

# Ethics Paper

Wilson Schwegler

`schwegler@oxy.edu`

Occidental College

## 1 Introduction

The development of self-driving vehicles and advanced driver assistance systems has driven large advancements in traffic safety technology. Traffic sign recognition is a critical aspect of these advancements. The topic for my senior comps project is implementing a Convolutional Neural Network (CNN) for traffic sign recognition. While traffic sign recognition holds immense promise for improving road safety, achieving a fully ethical implementation presents multiple challenges. This paper will explore several key ethical concerns surrounding my senior comps project.

## 2 Background

At the core of traffic sign recognition is a complex algorithm inspired by the human brain called a Convolutional Neural Network (CNN). These networks learn by processing information through interconnected layers of artificial neurons. CNNs are a deep-learning algorithm that takes an image as an input and assigns importance to various aspects of the image. By analyzing training data consisting of labeled images, CNNs can differentiate and identify different images Sumit (2018) [7].

## 3 Data Bias

One major ethical concern surrounding traffic sign recognition utilizing a CNN is data bias. Training data plays an important role in creating the model and shaping its performance. When the training data does not truly represent the population that your model is predicting then bias can be prevalent in the model Google (2024) [2]. Data bias is a major issue in the machine learning community and has had major real-life negative effects on people. For instance, training data for facial recognition over-represents white people, creating errors when attempting facial recognition on people of color IBM Data and AI Team (2023) [3].

In creating a CNN for traffic sign recognition, bias can be in the data set if the data primarily reflects specific regions, areas, weather conditions, variations in traffic sign design, or insufficient representation of warning signs on

low-volume roads. Biased data could lead to lower recognition accuracy for certain demographics of drivers, such as those driving in underrepresented areas or conditions. This raises concerns about fairness and potential safety risks, especially for lower recognition accuracy. If the model were to struggle to recognize stop signs in the snow then it could lead to accidents in the weather condition where accidents are already more likely. This could potentially do more harm than good as people who have developed a reliance on the CNN will be at risk in conditions where there is lower recognition accuracy.

When utilizing a pre-made dataset it can be difficult to verify its composition and to identify potential biases like those listed earlier. For instance, a model trained in one source like a city will struggle to recognize road signs for drivers in rural areas where their roads contain warning signs. This raises another point that pre-made data sets may not reflect the diversity of real-world traffic signs, leading to lower recognition accuracy.

### 3.1 Issues With Addressing Data Bias

Addressing data bias in my senior comps project presents several major issues. Firstly, acquiring a comprehensive dataset that encompasses the intended demographic will be challenging. Secondly, even with a large dataset, identifying and mitigating bias is a complex and difficult task.

## 4 Accessibility

While traffic sign recognition could potentially provide a safer driving experience, accessibility concerns threaten to exclude and leave some people behind. Language barriers pose a significant challenge for drivers unfamiliar with the language displayed on the traffic sign. While drivers may be familiar with universal traffic signs, a lack of language familiarity can be dangerous when driving in foreign countries RoadTrafficSigns (2012) [5]. Creating traffic sign recognition that will translate the traffic sign to the users' first language is crucial to allow access to all groups, as without it, major groups of people will not have access to utilize the technology.

Furthermore, economic disparities are a major factor in unequal access. Training a CNN model can require a large

amount of computing power which can also create a barrier for entry. After the model is created smaller companies or developing nations might lack the necessary resources to train the model on data pertaining to them. This may leave them having to be reliant on the pre-trained model that may not be best suited for their country or may contain biases. This also raises the concern that people who live in rural areas who may be suffering from the algorithm bias may not be able to eliminate the bias due to the inability to retrain the model. In the US the digital divide is still very prevalent. From 2019 to 2021 the share of Americans that have home broadband or a smartphone have not changed significantly. Americans with lower incomes are also more likely to be relying on their smartphone to replace a computer, leaving them struggling to complete tasks intended for a computer Vogels (2021) [8]. This puts Americans who are lower income or living in rural areas in a position where they do not have access to work on or have access to a sufficient model that will provide the most amount of road safety. This creates unequal access to an accurate model between wealthy and low-income Americans. Lastly, traffic sign recognition is useless to people who are unable to drive. A project like this intended to improve public safety as a whole will leave out a group of people who do not drive. While they may benefit from the project through overall road safety when riding as a passenger, they will not benefit from the project directly.

#### **4.1 Issues With Addressing Accessibility**

Achieving accessibility in my project presents significant challenges. One major issue would be integrating multiple languages within the model. This would require additional processing power that may not be available or compromise the effectiveness of the system. Another challenge is addressing the economic disparities. This major issue seems like it may be too vast for me to overcome and address in my senior comps.

### **5 Power Concentration**

The concentration of power surrounding traffic sign recognition presents a complex ethical issue. While this technology promises to enhance road safety, it is important to understand who ultimately benefits. An over-reliance on traffic sign recognition could lead to a power imbalance between the car manufacturers and the creator of CNN as they have a significant influence over driver behavior. The creator would be able to dictate the parameters of safe driving as well as have the ability to collect vast amounts of data from the users. This collected data could be used for targeted advertising or be sold to a third-party company, making it virtually impossible to separate your personal real life

from your online footprint.

In the tech industry, it is commonly known that knowledge is power. This would concentrate their power as they now have access to this large amount of data that can be sold to other tech companies or utilized for profit through data-driven decisions Miller (2021) [4]. The various amounts of data collected by a traffic sign recognition system could potentially shift power even more. Potentially identifiable location data and driving patterns shift power towards authorities as this information can be used for surveillance. This would concentrate power for authorities as their already existing ways of surveillance through road cameras and other means now become even more effective. Moreover, the act of outsourcing some driving functions could lead to a societal shift where responsibility for accidents becomes a complex legal issue, shifting the power to the creator of the CNN.

#### **5.1 Issues With Addressing Power Concentration**

My project, by itself, has limited influence on the overall power dynamics within traffic sign recognition. However, it is important to consider how my work could contribute to the issue. Advocating for strong data privacy regulations and user consent protocols can help prevent misuse of data. My project could also serve as a starting point for research into open-source and collaborative development. This would help decentralize power and promote innovation. However, it is worth noting that my project will have very minimal if any influence on the power structures that currently exist.

### **6 Transparency and Explainability**

Transparency and Explainability are important ethical concerns for a traffic sign recognition system. CNNs are often called black boxes where there is no traditional rule-based program with clear decision-making logic. They are very good at recognition by identifying relationships within data, however, there is no reasoning behind how and why they made a classification. This is a major issue in the field of machine learning. It is a big deal because it makes it difficult to fix systems when they produce unwanted outcomes. The infinite number of permutations makes it virtually impossible to know if the system is robust enough to handle every situation. This can make it difficult to trust these systems especially when it comes to making decisions about human safety as they can reflect biases from our human world Blouin (2023) [1].

The issue of safety validation is also very prevalent in traffic sign recognition. The inability to fully assess the robustness of the model due to a lack of reasoning makes it

difficult to determine how the model will perform in real-world scenarios. This can lead to deploying a system that performs well in controlled environments but fails in real-world use cases and causing harm.

The lack of transparency raises concerns about a traffic sign recognition system's accountability. This can especially be an issue in the event of an accident or a misinterpretation. The inability to see how a CNN identified an image will make it difficult to understand what caused an error and fix it. This is especially an issue if the error caused a crash as it will be difficult to understand if the issue is actually fixed when a solution is implemented. This can also be an issue when determining if there is bias in the training data. Without insight from the model, it is difficult to pinpoint biases and correct them.

### 6.1 Issues With Addressing Transparency and Explainability

My project faces major limitations in achieving perfect transparency and explainability. While utilizing techniques like attention maps where the tokens' weighting is visualized so we can see which ones received the most attention during processing can offer some insight, a comprehensive understanding of its reasoning will be unknown. This presents challenges in debugging the models and identifying biases.

## 7 Privacy, Security, and Consent

Privacy, security, and consent are other important ethical concerns for traffic sign recognition. This is also a major issue in the machine learning community and is one of the primary concerns of AI. The continuous collection of data from these systems can cause sensitive information to be exposed. Another concern is group privacy as AI can draw patterns from large groups of people and lead to stereotyping and a bias algorithm Sullivan (2023) [6]. Systems like this rely on collecting data from the real world, whether it is video or images. These pieces of data can include potentially identifiable location information or reveal one's driving patterns. This sort of data collection raises some major concerns over privacy. If user content for data collection is not explicitly obtained or the data is used beyond traffic sign recognition like surveillance by the authorities, then there are major privacy, security, and consent concerns. This shows how necessary it is to have a robust security system in place to ensure that sensitive data is safe from unauthorized access or breaches. A cyberattack on a traffic sign recognition system could compromise user privacy and disrupt traffic. Obtaining clear content from the user is crucial as without it, the ethical issues will lead to public distrust.

### 7.1 Issues With Addressing Transparency and Explainability

My project will need to address these issues throughout the development and deployment stages. One major challenge will be ensuring the dataset utilized is anonymous. This will be a difficult and possibly impossible task as I will be using a pre-made data set with many entries of data. Security is another concern if I am storing user data. If so, I must ensure that a robust encryption method is in place to protect user data. Furthermore, obtaining clear and informed user consent is critical to ensure full transparency.

## 8 Conclusion

Traffic sign recognition using a CNN offers immense potential for improving road safety, however achieving a fully ethical implementation remains a complex and difficult challenge. This paper explored some major ethical concerns including data bias, accessibility, power concentration, transparency, privacy, security, and consent. Addressing these concerns seems to be a major task that will be difficult and in some cases impossible to overcome. Concerns like power concentration and erosion of privacy present fundamental challenges that cannot be changed in the scope of my project. The road to a fully ethical traffic sign recognition system is long, however by acknowledging and actively addressing these concerns, may still be difficult to address all ethical concerns for this project and complete it ethically.

## References

- [1] Blouin, Lou. *AI's mysterious 'black box' problem, explained*. 2023. URL: <https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained>.
- [2] Google. *Bias in Machine Learning*. 2024. URL: <https://newsinitiative.withgoogle.com/resources/trainings/bias-in-machine-learning/>.
- [3] IBM Data and AI Team. *Shedding light on AI bias with real world examples*. 2023. URL: <https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/>.
- [4] Miller, Jon. *Knowledge Is Power: How Data Is Feeding Disruption*. 2021. URL: <https://www.forbes.com/sites/forbesbusinesscouncil/2021/11/10/knowledge-is-power-how-data-is-feeding-disruption/?sh=7e20ed3d1515>.

- [5] RoadTrafficSigns. *Signs and Language: A Universal Stop*. 2012. URL: <https://www.roadtrafficsigns.com/traffic-sign-universal-language#:~:text=While%20drivers%20might%20be%20familiar,foreign%20countries%20without%20ideographic%20signs..>
- [6] Sullivan, Morgan. *Examining Privacy Risks in AI Systems*. 2023. URL: <https://transcend.io/blog/ai-and-privacy>.
- [7] Sumit, Saha. *A Comprehensive Guide to Convolutional Neural Networks*. 2018. URL: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>.
- [8] Vogels, Emily. *Digital divide persists even as Americans with lower incomes make gains in tech adoption*. 2021. URL: <https://www.pewresearch.org/short-reads/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/>.