Self-driving cars are a new but promising technology that has been becoming increasingly prevalent within the last few years. This technology stands to improve our current transportation standards as well as reduce traffic and traffic related incidents but does not come without some potential ethical challenges that become very prevalent in the development and usage of such devices. In the case of an incident where an automotive accident has become unavoidable, the algorithm controlling the vehicle must decide what the best course of action is. The primary ethical concern of this lies in the potential for a biased algorithm that makes incorrect decisions with a given scenario, leading to events that could disproportionately endanger people based on race, gender, or socioeconomic status.

As self-driving technology continues to grow and become more prevalent within our daily lives, it is important that we for ethical safeguards to guarantee the minimization of risks. Without policy, we run the risk of harming vulnerable communities and creating distrust within societal groups. It is the responsibility of engineers, policymakers, and the companies developing these products to ensure that these products are put through due process and tested properly. By exploring these concerns through engineering ethics, we can ensure that the technology is developed in a manner that accomplishes its potential while also keeping users, pedestrians, and stakeholders as safe and accountable as possible.

This paper will present an analysis of algorithmic bias within self-driving technology within current engineering practices and address ethical dilemmas that it raises. To do so, apply ethical principles presented within this class and apply them to propose actionable ethical solutions. These solutions strive to both mitigate risk and bias while optimizing performance and application potential. This paper strives to emphasize the critical role of engineering ethics in developing technologies to reduce bias and ensure quality and responsible development.

The rapid development in self-driving technology has been driven forward by advances in artificial intelligence, machine learning, and sensor technology. Companies such as Tesla, Waymo, and Cruise are among those that have released several autonomous vehicles designed to reduce the number of human error related incidents, which account for 94-96% of all automotive incidents (TRAFFIC SAFETY FACTS, 2017). Despite these promising advancements, reliance on artificial intelligence and machine learning introduces ethical dilemmas, like the presence of algorithmic bias. Algorithmic biased often stems from bias present within the training data used, resulting in a biased representation of societal-economic groups that result in unfair and unsafe driving decisions made by the technology.

For example, it has been found that in state-of-the-art object recognition systems, like those found in self-driving cars, are less likely to identify pedestrians with darker skin tones, even when controlling common factors that could impede recognition across the distinct groups (Hao, 2019). This could lead to slower reaction times that affect the response of the algorithm, increasing the risk of an accident. This is caused by a misrepresentation of members of this demographic within training data used in these models, creating a bias in the algorithm that could be accounted for by engineers. Secondly, the decision-making models behind these technologies

are often unclear, leading to life-critical decisions that can be explained to neither users nor stakeholders.

These ethical dilemmas are becoming increasingly prominent now that self-driving vehicles are moving from experimental trials to commercial deployments. The inhibited ability to deal with bias within self-driving algorithms will lead to disproportionate harm to the biased groups, breaching the ethical responsibility that engineers hold to the users of their product and their stakeholders. Additionally, such events could lead to negative public opinions of the technology, inhibiting innovation in the field and delaying adoption of the technology depriving society of its potential benefits.

The application of consequentialist, deontological, and virtue ethics practices allows for the improvement in the developmental process for algorithms in self-driving vehicles, allowing us to examine the cause of biases within the process, dive into the ethical issues behind them, and identify viable solutions. These schools of thought work together collectively to universally apply the maximum number of societal benefits, optimize moral obligations, and encourage professional accountability in the development and deployment of these systems. By examining the problems through each of these approaches, engineers can better understand the trade-offs and benefits of the decisions made and the methods used to deliver them. Using this information, developers can align themselves with ethical principles while providing a deliverable product that provides societal benefits and promotes public trust in the product.

Using a consequentialist line of thought, engineers are pushed to maximize the safety of the deliverable of our work for all users. It encourages developers to critically evaluate their design's impacts, including the algorithms, object detection methods, and sensors used in the process. It also requires that they evaluate the data that is used to train their models, such as identifying potential biases that may be present within the data to prevent, and biases being carried over to their produce. Identifying these issues early in the process could lead to easy changes or adjustments in the development of the product that could produce a better result.

Through deontological principles engineers are also encouraged to emphasize the importance of developing products that promote safety to users and avoid causing harm wherever possible, promoting equality among all groups. Regardless of additional requirements, complexity wise or cost, it is important that the utmost priority is placed on the safety and fairness of the product, to users or otherwise. To quantify a quality product, it is important for the product to be developed in a non-biased way, and to be developed in a transparent and non-construed way that is easy for both users and stakeholders to understand. Lastly, it is important that self-driving cars are developed and deployed in a way that respects the inherent rights of all people, not only the ones that stand to profit or benefit from the use or sale of the product.

Virtue ethics supplements these approaches by placing strong emphasis on the professional integrity and social responsibility of the engineers. If challenges developers to live up to values like accountability, fairness, and a pursuit of excellence so that their work reflects the ideals of society. In the context of self-driving cars, virtue ethics points to the need for

proactive identification and addressing of biases, even when difficult or resource consuming. Merging these approaches, the analysis moves beyond the question of immediate technical problems to creating an ethical culture in engineering practice where fairness and equality go together with innovation and efficiency. This comprehensive approach helps ensure that self-driving cars realize their potential as transformative, inclusive, and socially responsible technologies.

There are a variety of issues associated with algorithmic bias in autonomous vehicles, with the first being biased training data leading to unfair decision making by the algorithms. For example, a dataset that underrepresents a certain demographic may lead the autonomous systems to recognize individuals from those groups. Another problem associated with self-driving cars is the "black-box" problem, which states that it is hard for engineers or stakeholders to identify why an AI model made a certain decision in a situation, making it difficult to identify or diagnose errors. Finally, it is difficult to assign blame when accidents occur, or harm was done due to a biased algorithm. Assigning this blame, whether it be to engineers, companies, or the program itself, is a complicated issue.

Several solutions may help resolve such ethical challenges. One potential solution is applying diverse and representative datasets to reduce bias. Such demographics and scenarios should be considered in the training data if an engineer wants AI systems to do better in diverse contexts. However, this method has several resource implications for developing, annotating, and validation of large datasets, while issues of privacy and consent of source data are raised.

Second is the independent auditing of third-party algorithms for fairness: independent audits would provide more objectivity in the conformance of AI systems with ethical standards. While accountability and increased public trust arise out of this approach, firms may also resist this approach because it can raise questions of ownership or increase compliance expenses.

Another resolution involves designing AI systems to be transparent and interpretable. In other words, more transparent models enable others to understand how certain decisions have been made, thus allowing them to find and reduce the bias more effectively. Often, a trade-off is made between transparency and performance because highly interpretable models may be less complex which may result in a decrease in performance or accuracy.

Finally, government mandates or regulations for the development of self-driving vehicles could be instated to establish a standard procedure. Regulations could establish a minimum criterion that companies would be expected to comply with to ensure the integrity and transparency of the product. However, this option offers strict guidelines and repercussions that might inhibit creativity within the field or put pressure on smaller organizations with fewer resources.

While each of these solutions has its own strengths and weaknesses, they should be seen in terms of their costs, feasibility, and support from stakeholders. A proper balance of each of these factors allows for the product's development to minimize bias and increase public support for the

technology. An implementation of one or potentially a combination of these solutions could deliver a vast improvement to the development of the technology, as well as improve the usage and support for it.

This paper's recommendation to solve the ethical concerns brought forward would be mandating third-party auditing for self-driving object recognition algorithms. This resolution makes sure that before being deployed on the road, self-driving systems are independently assessed against criteria of fairness, safety, and ethics. Audits by third-party players help discover the biases of AI models, review the representativeness of the data used in the development of those models, and shed light on decision-making. Although not without its challenges, this approach provides a critical balance between ethical considerations and the need for speed in the innovative cycle of the autonomous vehicle industry.

Mandatory third-party audits apply to all three ethical frameworks: consequentialism, deontology, and virtue ethics. The consequentialist would say that audits maximize benefit for society by preventing bad outcomes from bias decision-making processes and ensuring that self-driving vehicles provide fair safety outcomes for all users. Fair and unbiased AI systems build public trust in the technology necessary for wide-scale deployment of self-driving vehicles and all the societal benefits that come with them, such as fewer traffic incidents and reduced traffic on the road.

From a deontological point of view, mandatory audits are a guarantee of fairness and transparency, which should already be present in the design if the engineers and companies are acting ethically. This third-party testing ensures that the systems used in the self-driving vehicles meet ethical standards in a way that protects the rights of all people. Audits also provide a check to the “black-box” problem, ensuring that the system acts in a way that protects the inherent rights of all individuals and questions potentially negative outcomes.

This solution also helps to reinforce Virtue Ethics, as it helps to ensure that engineers and companies are adhering to virtues such as honesty, transparency, and respect for autonomy. The requirement for third-party audits will enforce that developers are acting virtuously. This will also help to gain public trust within the industry, helping it stand out among the autonomous vehicle field and improve public usage and support.

There are many different strengths to a mandate of third-party auditing. It promotes fairness by the systematic identification and mitigation of algorithmic bias, ensuring the safe and fair operation of autonomous vehicles across diverse populations. It also increases transparency and accountability, which increases public trust in the field, which is an important aspect of trying to deploy new and up-coming technology. Finally, standardizing auditing practices provides an industry standard that lowers the chances of ethical issues arising within the industry.

This recommendation does not come without its own challenges, however. Third-party audits come with additional costs, which will raise development costs. Smaller companies operating within the industry may be forced out due to this increased cost, as well as increasing the price

for the consumer. Companies may also resist audits for fear that the exposure of their intellectual property could lead to competitive disadvantages. To address these weaknesses, there could be additional policy, or a regulatory body implemented to subsidy smaller companies and regulate the process. They could offer incentives to companies to ensure competition and innovation, as well as ensure that proprietary information is protected throughout the process.

Mandatory third-party auditing is a practical and ethically robust solution to the problems of self-driving vehicles, such as algorithmic bias and the “black-box” issues. By addressing fairness, transparency, and accountability, it aligns with core ethical principles and directly responds to concerns. The process of implementing such a solution offers many benefits such as assurance of qualities such as safety, equality in representation, and increasing public trust. Although it will need careful consideration to minimize cost increases and overcome resistance, it sets a great precedent for ethics with the autonomous technology field. This ensures self-driving cars can be responsible and inclusive while servicing the public good.

References

Hao, K. (2019, March 1). *Self-driving cars may be more likely to hit you if you have dark skin*. MIT Technology Review. https://www.technologyreview.com/2019/03/01/136808/self-driving-cars-are-coming-but-accidents-may-not-be-evenly-distributed/

TRAFFIC SAFETY FACTS. (2017). Nation Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812456