The Science Behind the Guilty Brain and the Future of Neuroprediction

Violet Park

Abstract

As artificial intelligence and neuroimaging advance, society must consider the extent to which predictive algorithms should be applied to everyday life, such as within the legal system. By measuring how "guilty" you feel according to your guilt-related brain signature (GRBS), neuroprediction in court can be utilized for the benefit of humanity and permanently alter the way the justice system runs. Despite the advanced technology we've been able to formulate thus far, we run into obstacles when it comes to bias, privacy concerns, and consent. As society progresses toward the future, it's crucial not to overlook the evident concerns and reflect on how far we're willing to let technology take us.

Introduction

Whether it's lying to someone or secretly stealing a cookie from the cookie jar, we've all felt guilt. Guilt often haunts us and can sometimes push us to confess the truth. But what if there was a way to measure guilt? Knowing whether someone is guilty or not could help us decide whether those accused of a crime truly did commit it. We would be able to ensure those who are guilty pay for their crimes and those who are innocent are proven so. To know how to measure guilt, we first have to define what guilt even looks like.

What Does A "Guilty" Brain Look Like?

Before examining how neuroprediction and other methods of predictive algorithm work, we need to understand how to measure guilt and the science behind it. Scientists have identified what is known as a guilt-related brain signature, or GRBS, that is expressed when one feels responsible for the harm of another (Hongbo et al., 2020). GRBS serves as a key biomarker, which is some measure that indicates a condition like a disease or infection. GRBS is present in conditions of physical pain and emotional memories, which means it has generalizability and can be applied to not only a single sample but in a wide spread of studies. In addition to GRBS, self-reported guilt has often been associated with activation of the anterior cingulate cortex (ACC). ACC was activated whenever the individual was perceiving another's suffering or had the knowledge that their actions were causing the suffering of others. Although there is belief that guilt may be found in more than a single voxel of the brain, associations between the ACC and guilt can be a great starting point for understanding what parts of our brain are encoded by guilt.

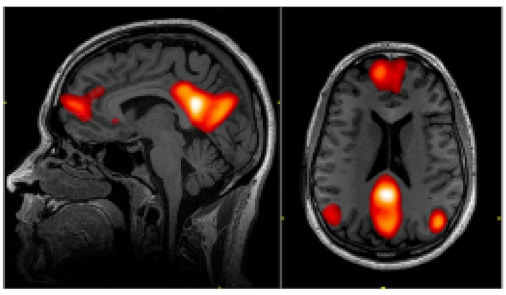

Figure 1. Displays how guilt-related brain signatures are expressed in the brain via electroencephalography (EEG) signals

Current Methods of Neuroprediction

Neuroprediction is the ability to predict human behavior by utilizing neurocognitive data. It can be applied to anticipating recidivism, which is the tendency of a convicted criminal to reoffend (Tortora et al., 2020). The only consistent methods of neuroprediction today are derived from fMRI scans and neuroimaging with AI. These fMRI scans can analyze activity of the ACC,which is in charge of impulse control and error processing. Based on ACC activity, scientists have found that the probability that offenders with low ACC activity would be arrested was approximately double compared to offenders with high activity. Previous fMRI data has shown to be useful in predicting the completion of substance abuse treatment within a prison inmate population using event-related potentials (ERPs) and functional network connectivity (FNS), which identified "neural fingerprints" that predicted cocaine abstinence during treatment (Elliott et al., 2020). However, there are drawbacks to measuring ACC activity alone. There is a likelihood that guilt is stored in more than one part of the brain resulting in an inaccurate measure of guilt. Essentially, neuroimaging identifies potential neurocognitive markers and combines it with statistical machine learning methods to create a multi-voxel pattern analysis (MVPA). MVPA has been readily used in healthcare for years to determine differences between healthy and diseased brains, but it has also been used to measure other factors such as the intention to perform one task over another, sequential stages of task preparation, and lie detection. Unlike examining ACC activity, which compares experimental conditions to identify which brain regions are activated by particular tasks, MVPA looks at patterns of brain activity to decide what subjects are looking at or thinking about like a mind-reading or brain-reading technique. In addition, MVPA has the ability to read the brain in the domain of visual perception by looking at how experiences are encoded in the brain. This is done by training a deep neural network to perform visual image reconstruction from the brain and decode visual content of dreams.

Ethical Issues and the Legal System

Neuroprediction has been a rising topic in the legal system as it could be key in deciding criminal sentences, parole, use of the death penalty, and discharge. The risk assessment

analyzes characteristics about the individual from criminal history, drug use, job history, childhood abuse, and more, to determine their risk of recidivism. Although there are high hopes for utilizing neuroprediction in court, current risk assessment displays poor to moderate accuracy with more than half of individuals targeted as high-risk being misidentified. Ethical issues associated with the use of neuroprediction include bias, privacy, and consent and coercion. In the past decade, there have been numerous cases in which race or gender has often played a role in misidentification, for example, Amazon Rekognition software incorrectly matched members of Congress with people who had been charged with a crime. The facial recognition software disproportionally wrongly identified African American and Latino members of Congress. Similar algorithms such as Predpol in 2016 unfairly targeted certain neighborhoods with a high proportion of people from racial minorities regardless of effective true crime rates. Regardless of the algorithm, those against neuroprediction argue that identifying guilt will always be as biased as the people who use it. For instance, AI trained on data such as criminal files may reflect biases on part of police officers, prosecutors, or judges.

Another pressing issue involves privacy. Researchers have found that neurodata can be used to screen job applications or lead to the commercialization of medical records. If data collected by neuroprediction falls into the wrong hands, privacy of individuals can become easily breached and spread to major corporations or to the general public. In the courtroom, use of machine learning methods can often lead to discussion about consent and coercion. Although there are arguments that algorithms designed to identify high-risk and low-risk offenders could be utilized for legal decisions, performing cognitive violations by forcing people to undergo scans without consent for sentencing becomes a concern. In addition, neuroimaging can exert a “seductive allure” that makes jury and judges overestimate accuracy of neuroscientific images making it misleading and create cognitive biases in the evaluation of evidence. With the endless concerns regarding the use of neurodata, we must evaluate whether the benefits of neuroprediction overpower the costs

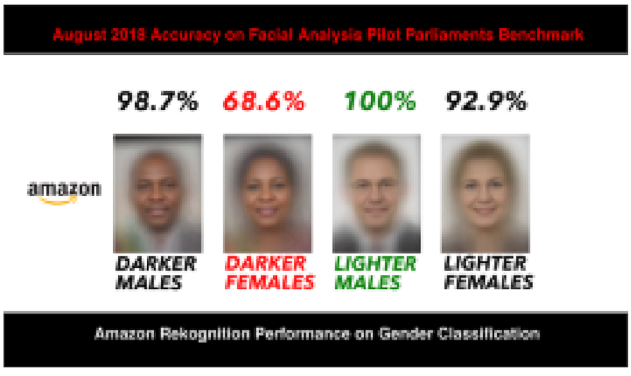

Figure 2. Analysis of Amazon Rekognition software’s accuracy presents gender and ethnic bias

Conclusion

As researchers continue to advance the art of neuroprediction, we must take into consideration the negatives that come with it. The use of neurodata is a powerful tool that would allow us to measure “guilt” in the brain and analyze what one is likely thinking. A program like this would give us the opportunity to prevent crimes before they occur, as well as assess the risk of a criminal to help determine their sentence. Despite these benefits, concerns regarding bias, privacy, and consent and coercion as well as the current inaccuracy of neuroprediction bring into light whether we should even continue to work on improving neuroprediction. If implemented into our society, how could we ensure that racial and gender bias are not factors in determining guilt and what can we do to eliminate coercion of offenders to agree to the programs? Before we fully develop neuroprediction, let’s first consider as a society how far we’re willing to take technology.

References

1. Buolamwini, J. (2019, April 24). Response: Racial and gender bias in Amazon rekognition-commercial AI system for analyzing faces. Medium. https://medium.com/@Joy.Buolamwini/response-racial-and-gender-bias-in-amazon-rekognition commercial-ai-system-for-analyzing-faces-a289222eeced
2. Buxton, R. B. (2013, September 4). The physics of Functional Magnetic Resonance Imaging (fmri). Reports on progress in physics. Physical Society (Great Britain). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376284/
3. Claydon, L., Catley, P. (2022, September 13). If a brain can be caught lying, should we admit that evidence to court? here's what legal experts think. The Conversation. https://theconversation.com/if-a-brain-can-be-caught-lying-should-we-admit-that-evidence-to-co urt-heres-what-legal-experts-think-80263
4. Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What Is the Test-Retest Reliability of Common Task-Functional MRI Measures? New Empirical Evidence and a Meta-Analysis. Psychological Science, 31(7), 792–806. https://doi.org/10.1177/0956797620916786
5. Green, S. (2012, October 1). Guilt-selective functional disconnection of anterior temporal and subgenual cortices in major depressive disorder. JAMA Network. https://jamanetwork.com/journals/jamapsychiatry/article-abstract/1171078
6. Steele, V. R., Maurer, J. M., Arbabshirani, M. R., Claus, E. D., Fink, B. C., Rao, V., Calhoun, V. D., Kiehl, K. A. (2017, August 1). Machine learning of Functional Magnetic Resonance Imaging Network Connectivity Predicts Substance Abuse Treatment Completion. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. https://www.sciencedirect.com/science/article/pii/S2451902217301210
7. Tortora, L., Meynen, G., Bijlsma, J., Tronci, E., Ferracuti, S. (2020, January 31). Neuroprediction and A.I. in Forensic Psychiatry and Criminal Justice: A neurolaw perspective. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00220/full

8. Yu, H., Koban, L., Chang, L. J., Wagner, U., Krishnan, A., Vuilleumier, P., Zhou, X., Wager, T. D. (2020, February 21). Generalizable multivariate brain pattern for interpersonal guilt. OUP Academic. https://academic.oup.com/cercor/article/30/6/3558/5735622