Writing Assignment: Machine learning meets the real world

Seyedehnaghmeh Mosaddeghi
Chalmers University of Technology
Gothenburg
seymos@student.chalmers.se

Charitha Madapatha
Chalmers University of Technology
Gothenburg
charitha@chalmers.se

Abstract

This essay briefly analyses and summarizes Andrew Silver's article ”Why AI Makes It Hard to Prove That Self-Driving Cars Are Safe?” published in IEEE Spectrum. The focus is on the challenges of ensuring the safety of self-driving cars, with attention to the issues of machine learning models and their impact on safety. The essay also gives real-life examples of accidents involving self-driving cars to emphasize the importance of these problems. Moreover, the essay studies potential solutions such as combining machine learning and rule-based systems and establishing standardized datasets and evaluation metrics while summarizing our views and conclusions.

1 Introduction

Self-driving cars have been a popular topic in the automation industry for many years. They have several benefits from different aspects, including reducing traffic congestion, improving fuel efficiency, reducing the number of accidents caused by human error, providing accessibility for the elderly or individuals with a disability to have their cars, and most importantly, constraining individuals from breaking laws. For example, one would not be able to exceed the speed limit, which may put other 3rd parties at risk.

Despite all these advantages and all the advances that have been made, as we get closer to self-driving cars as a reality, we observe that they could have some risks. Several concerns are regarding how safe and reliable autonomous cars can be substituted with ordinary cars with humans as drivers. One of the most critical issues that some researchers argue about is using artificial intelligence in self-driving cars. They introduce new challenges related to testing and verifying the safety of these systems to ensure that the system is safe in all possible scenarios. We will discuss the challenges related to testing and verifying the safety of AI-based self-driving cars and the regulatory challenges that must be addressed. Furthermore, we will explore some proposed solutions to these issues and evaluate whether they are meaningful.

2 Problem Statement

In the article, it is described how self-driving cars depend heavily on AI algorithms to make critical decisions and operate on roads. While AI has demonstrated tremendous potential in various applications, it also has limitations and shortcomings that need to be addressed. The article specifically highlights the following concerns:

- Reliability of AI algorithms
- Bias in training data
- Difficulty in testing

2.1 Reliability of AI algorithms

The operating mechanism of machine learning is to develop computer programs that may self learn using human-annotated examples. This methodology creates a challenge known as the black box problem, as the algorithm's decision-making process is not transparent. Philip Koopman believes that machine learning presents a problem in that it is challenging to establish requirements, making it difficult to ensure the safe operation of self-driving vehicles. He said there are infinite things that algorithms may learn from data, and we are unaware of those. Taking Google researchers as an example, the model was trained to detect dumbbells in images, but it only detected dumbbells when there was an arm in the image.

2.2 Limited data and Bias in training data

Koopman stated that most autonomous cars are being trained on train and test data sets that are very similar; he believes that the cars memorize the data rather than learn the data. Another challenge with data is that Autonomous cars are trained on datasets to learn from various situations. However, these datasets may not always be fully representative of all the possible situations on the road. This issue creates a problem since self-driving cars must navigate different environments with varying conditions. The reliance on a biased dataset may lead to dangerous decisions in self-driving cars, potentially resulting in accidents. For instance, if an AI system is trained on a dataset primarily comprising images of pedestrians in summer, it may not recognize pedestrians with hats in winter as efficiently. As a result, the self-driving car may not respond adequately to a person with a hat in its path, leading to undesirable outcomes. Therefore, ensuring that the dataset used for training AI systems is diverse and unbiased is essential to minimize the risk of biased behavior in self-driving cars.

2.3 Difficulty in testing

Testing the safety of self-driving cars in all possible scenarios is a challenging task because there are infinite possible situations that a car could encounter while driving. For example, a self-driving vehicle may encounter unexpected obstacles on the road, such as explosions, or it may face a plane crashing in front of it.

3 Real-Life Examples

In 2019, it was reported that widely-used object recognition algorithms in self-driving cars and other applications exhibited lower accuracy in identifying darker-skinned individuals and women compared to lighter-skinned individuals and men. The reason behind this bias is often the training data used to develop the algorithms, which may need to be more diverse to represent all demographics accurately. As a result, the algorithm may not recognize individuals from other demographics as accurately. This can lead to potential safety issues in self-driving cars where the AI system may fail to detect pedestrians or other objects on the road if they belong to an underrepresented demographic group (1).

Another example happened in March 2018, when an autonomous Uber car killed a pedestrian crossing the street outside a designated crosswalk. This accident was the first pedestrian death that happened by an autonomous car. Investigations about the accident revealed that the car was moving at about 40 miles/hour when it hit the pedestrian walking with her bicycle on the street. He said it did not appear the car had slowed down before impact and that the Uber safety driver had shown no signs of impairment. The weather was clear and dry. Later they stated that the vehicle's AI system had detected the pedestrian six seconds before the collision but failed to identify her as a pedestrian due to a flaw in the system's object recognition software(2).

4 Proposed Solutions

4.1 Solutions for the reliability of AI algorithms

Researchers and engineers are exploring various solutions to solve the problem of the reliability of AI algorithms in self-driving cars. Phillip Koopman stated that car production companies should explain the features of their algorithms and give reasons for why their simulations are safe for evaluating the safety of self-driven cars. Companies should think and plan to develop AI algorithms that are more transparent and interpretable to create models that show the decision-making process of the machine learning algorithm more clearly. This may support to recognize potential issues and ensure that the algorithm works as intended. For instance, one approach is to use explainable AI, which can provide insights into the algorithm's decision-making process, allowing humans to detect and correct any potential biases. However, it is unclear whether this solution would be enough to ensure the safety of the autonomous vehicle. He believed that another solution is to make automakers demonstrate the safety of their systems to an independent agency.

4.2 Solutions for data limitation and Bias in training data

It is suggested in the article that to fix the problem with limitations in the training data size, companies such as Google can train their cars for a limited and small area and also use them in the same area, not in a new area. Although it creates limitations for companies, it reduces the risk of using self-driving cars. The article suggested that one solution for solving Bias in data is to provide more diverse training data that includes a wide range of scenarios and examples to avoid overfitting and to reduce the risk of the algorithm making incorrect assumptions or predictions. It is mentioned that the data should include a variety of scenarios and demographic groups to avoid any implicit biases. For example, researchers can incorporate images of people from different ethnicities and genders to ensure that the algorithm is trained to recognize all individuals.

In addition to what was discussed in the article, adversarial training can help the algorithm learn to recognize and overcome biased data. Adversarial training involves exposing the algorithm to modify data that attempts to trick the system into making incorrect decisions. By exposing the algorithm to such data, researchers can ensure that it is trained to recognize and overcome biased data. Another solution is incorporating human error in the training process, which can help identify potential biases and ensure the algorithm is fair and unbiased. Explainable AI can also provide insights into the algorithm's decision-making process, allowing humans to detect and correct potential biases.

4.3 Solutions for difficulty in testing

The car companies can create virtual environments that simulate real-world conditions, i.e., digital twinning and allow developers to test their autonomous systems in a safe and controlled environment. It can also be mentioned that closed-track (controlled) testing that involves testing autonomous cars on a closed track, such as a test track, a private facility, where the conditions can be controlled and monitored may also carried out. Moreover, as the final step, these cars need to pass the testing being carried out in the real-world in outside environments, i.e., the ones being tested on public roads and highways, where the conditions are unpredictable and can vary significantly from different aspects. It is suggested that a combination of these testing methods can provide a comprehensive approach to testing autonomous cars. However, all of these tests are costly, according to Alessia Knauss of Chalmers University of Technology. Still, these approaches can help identify potential issues and ensure the system is robust enough to handle various scenarios. For instance, testing in different weather conditions can help ensure the system can operate safely in all climatic and weather conditions. In addition, using advanced sensor technology can help identify potential dangers on the road. Lidar sensors, for example , can create a detailed 3D map of the surrounding environment, while cameras can detect visual signs such as stop signs and pedestrians. The use of sensor fusion, which combines data from multiple sensors, can provide a more comprehensive view of the road and improve the system's decision-making process, leading to an increase in the system's accuracy.

5 Personal Opinion

We believe that self-driving cars has a lot of potential in reshaping the traditional transportation system. However, it is important that the vehicles are safe and trustworthy not only for the car drivers and passengers, but also to the general public who use the roads. More co-ordinated, and transparent AI algorithms will help these systems to evolve as more safe ones. We believe that a lot of research work should be carried out with collaborations across many fields while involving both academia and industry together, for such succeful evolution.

6 Conclusions

We would like to conclude that using AI in self-driving creates challenges for their safety and reliability. As we see, the reliability of AI algorithms, bias of the trained data, and testing difficulties are three key things highlighted in the article. Vigorous testing and validation should be carried out to ensure the reliability of AI algorithms in the future. Although there are challenges to overcome when ensuring the safety of such self-driving cars, the developing technology trends may have the potential to overcome such worries. Further research and development investments will make sure that such trends get to evolve in a good direction.

References

[1] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Conference on fairness, accountability and transparency*. PMLR, 2018, pp. 77–91.

[2] "Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian," https://www.ntsb.gov/investigations/Pages/HWY18MH010.aspx, accessed: 2023-03-15.