AI Technology and Bias in Detecting Psychiatric Disorders

Kennedy A. Flowers

The brain's frontal and temporal lobes are responsible for our social skills, memory, behavior, audio, emotional, and language processing. Schizophrenia is a psychiatric disorder that impairs a person's auditory and cognitive processing and affects their thoughts, behavior, and actions (NYU Langone Health, n.d.). Schizophrenia is also associated with a reduction in the size of the brain's frontal and temporal lobes, which can impact stimuli processing and cause hallucinations (Cleveland Clinic, 2024). Schizophrenia affects a disproportionate amount of Black men (especially those with other mental conditions) compared to other ethnic and racial groups (Gara, 2019). While this trend can be partially explained by genetic and environmental factors such as family history, childhood health complications, urbanicity, and substance abuse, a prominent reason that Black men are diagnosed with schizophrenia is due to racial bias present in the AI (Artificial Intelligence) technology used to aid physicians in detecting and diagnosing psychiatric disorders, as well as systemic racial bias present in America's healthcare system.

Keywords

Artificial Intelligence (AI) • Mental Health • Epigenetics • Intersectionality • Schizophrenia

Introduction

Artificial Intelligence and Mental Health

AI (Artificial Intelligence) technology has seeded its place in the healthcare field as a benevolent asset to physicians and scientists examining psychiatric disorders. Presently, AI technology is used to aid neuroscientists in testing hypotheses and physicians in processing imaging data and diagnosing diseases via machine learning and deep learning, respectively (Glaser et al., 2019). **Machine learning** allows physicians and scientists to analyze large sets of data, including functional magnetic

University of Michigan, kflow@umich.edu

doi: 10.3998/ujph.7619

Conflicts of interest:
The author has no conflicts of interest to disclose.

resonance imaging (fMRI) data used to diagnose psychiatric disorders (Gur, R. E. & Gur, R. C., 2010). **Deep learning** is used to identify exactly how one area of the brain relates to another and how stimuli affect the processing and firing of neurons in those areas (Glaser et al., 2029). When trained on behavioral tasks, previous research has shown that deep learning networks can model the human brain quite well and allow scientists to gain a deeper understanding of the brain and how it functions. In a similar vein, AI-powered *psychiatric* resources can improve a patient's accessibility to mental health resources. They could even help detect and treat psychiatric disorders, such as autism spectrum disorder, schizophrenia, bipolar disorder, and more. However, all patients do not enjoy these monumental developments equally. Time and time again, we see disastrous results when patients are misdiagnosed due to an inherent racial bias present in AI-based algorithms. Without proper treatment, untreated patients may turn to illegal substances to numb their pain, creating a spiral of abusive, toxic, or fatal events. This harmful trend is one excessively felt by minority patients, who are frequently misdiagnosed and mistreated in America's healthcare system.

As previously stated, recent developments in AI software are intended to combat these issues and improve access to quality mental health resources. Still, if AI technology disproportionately misdiagnoses minority patients, its usefulness diminishes significantly. In this article, I examine the history of psychiatric disorders in minority communities and analyze the bias involved in AI-powered psychiatric systems.

Body

What schizophrenia is

Schizophrenia is a chronic brain disorder impacting 1 in every 300 people (World Health Organization, 2022). The onset or predominant stage of this disorder is categorized as a mild form of psychosis in which a person may voice abnormal beliefs or behaviors, hallucinate, experience delusions or catatonic (sporadic or withdrawn) behavior, or cannot perform daily activities, such as keeping up with their hygiene. Before someone is officially diagnosed, a doctor performs a physical exam, looking through their medical and family history before having them describe their experiences. Blood tests, such as an enzyme-linked immunosorbent assay (ELISA) and complete blood count (CBC), as well as an fMRI test, rule out other psychiatric disorders (e.g., major depressive and bipolar disorders) and test biomarkers and image abnormalities that can indicate schizophrenia (Lai, 2016). Blood-based biomarkers include a reduction in brain-derived neurotrophic factor (BDNF), a protein and neurotrophic growth factor that encourages the growth of new neurons, and an increase in white blood cells, such as T lymphocytes and monocytes (Dakshinamurti, 2005). fMRI abnormalities include a decreased activation of the prefrontal cortex and an increase in amygdala activation for negative emotions (fear and anger) compared to positive emotions (pleasure and happiness) (Lai, 2016). Affecting 24 million worldwide (World Health Organization, 2022), active schizophrenia causes those afflicted to suffer hallucinations, speech impairment, and cognitive issues (Cleveland Clinic, 2024). Furthermore, unhealthy coping mechanisms, such as smoking and substance abuse, increase feelings of stress and anxiety (The Recovery Village, 2018) and worsen symptoms such as hallucinations, which in turn can trigger psychotic episodes (National Drug and Alcohol Research Centre, 2011). In addition, unhealthy coping mechanisms can interfere with and lower the effectiveness of antipsychotic medication and can exacerbate a patient's feelings of helplessness and frustration. This creates a negative feedback loop: negative coping mechanisms

lower the effectiveness of antipsychotic medication and/or positive coping strategies (e.g. therapy and rehabilitation) and increase feelings of anxiety and stress. In response, the patient relies more heavily on the aforementioned unhealthy coping mechanisms, and the cycle continues.

Epigenetics

In 2019, a study of 1,647 patients at a community behavioral healthcare clinic was conducted at Rutgers University (Gara et al., 2019). The study, which examined patients' medical records, investigated the use of depression screening when assessing and diagnosing schizophrenia in new patients. The study, led by psychiatry professor Michael Gara, found that "There [was] a tendency for clinicians to overemphasize the relevance of psychotic symptoms [while overlooking] symptoms of major depression in African-Americans compared with other racial or ethnic groups" (Gara et al., 2019). Unfortunately, this tendency is the reason that many black patients are often misdiagnosed and mistreated for schizophrenia when they are suffering from a major depressive disorder. This experience is especially common in Black men, who are often racially stereotyped as being as "threatening" or "irrationally aggressive". These stereotypes, whether we like them or not, flow through our brains subconsciously. This is the reason why many clinicians failed to weigh mood symptoms effectively when making their diagnoses in this study and is reflective of a trend throughout America.

The root cause of this trend is tied to *epigenetics* and *intersectionality*. Epigenetics is the study of how a person's environment can modify gene expression without changing DNA sequences (Centers for Disease Control and Prevention, n.d..). These changes may be passed down generations, and some have explored epigenetics in conversation with the idea of *intergenerational trauma* in Black Americans as the result of a repeated history of extremely traumatic events (such as slavery, torture, serial rape, etc). Transgenerational inheritance of trauma via modifications to gene expression (Youssef et al., 2018) causes many Black children to experience symptoms of PTSD concerning certain topics or situations from the moment they are born. Trauma is passed through DNA and/or gene transcription (Youssef et al., 2018), the process of making an RNA copy of a DNA sequence during egg fertilization (Centers for Disease Control and Prevention, n.d.). When a body experiences trauma and/or repeated, long-term traumatic events, the body may respond by altering gene expression, and some of these changes may be passed down to subsequent children. While the gene sequence remains the same (i.e., there are no physical changes to DNA), the way that a body *reads* and *responds* to an individual's DNA is changed and passed down through generations (Johnson 2023). This trauma has both a biological and physical effect on the body and can cause chronic stress and hypertension, among many other health issues (Klengel & Binder, 2015). While epigenetics as a whole is concerned with a variety of alterations in gene expression through chemical modifications of DNA, the study of intergenerational trauma and its effect on epigenetics is an emerging area of study in the field, and along with environmental differences (determined by family dynamics, socioeconomic status, regular exposure to stressful situations, etc.) could be linked to an increase in the risk of developing schizophrenia.

Dr. Joy DeGruy puts the concept of Epigenetics into a racial context in her book, *"Post Traumatic Slave Syndrome: America's Legacy of Enduring Injury & Healing."* Dr. DeGruy defines P.T.S.S. as "a condition that exists as a consequence of multigenerational oppression of Africans and their descendants resulting from centuries of chattel slavery" (DeGruy, 2005, p.168). This condition is not a new idea: scientists around the world have been researching ancestral trauma for years. Slavery

perpetrated the idea that Blacks were genetically inferior to Whites, leading to institutionalized racism that further fuels this idea (DeGruy, 2005, p.41).

For the Black community, hundreds of years of unaddressed, inherited trauma still impact our day-to-day lives. Being Black in America means carrying a heavy burden: inherited PTSD, caused not by our parents or grandparents or their behaviors but by the trauma that our enslaved ancestors endured that fundamentally altered the genes of Black Americans. This chronic stress, coupled with subconscious biases in the healthcare system, led to the harmful trend found in the Rutgers study. Repeated experiences of neglect, ignorance, and apathy by physicians and other medical professionals in the healthcare space create two outcomes: 1) a negative reputation surrounding Medical professionals and their tendency to misdiagnose and mistreat Black patients that is passed via horror stories through friend groups and circles of loved ones, which in turn leads to 2) increased distrust of physicians and other Medical professionals by Black patients, which may lead these patients to experience fear and/or anxiety around visiting the doctor's office and discourage them from seeking Medical care unless there is a dire emergency.

Intersectionality

Many people do not belong to a single minority. The term intersectionality is an all-encompassing term that accounts for the existence of "double" or "triple" minorities, or people who belong to multiple minority groups (United Nations Network on Racial Discrimination and Protection of Minorities, n.d.). Intersectionality is important because it helps explain the trends in mental health diagnoses that we often see with African-American men and women and people of color in general. Various socioeconomic factors, like poverty, ancestral and lived traumas, and stereotyping that affect people of color (particularly African-Americans) influence the negative stigmas in healthcare, making Black Americans more reluctant to seek treatment and care for their mental (and physical) health.

Artificial Intelligence (AI) and Mental Health

Researchers use AI to analyze electronic health records (alongside blood tests and brain images), questionnaires, voice recordings, behavioral signs, and more alongside supervised machine learning, deep learning, and Natural Language Processing (NLP) models to gain a deeper understanding of a patient's mental and physical state (Varnosfaderani & Forouzanfar, 2024). Additionally, virtual therapists like Woebot, Replika, and Wysa offer personalized therapy by acting as AI "chatbots" where patients can describe their feelings to the therapists (Haque & Rubya, 2022). AI chatbots aim to make care more accessible and affordable for patients and increase the efficiency of existing systems by automating administrative tasks, such as making appointments and delivering health education (Varnosfaderani & Forouzanfar, 2024). However, the potential **bias** in AI systems can render chatbots and virtual therapists insufficient for providing adequate social and emotional support for patients with psychiatric disorders and decrease their quality significantly.

AI and Implicit Bias

Researchers have discovered that current NLP models in leading AI-assisted mental health resources may widen existing health inequalities. Popular language learning algorithms include

GloVe and Word2Vec, which assess psychiatric terms and demographic labels. Researchers examined these NLP models by analyzing how the algorithms related the terms to the labels and found distinct religious, racial, gender, nationality, sexuality, and age-related biases embedded in the programming of both GloVe and Word2Vec (Straw & Callison-Burch, 2020). The algorithms directly correlated Black patients with a schizoaffective disorder diagnosis *without* any patient input, suggesting an implicit bias present in these models. Racial bias in medical settings is one explanation for *why* the differences in the effectiveness of AI-powered psychiatric resources for the Black patient population are often overlooked and unaddressed. The implicit biases in these NLP models are not inherently permanent, however. A major reason for these biases is the lack of minority representation in the early trials of these systems (Straw & Callison-Burch, 2020). Key developers of these models can also influence the biases in these algorithms by implementing their subconscious prejudices.

Conclusion

Removing biased practices and integrating equitable solutions can be achieved by employing one of three de-biasing techniques: *data manipulation, fine-tuning biases,* or *increasing representation in early development teams* (Nazer et al., 2023). **Data manipulation** involves using data augmentation algorithms that will increase the weight of the limited data available for minority patients, which can even out the gaps in data between patient groups. Data from particularly large sets can be removed if found to contain biases as well, balancing out the overall system. **Fine-tuning biases** involves re-purposing an existing model using a balanced, unbiased dataset, thus ridding the existing model of inherent biases. The third technique is based on the idea that the existing biases in these models are influenced by those who created the models. The amount of diversity present in the teams that make these models has been found to correlate directly with the amount of biases present in the systems created. Therefore, **increasing the diversity** of the teams working on these systems will help to create a balanced set of data to be used by these models.

Artificial Intelligence is used in a multitude of ways to aid physicians in diagnosing schizophrenia and many other psychiatric disorders. AI technology in the form of Natural Language Processing models is used to assess and interpret data collected by physicians and neuroscientists alike (Straw & Callison-Burch, 2020). However, researchers at the National Institute of Health found that many popular models, such as GloVe and Word2Vec have been found to contain inherent racial, gender, and age biases (Straw & Callison-Burch, 2020). The data inputted into these systems contains racial markers and is the basis of the predictive strategies these systems use to diagnose patients and associate certain physiological features with certain races. In practice, this is not necessarily a harmful feature of AI-powered diagnostic technology. Understanding health trends in relation to marginalized groups can provide evidence backing Public Health initiatives aimed at correcting these issues. However, this feature can also perpetuate harmful biases. In this case, both the GloVe and Word2Vec models diagnosed patients under the "African-American" label with a schizoaffective disorder diagnosis solely based on their ethnicity label (Straw & Callison-Burch, 2020). This unfortunate model defect leads to a disproportionate amount of Black men who are diagnosed (or simply misdiagnosed) with schizophrenia. To rid these models of their inherent biases the data processed by these models must be balanced.

References

Centers for Disease Control and Prevention. (n.d.). *Epigenetics, health, and disease*. https://www.cdc.gov/genomics-and-health/epigenetics/index.html

Cleveland Clinic. (2024). *Living with schizophrenia*. https://health.clevelandclinic.org/living-with-schizophrenia

Columbia University Irving Medical Center. (2015). *Untimely deaths in people with schizophrenia*. http://www.cuimc.columbia.edu/news/untimely-deaths-people-schizophrenia

Dakshinamurti, Krishnamurti (2005). Biotin — a regulator of gene expression. *The Journal of Nutritional Biochemistry, 16*(7), 419–423. https://www.sciencedirect.com/science/article/pii/S0955286305000884

Diagnosing schizophrenia. (n.d.). NYU Langone Health. https://nyulangone.org/conditions/schizophrenia/diagnosis

Gara, M. A., Minsky, S., Silverstein, S. M., Miskimen, T., & Strakowski, S. M. (2019). A naturalistic study of racial disparities in diagnoses at an outpatient behavioral health clinic. *Psychiatric Services*, *70*(2), 130–134. https://doi.org/10.1176/appi.ps.201800223

Glaser, J. I., Benjamin, A. S., Farhoodi, R., & Kording, K. P. (2019). The roles of supervised machine learning in systems neuroscience. *Progress in Neurobiology, 175*, 126–137. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8454059/

Gur, R. E., & Gur, R. C. (2010). Functional magnetic resonance imaging in schizophrenia. *Dialogues in Clinical Neuroscience, 12*(3), 333–343. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181978/
Guidance note on intersectionality, racial discrimination & protection of minorities. (n.d.). https://www.ohchr.org/sites/default/files/documents/issues/minorities/30th-anniversary/2022-09-22/GuidanceNoteonIntersectionality.pdf

Haque, R., & Rubya, S. (2022). An overview of chatbot based mobile mental health applications: Insights from app description and user reviews (Preprint). *JMIR MHealth and UHealth*, *11*(11). https://doi.org/10.2196/44838

Johnson, S. (2023). Understanding epigenetics: How trauma is passed on through our family members. *Arkansas Advocate*. https://arkansasadvocate.com/2023/07/05/understanding-epigenetics-how-trauma-is-passed-on-through-our-family-members/

Joy DeGruy Leary, & Robinson, R. (2005). *Post traumatic slave syndrome: America's legacy of enduring injury and healing*. Joy DeGruy Publications Inc.

Kaur, A., Basavanagowda, D. M., Rathod, B., Mishra, N., Fuad, S., Nosher, S., Alrashid, Z. A., Mohan, D., & Heindl, S. E. (2020). Structural and functional alterations of the temporal lobe in schizophrenia: A literature review. *Cureus, 12*(10), e11145. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7689947/

Klengel, T., & Binder, Elisabeth B. (2015). Epigenetics of stress-related psychiatric disorders and gene × environment interactions. *Neuron*, *86*(6), 1343–1357. https://doi.org/10.1016/j.neuron.2015.05.036

Lai, C.-Y., Scarr, E., Udawela, M., Everall, I., Chen, W. J., & Dean, B. (2016). Biomarkers in schizophrenia: A focus on blood-based diagnostics and theranostics. *World Journal of Psychiatry, 6*(1), 102–117. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4804259/

Nazer, L., Zatarah, R., Waldrip, S., Janny, X. C. K., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, *2*(6), e0000278–e0000278. https://doi.org/10.1371/journal.pdig.0000278

National Drug and Alcohol Research Centre. (2011). *NDARC psychosis final*. https://ndarc.med.unsw.edu.au/sites/default/files/ndarc/resources/NDARC_PSYCHOSIS_FINAL.pdf

ScienceDaily. (2019). African-Americans more likely to be misdiagnosed with schizophrenia, study finds. *ScienceDaily*. http://www.sciencedaily.com/releases/2019/03/190321130300.htm

Straw, I., & Callison-Burch, C. (2020). Artificial intelligence in mental health and the biases of language-based models. *PLOS ONE, 15*(12), e0244094. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC7745984/

The Recovery Village. (2018). *Anxiety and substance abuse*. The Recovery Village. https://www.therecoveryvillage.com/mental-health/anxiety/substance-abuse/

Varnosfaderani, S. M., & Forouzanfar, M. (2024). The role of AI in hospitals and clinics: Transforming healthcare in the 21st century. *Bioengineering*, *11*(4), 337. https://www.mdpi.com/2306-5354/11/4/337

World Health Organization. (2022). *Schizophrenia*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/schizophrenia

Youssef, N., Lockwood, L., Su, S., Hao, G., & Rutten, B. (2018). The effects of trauma, with or without PTSD, on the transgenerational DNA methylation alterations in human offsprings. *Brain Sciences*, *8*(5), 83. https://doi.org/10.3390/brainsci8050083