AI Misconduct and Prevention in Biomedical Research

Olayori Olasode
251215427

MEDSCIEN9505
Dr. Drysdale
July 17th, 2025

Interdisciplinary Medical Sciences
Schulich School of Medicine and Dentistry
Western University

Assignment Acknowledgement

To maintain transparency and uphold academic integrity, I acknowledge that I used generative artificial intelligence tools, specifically ChatGPT, to assist with the topic idea, editing and reviewing my work for grammar and clarity. This assignment has a Turnitin originality score of 14%.

Part One: Overview

Artificial Intelligence (AI) is quickly transforming medicine, with uses growing in diagnostics, treatment planning, research, and public health. Its ability to process large and complex datasets drives breakthroughs where traditional methods fall short. AI helps to speed up scientific discoveries and healthcare advances (Buch et al., 2018).

In clinical practice, AI is increasingly used to define medical tasks, assist in clinical decision-making, and improve care in resource-limited settings. For example, AI algorithms in dermatology have shown superior sensitivity and specificity in detecting skin lesions, often matching or exceeding the accuracy of experienced dermatologists. In radiology, AI systems help optimize imaging protocols to decrease unnecessary radiation exposure, while radiologists continue to guide treatment, demonstrating that AI acts as a tool rather than a substitute. Although powerful, AI lacks human qualities like empathy, ethical reasoning, and the ability to adapt to unexpected clinical nuances, emphasizing the need for human oversight (Buch et al., 2018).

In the field of biomedical research, AI plays a vital role in improving research accuracy, decreasing manual workload, and enabling complex modelling that supports drug discovery, genomics, and personalized medicine. AI also assists in strategic health planning and epidemic surveillance, providing real-time insights that can enhance disease control efforts and public health interventions (Sharma et al., 2024).

As AI continues to evolve, its integration into healthcare and biomedical research presents both exciting opportunities and ethical considerations. Ensuring transparency, fairness, and accountability in AI systems will be crucial to maximizing their benefits while safeguarding human dignity and trust in medicine (ai-admin, 2023).

Despite its transformative potential, ethical misuse often follows, and generative AI is no exception. The use of AI to assist in the production of scientific manuscripts has surged in recent years, but this rapid growth has also led to increasing abuse (Tang, 2024). While AI can support writing efficiency, it is now being used to generate entire papers from scratch, which raises concerns about authorship integrity, plagiarism and dilution of academic standards.

In the ongoing epidemic of “publish or perish” in biomedical research, the incentive structure has shifted away from quality over quantity, pressuring researchers to meet publication quotas at the expense of scholarly rigor (Tang, 2024).

Part One: Strategy To Commit Ethical Misconduct with The Use Of AI In Mental Health Diagnostics Via Digital Phenotyping

Introduction:

Mental health disorders, especially depression, anxiety, bipolar disorder and schizophrenia, often go undetected until symptoms become severe (Welch & Carson, 2018). With traditional diagnostics relying heavily on self-reports and clinical visits, early warning signs can be missed.

Digital phenotyping refers to the moment-by-moment quantification of individual human phenotypes using data that is generated from their smartphones and other connected devices, forming their behaviour and other factors that involve health (Huckvale et al., 2019). In conjunction with AI, it can be used to enable the analysis of the stream of data to identify patterns associated with emerging mental health conditions.

Data Without Consent

In collecting behavioural data from users like screen time, typing patterns, and voice recordings, without their full knowledge or explicit consent, researchers or developers risk violating principles of informed consent. This breach occurs when individuals are unaware of the extent, purpose or sensitivity of the data gathered, thereby undermining their autonomy and ability to make informed decisions about their participation. These practices are ethically concerning, particularly in mental health and biomedical contexts. In doing this, burying permissions in the fine print or a lengthy Terms of Service document will bypass true informed consent.

This approach leverages a subtle yet powerful shift in how data is collected, embedding permissions in complex terms and conditions that users normally skip. By bypassing explicit consent, this plan enables continuous, real-time data collection from large, diverse populations, which will later produce a massive dataset that can be quickly fed into AI systems to generate appealing results.

Building Predictive Models with Biased or Incomplete Data

Training AI algorithms on non-diverse datasets skewed toward specific demographics, such as white individuals, men, or urban populations, neglects the critical need to test generalizability across age groups, ethnicities, and socioeconomic backgrounds. While using biased or incomplete data may simplify model training and yield faster, seemingly accurate results within the narrow scope of that demographic, it compromises broader applicability. This approach can lead to quicker publications, funding, and media attention compared to more rigorous efforts aimed at building inclusive and generalizable models. Ultimately, by showcasing high-performance metrics without addressing real-world diversity, such strategies present an illusion of success while bypassing essential ethical and scientific standards.

Inflated Results Without Peer Validation

By leveraging widespread trust in artificial intelligence, researchers can overstate the diagnostic capabilities of digital phenotyping for complex mental illnesses like schizophrenia and depression, presenting it as a revolutionary advancement. Such claims can prematurely fast-track publications, attract awards, and generate institutional or media attention. In truth, AI in this context may only detect surface-level behavioural indicators, such as changes in speech patterns, phone usage, or sleep habits, which are not sufficient for a reliable diagnosis. When these findings are based on small, homogenous samples or involve cherry-picked data, the results may appear highly accurate within a limited scope but fail to generalize to broader populations. Publishing such work without robust clinical validation or peer-reviewed replication creates a false sense of innovation. This practice not only misleads clinicians, patients, and policymakers but also erodes scientific integrity and compromises ethical principles of transparency, justice, and patient autonomy in mental health research.

Overall, each unethical roadmap compromises the core ethical principles of autonomy, justice, and transparency that underpin responsible research and clinical practice. What begins as a promising innovation, such as AI-powered digital phenotyping, can quickly evolve into a mechanism of manipulation, exploitation, and harm when misused. Ironically, these unethical approaches often achieve more visibility and success than ethically sound models, as they bypass rigorous oversight in favour of speed, sensationalism, and short-term gain. Without intervention, these dynamic risks are reshaping the future of mental health diagnostics in profoundly harmful ways.

Part Two: Countermeasure

To combat the unethical use of AI and digital phenotyping, including data collection without consent, biased modelling and unvalidated claims, a proactive and enforceable regulatory framework is important. Mental health diagnostics are personal and sensitive, as misuse not only erodes public trust but can cause direct harm to individuals misdiagnosed or overlooked by flawed algorithms. Addressing these threats required standardized ethical safeguards, clinical oversight and accountability mechanisms.

Implementation strategies:

1. Mandated Informed Consent Protocols
    a. Regulatory bodies such as Public Health Agencies and Research Ethics Boards must require the use of clear, standalone informed consent forms for digital phenotyping to separate from the general app or device Terms and Services.
    b. Consent forms must mention specifics on the data included, like screen time, typing patterns, and audio through phone calls.
        i. The forms should also include how these metrics will be analyzed, stored and the intended use, including whether it will support research, diagnostics or both.
    c. To ensure ongoing transparency, periodic re-consent should be mandated, particularly when datasets evolve or are repurposed for new uses. This ensures that patients remain informed and in control of their data.
2. Diversity Audits and Dataset Disclosure
    a. Before publication or deployment, AI models must undergo independent data audits to assess demographic diversity and generalizability.
        i. Journals and funding agencies should mandate transparency reporting of population characteristics (i.e, race, gender, age).
            1. Researchers must also define a clear exclusion/inclusion criterion and report model performance across demographic subgroups.
        ii. This method ensures accuracy across diverse populations and accounts for cultural, economic and social factors that influence mental health outcomes.
    b. Developers must pre-register AI diagnostic tools detailing data sources, a preset sample size, and a study hypothesis. This process helps prevent publication bias and discourages cherry-picking results, thereby enhancing reproducibility and scientific integrity.
3. Clinical Validation Before Diagnostic Claims
    a. All AI-based mental health diagnostic tools must undergo independent clinical trials before any diagnostic claims are publicly made or marketed.
    b. The validation process should involve review by mental health professionals, ethicists and statisticians to ensure that the tools align with current diagnostic standards, such as the up-to-date DSM (Diagnostic and Statistical Manual).
        i. The model also ensures the demonstration of robust performance across diverse populations to avoid demographic bias and ensure that these findings are statistically sound with no reliance on misleading correlation.
4. System Maintenance and Whistleblower Protection
    a. Ongoing maintenance of AI diagnostic systems should be a mandatory condition for federal or institutional funding. This includes regular performance evaluations, updates to reflect new clinical guidelines, and mechanisms for ethical oversight.
    b. Institutions must also establish anonymous reporting systems that allow researchers and employees to safely report instances of data manipulation, unethical conduct, or misuse of AI diagnostic tools.
        i. Protection for whistleblowers must be legally and institutionally guaranteed to encourage transparency and accountability.

Overall, these countermeasures reinforce the ethical foundations of transparency, autonomy, and accountability in AI-driven mental health research. By preventing the misuse of digital phenotyping, they help restore public trust, promote equitable data practices, and protect patients from potential harm caused by biased or unvalidated tools. Ensuring diverse representation in datasets, requiring rigorous validation, and supporting whistleblower protections all contribute to a research environment that prioritizes clinical relevance over sensationalism. Ultimately, these measures shift incentives toward meaningful, reproducible progress and set a lasting precedent for the responsible and ethical use of AI across the healthcare system.

References

ai-admin. (2023, December 5). *AI for Social Good*. AI for Social Good. https://aiforsocialgood.ca/blog/artificial-intelligence-enhancing-human-lives-a-revolutionary-era-of-possibilities

Buch, V. H., Ahmed, I., & Maruthappu, M. (2018). Artificial intelligence in medicine: current trends and future possibilities. *British Journal of General Practice*, *68*(668), 143–144. https://doi.org/10.3399/bjgp18x695213

Huckvale, K., Venkatesh, S., & Christensen, H. (2019). Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. *Npj Digital Medicine*, *2*(1). https://doi.org/10.1038/s41746-019-0166-1

Sharma, A., Al-Haidose, A., Al-Asmakh, M., & Abdallah, A. M. (2024). Integrating Artificial Intelligence into Biomedical Science Curricula: Advancing Healthcare Education. *Clinics and Practice*, *14*(4), 1391–1403. https://doi.org/10.3390/clinpract14040112

Tang, B. L. (2024). AI is revolutionizing biomedical research, but are there any negatives? *PubMed*, *23*, 1226–1228. https://doi.org/10.17179/excli2024-7761

Welch, K. A., & Carson, A. J. (2018). When psychiatric symptoms reflect medical conditions. *Clinical Medicine*, *18*(1), 80–87. https://doi.org/10.7861/clinmedicine.18-1-80