for mental health disorders.

When processing mental data, it is essential to obtain explicit consent from data subjects, particularly when commercial interests are at stake [15]. The existing weak consent frameworks in consumer neurotechnology highlight the need for more robust measures to ensure individuals are adequately informed about the use and processing of their mental data.

Algorithmic Opacity

Algorithmic opacity in AI-driven psychological profiling is a significant concern, as the lack of transparency in deep learning algorithms complicates the understanding of their decision-making processes. This phenomenon, often described as 'black boxes,' raises issues of accountability and trust in AI applications. To address these challenges, it is crucial to develop interpretable AI models and establish robust governance frameworks that promote transparency and accountability in psychological profiling [16].

Data Minimization

The General Data Protection Regulation (GDPR) emphasizes that personal data should only be collected as necessary for its intended use, which helps reduce the risk of misuse or unauthorized access. AI systems typically require large amounts of data, especially during the training phase to identify patterns and correlations. The principle of data minimization mandates that data processing must be limited to what is necessary for achieving specific objectives.

Bias Amplification

Bias amplification in AI-driven psychological profiling stems from inherent biases present in healthcare data and societal systems. These biases can significantly influence AI algorithms, leading to disparities in healthcare outcomes [17]. Implicit biases in clinical decision-making, along with demographic factors such as gender and ethnicity, can skew AI predictions, resulting in biased profiling.

The lack of transparency surrounding AI algorithms complicates the identification and correction of biased results, allowing these biases to persist unchallenged. The ability of AI systems to identify patterns within large datasets can lead to the stereotyping of specific groups, exacerbating and amplifying existing biases throughout the profiling process.

LEGAL PROTECTIONS AND DEFICIENCIES

GDPR

The General Data Protection Regulation (GDPR) represents a significant advancement in data protection legislation within the European Union, focusing on individual rights to privacy and data security. It applies not only to personal data processing within EU boundaries but also to non-EU entities targeting individuals in the EU. This regulation impacts US companies that interact with or monitor EU individuals, emphasizing informed consent and transparency in data handling.

When considering the application of AI for profiling, challenges arise in meeting consent and specificity requirements [18]. The application of AI technologies in healthcare creates new privacy and

confidentiality issues. The GDPR is crucial in protecting individual privacy in this sector, ensuring patient data remains secure from cybersecurity threats.

CCPA

The California Consumer Privacy Act (CCPA) plays a significant role in safeguarding the privacy and data rights of individuals within the state. Mirroring the precedent set by the GDPR, the CCPA enhances consumer privacy protections and advocates for transparency in data collection procedures. Under the CCPA, individuals are entitled to be informed about the collection of their personal information, its purpose, and whether it is shared with third parties for commercial gain.

Furthermore, protecting patient data within the healthcare system is paramount due to the potential risks associated with AI integration, including cyber threats. The CCPA addresses these issues by imposing strict measures to secure sensitive health data and prevent unauthorized access or breaches.

Additionally, there is a need to balance AI-driven innovation with individual rights and societal values [18]. The CCPA ensures that data processing activities adhere to consumer rights and mitigate risks linked to AI technologies.

PROPOSED REGULATORY FRAMEWORKS

Data Sovereignty

Data sovereignty includes ownership issues where personal data ought to be managed by users rather than platform providers. Various interpretations of data sovereignty exist, including national jurisdiction over data flows, Indigenous data sovereignty, and patient data sovereignty which highlight the need for a thorough examination to clarify the definition of data sovereignty and distinguish it from related concepts like cyber sovereignty [19].

Discussions surrounding data privacy emphasize the importance of protecting sensitive information, such as medical histories and behavioral data, especially when artificial intelligence (AI) is utilized in mental health services.

Establishing clear guidelines involves granting people the right to refuse AI profiling based on publicly available data without their informed consent. By specifying these rights and responsibilities, regulatory frameworks can address privacy concerns associated with AI technologies and promote responsible usage of personal information.

Privacy

The ease of access to automated personality assessments raises significant concerns about individuals' privacy and self-governance [20]. Users may unknowingly share information online that can be used for psychological profiling without their consent, leading to potential privacy violations and manipulation of opinions and behaviours.

Privacy concerns arise from the opacity surrounding data collection and use, as many users are unaware of their digital footprints. It is vital to create mechanisms that allow users to access, engage with, and manage their data effectively, promoting privacy rights and addressing ethical issues associated with AI.

Accountability

Research studies highlight the vital role of responsibility in AI-driven psychological profiling, particularly regarding liability and accountability in mental healthcare. The use of AI systems presents difficult issues regarding who has responsibility for unfavorable results resulting from AI advice. To protect patient rights, privacy, and well-being, it is essential to establish clear guidelines and regulations governing the development and use of AI in this field. Different diagnosis and treatment options may

result from preexisting biases in training data. For all patients to get fair and equal care, it is imperative to recognize and address these prejudices.

RESEARCH GAPS

Current literature highlights several areas for further exploration in AI-driven psychological profiling, particularly focusing on enhancing techniques to combat the evolving strategies of cybercriminals. Additionally, investigating bias management in AI systems is crucial and should extend beyond traditional domains like business and accounting. Future inquiries should encompass AI bias in computer science and medicine, consumer bias in e-commerce, as well as gender and racial biases linked to job automation [21]. Moreover, data security concerns regarding social data bias, ethical considerations in tourism, and bias issues in health insurance product development warrant thorough examination.

CONCLUSION

Exploring emerging methodologies such as topic modelling and meta-analysis in AI applications could yield valuable insights into critical areas needing attention. Understanding how collaboration among authors from various disciplines affects different types of reviews can also provide significant insights. Furthermore, assessing review quality is vital for establishing enhanced standards and practices in AI research.

Collaboration between researchers, policymakers, and practitioners is essential to successfully navigating these challenges. Creating robust frameworks that balance innovation with ethical standards will be essential when integrating AI into psychological profiling practices

REFERENCES

1. Uhlenbusch N, Löwe B, Härter M, Schramm C, Weiler-Normann C, Depping MK. Depression and anxiety in patients with different rare chronic diseases: A cross-sectional study. *PloS one*. 2019 Feb 20;14(2): e0211343.
2. Ramon Y, Farrokhnia RA, Matz SC, Martens D. Explainable AI for psychological profiling from behavioral data: An application to big five personality predictions from financial transaction records. *Information*. 2021 Dec 13;12(12):518.
3. Thieme A, Hanratty M, Lyons M, Palacios J, Marques RF, Morrison C, Doherty G. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction*. 2023 Mar 17;30(2):1-50.
4. Wang B. Psychological AI: A critical analysis of capabilities, limitations, and ramifications. *Applied and Computational Engineering*. 2024 Mar 28; 53:205-12.
5. Thakkar A, Gupta A, De Sousa A. Artificial intelligence in positive mental health: a narrative review. *Frontiers in Digital Health*. 2024 Mar 18; 6:1280235.
6. Zhou S, Zhao J, Zhang L. Application of artificial intelligence on psychological interventions and diagnosis: an overview. *Frontiers in Psychiatry*. 2022 Mar 17; 13:811665.
7. Ahmed A, Agus M, Alzubaidi M, Aziz S, Abd-Alrazaq A, Giannicchi A, Househ M. Overview of the role of big data in mental health: A scoping review. *Computer Methods and Programs in Biomedicine Update*. 2022 Jan 1; 2:100076.
8. Garcia-Ceja E, Riegler M, Nordgreen T, Jakobsen P, Oedegaard KJ, Tørresen J. Mental health monitoring with multimodal sensing and machine learning: A survey. *Pervasive and Mobile Computing*. 2018 Dec 1; 51:1-26.
9. Jeyaraman M, Balaji S, Jeyaraman N, Yadav S. Unraveling the ethical enigma: artificial intelligence in healthcare. *Cureus*. 2023 Aug;15(8).
10. Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim HC, Paulus MP, Krystal JH, Jeste DV. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2021 Sep 1;6(9):856-64.
11. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, Jeste DV. Artificial intelligence for mental health and mental illnesses: an overview. *Current psychiatry reports*. 2019 Nov; 21:1-8.

12. Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*. 2023 Sep 22;23(1):689.
13. Kong Y, Ding H. Tools, Potential, and Pitfalls of Social Media Screening: Social Profiling in the Era of AI-Assisted Recruiting. *Journal of Business and Technical Communication*. 2024 Jan;38(1):33-65.
14. Ienca M, Malgieri G. Mental data protection and the GDPR. *Journal of Law and the Biosciences*. 2022 Jan;9(1): 1-06.
15. Peters H, Matz SC. Large language models can infer psychological dispositions of social media users. *PNAS nexus*. 2024 Jun;3(6): page 31.
16. Varsha PS. How can we manage biases in artificial intelligence systems—A systematic literature review. *International Journal of Information Management Data Insights*. 2023 Apr 1;3(1):100165.
17. Sullivan Y, de Bourmont M, Dunaway M. Appraisals of harm and injustice trigger an eerie feeling that decreases trust in artificial intelligence systems. *Annals of Operations Research*. 2022 Jan; 308:525-48.
18. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. In *Artificial intelligence in healthcare 2020* Jan 1 (pp. 295-336). Academic Press.
19. Hummel P, Braun M, Tretter M, Dabrock P. Data sovereignty: A review. *Big Data & Society*. 2021 Jan;8(1):2053951720982012.
20. Olawade DB, Wada OZ, Odetayo A, David-Olawade AC, Asaolu F, Eberhardt J. Enhancing mental health with Artificial Intelligence: Current trends and future prospects. *Journal of medicine, surgery, and public health*. 2024 Apr 17:100099.
21. Raman R, Pattnaik D, Hughes L, Nedungadi P. Unveiling the dynamics of AI applications: A review of reviews using scientometrics and BERTopic modeling. *Journal of Innovation & Knowledge*. 2024 Jul 1;9(3):100517.