Letter of Transmittal

Adrian Sam Daliri
4700 Keele St
Toronto, ON, M3J 1P3
adriansd@my.yorku.ca
2024/03/10

Dr. Milad Shakeri Bonab
ENG 2003 Professor
Lassonde Engineering
4700 Keele St
Toronto, ON, M3J 1P3

Dear Professor,
Please find attached the technical report titled “Humans and Autonomy: The Risks associated with autonomous systems and possible solutions” for the ENG 2003 Term Project one. This technical report examines the risks to public safety propagated by the use of autonomous systems. This report provides potential solutions to mitigate these risks before widespread adoption of autonomous systems such as autonomous vehicles.

This report goes in depth about the risks Artificial Intelligence systems pose such as decision failure when exposed to uncertainty, cybersecurity threats, sensor malfunctions, ethical concerns and job displacement. It discusses the lack of a standardized framework to test autonomous systems leaving them prone to failure and exploitation.

This report highlights the importance of having Industry standardized testing for these systems, ensuring machine-human collaboration, upgrading existing cybersecurity measures and establishing clear ethical guidelines for the use and adoption of autonomous systems.

This report was created in hopes of changing the status quo of the autonomous system industry so that this technology can be used ethically and responsibly without putting the public at risk.

Thank you for your consideration

Sincerely,

Adrian Sam Daliri
Software Engineering Student
York University

Humans and Autonomy:

The Risks associated with autonomous systems and possible solutions

Adrian Sam Daliri

ENG 2003

2024/04/02

Executive Summary

Report Title: Humans and Autonomy: The Risks associated with autonomous systems and possible solutions

Overview: This report provides an insight to the risks autonomous systems create and how we can mitigate these issues before mass-adoption of these systems.

Problem Summary:

- Autonomous systems pose several risks such as decision failure when exposed to uncertainty, cybersecurity threats, sensor malfunctions, ethical concerns and job displacement
- The lack of a standardized framework to test autonomous systems leaves them prone to failure and exploitation.

Solution Summary:

- There must be an Industry standardized testing for these systems
- Moving away from fully autonomous systems to machine-human collaboration will ensure the safe integration of Artificial Intelligence into existing systems

Adrian Sam Daliri
Software Engineering Student
York University

Table of Contents

1. Introduction

Autonomous systems have the potential to revolutionize society, autonomous vehicles are one of the biggest applications of this technology which can transform transportation by offering increased safety, efficiency, convenience all while reducing traffic which can drastically improve commute times, shipping deliveries and any other road dependent service. WIth these benefits also comes several risks, dangers and challenges that must be mitigated before mass adoption of autonomous systems like autonomous vehicles. This report identifies the risks associated with autonomous systems with a focus on autonomous vehicles and the threats they pose to safety and jobs, their vulnerability to exploitation and the lack of industry standardized testing frameworks for these systems. The goal is to expose these dangers and mitigate them before widespread adoption of autonomous systems. This report begins by delving into the risks posed when uncertainty is introduced to autonomous systems and how that affects their decision making abilities. Then it explores how these systems can be exploited to make erratic driving decisions. Furthermore, it discusses how fully autonomous systems have been hailed as the future of technology but with recent industry shifts there might be a plateau in development. This report also highlights how with the automation of tasks such as driving using Artificial intelligence, there is also a displacement of jobs that otherwise could be completed by humans, such as public transport and trucking. Finally this report touches on the lack of a standardized testing framework to ensure the safety of autonomous systems.

2. Background

Autonomous systems are taking the world by storm in the form of Large language models like OpenAI's ChatGPT or Autonomous Vehicles with extensive self-driving capabilities. These systems use artificial intelligence models to make decisions based on a wide range of inputs they receive[2], [5]. These systems then make decisions based on their certainty and the likelihood of a decision being the correct one[5]. Autonomous systems such as Autonomous Vehicles have many benefits such as eliminating human error, improved safety, reduction in traffic, and an overall improvement in convenience and efficiency. The elimination of human error has the potential to significantly reduce the number of car accidents and the related fatalities as human error is the leading cause of these collisions/fatalities[1]. Autonomous vehicles could also improve traffic as these vehicles can coordinate with other vehicles and make decisions logically removing human error and emotion[1]. These numerous benefits have pushed the industry towards developing self-driving vehicles and pushing the technology as the definite future[4]. Nonetheless, these autonomous systems have their own flaws that must be addressed before the widespread adoption of these systems. One of the major flaws within these systems is they make decisions and act upon certainty but roads have high levels of uncertainty as there are hundreds of factors affecting these conditions. Due to this low level of certainty autonomous systems are going to make incorrect and potentially fatal decisions as they have no certain conditions to base their decisions off of[2]. Autonomous vehicles also have the added threat of breaches in cybersecurity and although they have a reduction in human error there is an introduction of error from the machine and its own vulnerabilities that can be

maliciously exploited[3]. In addition to these Software failures, sensor malfunctions and other system failures can put the driver and the people around them at risk and due to the hands off nature of autonomous systems, the driver may be occupied at the time of these failures leading to an otherwise avoidable situation. This leads to the ethical dilemmas that arise when dealing with these systems as with these A.I. systems are faced with unavoidable accidents; they may need to make decisions about whether to prioritize the driver and the vehicles, pedestrians, other vehicles and different people of various backgrounds which can lead to unethical and biased decisions and a gray area of who is liable for the accident[2]. In addition to these ethical/safety issues there is also a threat to many jobs, especially in the transportation and delivery industries. Due to the potential automation of vehicles with autonomous systems many jobs are at risk as those who drive these vehicles perform deliveries and other logistics will be at risk to be displaced by cheaper more efficient systems which can have devastating effects on the economy[1]. With all these issues and uncertainty there is also a lack of a formalized legal and regulatory framework for autonomous systems to abide by which will halt widespread adoption of these systems until mitigated. Overall, where Autonomous systems hold the potential to revolutionize transportation they also have their own risks and challenges associated with them which must be addressed before deployment.

3. Main Topics

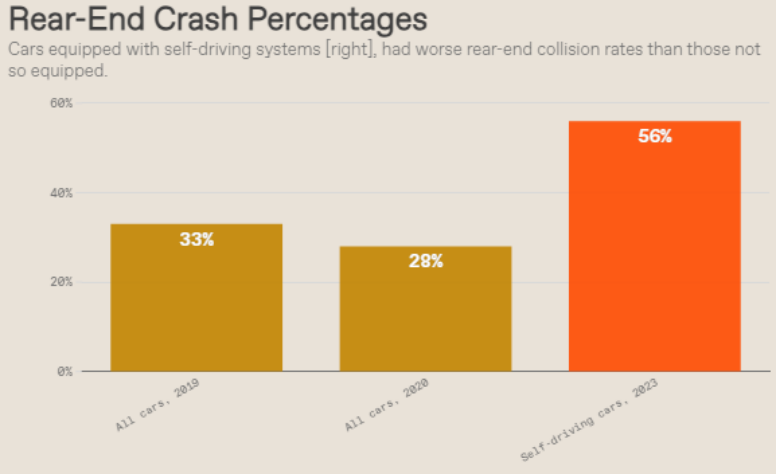
1) Safety Concerns

1.1) Risks and challenges associated with autonomous vehicles

There are several risks and safety concerns associated with autonomous vehicles and their widespread adoption. Some of these stem from the fact that autonomous systems rely on Artificial Intelligence to make their decisions[2]. The issue that this causes is that since A.I. systems make decisions based on certainty, placing them on roads which are extremely volatile and uncertain due to varying road conditions and driving habits of other drivers leading to poor judgment by these systems putting drivers and pedestrians at mortal danger[2].

1.2) Software malfunctions and system failure

There is a fallacy that due to the fact that human error is removed from driving in autonomous vehicles, this neglects that fact that human error is introduced in the creation of the A.I./autonomous systems that autonomous vehicles rely on[2]. This causes a much larger issue as Software bugs are often realized much later on after mainstream adoption which may lead to several fatalities before these issues are caught[1],[2]. These system failures may be caused by irregular detection of roadside objects as obstacles causing the system to make dangerous and erratic decisions when there is no threat present[3].



Rear-End Collisions caused by Self driving cars in 2023 [2]

2) Cybersecurity Threats

2.1) Software Bug and Maintenance

As with all software, maintenance is just as important as development as once exploits are realized they can be exploited for ill intent leading to potentially fatal consequences[2],[3]. Objects can be placed roadside to perform a DoS(Denial of Service) attack making it extremely easy for someone to perform such an attack with zero technical experience or with the need of any tools[3].

2.2) Exploits and Vulnerabilities

These vulnerabilities can also be taken advantage of through backdoors or privacy leaks within the software that autonomous vehicles rely on[2]. The large scale of such an autonomous software leaves plenty of opportunities for exploitation and errors which may be built upon leading to system critical errors[1],[2].

Autonomous Vehicle holding up traffic due to software malfunction(Rear door slightly open)[2]

3) Ethical Dilemmas

3.1) Life or death decisions

The uncertainty in life or death situations has led many autonomous car manufacturers to make the car halt whenever it is unable to make decisions within a given level of certainty[2]. The issue with this is the gray area of who is liable in the event of an accident, whether it is the responsibility of the driver to remain aware of their surroundings; whether it is the responsibility of the software engineers who developed these models or the car companies for promising fully autonomous vehicles past their capabilities[1],[2],[4].

3.2) Prejudice in autonomous systems

Another issue that raises ethical concerns is when in a situation where a collision is unavoidable and the vehicle has to decide between harming the driver or another road patron and if it decides to harm someone else how does it decide who is suitable[1]. These models can have bias and harm innocent people based on their appearance and correcting this issue may lead to overcorrection causing the vehicle to target another group of people which is highly unethical[4],[5].

4) Legal/Regulatory Challenges

4.1) Lack of Industry Standards

The lack of legal/regulatory frameworks in addition to the absence of industry standards for testing autonomous vehicles leads to various issues being ignored as they are either ignored due to a lack of legal responsibility or due to undertesting due to tight deadlines and pressure to release software to production before completion[2].

5) Job displacement and Implications on the Economy

5.1) Effects on Transportation Industry

In addition to the risks posed to public safety, autonomous vehicles also pose a risk to the transportation industry and those working in it. If autonomous vehicles are created without any safety issues, then this has the potential to displace thousands of jobs which can have disastrous effects on the economy as a whole[1],[5].

4. Discussion

With these many issues plaguing autonomous systems it seems we must either accept conditions as they are or cease all adoption of autonomous systems but this is not the case as there is a solution which helps alleviate most of these issues if implemented correctly and that is machine-human collaboration. Machine-human collaboration combines the automation of many mundane tasks through the use of AI and the safety of a human overseeing the operation to greatly improve decision making. By implementing machine-human collaboration we are able to mitigate safety risks by having a person such as the driver of an autonomous vehicle cognizant of the vehicle's actions. This allows the driver to interrupt the system and correct dangerous behavior or maneuver through uncertain situations the autonomous system may be unable to

handle. Furthermore, this aids in mitigating cybersecurity risks as by having a person monitoring the system at all times. They can help the vehicle avoid these exploits as at any point there is a breach in the system it can be deactivated and the driver can take control of the vehicle. This leads to an idea for the industry standard which can be that vehicles must be monitored by a human at all times when autonomously driving. This reinforces safety and allows AI systems to be a tool rather than being wholly relied on. Finally, by using autonomous systems as a tool, those working in the transportation industry can automate many of their tasks and work alongside these systems to make their jobs easier rather than replacing them. Machine-human collaboration allows for autonomous systems to have a real use case in our day to day lives rather than remaining a dream for future generations. Through machine-human collaboration we can receive the best of autonomous systems while mitigating the many issues that plague them to this day.

5. Conclusion

The risks associated with autonomous systems, such as autonomous vehicles, are apparent and must be mitigated before the widespread adoption of these systems. These issues range from safety concerns to cybersecurity threats to job displacement to regulatory challenges among other issues. With all these issues a possible solution exists in machine-human collaboration. This allows for the use of autonomous systems in a safe and ethical manner allowing for the benefits of AI to be combined with the safety of human oversight. It can help by allowing humans to make the decisions and maneuvers where autonomous systems fall short. Autonomous systems can be used as a tool to aid us on a day to day basis improving transportation, logistics, traffic and many other systems that suffer from human error by combining machine intelligence and human intelligence. Furthermore, it is important to still establish industry standards for developing and testing these systems to ensure their safety and have machine-human collaboration as a further safety precaution. In conclusion, while current autonomous systems have a variety of issues which make them infeasible for widespread adoption right now, with industry standards being set and machine-human collaboration autonomous systems can become one of the most useful tools at our disposal.

References

[1]J. Wang, L. Zhang, Y. Huang, and J. Zhao, “Safety of Autonomous Vehicles,” *Journal of Advanced Transportation*, Oct. 06, 2020. https://www.hindawi.com/journals/jat/2020/8867757/

[2]Mary L. Cummings, “What Self-Driving Cars Tell Us About AI Risks - IEEE Spectrum,” *spectrum.ieee.org*, Jul. 30, 2023. https://spectrum.ieee.org/self-driving-cars-2662494269

[3]B. Bell, “Autonomous vehicles can be tricked into dangerous driving behavior,” *University of California*, May 26, 2022. https://www.universityofcalifornia.edu/news/autonomous-vehicles-can-be-tricked-dangerous-driving-behavior

[4]J. Herrman, “Apple, Tesla, and the Dying Dream of Self-Driving Cars,” *Intelligencer*, Mar. 05, 2024. https://nymag.com/intelligencer/article/apple-tesla-and-the-dying-dream-of-self-driving-cars.html

[5]J. Ondruš, E. Kolla, P. Vertaľ, and Ž. Šarić, “How Do Autonomous Cars Work?,” *Transportation Research Procedia*, vol. 44, pp. 226–233, 2020, doi: https://doi.org/10.1016/j.trpro.2020.02.049.

Revision Summary

After receiving the feedback from my peers I was able to reflect on this and improve my report accordingly as well as reflecting on my work by myself. Some feedback I received several times was to improve the formatting of the report. After having done this I found that this made my report much more succinct and allowed it to be much easier to follow with bolded sections and subsections. Another piece of feedback I received was to make sure I was clearer in what the purpose of my report and what my proposed solution was. After having done this I was able to make it much more clear what my solution is and what the issue I am defining is as well as the overall goal of the report. Finally I also received feedback that I had too many run-on sentences and that I should use more punctuation to separate different points. This is something I often struggle with so I made sure to go through the report and thoroughly revise and update my report to resolve these issues to create a better report overall.