AI-Powered Ethical Decision-Making Frameworks for Autonomous Vehicles

By Dr. Pascal Fua

Professor of Computer Science, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

1. Introduction

The distinguishing features of this research with respect to the existing literature and the stated objectives include the development of multi-tier decision-making criteria of AV that are adaptable in nature and context-aware with respect to moral, legal, technological, and situational considerations, and operationalize the framework based on a deep-learning-based model of AV decision-making to demonstrate the feasibility and effectiveness of the proposed framework in real-time traffic conditions [1]. In order to realize the objective, this research hosts three significant outcomes. Firstly, the individual and contextual factors that need to be addressed on ethical AV decision-making from all the related fields will be articulated. Secondly, an ethical decision-making framework that allows a road-field user of non-ethical AV extend further, based on the ethical consideration of various state-of-the-art ethical decision-making frameworks will be suggested. Frailties of one might be mitigated by operationalizing another ethical decision-making approach and provisions may be made for fallback infrastructure for safe and better traffic conditions. Lastly, the multi-tier ethical decision-making criterion based on moral, legal, technological and situational considerations will be developed and then improvised with respect to the required scenario. Finally, an AV car-following model was developed using the above proposed ethical frameworks and parameters were tuned to understand the behavior of the AV following these proposed models in typical stop and go scenario.

Fully automated vehicles (AVs) have raised massive public interest, development of greenfield infrastructures, and enforcement of new policies. However, the ethical decision-making mechanisms that govern the decision-making of AV in ambiguous situations, which are non-linear in nature, is less studied [2]. In addition, there are alarming facts with respect

to availability and accessibility to ethical datasets, lack of transparency and interpretability of AI-based decision-making models, biases and non-accountability pertaining to non-ethical AVs, inclusion of ethical harvards in decisionmaking that may lead to conflicts with pre-existing traffic. In addition, these ambiances may cause a negative behavioral shift among humans. A deepened investigation with respect to ethical considerations of various state-of-the-art ethical decision-making frameworks to operationalize an ethical autonomous vehicle is not comprehensive [3]. To fill this research gap, the following objectives are identified to develop a unified multi-tier ethical decision-making framework.

1.1. Background and Significance

The research method of work investigates the state-of-the-art research to systemize and synthesize social concepts, insights, and knowledge from self-driving car projects. We examine the relevance of social science seachings and features from current projects and we addan all-embracing ethical insight to the existing concepts in philosophy. As a major result of our review, we present a conceptual framework based on an autonomous car. This includes the concepts of ethics, liability and trustworthiness, and provides anda orientation to future research projects in the field of autonomous driving issues.

Research on ethical issues surrounding autonomous vehicles, especially from an artificial intelligence and machine learning point of view, is also increasingly popular. Among the abundance of papers, some address the human's perspective and needs behind trusting a self-driving car and try to conceptually model these considerations, often linking them to psychological and neurobiological processing [4]. Although promising environments for self-driving mobility services—i.e., no human drivers and fully automated as well as predictable street scenarios—might increase the trust of potential customers, feelings of trustworthiness must be perceptible and comprehensible to every individual. This requires embedded cognitive alike and thus learnable behavior to achieve an adaptive street communication competency.

1.2. Purpose and Scope

Automated driving differs from driving by humans in one major aspect—when a situation requires a moral judgment in an accident, the car does not merely act in order to avoid the accident or minimize its consequences (a purely technical question), instead the car is required to compare injuries and to make a moral decision resulting from the collective consequences

of the entire situation. The car does not obey traffic rules without testing and optimizing its behaviors in various scenarios. Therefore, it needs to be trained ver scours of human experiences before testing its behaviors from different viewpoints. In summary, there is a big difference between technical and ethical rules of driving [5].

Autonomous vehicle (AV) technology has been advancing rapidly in terms of technological capabilities, with major technology companies and automotive companies heavily investing in AV technologies to have them ready for widespread deployment… With these technologies, self-driving cars need to make decisions about routing, experienced by self-driving cars and critical consequences in the case of an accident. In addition to these physical and physiological effects, we have the autonomy challenge, who is responsible for the moral judgment of such situations.

2. Ethics in Artificial Intelligence

Autonomous vehicles (AVs) are expected to revolutionize many industries, including transportation. As different researches may state that autonomous architecture behaviors present countless ethical actuations, they are adopted to cohere to objectives modified from ethical theory. Utilitarianism stands out as the strongest moral theory that incorporates machine learning (ML), and its briefings should be preferred so that society can maximize the benefit. However, adopted material moral philosophies oblige the orders of KPSS, namely the loyalty of one's logical rigidity, their integrity and made-upness, the righteousness' t and measuring and consistent agreement with Holden's Analysis and Holder's Decisions. The confidentiality of Human Association junctions is of considerable importance, delaying the development of values such as justice, element and impartiality, especially in the corridors of work that are susceptible to bias. In our story, we manufacture that the attention of legible Pisanship, fairness and the justiceless of the values are underrepresented. And also, whether the executed automatisms should be executed in the delivered product. We hope to contribute a platform that allows the chance of associative development, as well as theme compatible with the chosen model of coop to be perpetrated care may be constructed. [2]

As artificial intelligence (AI) has become a vital part of our everyday lives, responsible AI by design should be emphasized in order to provide the advantages of AI technology with minimal risk. Particularly, AVs are likely to create a radical improvement in road-prevention graphis, in fact, at the same time a greater understanding of how these vehicles make decisions

is also essential, especially when accidents are under consideration. It is clearly important to understand that we have taken a thorough look at the standardization, legalization, and policies that are needed to verify and validate the ethical conduct of autonomous systems. In order to decoct this discomposing problem, it must be engadge to the concerned authorities, and laws should be organized subsequent to the policies on the basis of the values of countries so that agreements are adhered to when designing and regulating the algorithms. Lastly, we consider the hiring of autonomous Connexion officials as part of the debate, since the clear effects on the valuing to be convicted are quantifying cause of attention. In essence, ethicists and specialists are demand to actively participate in the politics and also in the decision support in order to secure the well-being of individuals, institutions and establish. [1]

2.1. Ethical Principles and Theories

An ethical theory pivots on principles that deduce from its root premises to reach plausible conclusions concerning the fundamental ethical responses by which autonomous vehicles should be programmed to react in any given life-threatening traffic situation [1]. Another does not rely on principles but on a methodology analogous to that of an ethical bystander observing an event. It is proposed that a self-protection system is crucial for autonomous vehicles to resolve a so-called "Heath riddle." An ethical deployment strategy of machine learning (ML) utilitably underlines a hybrid decision scheme, consisting of ML-derived ethical principles and on-line principled decisions in life-threatening traffic situations that depend on the vehicle's own assessment of its environment.

With the rise of autonomous vehicles, there has been an increased interest in the problems of ethical decision-making [6]. It has been assumed that a vehicle may have to choose between various actions in a life-threatening traffic situation. It is claimed that programming in these vehicles will always be an ethical issue and that agreement should be sought on how to act in such situations in which each action has both bad and good results creating dilemmas that may affect the decision of the vehicle. Philosophical views on the approach to these kinds of moral predicaments in traffic have been presented. A critical view of the formulated principles and the theories, a more specific approach to autonomous vehicles, and a view of the ethical issues in traffic have been prepared.

2.2. Challenges in Implementing Ethics in AI

[2]The ethical challenges of AI are linked, but also extend beyond algorithms to include technical challenges and political and business constraints. An illustrative example of this situation are autonomous vehicles. One typical consideration in the discussion of the implementation of ethics and regulations in autonomous vehicles is the optimal behavior in cases that lead to a casualty. Therefore, the development of ethical AI seeks to respond to multiple challenges that must also consider the ethical behavior of the user.[4]Currently, the discussion about trustworthy AI is mostly dominated by representational perspectives on ethics, while the evaluation processes place more emphasis on performance aspects of AI. Aim of this position paper is to bridge this gap by offering a coherent framework to implement and evaluate trustworthy AI. To this end, we present a new, ethical decision-making (EDM) tool specifically aimed at integrating ethical considerations during development. As the scope of our approach is broader than just self-driving vehicles, we show how it can be applied to any AI application.

3. Autonomous Vehicles and Ethical Dilemmas

However, developing a universally accepted ethical framework for all situations for all the cultures of the world constitutes a complex puzzle whose solution seems practically impossible. S. Wolf's survey suggests that the existing ethical guidelines likely reflect a cultural bias, and different societies can come up with different solutions for them. In this framework, it is argued that "ethicists and traffic scientists should recommend a range of speed for each road" instead of speed limits and AMVs should make the final ethical decision [7]. Given the huge number of combinations of ethical problems that could occur in traffic situations dozens of times per second, it would be extremely challenging to develop an ethical framework that could address all combinations. In this case, AMV developers and engineers should aim to design an ethical framework that aims to resolve new ethical problems by machine learning and artificial intelligence learning from experienced ones rather than entering these rules into AMVs one by one.

[8] Challenges are gradually being overcome in the development of self-driving cars. At this stage, the relating legal and ethical aspects, especially ethical considerations, appear to be largely neglected [9]. If autonomous vehicles begin to operate on public roads in greater frequency, we generally expect higher traffic safety levels due to the elimination of human

errors, which are responsible for at least 90% of all road accidents. Nevertheless, the emergence of new dilemmas and challenges arising from the transition of the responsibility from the human drivers to the owners, developers, and manufacturers has often been overlooked. Since autonomous vehicles will make decisions in complex and chaotic real-world situations, involving the possible sacrifice of individuals for the benefit of others, many ethicists and researchers agree that these challenges should be addressed urgently by the relevant stakeholders.

3.1. Overview of Autonomous Vehicles

Autonomous vehicles (AVs) are the ultimate outcome of a series of cutting-edge innovations in various domains such as vehicle mechanics, electronic control systems, communication technologies, artificial intelligence, and human–machine interfaces. These innovations have been called revolutionary because they may lead to significant transformations in current personal and commercial transportation systems [10]. Human decision-making biases in the moral dilemmas of autonomous vehicles (AVs) can have significant implications for the establishment of moral frameworks for AVs. A universally accepted moral code is indispensable for developing compliant AV operation software to be implemented in real-world vehicles [11]. The rapid advancements in these sectors hold the promise for AVs exhibiting extra-ordinary benefits, i.e. improved travel safety, greater access to transport services following removal of key restrictions on drivers like age, sickness and disabilities, emissions reduction, congestion mitigation, and improved fuel efficiency. On the other hand, the general adoption of several critical-system-controlled AVs, as already argued by Michael J. Shafer (calling the mode-avoiding virtues favored on the grid and on Trolley-world too high a price for our robot cars to pay), obstacles to realizing those revolutionary benefits may emerge by way of moral dilemmas or character logically indefensible decisions in AVs conflicting with societal ethical standards. This article gives a recent picture of the moral literature concerning the universally acceptable standards or rules of ethical law as opposed to ethical advice for a compliant AI in AVs.

3.2. Ethical Dilemmas in Autonomous Driving

AI-powered decision making for self-driving vehicles gathers massive investments from industries building increasingly more reliable technologies [12]. From the correct identification of a stopped school bus to an ethical decision constrained to Syrian soldiers' lives, a new benchmark of autonomous vehicle decision-making is extracted. Employing 42

standard real-world traffic scenes generated from multiple primary databases, a novel benchmark is proposed by developing a survey regarding hybrid traffic scenes with two or more traffic participants, including vehicle–vehicle/vehicle–pedestrian interactions in terms of the total, violent, and reverse performances evaluated by a state-of-the-art human annotated dataset. To our knowledge, although the authors of firstly studied multi-agent reinforcement learning, the benchmark and the ratio of different states' occurrences are different.

Various studies have explored ethical aspects and decision-making in autonomous vehicles (AVs) [7]. A recent work has proposed a deep reinforcement learning (DRL) method to extract and model trajectories generated to ensure a distributed planning of AVs. However, the formulated problem assumes the uniformity and impartiality of the societal ethical values and moral principles, neglecting the opposite outcomes for the society elements [2]. This research seeks a comprehensive approach and proposes an Autonomous Vehicle Ethical Decision Making framework (AVEDM) that includes the optimal trajectory planning for AV, considering the criticality of human-driven vehicles' (HDVs) behaviors leading to the potential violation of traffic regulations, societal values, and ethical principles. With the advantage of expert experience in teaching a self-driving decision-making engine, an interpretable rationality-based variant of supervised deep reinforcement learning to exhibit a reliable behavior outperforms other baseline violent and total unrestrained tri-modular quantifiable and relatively recent approaches.

4. Existing Ethical Frameworks for AI in Autonomous Vehicles

AI-dependent ethics also matters for AVs as they are "learned" systems. These systems do not have a one-to-one relation between input signals and output signals or instructions. Rather, they optimize a set of parameters - derived from data inputs - to achieve a high accuracy on the training data. Thus, these developing processes require decisions about what to include into the learning process and about the structure and content of the input data [13]. Alongside, value matters for AI as well because this is a crucial matter in the debate about the societal well-being enhanced (or endangered) by the AI systems. For example, if patient values on ethical topics are involved into pattern recognition technologies, questions of ethicality are raised, further, it is a question, whether "ethical" was judged in the ideal value system and

conforming to ethic-based instruments or in the actual value system and conforming to survey-based information [1].

Encoding ethical behavior in autonomous vehicles (AVs) can be achieved by applying ethical goal functions to steer the operation of AVs and ensure they align with societal values [6]. This involves capturing societal values and translating them into AV behaviors. This approach could potentially lead to incorporating a wide range of ethical alternatives, including those that express deontological as well as (possible) non-consequentialist requirements. Moreover, discussions on the programming of ethical concepts and the ethical behavior of digital autonomous systems have recently received significant attention. For example, under the heading of the trolley problem, questions regarding whether automation should be used to determine the choices between different life-endangering consequences have been addressed by scholars and the public.

4.1. Utilitarianism-Based Frameworks

With regard to traffic milestones, the first utilitarian ethics rule uses braking only as the measure of control. The expected future distributions are taken in consideration in both a tradi-tional and a new way. A careful trajectory presents the method used in which trajectory is divided into the assessment of impact and the evaluation of the trajectory alternative. This work does not suggest that the expected velocity after the avoidance riding has been used to consent to the new approach of using both the existing and the terminal speed in weight assessment. The percentage of distribution of probability in EV can be evaluated in spontaneous decision-making by present drivers.

Utilitarianism is frequently used as an ethical principle to make decisions in AV driving. This approach maximizes the wellbeing of all affected entities to make decisions in AVs, as described in. From an equality perspective, utilitarianism tries to maximize the happiness for everyone. Utilitarian decision-making is thought to ethically solve unavoidable accident situations because it considers the most affected and most weakly protected person and minimizing the number of injured persons overall.

4.2. Deontological Frameworks

It is possible to consider the deontological rules that have been suggested for robots2 in order to support deciding about how to program AVs. The traditional sources of deontological approach are religion, culture, and human experience. Deontological reasoning emphasizes

the importance of considering respect for human rights and dignity. Many Kohlbergian moral psychology theorists also propose the rights of individuals as a priority in deontological judgments. The commitment to follow specific rules which have been established based on moral reasoning or intuition is another component of the deontological approach. Since it is important to approach ethical conflicts in a principled manner, establishing rules in advance is necessary [14].

Previous sections have explained how utilitarian principles can be ethically justified. They should be considered in designing emotional responses when generating behavior for autonomous vehicles. However, users' attitudes and beliefs are not entirely fundamental in moral decision-making for AVs. Deontological ethics are concerned with specific rules, duties or laws that are universally valid, such as Asimov's Three Laws. The literature on deontological principles for AVs is not as developed as those of utilitarian principles. Furthermore, there is a lack of agreement on a standard list of deontological rules that should be used to inform the programming of AVs [3].

4.3. Virtue Ethics-Based Frameworks

Different ethical theories are available to guide AVs. Commonly mentioned are utilitarian, deontological and virtue ethics. Each developmental approach has its pros and cons. But from a social perspective it is important to recognize harm that might be caused by AV models that simply enforce popular norms— such as those that disadvantage certain racial or socio-economic groups. While some speculative AV theory is useful, existing vehicles are programmed to follow specific guidelines that lead to ethical decision-making. It is especially important for these vehicles to be trained ethically, and operate in a way that people in each surrounding society broadly find acceptable. It is possible that certain principles proposed directly by deontological ethics, like "autonomy of individuals" or "the duty of assistance," will be incorporated even where those are not principles typically used to govern human drivers. Using orientation from a moral stakeholder that internalizes progressive advancements, our paper explores what it might look like and how to look deeper to integrate virtue ethics into AV design [15].

For all of its potential benefits, automating transportation services through autonomous vehicles (AVs) raises significant ethical concerns. Morally, AVs must be able to ensure not just the safety of their passengers, but also that of pedestrians and human drivers. Once the

technology spreads broadly, moral considerations will also set the tone for the automotive market. Though some may argue that, "the market will decide" and that consumer popularity will lead to the most ethical choice, it is essential that AVs are – in keeping with the principle of ethical market modeling – initially designed and regulated with full consideration of public safety [16]. As the first automated vehicle crashes in the mid 2010s demonstrated, AVs are likely to lack the moral wisdom of a human driver, raising the question of whether we want vehicles judgments to be affected by elements of popular society, or governed by ethical values.

5. AI-Powered Ethical Decision-Making Frameworks

However, despite attaining this significant progress for instance, through AI and data systems integration, it is likely that ethics of usage will broaden in terms multi-faceted evidenced-based decision-making from which applicable strategy solutions will be extracted for the development of both autonomous systems and the direction of policy. The role of AI in the management of human behavior is expected to increase and an automated means of managing the movement of people deviating from the desired norm is expected to come into use. It can be argued that such a significant shift from individuals to government decisions that have the means to carry out a forceful change of individual behavior could breach existing standards or be seen as a more substantial limitation on human freedom. Autonomy is expected to be one way to preserve transcending the usage of AI in surveillance and enforcement. Creating a legal framework that limits the use of AI to act on norm deviations with the intention of protecting a range of human rights, such as the right to privacy, freedom of movement and thought, and effective legal protection against abuses of power by public authorities, could substantially rebalanced possible negative shifts in human and social relations due to the utilization of AI. The EU's Strategic Action Plan on AI appreciated this need and advised development that should be the use of ethical-legal frameworks with clear responsibilities and obligations for usage of AI. In addition, efforts should be made to agree on a legal framework on the ethical boundary for AI decisions.

As with any ground-breaking technologies, the proper use and protection of artificial intelligence (AI) for socially beneficial ends is currently in the early stages of development, and ethical implications should be continually studied [2]. IVs provide numerous social, environmental, and economic benefits, such as reducing traffic accidents, improving road

safety, avoiding congestion and providing mobility service for special user groups. It is expected by policy and decision makers to speed up market introduction and mass inlectrication of AVs. The widespread introduction of fully AVs into the urban environment will significantly improve safety levels much farther than the current level reliably assumed for highly automated driving (HAD), and the improved traffic efficiency will have a positive environment al effect [17].

5.1. Components of an Ethical Decision-Making Framework

The fourth component is some sort of personal values intrinsically associated with one's sense of moral obligation or commitment to ethical action that generates normal feelings of compromise in between conflicting decision maxims principle. This is typically different for individuals, but car manufacturers could introduce their own sets of typical or average values to make the default recommendation without the hassle of specific individual customer engagement. The last component is the personality of the owner, that plays a significant role in auto_BEL. It may affect trust in the positions in the direction suggested by the BEL trend, explorative aspects and cautiousness when driving, and the engagement of assistants/lights/sound systems. Keep in mind that personality affects attitude and behavior and ethical framework with personalization can produce selection problems for users. This is where EDM deals with various modes in which people will ope Their overall aim is to explore how ethical allocations in EDM work with other moral dimensions as well as taking intrinsic culture into account where appropriate.

It is crucial that an EDM comes in the form of a clear decision table, as this allows the designer to be clear about the conditions of each decision and prevents double standard issues. The next component of an EDM would a set of ethical principles associated with those preferences such as 'impact avoidance', 'expected utility', 'equal division', 'lexicographical avoiding harm', 'Nash bargaining' or some other newly-emerging notions. The third component is a set of beliefs that predict the likely consequences of potential actions and describe the probabilities of environmental events when interacting with the main vehicles or autonomous car driving conditions. The probabilistic nature of this framework that can be updated in the light of evidence is initially derived from our auto_BEL model.

5.2. Machine Learning Algorithms for Ethical Decision-Making

Deep learning has shown potential in AI ethical decision-making tasks, such as learning to reason about the fairness of algorithms. However, the reliance on large amounts of labeled data and the lack of interpretability are limitations. The moral AI literature demonstrates that human cognition has the ability to learn from sparse examples and rationalize moral decisions. To ensure ethics-compliance in AI decisions, it is critical for engineers to leverage recent advances in ML, such as transfer learning, model-based reinforcement learning and learning from human feedback. The autonomy with which AI plays a role in decision-making highlights the need for interpretability in models when reasoning about ethical judgements. This poses a specific challenge if competitor companies leverage black box personalized models for autonomous decision-making tasks. In order to ensure ethical decision-making with AI, engineers have to combine the benefits of black box models with interpretable AI, and, if necessary, with the provision of accountability settings to aggregate normative human moral values and preferences.

[18] [19] An ethical decision-making framework must aim at structurally empowering AI to receive and process the most relevant data. This implies that any ethical decision is subject to the training data, algorithms and policies that define the behavior of the AI. Since the way in which machine learning (ML) algorithms are fed information shapes ethical decision making, recent advances in machine learning algorithms for ensuring ethical decision-making in AI need to be analyzed.

6. Implementation Challenges and Considerations

At the outset, it seems evident that the world is yet to arrive at universally applicable ethical principles for AI and it seems these ethical parameters are not uniform in much less connected, diverse areas. So too are they not for the automotive vehicles though it has enjoyed some measure of global rules and guidelines set by United Nations and its subgroups like WP.29, NHTSA, and NCAP. Furthermore, what adds to the complexity is the market-driven nature of AI and autonomous automotive vehicles.

As with any novel technological applications, the implementation of AI-powered ethical decision-making frameworks in autonomous vehicles is plagued with certain obstacles and concerns that concern the sharing of responsibility [20]. There is also the broader implementation issue of a standardised ethical decision-making framework for the

automotive sector, as a whole unit. This shared responsibility issue is important to the social acceptance of any autonomous vehicle as it would have far-reaching consequences leading from the legal repercussions to the sound confidence of the public in the established standing of any policy making [13].

6.1. Technical Challenges

By stride in this respect, we mean vehicle intelligence which could be categorized as recognized, prospected and actionable. Recognized intelligence encompasses the capability of the cars in identifying pertinent occasions autonomously, envisioning the probable avocations of associated entities in the instant future and on the basis of that having an estimate of the probable future vitality of the collision with the various alternatives. The prospected intelligence encompasses the capacity to envisage the all-round consequences of these alternative eventualities. The logistical challenges for these recognized and prospected intelligences to operate far from the edge are conspicuous. In addition to the complexity and time-extension of the involved calculations, inaccuracies in the actuating environment, the unavailability of complete environmental information and the high degrees of dynamism of those environments are also the snags that must be overcome. To top it, the actions to be taken by occupants and pedestrians in crashes or near-crashes very often deviate substantially from the expected-norms of humans. The inability of the recognized and prospected intelligent modules in proactively bringing out the earnest desires assigned to the occupants of the autonomous cars and the pedestrians largely mean that all the ethical challenges of pedestrian crash will have to be addressed within the action intelligence [21].

Huge strides have been made in creating autonomous vehicles (AVs). As a result, we should soon—hopefully—see a widespread usage of automated vehicles on our highways, at the latest. By which time, technical and ethical challenges surrounding the decision-making process for autonomous vehicles with regard to ethical dilemmas must be resolved [22]. This subsection delineates a few of the conceivable technical challenges and suggests that an automatic ethical preconditioning method may be utilized to overcome these and create ethically advanced VA systems.

6.2. Regulatory and Legal Considerations

Certain AI systems are starting to be regulated due to their strong ethical and societal impact. In the European Union (EU), proposed regulation in the Artificial Intelligence Act aims to

create a common legal framework for the development, placement on the market and use of artificial intelligence systems across the EU, while ensuring comprehensive and effective protection. The Act has provisions on the broad use of artificial intelligence in multiple areas, from the healthcare and transport sectors to parts of the business activities and public sector processes. The Act has a direct impact on how AI will be developed and regulated in the context of autonomous systems and AV, with critical implications for the research carried out in this book, as modern AI powered AVs are thought to be covered by it. It is thus critically important to present in this section the existing AI regulations and further proposed regulations that are in the process of implementation.

A potential barrier to the application of AI-powered decision-making models in the AV domain has been the lack of regulatory and legal frameworks [9]. Concerns have emerged with respect to the contradicting regulatory and operational requirements between modern vehicles and AI software. The interaction between these two systems has been observed to complicate traditional vehicle development divisions, such as vehicle design authority, vehicle type approval, and driving and operational requirements [6]. This has created both legal and regulatory compliance challenges for AI powered systems. This chapter provides an overview of the current regulatory and legal landscape surrounding autonomous systems and AI, and then discusses the incorporation of AI powered ethical decision-making models in control systems [13].

7. Case Studies and Applications

[23] As summarized by Ginn et al., there are currently only a small number of AI safety projects available within the public domain, which include the Moral Machine, HRI-Colecta, and Local Relative Prospective Reach. The majority of ongoing research is concerned with driving psychology and ergonomic design of the user interface, robotics, truck driving, and social influence groups. However, the public is not familiar with the wider implications of a new class of robot vehicles that have no driver and convey no passenger, i.e., truck- and city-taxis, sea-, land- and air-freighters, aerial survey drones, agriculture drones, and industrial robots, which might accompany that robot class.[1] According to OpenAI, the scientific community and industry adopted the perspective or methodological principle-compatible with contextual embedding, guided by fairness, reliability, privacy, safety, security, inclusivity, and relevance. It is necessary to promote trustworthy AI. These objectives should

be achieved in the successive stages of introducing a new technology-based system into the real-world. The failure to achieve responsible design, deployment, and human–AI interaction impose a large number of risks and challenges. As an example, ethical concerns are increasing with widespread AI development. Programming error, structural flaws, malpractice, no AI-free zones, bad governance, no root-assessment, mix-up of the social and machine mission, value-mismatch, moral agency, accountability, transparency, fairness, and hidden risks are critical issues to be tackled in Ethically Aligned Design.

7.1. Real-World Examples of Ethical Dilemmas in Autonomous Vehicles

For instance, in some situations, AV driving styles are regulated by a set of vehicle motion planning laws and regulations. However, it is desirable to have AVs adapt to changing CA—the CA of vehicle motion sometimes contravenes CA such as traffic laws, required safety distance, obligation to end a non-dangerous situation, right-of-way; in such situations, safe and comfortable motion is performed, and law-breaking behavior is furthermore avoided if possible. Additionally, the higher the deviations in the values of an AV's evaluative objective function and the achieved objective function in the case of the CA, the less the CA is favorably ranked. The handling of unexpected encounters such as ambiguously observed traffic signs via adaptation of CA is attested through grounding numerically on real-world data.

Autonomous vehicles (AVs) are garnering widespread attention in recent years given their potential to significantly change our transportation systems and impact society [1]. While the promised advantages of AVs look appealing, their introduction also triggers a debate on their societal and individual ethical impact. For example, who should an AV car save in an unavoidable traffic situation or what kinds of behaviors should be rewarded and penalized in the operation of an AV? These questions cannot be addressed merely by adopting black-and-white ethical decisions, such as "save the passenger at the expense of pedestrian(s)" or "by default, follow traffic rules that implement traffic regulations" [5]. Rather, a more nuanced ethical framework is required to help AVs manage and resolve real-world ethical dilemmas appropriately [18].

8. Future Directions and Emerging Trends

As AVs continue to mature, EAD with a particular focus on morals has evolved into an interesting point of continued focus within the autonomous systems community to better undertake the science, engineering, and design of machine interfacing AI based technology,

some of which is highlighted in find fault in AV decisions. It becomes apparent that there remains a critical area where AV designers have incorrectly aligned engineering practice with societal expectation, thus generating a requirement to understand what constitutes an EAD for industrial AI developers of AVs, particularly in their use of ML AI technologies. The human ability to be moral may then be compromised by the indirect effects of unethical decision-making, for example, when a pedestrian avoids crossing in front of a dangerous car, since they detected that the AV system inside the vehicle was acting obnoxious.

[4]The perspectives of AI experts can differentiate between machine learning technologies that are safe and unsafe for the general public, emphasizing the need for rigorous testing and ethical scrutiny in the automotive sector. While engineers viewed AI as a predictable tool requiring social communication and professional regulation, design experts viewed AI as a more unpredictable technology that pushes power relationships. The automotive sector has increasingly embraced AI, exploring human skills that can work in harmony with autonomously operating vehicles. Simultaneously, this has also raised worries about a declining role for humans driving transportation vehicles, both skillfully and ethically.[18]There is also a growing interest in how decision-making in these intelligent vehicles should be controlled. It has been found that the interplay of sensor faults, world model uncertainty, and collaboration requirements can often lead to situations where the vehicle controllers need to make splits-second (race condition) style decisions while being stranded in a local optima that has strong safety boundary constraints (for example, a decision to slow down to avoid an obstacle might lead to a reduction in a possibly non-recoverable high altitude for a takeoff, again leading to an unavoidable crash). The responses to such 'fast-time tactical decision-making' in high-dimensional uncertainty space should be balanced between safety and ethical considerations for clear human–AI operator intervention. The authors believe these aspects to be largely ignored in current discourses in the research community for which effective decision-making paradigms are needed in introducing a modicum of ethical reasoning. Explain the theory and practice of tactical autonomous decision-making of intelligent vehicles and integrate the ethical aspects of autonomy. The resulting perspective will show that advances in research on artificial intelligence and autonomy pose many complex challenges via their technology, on several fundamental technical and safety levels, but also in relation to a number of complex ethical and decision modeling challenges regarding their transport.

8.1. Advancements in Ethical AI

However, even considering these risks, the technology is here to stay due to the immense potential benefits it holds. Governments and companies should see accidents not as a reason to slow down the progress of AI, but as a reason to step up attention and funding to appropriate safety assurance efforts. These will be needed until the future technology has matured to the point that it was previously reached mobile and stationary machines, achieved by millions of years of evolutionary progress with a 1:1 ratio of successes and failures. More ambitious visions of intelligent AI-based transportation technologies further include the networking of autonomous vehicles within a traffic system which coordinates at the micro level the vehicles on the road and at the macro level optimises overall city traffic flow and minimises travel inconveniences. All these promising ideas ultimately shall increase the safety, efficiency, and driving convenience for the urban society.

To harness the potential of AI in autonomous vehicles safely, major advancements have been made recently to enhance several critical areas. This includes approaches for more sound decision-making, trying to understand and anticipate the different kinds of adversarial attacks vehicles may be subject to, and developing robust measures to lock, defend and adapt the learned policies under conditions of multiple distributional shifts [10]. In ethical terms, the processing of AI-derived transportation data has also advanced becoming more complete and will likely continue being developed in the forthcoming years. Despite the many positive developments in the area of AI-supported autonomous vehicles, future safety incidents and accidents involving this new technology may still occur.

8.2. Integration of Ethical Frameworks in AI Development

We see a general consensus that the definition of ethical dilemmas, values and norms and decision making [24] processes regarding AI are necessary to inform and guide AI development. Superficially, this might be taken to mean that AI must obey rules established by an ethical framework representing a particular culture. For many practical purposes, it might be possible to base partially autonomous ethical reasoning on existing professional, legal or cultural frameworks. However, the ethical problem is exacerbated when ethical domains either provide no ethical guidelines or conflict.

Building AI systems raises ethical issues [25]. Critical practical and theoretical questions are, for example, related to the ways in which decisions are reached by AI systems which have

direct impact on human lives [2]. This is particularly true in high-stakes decision-making scenarios like autonomous vehicles and healthcare. In order to restrict the adverse impact of AI systems and to provide guidance for AI designers and developers, it is necessary to develop ethical and value-based frameworks for AI decision-making.

9. Conclusion

In the past decade, there has been a surge in the development and deployment of AI algorithms aimed at creating so-called autonomous driving. These systems, which function mostly as self-driving cars and taxis, become increasingly capable of fulfilling this task. However, their deployment also raises fundamental ethical questions: Who decides whom an autonomous car is going to harm when an accident is unavoidable or just likely? Which attention to draw towards avoiding or causing accidents? We therefore proceed to also investigate what makes someone decide for which of the potential accident participants a car should be optimized [13]. For the purposes of this investigation we have deployed its novel autonomous driving agent system, ShifteR, which shall hence refer to the machine learning agents used in the continuous-action reinforcement learning problem setup using Markov decision processes. .

Encoding ethical behavior in AVs through the application of ethical goal functions can ensure that societal values steer the operation of AVs [6]. This is crucial in delivering on societal expectations and integrating AVs safely and ethically. While it is important for AVs to perform ethically, a need exists to ensure that the ethical decision-making will satisfy societal expectations. In this work, we seek to address this by deriving a set of norms to fulfil ethical being intended for implementation in AVs (EthNavi) [5]. This framework is inspired by Asilomar AI Principles and covers diverse topics, including safety, maneuvering, and social behavior. Initial simulation results show how EthNavi allows various stakeholders to express their preferences and how these preferences impact AV behavior. Ultimately, EthNavi yields a large decisional space in which external stakeholders can prioritize ethical objectives, and the decisions of AVs can be aligned such that they satisfy the social and safety objectives and the legislative framework.

9.1. Summary of Key Findings

Machine learning (ML) is one of the most popular and most frequently used methods for developing artificial intelligence (AI) based systems. However, the interpretation and

explainability of ML systems in safety-critical applications such as automotive domain are the primary concerns to guarantee safety of AI-based systems. It is difficult to explain in a human-understandable language the decision processes of ML models due to their black-box behaviors. Recently, many interpretability methods have been developed and evaluated in different test environments to explain the decision process of the ML models. Additionally, explainability methods have been evaluated in different test environments to evaluate the accuracy of the ML models. In this study, a literature update review and comparative analysis of the interpretability methods were given in order to provide ethical solutions for AI-driven AVs. According to the findings, Decision tree, Support vector machine (SVM), Random forest, Artificial neural network (ANN), and Logistic regression are the most prominent ML sub-databases of IEEE Xplore and Access Science/ McGraw Hill databases. The test environment has also been selected to meet two main requirements of its features in order to find the most effective implication method [26].

In this study, we used a mixed-methods research approach by focussing on the stakeholders and a number of guiding questions in order to develop an ethical decision-making framework for future autonomous vehicles. The stakeholders were the automotive industry, the large public, the legal entities, and governmental institutions. The well-known ethical theories such as utilitarian ethics, deontological ethics, virtue ethics, and pragmatic ethics were used to consider ethical uncertainties [2]. Furthermore, these theories were combined with procedural ethics and axiology to resolve the identified ethical dilemmas and their potential consequences. As a result, proactive solutions were presented for the participants. The methods chosen for the research were Duke-UNC Functional Social Support Questionnaire (DUKE-UNC), Interpretative Phenomenological Analysis (IPA), and Rule-based Reasoning Map (RR-map) and CAFE. Due to the insufficient number of current stakeholders in the automobile industry, the results suggest that young people (trained) should be involved in extra training to ensure that they can take on the roles of tomorrow [27].

9.2. Implications for the Future

Societal alignment would mean all the though effort, a technical system would still be faced with the same requirements of crisis situations, changing norms, and edge cases as those for societal alignment. In addition, the ability for the other agents in the system, such as the car, to be transparent about the methods it uses to make its decisions would allow better coordination between other agents, humans and robots, on the road to become better road

citizens. Another stepping stone to resolve such problems and ease the difficulties of the societal alignment process is meaningful legal and societal regulations developed with the goal of ethical alignment in mind. Any development of meaningful laws and regulations from the vehicle perspective needs a thorough conceptual ethical model, which we have laid the foundation for in this paper. Here, AVs are looked at from the inside out, meaning one starts with the specification of the mathematical framework predictive solution for the agent's decision-making as well as the mathematical model for the agent's internal norms or utilities [13].

The use of ethical goal functions as a means to can ensure that decisions made by autonomous systems such as AVs align with values in line with societal norms [6]. This would provide a test bench environment for AVs to use to develop and test their decision making without legal risk, contrasted with ensuring that they always operate as expected in the real world. Currently, however, the translation of ethical concepts and principles derived from a general society/agreement on 'ethical desiderata' design to social norms and values in the automotive sector is rarely stated substantively and transparently in research publications. The 'ethical alignment' of AVs, i.e. the field directly concerned with the issue of how ethical decisions can be integrated and controlled for, is also approached from a largely technological point of view, focusing on the verification and validation of the AI systems that make these decisions [5].

References:

1. [1] V. Vakkuri, K. K. Kemell, J. Kultanen, M. Siponen et al., "Ethically Aligned Design of Autonomous Systems: Industry viewpoint and an empirical study," 2019. [PDF]
2. [2] L. Luxmi Dhirani, N. Mukhtiar, B. Shankar Chowdhry, and T. Newe, "Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review," 2023. ncbi.nlm.nih.gov
3. [3] J. Rhim, J. H. Lee, M. Chen, and A. Lim, "A Deeper Look at Autonomous Vehicle Ethics: An Integrative Ethical Decision-Making Framework to Explain Moral Pluralism," 2021. ncbi.nlm.nih.gov
4. [4] G. Jonelid and K. R. Larsson, "Finding differences in perspectives between designers and engineers to develop trustworthy AI for autonomous cars," 2023. [PDF]
5. [5] T. Holstein, G. Dodig-Crnkovic, and P. Pelliccione, "Ethical and Social Aspects of Self-Driving Cars," 2018. [PDF]

6. Tatineni, Sumanth. "Customer Authentication in Mobile Banking-MLOps Practices and AI-Driven Biometric Authentication Systems." *Journal of Economics & Management Research. SRC/JESMR-266. DOI: doi. org/10.47363/JESMR/2022 (3)* 201 (2022): 2-5.
7. Vemori, Vamsi. "Evolutionary Landscape of Battery Technology and its Impact on Smart Traffic Management Systems for Electric Vehicles in Urban Environments: A Critical Analysis." *Advances in Deep Learning Techniques* 1.1 (2021): 23-57.
8. Leeladhar Gudala, et al. "Leveraging Artificial Intelligence for Enhanced Threat Detection, Response, and Anomaly Identification in Resource-Constrained IoT Networks". Distributed Learning and Broad Applications in Scientific Research, vol. 5, July 2019, pp. 23-54, https://dlabi.org/index.php/journal/article/view/4.
9. [9] S. Paiva, M. Abdul Ahad, G. Tripathi, N. Feroz et al., "Enabling Technologies for Urban Smart Mobility: Recent Trends, Opportunities and Challenges," 2021. ncbi.nlm.nih.gov
10. [10] D. Garikapati and S. Sudhir Shetiya, "Autonomous Vehicles: Evolution of Artificial Intelligence and Learning Algorithms," 2024. [PDF]
11. [11] D. A. Frank, P. Chrysochou, P. Mitkidis, and D. Ariely, "Human decision-making biases in the moral dilemmas of autonomous vehicles," 2019. ncbi.nlm.nih.gov
12. [12] D. Hettiachchi, K. Ji, J. Kennedy, A. McCosker et al., "Designing and Evaluating Presentation Strategies for Fact-Checked Content," 2023. [PDF]
13. [13] A. Kriebitz, R. Max, and C. Lütge, "The German Act on Autonomous Driving: Why Ethics Still Matters," 2022. ncbi.nlm.nih.gov
14. [14] L. Li, J. Zhang, S. Wang, and Q. Zhou, "A Study of Common Principles for Decision-Making in Moral Dilemmas for Autonomous Vehicles," 2022. ncbi.nlm.nih.gov
15. [15] H. Yu, Z. Shen, C. Miao, C. Leung et al., "Building Ethics into Artificial Intelligence," 2018. [PDF]
16. [16] F. Riaz and M. A. Niazi, "Designing Autonomous Vehicles: Evaluating the Role of Human Emotions and Social Norms," 2017. [PDF]
17. [17] V. Dubljević, G. F. List, J. Milojevich, N. Ajmeri et al., "Toward a Rational and Ethical Sociotechnical System of Autonomous Vehicles: A Novel Application of Multi-Criteria Decision Analysis," 2021. [PDF]

18. [18] D. Haileselassie Hagos and D. B. Rawat, "Recent Advances in Artificial Intelligence and Tactical Autonomy: Current Status, Challenges, and Perspectives," 2022. ncbi.nlm.nih.gov
19. [19] R. Kim, M. Kleiman-Weiner, A. Abeliuk, E. Awad et al., "A Computational Model of Commonsense Moral Decision Making," 2018. [PDF]
20. [20] M. Karim Khan Rony, M. Rina Parvin, M. Wahiduzzaman, M. Debnath et al., "“I Wonder if my Years of Training and Expertise Will be Devalued by Machines”: Concerns About the Replacement of Medical Professionals by Artificial Intelligence," 2024. ncbi.nlm.nih.gov
21. [21] T. LaCroix, "Moral Dilemmas for Moral Machines," 2022. [PDF]
22. [22] R. Martin, P. Kusev, J. Teal, V. Baranova et al., "Moral Decision Making: From Bentham to Veil of Ignorance via Perspective Taking Accessibility," 2021. ncbi.nlm.nih.gov
23. [23] Y. H. Weng and Y. Hirata, "Design-Centered HRI Governance for Healthcare Robots," 2022. ncbi.nlm.nih.gov
24. [24] V. Vakkuri, K. K. Kemell, M. Jantunen, and P. Abrahamsson, "“This is Just a Prototype”: How Ethics Are Ignored in Software Startup-Like Environments," 2020. ncbi.nlm.nih.gov
25. [25] P. Radanliev, O. Santos, A. Brandon-Jones, and A. Joinson, "Ethics and responsible AI deployment," 2024. ncbi.nlm.nih.gov
26. [26] S. Malik, M. Ahmed Khan, H. El-Sayed, J. Khan et al., "How Do Autonomous Vehicles Decide?," 2022. ncbi.nlm.nih.gov
27. [27] V. Vakkuri, K. K. Kemell, and P. Abrahamsson, "AI Ethics in Industry: A Research Framework," 2019. [PDF]