Research Output Journal of Biological and Applied Science 3(1):68-73, 2024

ROJBAS Publications **ISSN: 1115-6120**

https://rojournals.org/roj-biological-and-applied-science/

Ethical Considerations in the Development and Implementation of Artificial Intelligence for Autonomous Vehicles

Mugabo Kalisa G.

Faculty of Engineering Kampala International University Uganda

ABSTRACT

The integration of artificial intelligence (AI) in autonomous vehicles presents numerous opportunities for enhancing transportation efficiency and safety. However, it also raises critical ethical issues that must be addressed to ensure responsible development and deployment. This paper explores the ethical considerations in AI for autonomous vehicles, focusing on three primary ethical frameworks: utilitarianism, deontology, and virtue ethics. Each framework provides unique perspectives on how AI should be designed to make morally sound decisions, especially in life-and-death scenarios such as the Trolley Problem. Furthermore, the paper discusses the challenges of responsibility and accountability in AI systems and highlights regulatory and policy approaches to manage these ethical dilemmas. By examining these frameworks and challenges, the paper aims to contribute to the ongoing discourse on creating ethically sound AI systems for autonomous vehicles.

Keywords: Autonomous Vehicles, Artificial Intelligence, Ethics, Utilitarianism and Deontology

INTRODUCTION

Autonomous vehicles are self-driving vehicles that are capable of sensing the surrounding environment and navigating using sensors, cameras, radars, or GPS. They have the potential to be safer than human drivers. Manufacturers of autonomous vehicles rely on artificial intelligence (AI) for data processing speed, pattern recognition, and decision-making capabilities, and consider AI to be essential to the development and operation of autonomous vehicles. Technological advancement is helping them to continue to lean on AI for advancements in autonomous vehicles. AI nodes can process and learn from vast quantities of data to improve decision-making during a journey in less time than a human can, making the operation of autonomous vehicles more efficient. AI also broadens the possible uses for autonomous vehicles by giving them the ability to process, learn from, and store vast quantities of data. This has the potential to impact a wide range of industries [1, 2]. The process of further developing AI for autonomous vehicles, as well as currently implementing it, requires careful thought about the underlying ethical considerations. Some of the leading developers and operators of AI view it as crucial to address their associated ethical and social impacts. In this article, we describe the functions performed by AI systems in autonomous vehicles and discuss four areas in which the ethical considerations of AI systems in autonomous vehicles are not receiving sufficient attention: the capabilities of AI, AI failure, AI making life and death decisions, and social and individual impacts of autonomous vehicles [3].

ETHICAL FRAMEWORKS FOR AI IN AUTONOMOUS VEHICLES

The development and implementation of AI for ensuring the ethical behavior of autonomous vehicles could employ different ethical frameworks. The ethical frameworks in AI can be borrowed from the debating and research on ethical considerations in the application of robots and AI in different domains. In considering the specific use of AI, i.e. in AI for catalyzing autonomous vehicles, three dominant kinds of ethical frameworks offer a range of conceptual and methodological threads for researchers and experts engaged in the problem-setting of AI's use in autonomous vehicles. These three normative-ethical

concepts include: act-centered/utilitarianism; rule-based/deontological ethics and community/or agent virtue-based/virtue ethics. These three normative-ethical frameworks provide three starting points for developing the design and use of AI ethics for the purpose of the responsible deployment of fully autonomous vehicles [4]. The foundational principle for act-centered/utilitarianism is for the implementation of the AI in autonomous vehicles so as to be beneficial for society, i.e. the well-being of as many as possible and the reduction of harm. Each situation will be judged on its own actions and the consequences, including the moral perception of AI. Whatever ethical decision the AI makes in the case of 10 individuals being saved rather than 1 being saved, the principles that are obtained will be used or generated by the AI when making an appropriate decision-making in traffic. The second kind of normative-ethical framework can provide a more disposition/agent-focused system than the above. This ethics of rule follows deontological or duty-based ethics that sets ethical norms and generally agreed independent standards. Rules for the use of AI in autonomous vehicles could advise the following: always avoid harm; promote the well-being of others; ensure an acceptable quality of life for those AI could harm; and ensure autonomy through making ethical transparency and control possible. The third kind of ethical normative-ethical framework originates from ancient Greece and is called virtue ethics, alternatively community ethics or agent ethics. As in the rule-based ethics, virtue ethics incorporates a broader perspective on the ethics of autonomous vehicle behavior by AI. Virtue ethics shifts the emphasis from act-centered analyses and agent-centered and disposition-comparison to people-oriented ethics, whether they involved or affected. Virtues that AI should have include responsibility, accountability, respect, ethical integrity, and upholding dignity and human rights [5, 6].

UTILITARIANISM

Utilitarianism is a type of consequentialism that states that the best action is the one that maximizes utility, or overall societal happiness. This is often operationalized as health and longevity and is especially relevant for autonomous vehicles. These vehicles will behave in ways that may lead to accidents, just as human drivers can, but they will also adapt to avoid them. The outcomes of various styles of autonomous vehicle decision-making may involve different numbers of overall traffic fatalities, severity of individual harm, fairness to various stakeholders, future demands for resources and opportunity cost, and moral harm that the decision inflicts on others and on oneself during reflection [7, 8, 9]. Utilitarianism can be viewed as realist or idealist. As idealist utilitarians, informatics can have an antecedent preference for some idea of what is best for whatever agents or stakeholders are relevant, and then try to get there from the standpoint of a designer of algorithms. A realist viewpoint would ask how the preservation of injury-capability could lead on to more realized utility being produced for some particular set of stakeholders. The former view could see hopes dashed if decision-makers were prevented from doing what is thought to be the "most ethical" under built-in constraints. The latter view would appeal to empirical data, for example showing that decision-makers who have been prevented from issuing outcome-maximizing decisions have "lost welfare". A 'prioritarian' viewpoint would assume that the severely injured are the most morally entitled to relief, leaving effectiveness as converted to maximum good. These require careful specification if they are to be translated to action for car autonomy [10, 11].

DEONTOLOGICAL ETHICS

Deontological ethics, or deontology, refers to the normative ethical theory that emphasizes doing one's duty and abiding by moral rules when making decisions. A central focus of deontological ethics is the usefulness of rules in discerning the moral value of actions. The German philosopher Immanuel Kant was one of the primary proponents of deontological ethics, arguing that an action is morally right if it is performed out of respect for the categorical imperative, which is based on the notion of universal truth. In other words, the action is performed freely based on a universal truth that it is the right thing to do. Because it is based on a universal truth, it must also be the right thing to do for another, and therefore a good duty to have. Contrary to procedural ethics which assigns ethical value to the process of decision-making, deontological ethics assigns ethical value based on the moral norms inherent in the principle applied in a specific case [12, 13]. The three most prominent thinkers associated with deontological ethical theory include Thomas Hobbes, John Locke, and Immanuel Kant. Kant's main philosophical treatises include the Groundwork of the Metaphysics of Morals (1785), The Critique of Practical Reason (1788), and The Metaphysics of Morals (1797). Kant believed that moral principles or 'axioms' are applied a priori, independent of information or practical considerations. Kant created categorical imperatives - a test for ethical action, independent of consequences, that can be universally applied. Kant developed three formulations of the categorical imperative which were grounded in the nature of rationality as responsible for itself [14].

VIRTUE ETHICS

In contrast to the preceding images of right actions, virtue ethics focuses on the character and virtues of the agents who perform them. The fundamental question is less 'What should I do?' and more 'What kind of person should I be?' What kind of engineering teams, organizational workplace cultures, and societies are necessary for ethical and responsible uses of AI in private and public spaces? There is an increasing recognition that the effective design, development, and implementation of AI powered and ultimately autonomous vehicles will be influenced significantly by the moral character, virtues, and habituation of those creating, programming, welding, buying, interfacing, and otherwise interacting with them. Several scholars have also drawn parallels between machines inclining others to act and the way in which characteristics and expectations shape and influence the actions of human beings [15]. Virtue ethics has attracted increasing attention and generated much research in AI and robotics. Through examining the aspects of character, both virtues and vices (the lack or excess of virtues), it is possible to consider how engineers and machine learning practitioners should act rather than merely framing questions in terms of the architecture of an autonomous system. Research in virtue ethics also helps to fill the gaps in the existing literature on AI ethics that stems mostly from other ethical theories such as deontology or consequentialism. Instead of focusing on a priori principles that set standards of right conduct for particular kinds of situations as deontologists might, or focusing on outcomes of actions that maximize value as consequentialists might, virtue ethic analysis evaluates the moral character and virtues of the individual moral agent. Once considered, a moral agent enjoys significant latitude in responding to ethical challenges, inquiries, and dilemmas. Overall, attention to the character traits necessary to build and develop just, honest and ethical systems will assist engineers in determining ways to advance them. By focusing attention on the moral character and behavior of the agents creating AI technology, virtue ethic research is intended to support the development of a shared responsibility for building moral machines [16].

CHALLENGES AND DILEMMAS IN ETHICAL AI FOR AUTONOMOUS VEHICLES

Ethics reflects the moral principles that govern our behavior and relationships. Embedded within ethics are discussions focusing on right and wrong behavior, including virtues, equality, power, fairness, and reciprocity. Consequently, many challenges and difficulties are presented in defining the "right" behavior and decisions for autonomous vehicles (AVs), enabled by artificial intelligence (AI). The difference in an AV system acting based on a traditional deterministic decision set by a human programmer, and the various facets of AI that enable complex sensors and perception of the environment, reasoning, learning, and sophisticated decision-making, emphasize the need to define and develop AI "ethically." Most of all, a critical question is how AV AI "should" behave in an applied scenario. These difficult and complex dilemmas are commonly explored through thought experiments and discussions of (im)possible solutions, balancing fundamental philosophical questions about utilitarianism, deontology, virtue ethics alongside the practical and useful aspects that will guide the development of ethical AI systems that are socially acceptable and that will embrace human autonomy and agency [3]. The autonomous vehicle (AV) ethical "Trolley Problem" refers to different scenarios and situations depicting a moral dilemma in which the AV AI needs to make a decision based on an unexpected or rare event. The classical Trolley Problem describes a situation in which a runaway trolley car is hurtling out of control down a track, towards a group of five people. The only way to save those five people is to divert the trolley onto a different track, where there is only one person. A choice is forced onto the trolley operator: should they flip the switch and divert the trolley, saving five and killing one, or should they do nothing and preserve the one at the cost of the five? Research posits that there are emotional and social aspects related to ethical AI choices and designs as well, such as categorizing and dealing with a multitude of different individuals, their characteristics, aptitudes, and assets of society. The clustering and reductionist view might lead to even more social discrimination, and hence, to a larger societal controversy. Corporate social responsibility, and issues of power, accountability, and blame, intertwine with ethical considerations and discuss questions related to what AV AI should do, can do, may do, and how to deal with its actions [17, 18].

TROLLEY PROBLEM AND ITS VARIANTS

Way back in 1967, the philosopher Philippa Foot described a scenario as a "beginning of a new series of investigations" through a series of denials of different potential morally significant differences between alternative actions (those whose outcomes are being decided on), where the alternatives always involve trading off the lives of individuals. This wonderful piece of moral philosophy has become an iconic thought experiment in ethics and is widely used in contemporary discussions of the ethical issues raised by AI for autonomous vehicles [19]. Research on the use of the Trolley Problem in contemporary ethical debates (about autonomous vehicles, primarily) typically proceeds in one of two directions. On the one

hand, it is said that the very extremeness of the trolley problem and its variants make them less relevant for our moral lives and decisions. On this view, who will or will not survive freak trolley accidents is simply not something we need to consider much. Rather than pose a challenge to everyday moral reflection and decision-making, the trolley problem, this argument suggests, is just plain bizarre. However, much as Foot suggested, there is something about the trolley problem and its various scenarios that seems to provoke a strong moral intuition in those that consider it, and it is for this reason that its proponents seem to think it has some value. In this section, two of the trolley problem's variant scenarios will be discussed in some detail [20].

RESPONSIBILITY AND ACCOUNTABILITY

Responsibility and accountability are considered the core of ethics of AI in autonomous vehicles. Research in this section is primarily concerned with the question of to whom it attaches responsibilities in the context of incidents related to AI. Responsibility is often discussed in relation to accountability. However, the notion of accountability is defined in a way that disunites it from responsibility. It is proposed that accountability is rather about the enforcement of responsibilities or to create responsibility where it appears not to exist. While the focus in this section on responsibility and accountability is on actors and organizations held responsible or accountable for incidents, ethical reflections of attributing or obligating responsibilities to these are also included in the following subsections [21]. Responsibility and accountability depend on what type of legal person the researchers think should be held accountable or responsible for an AI-related incident when considering an ethical approach, such as the natural human operators, corporations, AI developers, or the AI system itself. If the natural person has to be held responsible or accountable for the AI-related incident, the difficulty would be in differentiating the natural individual and the AI. Although it is difficult to hold AI systems accountable or responsible because of their legal personhood, there are ethical concerns in attributing responsibilities or obligations to them, like the loss of agency in humans, the treatment of robot dignity, and moral hazard [22

REGULATORY AND POLICY PERSPECTIVES ON ETHICAL AI IN AUTONOMOUS VEHICLES

The emergence of artificial systems in the form of machine learning and decision-support and decision-making algorithms that drive the operation of vehicles is also the focus of attention, especially regarding the ethical and legal aspects involved. Although ethics is usually seen as a matter of personal reasoning, these systems have to fall under a general framework specified in advance of their operation. At present, there is no legislation that punishes the violation of ethical principles in terms of artificial intelligence [2324

ETHICAL AI AND AUTONOMOUS VEHICLES

Regulators and researchers are beginning to explore how to address the ethical development and deployment of artificial intelligence across various sectors, including autonomous vehicles. Given the numerous applications of AI, regulations that would govern the systems as a whole are not a viable option. An alternative that could be sought is "soft" and "hard" regulations, which would divide the AI into several classes for which different rules would apply. The World Economic Forum has published Policy Principles for Trustworthy AI and recently adopted a document on How to successfully adopt AI: Strategy regulation and action plan, which has both immediate and long-term implications in the development and deployment of AI. The European Union (EU) has developed a model of ethical AI principles that have resulted in the development and adoption of a recommended approach for AI that seeks to bring them down to the national level AI governance strategy and working groups responsible for the realization of those principles. The UN has developed Ethical aspects of data, technology, and AI and encourages stakeholder groups to consider the ethical issues associated with AI in all fields. In addition, it created a roadmap guidebook for developing guidelines on the development of AI in autonomous systems. The proposed concepts seek to formulate research issues intended to promote ethical AI themes for those systems. Here, we consider the regulatory approaches and topics that can be used to develop such normative frameworks, pointing out the most relevant issues from each approach [25].

FUTURE DIRECTIONS

This paper draws together insights from a range of fields and incorporates discussions from a diversity of stakeholders, including industry, the ivory tower, civil society, and government. We argue that a bottom-up approach is necessary in order to fully capture the wide range of ethical considerations at play in the development and deployment of AI in autonomous vehicles. Additionally, we have shown that the development and implementation of a truly ethical and socially-just implementation of AI will necessitate a re-thinking of technological development and adoption on a more fundamental level. While we have

drawn general conclusions from the case of autonomous vehicles, in our future work, we will identify specific directions that are needed for future research and training to meet these goals [26]. The paper presents and critically examines the ethical challenges associated with AI, both "narrow" (i.e., weak AI with a narrow range of competencies) and "general" or "strong" AI, the latter including potential super-intelligence AI. In particular, the paper examines the field of "ethics of AI for the digitization of mobility" and offers a typology of different perspectives on these issues. How to integrate human values in the AI system of autonomous vehicles for decision-making is a major open area to research for ethical principles and guidelines in AI development. This requires a lot of effort to interpret, integrate, and capture the adaptations of external human values for AI-driven decision-making in a demographic context of AVs. As of now, AVs already are legally entitled to make some human-based decisions for the safety of passengers and pedestrians alike. How can values perceived as intrinsic to humans be equally extended to the physical shared spaces as well as on a remote operational/control center's virtual space while only autonomous vehicles are concerned? In conclusion, the paper proposes a few equivocal ethical guidelines and future fields of research-active topics in this area of ethics for AVs [15].

CONCLUSION

The development and implementation of AI in autonomous vehicles necessitate a comprehensive understanding of ethical considerations to ensure socially responsible outcomes. This paper emphasizes the importance of applying ethical frameworks such as utilitarianism, deontology, and virtue ethics to guide the moral behavior of AI systems in autonomous vehicles. Addressing these ethical issues is crucial for gaining public trust and ensuring the safety and fairness of autonomous vehicle technologies. Future research should focus on refining these ethical frameworks and developing practical guidelines for AI developers and policymakers. Collaboration among industry, academia, and regulatory bodies is essential to create robust ethical standards and ensure the responsible advancement of AI in autonomous vehicles.

REFERENCES

1. Shi X, Wong YD, Chai C, Li MZ. An automated machine learning (AutoML) method of risk prediction for decision-making of autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems. 2020 Jul 1;22(11):7145-54. [HTML]
2. Blasch E, Pham T, Chong CY, Koch W, Leung H, Braines D, Abdelzaher T. Machine learning/artificial intelligence for sensor data fusion–opportunities and challenges. IEEE Aerospace and Electronic Systems Magazine. 2021 Jul 1;36(7):80-93. researchgate.net
3. Dubljevic V, List G, Milojevich J, Ajmeri N, Bauer WA, Singh MP, Bardaka E, Birkland TA, Edwards CH, Mayer RC, Muntean I. Toward a rational and ethical sociotechnical system of autonomous vehicles: A novel application of multi-criteria decision analysis. Plos one. 2021 Aug 13;16(8):e0256224. plos.org
4. Zoshak J, Dew K. Beyond kant and bentham: How ethical theories are being used in artificial moral agents. InProceedings of the 2021 CHI Conference on Human Factors in Computing Systems 2021 May 6 (pp. 1-15). [HTML]
5. Ferdman A. Bowling alone in the autonomous vehicle: The ethics of well-being in the driverless car. AI & SOCIETY. 2024. philpapers.org
6. Bertrandias L, Ben LO, Sadik-Rozsnyai O, Carricano M. Delegating decision-making to autonomous products: A value model emphasizing the role of well-being. Technological Forecasting and Social Change. 2021 Aug 1;169:120846. kent.ac.uk
7. Scarre G. Utilitarianism. 2020. utilitarianism.com
8. Roff HM. Expected utilitarianism. arXiv preprint arXiv:2008.07321. 2020. [PDF]
9. Smart JJC. Utilitarianism and its applications. New directions in Ethics. 2020. [HTML]
10. Keeling G. The ethics of automated vehicles. 2020. bris.ac.uk
11. Resnik DB, Andrews SL. A precautionary approach to autonomous vehicles. AI and Ethics. 2023. nih.gov
12. Queloz M, van Ackeren M. Virtue Ethics and the Morality System. Topoi. 2024. springer.com
13. Poff DC. Deontology. Encyclopedia of Business and Professional Ethics. 2023. [HTML]
14. Chikwado EP. Kant's Categorical Imperative: A Panacea For Politics Of Ethnicity In Nigeria. Oracle of Wisdom Journal of Philosophy and Public Affairs (OWIJOPPA). 2020;4(2). acjol.org
15. Umbrello S, Yampolskiy RV. Designing AI for explainability and verifiability: a value sensitive design approach to avoid artificial stupidity in autonomous vehicles. International Journal of Social Robotics. 2022. springer.com
16. Constantinescu M, Voinea C, Uszkai R, Vică C. Understanding responsibility in Responsible AI. Dianoetic virtues and the hard problem of context. Ethics and Information Technology. 2021 Dec;23:803-14. springer.com

17. LaCroix T. Moral dilemmas for moral machines. AI and Ethics. 2022. [PDF]
18. Martinho A, Herber N, Kroesen M, Chorus C. Ethical issues in focus by the autonomous vehicles industry. Transport reviews. 2021. tandfonline.com
19. Hyland T. Telling moral tales: exploring ways of enhancing the realism and explanatory power of ethical thought experiments. Open Journal of Social Sciences. 2020. scirp.org
20. Körner A, Deutsch R, Gawronski B. Using the CNI model to investigate individual differences in moral dilemma judgments. Personality and Social Psychology Bulletin. 2020 Sep;46(9):1392-407. sagepub.com
21. Dignum V. Responsibility and artificial intelligence. The oxford handbook of ethics of AI. 2020. [HTML]
22. Naik N, Hameed BZ, Shetty DK, Swain D, Shah M, Paul R, Aggarwal K, Ibrahim S, Patil V, Smriti K, Shetty S. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility?. Frontiers in surgery. 2022 Mar 14;9:862322. frontiersin.org
23. Brennan L. Ai ethical compliance is undecidable. Hastings Sci. & Tech. LJ. 2023. uclawsf.edu
24. Sartor G. Artificial intelligence and human rights: Between law and ethics. Maastricht Journal of European and Comparative Law. 2020 Dec;27(6):705-19. unibo.it
25. Jhurani J, Reddy P, Choudhuri SS. Fostering A Safe, Secure, And Trustworthy Artificial Intelligence Ecosystem In The United States. International journal of applied engineering and technology (London). 2023;5:21-7. researchgate.net
26. Veruggio G, Operto F. Roboethics: a bottom-up interdisciplinary discourse in the field of applied ethics in robotics. Machine Ethics and Robot Ethics. 2020. informationethics.ca

CITATION: Mugabo Kalisa G. Ethical Considerations in the Development and Implementation of Artificial Intelligence for Autonomous Vehicles. Research Output Journal of Biological and Applied Science. 2024 3(1):68-73