Ethical AI Framework Assignment

Name : Sridhar Ponnusamy

Student ID: 101387275

Ethical Framework in the cases of Autonomous vehicles

Introduction

In the future, autonomous vehicles (AVs) are expected to play a significant part in transport systems. With so many businesses racing to build the first completely autonomous car, detailed ethical frameworks are important to the success of these robots as well as the safety of humans. Ethical frameworks and principles from philosophy have long been used to characterize human behaviour. As driving is transitioning from human control to automated control, below ethical frameworks can be used to guide engineering design decisions for autonomous vehicles in a responsible manner. (Erina DuBois,2010).

There are numerous paradigms for dealing ethically with AI technologies. The following three frameworks are used to develop the new framework:

1. Ethical 8 Risk Zone Framework By IFTF
2. UK Data Ethics Framework
3. IBM AI Explainability Toolkits

Pillars of the framework:

1. Algorithmic Bias:

Autonomous vehicles that prioritise passenger safety would exhibit a moral bias according to the standards of the former position, while proponents of the second view would regard this behaviour as unbiased or appropriate. Additional considerations may also be relevant to the “all things considered” determination of whether an ethical bias in favour of passengers should be eliminated or corrected. For example, suppose an ethical bias in favour of passengers turned out (after empirical investigation) to be the only way to secure trust in autonomous vehicles. Since this technology is reasonably expected to reduce the current high levels of traffic-related fatalities, then there may be an “all things considered” ethical obligation to act so as to increase the likelihood of their adoption, even if that means using a “local” ethical bias (David Danks & Alex John London).

2. Fairness

In order to trust AI, it must be fair and impartial. As more and more decisions are delegated to AI, we must ensure that those decisions are free from bias and discrimination, it's vital that decisions made by AI are fair, and do not deepen already entrenched social inequalities. But how do we go about making algorithms fair? It's not as easy as it seems. **The problem is that it is impossible to know what algorithms based on neural networks are actually learning when you train them with data.** (C. Havens,2020).

3. Transparency:

People engaged in training machine learning models may not want their data or judgments about their data to be released, therefore transparency clashes with privacy. Furthermore, the general public, as well as authorities, may lack the technological expertise to comprehend and evaluate algorithms.

The AI Now Institute at New York University, which researches the social impact of AI, recently released a report which urged public agencies responsible for criminal justice, healthcare, welfare and education to ban black box AIs because their decisions cannot be explained. The report also recommended that AIs should pass pre-release trials and be monitored 'in the wild' so that biases and other faults are swiftly corrected (AI Now Report, 2018).

4. Accountability:

Accountability is the key to assuring AI's trustworthiness. Accountability assures that if an AI makes a mistake or causes harm to someone, someone may be held accountable, whether it's the AI's designer, developer, or the company that sold it. There must be a system for redress in the event of injury so that victims are fully compensated.

'How do decision-makers make sense of what decisions get made by AI technologies and how these decisions are different to those made by humans?... the point is that AI makes decisions differently from humans and sometimes we don't understand those differences; we don't know why or how it is making that decision.' (Janna Anderson and Lee Rainie,2018).

5. Privacy

Autonomous vehicles are largely experimental at this time, it remains unclear what type of personal information may be collected by these vehicles. Nonetheless, at a minimum, location data associated with a particular vehicle will be tracked and logged. Location tracking has already proven to be a lightning rod with respect to mobile phones. Some of the privacy considerations related to the use of autonomous cars are discussed further below. (Navetta et al, 2019).

a. Owner and Passenger Information

To validate authorised use, autonomous vehicles would most likely need to keep information about passengers. autonomous vehicles will most likely be able to recognise drivers, passengers, and their behaviours with a high degree of accuracy based on setting preferences and other information acquired while in use.

b. Location tracking

The privacy risks associated with the collection and use of location data create both individual and social concerns.

c. Sensor data

It's possible that autonomous vehicles might collect information about other drivers' driving habits, destinations, and other personal details without their knowledge or agreement.

6. Explainability

In human-machine interactions, explainability is defined as the ability of the human user to understand the agent’s logic (Rosenfeld & Richardson, 2019). The explanation is based on how the human user understands the connections between inputs and outputs of the model. According to (Doshi-Velez & Kortz, 2017), an explanation is a human-interpretable description of the process by which a decision-maker took a particular set of inputs and reached a particular conclusion. They state that in practice, an explanation should answer at least one of the three following questions:

- What were the main factors in the decision?
- Would changing a certain factor have changed the decision?
- Why did two similar-looking cases get different decisions or vice versa?

End-users and citizens must have faith in and be reassured by the autonomous system. They entrust their lives to the driving system and must so acquire trust in it(Ben Younes, Hedi, 2021).

7. Conclusion

The technological advancements in autonomous vehicles will be astounding, and they have already generated a lot of interest. However, before commercialization, legal difficulties and hazards involved with gathering and exploiting personal data, as well as numerous cybersecurity threats, must be thoroughly explored.

To support the astounding future of AI systems in our society, I believe that an ethical framework for autonomous vehicles must be carefully defined and investigated. Machine Ethics and Algorithm Bias, transparency and accountability, fairness, privacy, and explainability are all part of my suggested framework. I believe it addresses the most significant concerns about autonomous vehicles.

Work Cited

1. Erina DuBois. "Ethical Decision Making for Autonomous Vehicles | Dynamic Design Lab." *Dynamic Design Lab* |, 2010, https://ddl.stanford.edu/publication-research-theme/ethical-decision-making-autonomous-vehicles. Accessed 23 February 2022.
2. David Danks & Alex John London. "Algorithmic Bias in Autonomous Systems." *IJCAI*, https://www.ijcai.org/proceedings/2017/0654.pdf. Accessed 23 February 2022.
3. C. Havens, : John. "The ethics of artificial intelligence: Issues and initiatives." *European Parliament*, 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf. Accessed 23 February 2022.
4. "AI Now Report 2018." 2018, https://ainowinstitute.org/AI_Now_2018_Report.pdf.

5. Janna Anderson and Lee Rainie. “3. Improvements ahead: How humans and AI might evolve together in the next decade.” *Pew Research Center*, 10 December 2018, https://www.pewresearch.org/internet/2018/12/10/improvements-ahead-how-humans-and-ai-might-evolve-together-in-the-next-decade/. Accessed 23 February 2022.
6. Navetta, David, et al. “Privacy and Security Issues in Autonomous Cars.” *Cyber Defense Magazine*, 24 October 2019, https://www.cyberdefensemagazine.com/privacy-and-security/. Accessed 23 February 2022.
7. Rosenfeld & Richardson. *How can we make driving systems explainable? | valeo.ai blog*, 18February 2021, https://valeoai.github.io/blog/2021/02/18/explainable-driving.html#RosenfeldR19. Accessed 23 February 2022.
8. Doshi-Velez & Kortz. *How can we make driving systems explainable? | valeo.ai blog*, 18 February 2021, https://valeoai.github.io/blog/2021/02/18/explainable-driving.html#doshi2017accountability. Accessed 23 February 2022.
9. Ben Younes, Hedi. “How can we make driving systems explainable?” *ml-certified systems*, 18 Feb 2021, https://valeoai.github.io/blog/2021/02/18/explainable-driving.html. Accessed 23 Feb 2022