How are Autonomous Vehicles Programmed to have Judgement?

Benjamin Ancheta III

4/23/2024

Abstract: A future where autonomous vehicles (AVs) dominate roads is not far off. Properly constructed AVs could remove human error from the driving process, thus preventing millions of injuries and crashes [1,3]. This paper provides a literature review of the most prevalent methods of implementing “judgment” in AVs, including comparing different ways of programming control systems in AVs and exploring how to build ethical reasoning into AVs.

Introduction:

Autonomous vehicles (AVs) are self-controlled vehicles that take in and process input from sensors in order to navigate traffic. According to the World Health Organization (WHO), road traffic crashes kill 1.35 million people every year and cause up to 50 million injuries, with 94% of these accidents occuring due to human error [1,3]. Propagation of AVs would reduce vehicle accidents by eliminating human error, thus decreasing fatalities and injuries due to road traffic by a significant margin.

Yet how do AVs assess safety and danger? How do they make judgements in difficult moral scenarios? The objective of this literature review is to first give an overview on current algorithms for AV decision making, then outline various methods for implementing ethical reasoning in AVs, along with exploring the challenges and limitations of each method.

I - Regular Decision-Making Architecture

Current AVs struggle with unconventional situations such as crosswalks, roundabouts, and near-accidents. There are a few popular methods of programming AVs to deal with unconventional situations, which are: graph-based, optimization, and machine learning. The criteria we will use to evaluate these methods will be accuracy and interpretability [5,9]. Accuracy, of course, is important because if an AV makes the wrong decision an accident is more likely to occur [9]. Interpretability is important because if an AV makes a wrong decision that causes an accident, engineers need to be able to understand why the AV made the wrong decision in order to correct the fault in its decision-making process [5]. In this section, we will provide an overview of graph-based, optimization, and machine learning navigation models.

1) Graph-based navigation

Dynamic path planning is required in AV navigation as changes in the environment are happening continuously, and the rapidly exploring random tree (RRT) algorithm is the most commonly used for dynamic path planning. The RRT algorithm has two distinct key steps: tree expansion and path construction. First, the algorithm generates a random series of actions that could get from the initial state to the goal state. If this random series of actions avoids accidents, it is added to a decision tree. This decision tree grows larger as a series of potential actions to get to the goal state is generated many times. Next, the algorithm attempts to find a path from the start state to the goal state. The tree is regenerated repeatedly at high frequencies for obstacle avoidance and to account for changes in the environment.

Graph-based navigation is interpretable, as the results (decision trees) are able to be viewed by programmers after each run of the RRT algorithm.

However, some drawbacks to graph-based navigation are that the refresh rate of the decision tree can be too slow in situations where quick action is needed [5]. Additionally, when solving the graph to find a possible path from the start state to the goal state, the algorithm may be programmed to find a path that gives a short computational time in order to make a quick decision in situations that need quick action, but might not be the optimal path based on speed to traverse the graph (conservative decision) [6]. Many AVs making conservative decisions in an intersection could cause traffic jams which is suboptimal and inefficient for the road ecosystem. For example, consider a scenario where four AVs arrive at an intersection at very similar times. A conservative navigation model could wait for the other cars to go in order to ensure safety; and if all of the AVs are running off of the same model, they could all stay stopped, thus leading to a traffic jam.

2) Optimization-based navigation

Another solution for AV navigation in complex scenarios is to distill the scenario into an optimization problem. This distillation requires setting up a cost function, boundary conditions, and constraints [5]. The most common type of algorithms used in optimization-based navigation in AVs are genetic algorithms (GAs). GAs create an initial pool of random solutions, some of which might not work. The algorithm determines which solutions are best, based on a "fitness factor," and uses the found best solutions to produce new solutions, along with keeping the found best solution, and drops the solutions that aren't as good. In future rounds of selection, this process happens repeatedly. The final surviving solution after many rounds of optimization is the optimal solution. GA is a solution that can run very quickly, as the solution with the highest fitness factor can be selected at any round of optimization, so if a decision is needed right away, the process can be ended and a potential solution will be present [6]. However, a large disadvantage of GA and other optimization-based approaches is their computational cost; assumptions, which may be unrealistic, must be made in order to formulate an optimization problem and solve it efficiently [5,6].

3) Deep learning navigation models

Deep learning navigation models use artificial intelligence and large datasets to create a model that makes decisions by considering previous decisions, and they learn by discovering complex patterns in the data they are provided. The model is given example situations and told to make a decision on what to do. After a decision is made, the model is given positive or negative reinforcement in order to train it to make more accurate decisions (reinforcement learning) [5,9]. Deep learning is currently the most prominent model because if the model is given a situation that is similar to a previous experience, it is able to use the similarities instead of compute a completely new situation when making its decision on what to do, unlike graph-based and optimization-based methods which would have to regenerate the decision tree or equation to solve respectively [9].

However, deep learning is not without its drawbacks. Deep learning models require large amounts of data in order to make good decisions. For example, models such as ChatGPT (which feeds off of writing accessible by the internet) or financial arbitrage models have easy access to historical datasets. For AVs on the other hand, large datasets of potential scenarios would have to be created. The number of points of data needed to capture the many different possible scenarios that could present themselves on the road would need to be huge.

Additionally, deep learning models' decisions are notoriously unexplainable. Deep learning neural networks often consider dozens of different variables when making decisions, and unless transparency in decision making is designed into the model, it isn't possible to see why a model made a specific decision, which is important because researchers would need to be able to pinpoint the reason why a model makes an incorrect decision in order to properly apply positive or negative reinforcement to the model to ensure it makes the correct decision in future cases. Recent advances in deep learning navigation methods have involved creating explainable deep learning models by using feature attribution [8]. Simply put, feature attribution involves revealing the weights that each variable has on a model's decision, so that when given a situation, engineers can see why the deep learning model made the decision that it did.

II - What is Ethical Architecture?

If AVs are to be fully autonomous, they need to be able to replicate or improve on the human decision-making process. However, some decisions involve more than just applications of traffic laws and plotting the safest, most efficient path.

Consider a scenario where an AV must swerve right and hit a child, or swerve left and hit an 80-year old, and if the car stops it will cause a large accident that will kill at least three people. Some might say that hitting the 80-year old is a lesser evil; although we don't have time to discuss specific ethical frameworks to implement in this paper, it should be clear to most that it is important for an AV to be able to make a decision in scenarios where harm is unavoidable. If an ethical architecture wasn't in place, and the AV wasn't considering the impact of its actions on other entities on the road, the AV might obey its sensors, brake, and cause a huge accident. It's important for cars to have the capability to make ethical decisions in scenarios where harm is unavoidable, because we have a responsibility to cause the least harm to humans that will share roads with AVs.

The field of ethical architecture is still a very new field, with new research being published every year. In 2021, a simple ethical decision-making algorithm was created that can take in very limited parameters [10]. Contrary to the scenario discussed above, the algorithm weighed all human lives the same regardless of age, gender, economic status, or any other possible trait; the algorithm was also only ever presented with situations with two single individuals [10]. Much more research and development is needed to bring ethical algorithms to a level where they are able to be used in AVs, especially considering the fact that recently developed algorithms are only able to process highly simplified scenarios.

Challenges/Future directions

Generally, it has been agreed upon that deep learning methods are the most promising path to pursue for navigation of AVs [5,7,8]. However, deep learning models have many requirements for optimal performance including large datasets that cover edge cases and interpretability. Resources should be concentrated on creating comprehensive datasets, and future research should focus on creating interpretable models so that engineers will be able to tell whether a model is working well or not.

In regard to ethical architecture, more research is needed on developing ethical algorithms that can process the complex scenarios that occur on real roads. Additionally, after developing working ethical algorithms, it will be necessary to determine how best to integrate them with the vehicle navigation system [4]. There are three main possible methods to accomplish this: zero ethics coupling, parallel ethics coupling, and serial ethics coupling. Zero ethics coupling entails the ethical architecture simply observing the navigation architecture and storing ethical judgements. These ethical judgements would be used later by the navigation architecture developers to evaluate the validity of the navigation architecture's decisions [4]. Parallel ethics coupling involves the ethical architecture "taking over" control of the vehicle when a 'moral' decision is needed [10]. Finally, serial ethics coupling requires approval from the ethical architecture for any action to be taken by the AV.

Conclusion

In conclusion, the implementation of AVs on a widespread level is a future that is quickly approaching. It is important that AVs are brought onto streets being able to handle complex scenarios and behave ethically, so that they are a boon to society instead of a hindrance.

Acknowledgements

I would like to thank my ENGW 3302 peers, Matthew Urbano and Raimon Bleda-Vilalta, along with my instructor, Alison Stephens, for their feedback during the review period, which has been critical in bringing this paper from conception to completion. I used Elicit to find sources and OpenAI's ChatGPT to help pull potential passages that might be useful from those sources.

A huge thank you overall to my friends and family; I wouldn't be who I am today without all of you.

Works Cited

[1] World Health Organization. *Global Status Report on Road Safety 2018*; Technical Report; World Health Organization: Geneva, Switzerland, 2018.

[2] Fagnant, D.J.; Kockelman, K. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transp. Res. Part A Policy Pract.* 2015, *77*, 167–181.

[3] Singh, S. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey; Technical Report; National Highway Traffic Safety Administration: Washington, DC, USA, 2018.

[4] Narayanan, A. (2019). Ethical judgement in Intelligent Control Systems for autonomous vehicles. 2019 Australian & New Zealand Control Conference (ANZCC). https://doi.org/10.1109/anzcc47194.2019.8945790

[5] Sana, F., Azad, N. L., & Raahemifar, K. (2023). Autonomous Vehicle Decision-making and control in complex and unconventional scenarios—a review. Machines, 11(7), 676.

[6] R. Kala, On-Road Intelligent Vehicles: Motion Planning for Intelligent Transportation Systems. Oxford : Butterworth-Heinemann, 2016, pp. 1–525.

[7] K. Muhammad, A. Ullah, J. Lloret, J. D. Ser, and V. H. de Albuquerque, “Deep Learning for Safe Autonomous Driving: Current challenges and Future Directions,” IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 7, pp. 4316–4336, Jul. 2021. doi:10.1109/tits.2020.3032227

[8] L. He, N. Aouf, and B. Song, “Explainable deep reinforcement learning for UAV autonomous path planning,” Aerospace Science and Technology, vol. 118, p. 107052, Nov. 2021. doi:10.1016/j.ast.2021.107052

[9] S. Abbasi and A. M. Rahmani, “Artificial Intelligence and software modeling approaches in autonomous vehicles for Safety Management: A systematic review,” Information, vol. 14, no. 10, p. 555, Oct. 2023. doi:10.3390/info14100555

[10] H. A. Moreno, I. G. Carrera, R. A. Ramírez-Mendoza, J. Baca, and I. A. Banfield, Software Architecture Proposal for Navigation and Decisions of Autonomous Ground Vehicles in Controlled Environments with Unavoidable Collision Scenarios, vol. 347. Cham, Switzerland : Springer, 2022, pp. 270–278.