# Developing Wearable Headband for Enhancing Slow Wave Sleep in Older Adults

# Gender Bias and Discrimination in Machine Learning

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Biomedical Engineering

By
Justin Tran

[11/29/2023]

Technical Team Members:

Josiah Calhoun, Hannah Park, Devanshi Radadiya, and Keerthana Vijayaragavan

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for thesis-Related Assignments.

ADVISORS

MC Forelle, Department of Engineering and Society

Timothy Allen, Department of Biomedical Engineering

**[Introduction]**

There is often bias and discrimination in machine learning as a result of faulty data collection and training. Machine learning is a growing branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms in order to imitate the thinking process of humans (Brown, 2023). The term "machine learning" was first coined by Arthur Samuel in his research of creating a computer algorithm that can beat humans in checkers (Gabel, 2019). Since then, the world of machine learning has grown, integrating more itself into society. Use cases of machine learning can be seen through generative AI such as Chat GPT, speech recognition software like Siri, and anomaly detection algorithms used by banks. Interest in the growth of machine learning has allowed for innovation with the hopes of creating a more "smart" and efficient society.

With the growing dependence on machine learning and its application to various features in society, it is clear that measures for objectiveness and fairness should be in place to ensure that benefits of this cutting edge technology is shared by all.Although machine learning is often seen as a key tool for eliminating the human element in creating data-driven analytical solutions, machine learning has had a notorious history with bias and discrimination stemming anywhere from gender, race, and socioeconimics. One prime case of this bias can be seen through a study done in 2018 on Amazon's hiring algorithms. When processing the resumes of early candidates, it was discovered that when put through their resume scraping software, resumes mentioning certain words such as "women" were penalized, resulting in male candidates being inadvertently favored over their female counterparts (Gillis, 2023). Another such case of bias in machine learning can be seen through the Green Light Project (GLP) that took place in Detroit. The Detroit police department conducted a project that involved the installation of HD cameras

throughout the city in order to stream video surveillance footage straight into police database systems and into a facial recognition algorithm in order to fight crime and improve neighborhood security (Detroit PD, 2023). This project was designed to protect neighborhoods and to take human error out of the equation by implementing artificial intelligence for accuracy and efficiency. In reality, this project only perpetuated the problem by making false positive identification on people of color, and by creating over policing of certain areas majorly with African American/LatinX demographics (Rauenzahn, 2021).

It is evident that there are problems with machine learning and the effects of machine learning applications may have on already disadvantaged groups. That's why in the technical portion of this project, the data used to train the machine learning model used to read slow wave sleep signals in order to send acoustic stimulation must be checked and verified in order to avoid biases. In the STS portion of this project, there will be a focus on the deployment of machine learning in hiring algorithms in order to determine how biases and discrimination against certain groups are perpetuated within these programs.

**[Technical]**

The goal of the technical project is to develop of a wearable device for Alzheimer Disease (AD) patients that uses machine learning to process electroencephalography (EEG) signals as a means to determine the most optimal time to relay acoustic stimulation in the form of pink noise in order to improve memory retention. As of now, there are two ways to treat Alzheimer's Disease patients, through medicine or therapy. For the medicinal path, one may take drugs in order to slow or lessen the progression of Alzheimer's Disease by removing the plaque build up in the brain, or they can take drugs to treat the symptoms of Alzheimer's Disease by affecting chemicals that are involved in neural communication within the brain. (Alzheimers

Association, 2023). As for therapy, cognitive stimulation therapy (CST) and cognitive rehabilitation is utilized in order to stimulate the activity in parts of the brain in order to slow the down progression of AD and to improve cognitive ability. (Livingston, 2020).

The proposal for this project consists of designing a wearable device for AD patients that uses machine learning to process electroencephalography (EEG) signals as a means to determine the most optimal time to relay acoustic stimulation in order to improve memory retention and consolidation. This is a good solution because it is a completely non-invasive method to slow down the progression of AD, and only requires a one time payment instead of continuously paying for medication. Typically, plaque build up within the brain is cleared during slow-wave sleep, which is when cerebral spinal fluid (CSF) flows complementary to the blood pulses within the brain (Wafford, 2021). As people age, slow-wave sleep is reduced by nearly 80% starting from 20 to 50 years old. As a result, the plaque clearing property of slow-wave sleep with the help of blood and CSF oscillation is not fully utilized. Thankfully, acoustic stimulation can actually improve slow-wave activity, which inturn helps with clearing the plaque build up that impairs memory retention (Papalambros, 2017).

In the past, there have been many cases of bias within machine learning frameworks that is used within healthcare. The Framingham Heart Study cardiovascular risk score performed very well for Caucasian but not African American patients, which means that care could be unequally distributed and inaccurate. In the field of genomics and genetics, it is estimated that Caucasians make up about 80 percent of collected data, and thus studies may be more applicable for that group than for other, underrepresented groups (Igoe, 2021). The key actor of this product that is the difference between a functional medical device and an expensive alarm clock is the machine learning algorithm used to trigger the built in acoustic stimulation. This product will

address the problem of bias and discrimination in machine learning by taking into account that this is a medical device that will be used on patients of various ages, backgrounds, and ailments. As a result, the development team is fully aware of the need to carefully pick totally encompassing data to be able to accommodate for the widest range of possible end-users. Additionally, because this device is a non-invasive therapy for memory retention in AD patients, it may have other potential use cases such as early preventative measures for people without AD, or even for patients suffering from memory impairment through years of drug use. The byproduct of this fact allows the developing team to hopefully create a device that will not have bias and discrimination due to the fact that a fundamental constraint of the product is to be able to service everyone of any background, race, or gender.

**[STS]**

The field of machine learning as a whole is quite large, so I will narrow the scope of this research project to craft a clear research question: "To what extent does the deployment of machine learning models in hiring algorithms contribute to or mitigate racial and gender discrimination?" This question will become the crux of our inquiry, enabling a focused exploration of the sociotechnical intricacies at play in the biases and discrimination led by machine learning.

Evidence of bias and discrimination is plentiful within the realm of machine learning in resume reading algorithms. For example, it was revealed in 2015 that the AI recruitment tool used to screen candidates' resumes by Amazon was highly gender discriminative. The model deployed by Amazon was trained using the applications of candidates for the past 10 years. Due to the tech industry being a predominately male dominated field, the resulting resume screening algorithm gained preference for male candidates over their female counterparts. Additionally,

resumes containing words such as “women” or mentions of graduating from all-women colleges were penalized while resumes exhibiting words often found in the resumes of male engineers such as “executed” and “captured” were favored (Dastin, 2018).

In another study done by the University of Pennsylvania, there was evidence to suggest that white women and minority men are disadvantaged, especially when employers are seeking candidates from STEM majors (Kessler et al., 2019 ). In the study, researchers developed a model to digest applications of fabricated candidates, and simulate the evaluation of employer preferences. As a result, resumes with white sounding female names and names associated with being a minority scored lower than stereotypical white male names for roles in STEM. Additionally, on average a female with a grade point average of 4.0 would score about the same as the average white male with a GPA of 3.75. This just goes to show that the deployment of machine learning on resume reading algorithms can often illuminate bias and discrimination within the current hiring culture.

In order to better understand the problem and answer the question, the Actor-Network Theory (ANT), a pivotal framework within Science, Technology, and Society (STS) (Law, 2015 ) will be used to unravel the intricate interactions between human actors (users and developers) and non-human actors (machine learning models and algorithms) . Through this framework, the construction of sociotechnical systems, the influence of power dynamics on their design and deployment, and the emergence of bias and discrimination in resume reading software as unintended consequences that will be examined.

Research for this topic will be done through extensive literature review and interviews with companies utilizing resume screen technology. By focusing on the statistical analysis of resume screening algorithms and current regulations and guidelines for machine learning

deployment in hiring practices, our research is poised to illuminate how these algorithms can either perpetuate or mitigate the racial and gender discrimination.

**[Conclusion]**

In the pursuit of addressing bias and discrimination in machine learning, our research project combines technical and sociotechnical aspects to offer a comprehensive and responsible approach. Our primary technical deliverable is the development of a wearable device capable of monitoring slow wave sleep signals using machine learning and delivering acoustic stimulation in the form of pink noise. This device has the potential to significantly improve memory retention in Alzheimer's disease patients, thereby enhancing their quality of life and reducing caregivers' burden. The positive change stemming from this technical innovation will directly benefit patients, caregivers, and healthcare professionals, allowing them to better understand and manage the cognitive challenges associated with Alzheimer's disease.

In tandem, our sociotechnical deliverable focuses on the societal and ethical implications of deploying machine learning application in real world settings. We aim to address fairness and to understand the underlying factors that have allowed for biased and discriminatory applications to be released in the first place when downstream consequences that may reinforce perpetuated inequalities to already disadvantaged groups. This sociotechnical aspect of our research contributes to resolving the problem of bias and discrimination in machine learning by providing insights into the ethical and responsible implementation of AI-based large language model technologies. Our findings will be of interest to healthcare administrators, policymakers, ethics committees, and researchers, as they seek to establish guidelines and best practices for the ethical use of machine learning in the workplace and in healthcare. By combining technical innovation

with sociotechnical responsibility, our research strives to make a positive impact and advance the broader field of machine learning ethics.

The fast-paced growth of the field and market of machine learning requires that there is extra care into the development and deployment of these technologies. The synergy between these projects ensures that while we strive for technical innovation to increase efficiency in our lives, we do so in a way that minimizes potential biases and ethical issues. This integrated approach contributes to resolving the problem of bias and discrimination in machine learning by demonstrating a real-world application that embodies fairness and ethical use of machine learning technology. The combination of technical and STS research allows us to not only innovate but also to innovate responsibly, ensuring the well-being of vulnerable populations.

## References


Association, A. (2023). Medications for Memory, Cognition and Dementia-Related Behaviors. Alzheimer's Disease and Dementia. https://alz.org/alzheimers-dementia/treatments/medications-for-memory

Brown, S. (2023). Machine learning, explained. MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained

Dastin, J. (2018, October 11). Insight—Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/idUSKCN1MK0AG/

Gabel, F. (2019). Some studies in machine learning using the game of checkers. Artificial Intelligence for Games: Seminar. https://hci.iwr.uni-heidelberg.de/system/files/private/downloads/636026949/report_frank_gabel.pdf

Gillis, A. S. (2023). What is Machine Learning Bias? Definition from WhatIs. https://www.techtarget.com/searchenterpriseai/definition/machine-learning-bias-algorithm-bias-or-AI-bias

Igoe, K. J. (2021). Algorithmic Bias in Health Care Exacerbates Social Inequities—How to Prevent It. Executive and Continuing Professional Education. https://www.hsph.harvard.edu/ecpe/how-to-prevent-algorithmic-bias-in-health-care/

Kessler, J. B., Low, C., & Sullivan, C. D. (2019). Incentivized Resume Rating: Eliciting Employer Preferences without Deception. The University of Pennsylvania. https://faculty.wharton.upenn.edu/wp-content/uploads/2018/09/KesslerLowSullivan_Revision1.pdf

Law, J. (2015). STS as a Method. The Open University. http://heterogeneities.net/publications/Law2015STSAsMethod.pdf

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., … Mukadam, N. (2020). Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. 396(10248), 413–446. https://doi.org/10.1016/S0140-6736(20)30367-6

Papalambros, N. A., Santostasi, G., Malkani, R. G., Braun, R., Weintraub, S., Paller, K. A., & Zee, P. C. (2017). Acoustic Enhancement of Sleep Slow Oscillations and Concomitant Memory Improvement in Older Adults. Frontiers in Human Neuroscience. https://www.frontiersin.org/articles/10.3389/fnhum.2017.00109

Project Green Light Detroit. (2023). City of Detroit. https://detroitmi.gov/departments/police-department/project-green-light-detroit

Rauenzahn, B. (2021). Facing Bias in Facial Recognition Technology. The Regulatory Review. https://www.theregreview.org/2021/03/20/saturday-seminar-facing-bias-in-facial-recognition-technology/

Wafford, K. A. (2021). Aberrant waste disposal in neurodegeneration: Why improved sleep could be the solution. Cerebral Circulation - Cognition and Behavior, 2, 100025. https://doi.org/10.1016/j.cccb.2021.100025