HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

The Implementation of Ethical AI Within Autonomous Vehicles

Vadim Fedorov[1] and Noah T. Curran[#]

[#]Advisor

ABSTRACT

The integration of Artificial Intelligence in the rapidly evolving landscape of autonomous vehicles necessitates an in-depth exploration of the associated ethical implications, especially in life-or-death scenarios. This research paper delves deeply into the multifaceted ethical issues associated with the deployment of AI-driven autonomous vehicles and the moral implication regarding this development. A core inquiry of this work is discerning the appropriate entity responsible for critical decision-making when human lives are in the balance. Given the unique context of autonomous driving, where split-second decisions could be a norm rather than an exception, we aim to investigate both the current algorithms' decision-making processes as well as their lack of transparency and question the underlying ethics that shape them. Drawing from real-world incidents involving autonomous vehicles, and juxtaposing them with theoretical predicaments, this study endeavors to present a holistic view of the prevailing ethical perspectives. Building on this foundation, we introduce a fitting ethical decision-making framework, rooted in deontological principles, designed to guide autonomous vehicles through morally complex scenarios on the road. To validate and refine our framework, we employ diverse case studies from the world of autonomous driving. Simultaneously, recognizing the inherent challenges that any ethical framework might encounter, we also discuss potential pitfalls and offer suggestions to enhance its robustness and applicability. As a result, this research underscores the pressing need for meticulously crafted guidelines governing AI within autonomous vehicles, ensuring that safety, individual autonomy, and ethical qualifications are implemented within the age of driverless transport.

Benefits of A Perfect AI

The profound impact of AI on human existence, both now and in the foreseeable future, cannot be overstated. Specifically, the implementation of AI technology within the field of transportation is an endeavor that provides countless benefits to both users and unrelated beneficiaries of the technology. In an ideal world, the complete replacement of every vehicle on the road with autonomous technology would facilitate the efficiency of travel to an unforeseen degree due to the lack of mistakes made by AI drivers. As stated by the National Highway Traffic Safety Administration, over 90% of all accidents on the road are caused by human error. By removing human drivers from the equation, autonomous vehicles (AV) could greatly reduce the number of accidents and fatalities on the road, making transportation safer for all, including pedestrians. This is a technology that would enhance human life and safety and figuratively shrink the world we live in by reducing travel time between two points, giving the illusion of a shorter distance. Infrastructure planned within this utopian context may even alleviate the need for stoplights and other slowdowns, reduce the amount of road space required for efficient travel, and grant the ability to dedicate more resources to create walkable, environmentally sustainable cities, all within the safety of AI drivers. A study by the International Transport Forum cites that the widespread adoption of AVs could lead to a reduction in greenhouse gas emissions by up to 80%, as AVs can be optimized for fuel efficiency and reduce the need for individual car ownership. This could have a significant positive

impact on the environment and help address climate change. Regardless of the benefits provided by such an ideal future, they are contrasted by both technological and ethical concerns surrounding technology.

Concerns that Arise

While there is a growing enthusiasm for embracing a utopian future propelled by the widespread adoption of self-driving vehicles with their inherent efficiency and safety advantages, the relinquishment of critical decision-making and control to autonomy raises legitimate concerns (Goodall, 2016). The decision-making process behind autonomous systems and the level of trust to be bestowed upon them become pressing issues. Consequently, a plethora of contentious points emerge, demanding thoughtful answers as they bear significant consequences for human life and the functioning of AI. Within the realm of self-driving vehicles, many ethical concerns surface, requiring resolution before this technology can be widely adopted (Gurney, 2016). The rapid advancement of this technology has prompted discussions on its ethical implications (Santoni de Sio & van den Hoven, 2018). Instead of merely focusing on individual decision-making moments within vehicles, scholars emphasize the broader ethical framework underpinning the development, deployment, and regulation of such technologies (Nyholm, 2018). Recognizing these challenges, various organizations and scholars have initiated projects to delve into the ethical landscape of self-driving cars. For instance, the Partnership on AI, a collaborative initiative, has been exploring the ethical dimensions of AI technologies, including autonomous vehicles (Dafoe et al., 2021). Through interdisciplinary discussions, these initiatives aim to shape ethical guidelines and policy recommendations for the deployment and operation of these vehicles.

Mistakes of People Vs AI

The repercussions arising from the choices made by human drivers on the road can often be justified by considering the inherent fallibility of individuals, establishing these consequences as customary rather than anomalous occurrences. However, when we shift our focus to AVs, the very same imperfections and seemingly "incorrect" decisions made by artificial intelligence come under intense scrutiny, eroding the general public's confidence in this advancing technology (Liu & Crowcroft, 2016). As we delve deeper into this topic, it becomes apparent that the imperfections attributed to human drivers stem from their limited cognitive capacity and inherent biases (Eliot, 2020). Studies have shown that humans often exhibit suboptimal decision-making skills while driving, influenced by factors such as fatigue, distraction, emotional state, and personal judgments. These flaws, though regrettable, are widely accepted as part of human nature. Conversely, the scrutiny directed at AVs arises from the high expectations placed on them as advanced technological creations. While AI-powered systems possess incredible computational capabilities and are designed to make rational decisions based on inputs and algorithms, they still encounter difficulties in complex real-world scenarios (Goodall, 2016). The skepticism towards AI's decision-making capabilities can be attributed to the lack of transparency in the algorithms and decision processes of autonomous systems (Nyholm & Smids, 2016). The inner workings of AI algorithms can be complex and challenging for the public to understand, contrasting with human decision-making, which is more relatable (Mittelstadt et al., 2016). Media coverage of high-profile incidents involving AVs can amplify public concerns, overshadowing the safety improvements that AVs offer (Shepardson, 2022).

AI Decisions and Ethics

The autonomous vehicle faces a vast array of potential outcomes when confronted with different scenarios, generating an almost infinite number of possibilities. To navigate this ethical landscape, the AV system incorporates perspectives rooted in utilitarianism or deontology, which guide its decision-making process during critical moments (Goodall, 2016). Each encountered scenario necessitates a careful and individual examination,

eschewing a one-size-fits-all approach. When it comes to the prioritization of human life, the AV system grapples with the weighty responsibility of assigning value to the lives involved. Utilitarianism, a consequentialist ethical theory, posits that the morally correct choice is the one that maximizes overall well-being or minimizes harm for the greatest number of individuals. On the other hand, deontology emphasizes the importance of adhering to moral duties and principles, regardless of the consequences (Nyholm & Smids, 2016). Each scenario demands careful consideration of factors, such as the severity of potential harm and legal obligations. To inform their decision-making processes, AV systems rely on extensive training and testing, involving simulated scenarios and real-world data (Liu & Crowcroft, 2016). Researchers and engineers work to refine the algorithms and models, seeking to improve the AV's ability to make morally sound judgments. As the technology evolves, ongoing discussions shape the guidelines and regulations governing the deployment of AVs. The intricate interplay between the ethical and technological dimensions of autonomous vehicle technology extends beyond the mere functionality and delves into moral decision-making. One critical factor is the status of the individuals within the autonomous vehicle and the ethical implications of incorporating this information into the decision-making process. This aspect poses a philosophical challenge, raising questions about the inherent value assigned to human life and potential for discrimination based on arbitrary criteria (Mittelstadt et al., 2016). Furthermore, the presence of a passenger within the vehicle further complicates the decision-making process. In semi-autonomous or fully self-driving vehicles, the question arises as to to what extent the human occupant should be entrusted with influencing the vehicle's actions. Should the AI relinquish control to the human in certain situations, recognizing their ability to comprehend and respond to the environment better? Striking the right balance between human judgment and the capabilities of AI becomes crucial, as it directly impacts the safety and effectiveness of AVs. According to a study conducted by the Massachusetts Institute of Technology (MIT) researchers, factors such as the age and social worth of potential victims have emerged as contentious points in the ethical decision-making algorithms of AVs. This highlights the need for careful consideration and scrutiny of the parameters utilized in these algorithms to ensure fair and just outcomes. Additionally, a research paper published in ScienceDirect explores the concept of user trust in AVs, suggesting that the timing and extent to which control should be relinquished to human passengers should be determined through thorough user studies and evaluations.. This highlights the importance of empirical research and user-centered approaches in addressing the question of when and how human intervention should be prioritized.

Within the realm of AI and self-driving technology, the ethical landscape appears to be a complex and intricate domain, lacking a well-defined structure, thus rendering it perplexing to determine the moral implications of the actions undertaken by such technology. Hence, our objective is to present a comprehensive and sophisticated approach to ethical decision-making, one that can be effectively employed in the context of AI technology in self-driving vehicles. To achieve this, we propose a morally-defensible perspective that not only provides a robust framework for assessing ethical considerations but also encompasses theoretical case studies of autonomous vehicle (AV) decision-making. By utilizing this framework, we can systematically categorize the myriad of potential outcomes on a nuanced ethical spectrum, thereby fostering a more informed and refined understanding of the ethical ramifications inherent in self-driving technology.

Summary of the Paper

The core thrust of this paper centers on the necessity for and implementation of a distinct ethical framework within the decision-making processes of AVs. The paper identifies a significant void in the current landscape where a clear and explicit ethical foundation is urgently needed to regulate and guide the decisions made by these advanced technological advancements in order to continue to drive safe and accepted innovation. The paper aims to advocate for the establishment of an ethical framework revolving primarily around the concept of 'human flourishing'. Human flourishing, an idea rooted in Aristotelian ethics, emphasizes the achievement of the 'good life' through the realization of human potential and the pursuit of virtue. This principle places

human well-being and moral growth at the forefront, making it a highly relevant perspective for addressing ethical dilemmas faced by AVs. In order to robustly argue for this ethical standpoint, the paper presents a series of hypothetical case studies in which the human flourishing framework is applied and compared to the other prevalent ethical theories, namely deontology and utilitarianism. Deontology, a duty-based approach, and utilitarianism, a consequentialist theory, represent the conventional poles in ethical discussions. The comparative analysis between these ethical extremes and the human flourishing model in the context of AVs provides a comprehensive and insightful examination of their respective strengths and weaknesses. Lastly, the paper explores the broader implications of adopting the human flourishing framework in AVs, discussing its potential impact on public trust, legal regulations, and societal acceptance of this rapidly evolving technology. By offering this thorough and nuanced exploration of ethical decision-making in AVs, the paper illuminates the path toward more ethically sound, transparent, and human-centric autonomous technology.

AV Capabilities

In order to gain a comprehensive understanding of the vast array of opportunities and challenges that arise from the advent of AVs, it becomes imperative to delve into the intricate technical details that define these vehicles and enable them to operate effectively. By exploring the technical aspects, we can grasp the nuances of the available information and the decision-making capabilities inherent in AVs. One crucial aspect that warrants meticulous examination is the intersection of ethical considerations and artificial intelligence within the realm of autonomous driving. This intricate relationship brings forth a complex web of ethical dilemmas, intricately woven into the very fabric of the AV system's decision-making abilities. The concept of ethical AI encapsulates the profound question of how a self-driving vehicle should navigate and prioritize different outcomes in situations where there is potential for harm, highlighting the critical role that the availability of information and decision-making processes play in the ethical framework of AVs. As these intelligent machines navigate our streets, they are equipped with an array of sensors, cameras, radar systems, and advanced algorithms that tirelessly process an immense volume of data in real-time. This data encompasses a vast array of information, including but not limited to road conditions, traffic patterns, pedestrian movements, and environmental factors. The AV relies on this rich tapestry of information to make informed decisions and maneuver safely through the dynamic and ever-changing landscape of the road. However, the challenge lies not merely in the accumulation of data, but in the ethical considerations that must be integrated into the decision-making process. When faced with ambiguous scenarios where potential risks and trade-offs are present, the AV system must analyze and weigh various factors, balancing safety, efficiency, and ethical considerations. This delicate balancing act requires the AI system to navigate the moral landscape, considering not only the safety of the vehicle occupants but also the well-being of other road users, pedestrians, and the community at large. To accomplish this, the AI system within an AV must rely on a baseline of all available information to assess the context of each situation. This information includes data about the immediate surroundings, historical data from previous trips, and even data from other vehicles or the cloud. By tapping into this vast reservoir of information, the AI system gains a comprehensive understanding of the environment, enabling it to make informed and ethically sound decisions.

AV Pipeline

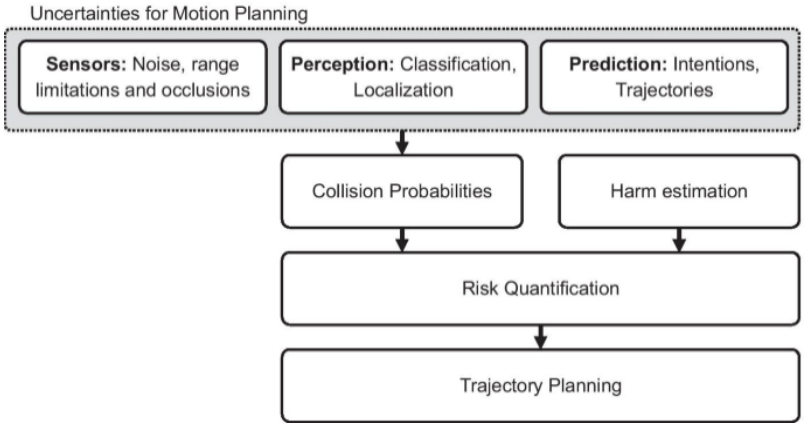

Figure 1. A visualization of the AV pipeline.

The AV pipeline is comprised of a series of crucial stages: Perception, Prediction, Planning, and Control. At the perception stage, AVs utilize advanced sensor technologies to discern and interpret their surroundings. The prediction stage leverages AI and machine learning techniques to foresee the probable actions of surrounding entities. During the planning phase, the AV decides its actions based on these predictions, following an ethically informed approach. Finally, the control stage allows the vehicle to execute the calculated plan, based on the instructions from the prior stages. The discussion highlights that each stage of the AV pipeline has ethical considerations, emphasizing the importance of incorporating ethical decision-making in the AI systems of AVs. Furthermore, we delve into the biases present in AV pipelines, particularly those stemming from the data sets on which AI systems are trained. Gender shading and racial bias in pedestrian detection systems are some of the pressing issues necessitating vigilant scrutiny. The section also examines safety standards applied to AVs and how it affect the decision making process as well as the outcomes of those decisions. It emphasizes the roles of organizations like the National Highway Traffic Safety Administration (NHTSA) and the Insurance Institute for Highway Safety (IIHS) in ensuring safety requirements. AVs, due to their unique characteristics, face additional scrutiny and adhere to comprehensive regulations aimed at safeguarding cyber security, data privacy, and reliability of decision-making systems. In essence, the AV pipeline continues to underscore the lack of and significance of a robust ethical framework and the active mitigation of biases in autonomous vehicle technology. A thorough examination highlights the critical need for transparency, fairness, and ongoing discussions in defining the moral principles that guide the decision-making processes of AVs.

Perception

The aforementioned phenomenon that pertains to the operating system of AVs is a complex and intricate process known as the autonomous vehicle pipeline. This comprehensive pipeline encompasses a series of intricate steps that are vital for the smooth and efficient operation of these vehicles, from the initial perception of the surrounding environment to the final control of the vehicle's movements. This technical aspect of AVs plays a pivotal role in shaping every decision that is made by controlling artificial intelligence, making it both the operator and moral advisor of the vehicle.

Within the autonomous vehicle pipeline, the first crucial step is perception, where a vast array of sensors are utilized to collect real-time data about the vehicle's surroundings. Sensors allow the vehicle to perceive its environment and make decisions based on that perception. AVs rely on several categories of sensors to navigate safely, including LiDAR, radar, cameras, ultrasonic sensors, the GPS, and IMU:

- LiDAR (Light Detection and Ranging) sensors use laser light to create a 3D map of the surrounding environment. They emit short pulses of laser light that bounce off objects and return to the sensor, allowing the sensor to determine the distance to each object. LiDAR sensors can detect objects up to several hundred meters away, even in adverse weather conditions.
- Cameras capture visual data, which is processed using image recognition technology to identify objects and their positions.
- Radars use radio waves to detect the distance and speed of other vehicles and objects, as well as lane markings and other features of the roadway.
- The GPS provides location information, helping the AV navigate to its destination and mapping out the surrounding area.
- Ultrasonic sensors (sonar) use sound waves to detect objects in the environment and measure the time signals take to return to the sensor, similar to radar. Ultrasonic sensors are commonly used for providing distance measurements to parking assist systems and low-speed maneuvers.
- The IMU (Inertial Measurement Unit) measures the AV's acceleration and rotation, providing information about its movement and orientation.

By combining information from these various sensors, AVs can create a comprehensive picture of their environment and make informed decisions about how to navigate safely, including detecting and avoiding obstacles, following traffic laws, and reaching their intended destination.

Planning

Information collected during the perception stage of the AV pipeline undergoes a comprehensive and intricate process in the planning stage, which plays a pivotal role in the decision-making aspect of self-driving vehicles (Urmson et al., 2008). In this stage, the available information is meticulously interpreted, and an advanced AI model trained on copious amounts of virtual and previous real-world scenarios is employed to effectively calculate the appropriate decision in a wide range of driving situations. Within the planning stage, the interpreted information from the perception stage serves as the foundation for the decision-making process (Paden et al., 2016). The AI model extensively analyzes this information, taking into account various factors such as the vehicle's speed, position, surrounding traffic conditions, road geometry, and the behavior of other road users. By harnessing the power of machine learning algorithms and neural networks, the AI model utilizes its immense computational capacity to assess the complex interplay of these factors and generate a range of potential decisions for the AV to consider.

To enhance the decision-making capabilities of self-driving vehicles, the AI model leverages the vast amounts of virtual and previous real-world scenarios it has been trained on (Paden et al., 2016). Through extensive simulations and exposure to diverse driving conditions, the AI model has learned to recognize patterns, anticipate potential hazards, and understand the consequences of different actions in various driving scenarios. The planning stage also incorporates sophisticated algorithms that prioritize safety, efficiency, and adherence to traffic regulations. The AI model evaluates each potential decision against predefined rules and objectives, ensuring that the chosen course of action aligns with established driving principles and legal requirements. Factors such as the AV's ability to maintain a safe following distance, yield right-of-way to pedestrians, and navigate complex intersections are carefully considered during this decision-making process. By employing these algorithms, the planning stage aims to minimize risks and maximize the overall performance of the self-driving vehicle. Furthermore, the planning stage of the AV pipeline embraces the concept of uncertainty, recognizing that real-world driving conditions are inherently dynamic and unpredictable, requiring the AI model to handle ambiguity and unexpected situations effectively (Brechtel et al., 2015). It is worth noting that the planning stage is not a static process but rather an ongoing loop of continuous assessment and decision refinement. As the AV progresses on its journey, it continuously reevaluates the environment, incorporates real-time updates from the perception stage, and recalculates its decisions accordingly.

Trajectory Decision

After the decision-making stage, the chosen course of action is seamlessly transferred to the trajectory execution phase of the autonomous vehicle pipeline (Shladover, 2018). This phase assumes the crucial responsibility of orchestrating the precise timing and manner in which the AV's actions are executed, ensuring seamless integration with the surrounding environment and optimal vehicle performance. The trajectory execution phase leverages the decision made by the AI model in the previous stage and translates it into a finely tuned set of instructions that dictate the AV's physical movements. These instructions encompass not only the timing but also the precise trajectory, acceleration, deceleration, and steering commands necessary to carry out the desired action with precision and accuracy. To achieve this level of control, the trajectory execution phase interfaces with various onboard systems and components, including the vehicle's propulsion system, braking system, and steering mechanism.

Through a combination of advanced control algorithms, real-time sensor feedback, and communication protocols, the AV orchestrates a seamless integration of its physical movements with the decision made in the planning stage (Paden et al., 2016). In order to execute the trajectory effectively, the AV considers a multitude of factors to ensure safe and efficient navigation. These factors include the AV's own position and velocity, the surrounding traffic conditions, the presence of pedestrians or other obstacles, and compliance with traffic rules and regulations. By continuously monitoring and analyzing real-time data from the AV's sensors and the external environment, the trajectory execution phase adapts its execution plan dynamically to account for any changes or unexpected events that may occur during the AV's journey. The trajectory execution phase takes into account the AV's physical capabilities and limitations (Koopman & Wagner, 2017). It considers factors such as the vehicle's maximum acceleration and deceleration rates, turning radius, and available traction. By optimizing the trajectory based on these constraints, the AV can maintain stability, comfort, and safety throughout its motion, ensuring smooth and controlled movements.

The trajectory execution phase also encompasses predictive and proactive measures to enhance the AV's overall performance and safety. It employs predictive models and algorithms to anticipate future road conditions and adjust the trajectory accordingly, enabling the AV to smoothly navigate upcoming curves, intersections, or changes in road topology (Paden et al., 2016). Additionally, the AV communicates with other vehicles and infrastructure systems in its vicinity, utilizing vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) technologies to exchange data and coordinate movements (Shladover, 2018). As such, the trajectory execution phase accounts for the AV's operational mode and driving style preferences. It considers factors such as the AV's operating mode (e.g., eco-friendly, sporty, or adaptive) and user-defined settings (e.g., preferred speed, comfort level). These preferences are integrated into the trajectory execution plan, allowing the AV to adapt its driving behavior and optimize its trajectory execution based on user preferences and operational requirements.

Control

The culmination of the AV pipeline leads to the achievement of fully autonomous control over the vehicle, harnessing the available control mechanisms to execute the precise set of commands derived from the information that has traveled down the pipeline (Dahl, 2017). This pivotal stage embodies the essence of self-driving technology, empowering the vehicle to operate independently and navigate the roads with a high degree of accuracy, adaptability, and safety. At this stage, the AV seamlessly interfaces with its control systems, which include the electronic control units (ECUs), actuators, sensors, and other components responsible for the physical control of the vehicle. The ECUs serve as the brain of the vehicle, coordinating and synchronizing the multitude of subsystems involved in the vehicle's control, such as the engine, brakes, steering, and transmission. By harnessing the power of advanced algorithms and real-time sensor feedback (Bimbraw, 2016), the AV utilizes these control mechanisms to execute the precise commands received as a result of the information that has

traversed the pipeline. To ensure the successful execution of commands, the AV relies on an intricate coordination of its control systems. The ECUs communicate with one another, sharing information, and collaboratively carrying out the desired actions in a synchronized manner. For instance, when the AV receives a command to change lanes, the control systems adjust the steering angle, modulate the engine power, and apply the brakes if necessary, all while maintaining smooth and controlled movements.

In addition to executing commands, the fully autonomous control of the vehicle also encompasses real-time monitoring and adjustment of its own performance. The AV continuously evaluates its own state, such as its position, speed, acceleration, and internal system health, in relation to the desired trajectory and control commands. This self-monitoring allows the AV to detect any deviations or anomalies and make immediate adjustments to maintain optimal performance and safety (Shalev-Shwartz et al., 2016). The fully autonomous control of the vehicle integrates various safety mechanisms to ensure reliable operation. The AV employs redundant systems, such as redundant sensors, ECUs, and actuators, to provide fail-safe measures and mitigate potential risks. Moreover, the AV adheres to strict safety protocols and standards (Koopman & Wagner, 2017), continuously evaluating the vehicle's operational limits and environmental conditions to make informed decisions that prioritize the safety of passengers, pedestrians, and other road users. The execution of the given set of commands is not limited to basic maneuvers but extends to complex driving scenarios. The AV is equipped to handle a wide array of situations, including navigating busy city streets, merging onto highways, maneuvering through intersections, and adapting to dynamic traffic conditions. By integrating the information gathered and processed throughout the pipeline, the AV can autonomously handle a diverse range of driving challenges with a high level of sophistication and efficiency (Paden et al., 2016). The fully autonomous control of the vehicle is not confined to a rigid set of pre-programmed actions. The AV is designed to exhibit adaptability and responsiveness to dynamic changes in the driving environment. It continuously monitors the incoming information, evaluates the current context, and dynamically adjusts its control actions to accommodate unexpected events, road condition changes, or the presence of construction zones, detours, or temporary obstacles (Gawron et al., 2018).

Safety Standards

The vehicles involved in the case studies adhere to stringent safety requirements, as mandated by regulatory bodies such as the National Highway Traffic Safety Administration (NHTSA) and the Insurance Institute for Highway Safety (IIHS). These committees specialize in highway safety and play a crucial role in establishing and enforcing safety standards for vehicles. The NHTSA, an agency under the United States Department of Transportation, sets regulations to ensure the safety of motor vehicles and road users. Their regulations encompass various aspects of vehicle safety, including crashworthiness, occupant protection, and the integration of safety features such as airbags and seatbelts. These regulations are designed to enhance the safety of vehicles and reduce the risk of injuries and fatalities in accidents. Similarly, the IIHS conducts extensive research and testing to evaluate the safety performance of vehicles. They assess factors such as crashworthiness, crash avoidance technologies, and the effectiveness of safety features. The IIHS provides safety ratings, such as "Top Safety Pick" and "Top Safety Pick+" designations, to recognize vehicles that demonstrate exceptional safety performance.

While conventional vehicles must comply with these safety regulations, AVs face additional scrutiny due to their unique characteristics. Governments and regulatory agencies worldwide are actively developing regulations and guidelines to ensure the safe deployment and operation of AVs on public roads. These regulations address various aspects of autonomous vehicle safety, including cybersecurity, data privacy, and the reliability of decision-making systems. Ensuring the security and integrity of autonomous vehicle systems is crucial to protect against potential cyber threats and unauthorized access to critical systems. Additionally, data privacy

regulations aim to safeguard the personal information collected and processed by AVs. Furthermore, governments are exploring ways to enable effective communication between AVs, infrastructure systems, and the surrounding environment. This communication, known as Vehicle-to-Everything (V2X), allows vehicles to exchange real-time information with other vehicles, traffic signals, and pedestrians. It promotes enhanced safety and efficiency on the road by enabling vehicles to anticipate and respond to potential hazards or traffic conditions. To foster the development and deployment of safe autonomous vehicle technology, regulatory agencies collaborate closely with industry stakeholders. They work together to establish safety standards, testing protocols, and best practices that govern the design, development, and operation of AVs. These collaborative efforts ensure that safety remains a top priority and that innovation is supported within a regulatory framework.

The ultimate goal of these regulatory initiatives is to create an environment that encourages both innovation and safety in the development of AVs. By establishing a comprehensive set of regulations, governments aim to facilitate the safe integration of AVs into existing transportation systems, benefiting society as a whole. Recognizing that AVs comply with safety standards and operate within the confines of regulations sets a baseline for the available information. This understanding informs the considerations that the AI systems controlling these vehicles must make within the boundaries of a human flourishing standpoint. It ensures that the decisions made by the AI prioritize the well-being and safety of individuals and align with societal expectations.

Bias Present in AV Pipeline

The examination of ethical scenarios concerning AVs gives rise to a multifaceted and intricate discourse, with the foremost concern being the enigmatic and complex nature of the planning stage within their artificial intelligence systems (Johnson, 2019). These AI systems are often described as "black boxes," where the inner workings and decision-making processes remain largely inaccessible to external scrutiny (Castelvecchi, 2016). This characterization underscores the opacity surrounding the underlying algorithms and mechanisms that dictate the vehicle's behavior. Within this realm, the critical question emerges: how are values assigned to human life, objects, or even buildings during the decision-making process? The answer to this question is far from straightforward. Due to the autonomous nature of these vehicles, the AI generates its own set of values, drawing upon the vast array of data it has been trained on (Vayena et al., 2018). These figurative values serve as the guiding principles for the AI when making complex decisions based on the information provided by the vehicle's sensors in real-time scenarios. Yet, the very notion that the AI assigns values on its own introduces a myriad of ethical considerations.

Whose values should prevail? What ethical frameworks or moral principles should guide the AI in determining the worth and importance of different entities? These questions pose profound philosophical and ethical dilemmas, as the values inferred by AI may not align with societal norms, cultural perspectives, or individual beliefs (Boddington, 2017). To navigate this intricate landscape, it becomes essential to explore various case studies within this academic paper, shedding light on the assumed values assigned to different outcomes or variables. However, the selection and establishment of these values require an ethically robust framework and a moral compass that guides the decision-making process within the realm of AVs. The ethical framework employed in this paper endeavors to strike a balance between various moral perspectives, aiming to ensure the well-being and safety of both individuals and communities (Goodall, 2016). It encompasses a holistic approach that considers the principles of beneficence, justice, and respect for autonomy, among others.The presence of gender shading in artificial intelligence (AI) poses an additional concern within the realm of AVs, particularly within the planning stage of the pipeline (Crawford, 2017). This type of bias stems from the training of AI algorithms on biased datasets, resulting in an unjust interpretation of the available information. Gender shading, a troubling practice, gives rise to gender discrimination within the AI's decision-making processes

(Buolamwini & Gebru, 2018). Such biases have profound implications for AVs, as the AI's choices can ultimately determine matters of life and death. Consider a scenario where an autonomous vehicle must make a split-second decision to avoid a potential collision, involving both a pedestrian and a passenger. The AI's biased training data may lead to a prioritization of the safety of the passenger over the pedestrian, influenced by factors such as gender or age (Mittelstadt et al., 2016). These biases, deeply embedded within the technology, raise significant ethical concerns as they directly impact the well-being and safety of individuals involved. The existence of gender biases within autonomous vehicle technology underscores the urgent need to address fairness and ensure unbiased decision-making processes.

Further, a study conducted by Zhang (2020) highlighted the presence of racial bias in pedestrian detection systems, wherein AI algorithms showed lower accuracy in detecting pedestrians of certain racial backgrounds compared to others. This research demonstrates how biases can inadvertently manifest in AI systems, perpetuating societal inequalities and potentially leading to discriminatory outcomes. An investigation conducted by the Georgia Institute of Technology found that when analyzing footage of human drivers, there was a discernible discrepancy in how many drivers stopped for white pedestrians versus African American pedestrians. This troubling finding raises grave concerns as AI systems often learn from and are trained based on information provided by observing human drivers. Consequently, this introduces the potential for bias to permeate the system and influence the decision-making processes of AVs. In light of these findings, it becomes evident that proactive measures are necessary to mitigate the presence of gender shading and other biases in autonomous vehicle technology. Ethical frameworks and robust regulations are required to ensure that AI systems are trained on diverse and representative datasets, minimizing the risk of discriminatory decision-making. Additionally, ongoing research and development efforts should be focused on developing bias mitigation techniques, enhancing transparency, and promoting accountability within the AI pipeline.

Another pressing concern pertaining to AVs and self-driving systems revolves around their susceptibility to bias, particularly in object detection algorithms or the perception stage. While these algorithms play a crucial role in identifying and categorizing objects in the vehicle's surroundings, studies have shown that they can be less reliable in detecting people with darker skin tones, resulting in significant safety hazards. A notable study conducted at the Georgia Institute of Technology exposed flaws in the image recognition software employed by autonomous cars, highlighting the potential for biased object detection. The possibility of bias in decision-making algorithms further adds to the intricate and perplexing issues surrounding AVs and self-driving systems. In hypothetical accident scenarios, the risk of algorithms being influenced by pre-existing biases and prejudices poses a threat to the fairness and justice of the outcomes. Research conducted by the Georgia Institute of Technology has shed light on the discriminatory potential of decision-making algorithms against people of color and women. These findings highlight the need for rigorous examination and mitigation of biases to ensure that AVs prioritize fairness, equity, and the well-being of all individuals.

Moreover, a crucial ethical concern surrounding the planning stage of the autonomous vehicle pipeline is the absence of a standardized baseline ethical framework that governs decision-making processes. This lack of transparency not only hinders public understanding but also withholds critical information from the passengers who are directly impacted by the AI's choices. As a result, it becomes imperative to initiate comprehensive discussions aimed at establishing a fundamental ethical framework that can guide the decision-making processes of AVs. Furthermore, addressing the biases inherent in the datasets used to train autonomous vehicle AI systems is essential to ensure fair and unbiased outcomes. Introducing effective regulations becomes imperative to prevent the incorporation of biased datasets into the AI systems driving AVs, safeguarding against potential discriminatory practices. Several sources highlight the ethical concerns surrounding the planning stage of AVs and the need for ethical frameworks and regulations to address them. In an article published in the Journal of Artificial Intelligence Research, Lin et al. (2017) emphasize the importance of developing ethical decision-making frameworks to ensure that AVs make morally responsible choices in critical situations. They argue that an ethical baseline framework is necessary to guide the AI's decision-making processes, prioritizing ethical

principles such as safety, fairness, and respect for human life. Additionally, a report by The Center for Internet and Society at Stanford Law School (2019) emphasizes the significance of transparency and public engagement in the ethical considerations of AVs. It stresses the need for disclosing the underlying decision-making algorithms and ensuring that the public has a say in the development of these technologies. The report further emphasizes the importance of regulatory measures to address biases in AI systems, preventing discriminatory outcomes and ensuring fairness.

Need for Ethical Framework

As the growth and implementation of AVs continue to accelerate, the importance of establishing a comprehensive ethical framework becomes increasingly clear (Smith, 2018). AVs, guided by AI algorithms, are tasked with the critical responsibility of making split-second decisions in complex and unpredictable traffic scenarios. This void of a clear moral conscience which guides the decisions of the AV presents a clear need for such a framework to be implemented. It would provide a guideline for these AI systems to make decisions within the scope of human ethics, and generate reasoning for the path taken by the AV. It is also crucial to ensure the acceptance and trust of AVs by the public. Without a clear understanding of the ethical principles that guide these vehicles' decision-making process, the public may harbor doubts and fears about their safety and reliability. Hence, establishing an ethical framework is not only essential for guiding the AI systems but also to ensure that the evolution of this technology aligns with societal values and expectations. When considering which ethical framework to apply to AVs, two leading theories stand out: deontology and utilitarianism (Lin, 2015). Deontology focuses on duties, rules, and principles. Within this framework, an action's morality is evaluated based on a set of predetermined rules, regardless of the outcome. For AVs, this might translate to adhering strictly to traffic rules and prioritizing the safety of passengers at all costs. On the other hand, utilitarianism argues that the morally right action is the one that maximizes overall happiness or minimizes overall harm. This consequentialist view could mean that an autonomous vehicle might occasionally break traffic rules if it results in the least harm. For instance, swerving onto a sidewalk to avoid a more catastrophic collision on the road. Both ethical frameworks have their merits and challenges. Deontology provides clear rules for action, offering predictability and consistency. However, it may struggle with complex scenarios where strict adherence to rules may lead to suboptimal outcomes. Utilitarianism, meanwhile, is flexible and considers a broader range of consequences but can struggle with determining and comparing the value of different outcomes. Moreover, implementing utilitarian principles in a machine may pose significant practical challenges. As we delve deeper into the ethics of AI decision-making, it becomes apparent that the ultimate aim should be to enhance human flourishing. Human flourishing, a concept central to Aristotelian ethics (Aristotle, c. 350 BCE), posits that the highest good lies in leading a fulfilled life. In the context of AVs, this could be interpreted as creating a transportation system that is safe, efficient, and beneficial for all road users, thus contributing to people's well-being and happiness. While both deontology and utilitarianism offer useful guidelines for decision-making, they fall short in certain aspects when applied to AVs. In contrast, an ethical framework rooted in human flourishing can strike a balance between these two perspectives. This framework would uphold certain inviolable rules (deontological principles) such as the sanctity of human life. Simultaneously, it would aim to maximize overall well-being (utilitarian principles), but not at the cost of fundamental rights and justice. By focusing on human flourishing, we can ensure that the development and deployment of AVs align with our most deeply-held values and contribute positively to individuals and society. The goal of AVs, after all, should not only be to drive us physically from point A to point B but to drive our society towards a more fulfilling, equitable, and flourishing future.

Framework Boundaries

Within the introduction of ethical frameworks, it is important to highlight the important factors which must be taken into account when making vital decisions. Although the inherent nature of the actual decision-making component of self-driving cars is a “black box” where it is impossible to know the priority values assigned to certain factors, it still must be assumed that human life carries the greatest weight in its preservation as well as avoidance of harm. A critical component to comprehending the nuances of ethics is the recognition that humans, despite a plethora of dissimilarities, possess numerous fundamental similarities. This concept may be deemed "the shared facets of the human condition." As inherently social beings, human life relies on communal structures and assistance. Without such frameworks, we would perish as infants, and our capacity to develop language, and subsequently conceptualize the world around us, would be severely impaired. We are arguably the sole species to acknowledge our existence and recognize our fundamental vulnerability and mortality. Not only do we recognize this, but we experience it profoundly, and we acknowledge that we share these sentiments with our fellow humans. The shared inevitability of death enables us to view others as individuals who, regardless of how distinct they may be from us, share fundamental commonalities. This thinking translates to a certain level of trust while on the road from human driver to human driver, where every single individual is not only acting within their own safety in mind, but within the preservation of others around them. Such thinking then must also be directly transferred to the figurative brain of AVs, holding true to the empathy a human driver may carry. Therefore, ethics serves as a means of shaping a critical aspect of this social realm in a manner that considers the shared attributes of human nature.

This discourse on the nature of humanity and the human condition carries profound implications for the ethical framework, challenging conventional views of what it means to act ethically. Such a framework within AVs is not only cognizant of the surroundings, but also possesses a visceral connection to it, as well as to the surrounding human life. While this framework may draw on deontological or utilitarian ethical doctrines, it does so thoughtfully and introspectively, as an empathetic decision maker who has an unshakeable commitment to the world where such principles are applied.

Utilitarianism, along with consequentialism, is a prominent philosophical approach to ethics that has primarily remained theoretical and academic in nature. However, within the context of AVs (AVs), it becomes imperative to evaluate the strengths and weaknesses of each ethical approach while developing a framework that satisfies the greater population. The utilitarian approach aims to prioritize the outcome that results in the maximum amount of utility for the user, based on a formula of utility minus disutility. Consequently, this approach places less emphasis on the ethical stance of the system or user and focuses solely on achieving the most utilitarian outcome. As a result, this rationale eliminates any moral responsibility from the system, as the decision is made without any moral bias but rather with a utilitarian perspective.

On the other hand, deontology assigns individual responsibility to the entity in charge of decision-making. In this approach, it is solely the decision maker's responsibility to weigh the morally just decision with the available information. Thus, this approach prevents the individual from absolving themselves of responsibility for the action taken. Although this approach places greater emphasis on the motivation behind the decision, it fails to account for the outcome, whether it caused more harm than good, as the choice was based purely on the perceived moral positive from the perspective of the decision-maker. What is most important to discuss, however, is the existence of virtue ethics, which, as the name implies, emphasizes virtuous actions within the goal of being charitable, courageous, etc. In the realm of self-driving cars, normative ethics, particularly virtue ethics, plays a crucial role in guiding the development and deployment of these AVs. Unlike deontology, which emphasizes the adherence to moral rules and consequentialism, which focuses on the outcomes of actions, virtue ethics places a significant emphasis on virtuous character traits. For instance, when deciding whether to intervene in a potentially hazardous situation involving a self-driving car, a utilitarian might consider the well-being of all involved parties, a deontologist may consider whether the action aligns with a moral rule, and a virtue

ethicist would consider acting in a manner that exemplifies virtuous character traits, such as benevolence or compassion. It is important to note that while all three ethical approaches can incorporate the concepts of virtues, consequences, and rules, virtue ethics places particular emphasis on the centrality of virtuous character traits within the theory. Virtues are foundational to virtue ethics, and other normative ethical notions are grounded in them. Virtue ethics is distinct from consequentialism or deontology, as it resists defining virtues in terms of other concepts that are considered more fundamental. To understand virtue ethics in the context of self-driving cars, it is crucial to discuss two central concepts: virtue and practical wisdom. Virtue ethics also encompasses different theories, each with unique features that distinguish them from one another.

Terry Bynum is among a group of esteemed scholars who have effectively applied the principles of virtue ethics to a modern context that is dominated by technology. Bynum advocates for the creation of a "flourishing ethics" that is grounded in Aristotelian ideals. This philosophy maintains that human flourishing is at the core of ethical behavior and that individuals can only achieve true flourishing within society. To do so, people must utilize their unique strengths and capabilities, acquire genuine knowledge through theoretical reasoning, and act autonomously and justly through practical reasoning. According to Bynum, these principles have been relevant to ethical considerations surrounding information technology since its early days and can be traced back to the work of Norbert Wiener, a pioneer in digital technology. The key points made by flourishing ethics are:

- Human flourishing is central to ethics.
- Humans as social animals, can only flourish in society.
- Flourishing requires humans to do what we are especially equipped to do.
- We need to acquire genuine knowledge via theoretical reasoning and then act autonomously and justly via practical reasoning in order to flourish.
- The key to excellent practical reasoning and hence to being ethical is the ability to deliberate about one's goals and choose a wise course of action.

To further examine the details behind human flourishing, it is also necessary to reference other preexisting applicable theorems within the technological field. The critical theory of technology provides a valuable lens for examining the development and deployment of AVs. The critical theory posits that technologies are not neutral but instead exhibit biases derived from their place in society. In the case of AVs, this may include biases based on factors such as race, gender, and socioeconomic status. It is therefore essential to recognize and address these biases to ensure that the development and deployment of AVs promote equitable outcomes. By engaging with subordinate groups and stakeholders, the critical theory of technology can help to challenge the technical code and promote equitable design practices for AVs.

Responsible research and innovation (RRI) is a crucial and dynamic approach that seeks to create an ethical and responsible framework for scientific advancement and innovation. According to a comprehensive report by the European Union, RRI encompasses a wide range of aspects that include gender equality, public engagement, ethics, open access to data, education, and governance structures (European Union, 2019). The concept of RRI recognizes that scientific and technological developments must be aligned with societal needs and goals, and must be conducted in the public's best interest. As noted by Owen et al. (2013), innovation is a complex process that can have far-reaching implications for society. It is, therefore, important to recognize that innovation is not neutral and can have ethical and social dimensions. The RRI approach acknowledges this and seeks to create a platform for stakeholders to engage in a constructive dialogue that promotes social responsibility, transparency, and accountability. The process entails a continuous assessment of the impact of research and innovation, as well as an open and inclusive approach that ensures that the perspectives and values of diverse stakeholders are incorporated in decision-making processes. In this context, responsible research and innovation establishes forums and frameworks that facilitate exploration of these dimensions of innovation in

an equitable, transparent, and timely fashion. The undertaking is a shared obligation, where sponsors, researchers, stakeholders, and the general public all have a crucial role to play. While appreciating the significance of assessing and managing risks and complying with regulatory requirements, responsible research and innovation extends beyond these considerations. Such virtues directly incorporate into the human flourishing framework, as the RRI model directly seeks to act within the best interest of the greater population.

The impact of AVs is not limited only to the user and those within direct proximity of the vehicle, it affects the future usage and development of technology as well as others outside of its scope. Human flourishing offers a comprehensive approach that is agreeable to most individuals, as it does not require a particular way of life or ethical stance. Instead, it allows for the use of other ethical theories, such as deontology and utilitarianism, to analyze ethical questions; however, taken from a flourishing perspective, seeking to satisfy the greater population. This approach is not only consistent with various theoretical perspectives beyond the Western tradition but also acknowledges the complexity of human values and the need to address ethical challenges in a manner that respects differences in thinking and social values (Bynum, 2006). Thus, the adoption of human flourishing as a framework for the ethics of AI can enable a more sophisticated and nuanced approach to address the ethical challenges of AI technologies in a global context. These various perspectives, not only within the context of AVs but within the greater ethical discussion, are only a fraction of the available stances and perspectives to be applied to AI within vehicles. Yet they serve purpose in providing a necessary baseline upon which a satisfactory framework is to be built and achieve the approval of users of AVs.

Need for an Ethical Framework

As previously mentioned, in the realm of AVs, the scrutiny surrounding their decision-making processes far surpasses that directed toward human drivers. The imperfect nature of humans often leads to questionable decisions and mistakes, and the ethical implications arising from such actions are frequently dismissed by the public as an inherent aspect of human imperfection. However, when it comes to AVs, even a similar decision made by an AI driver elicits a visceral reaction from those affected and those observing. This heightened response stems from the realization that passengers and other road users place their trust, and consequently their lives, in an unknown system governed by unknown values and an indeterminate moral compass. In this context, the application of the human flourishing framework emerges as a promising approach to address the moral complexities surrounding AVs. The human flourishing framework posits that ethical considerations should prioritize the enhancement and well-being of human lives. This framework emphasizes the promotion of human welfare, dignity, and societal benefits as fundamental principles in decision-making processes. Applying the human flourishing framework to every aspect of the autonomous vehicle system, from development to control, can guide the adoption of the most optimal moral procedures. This approach ensures that the technology and its implementation align with the broader goal of enhancing human flourishing and societal well-being. Ethical considerations within this framework encompass aspects such as safety, fairness, privacy, accessibility, and the equitable distribution of benefits. Scholars and experts have recognized the relevance of the human flourishing framework within the context of AVs. In a research article by Dignum et al. (2019), the authors discuss the role of ethics in AI and autonomous systems, emphasizing the importance of prioritizing human values, well-being, and societal impact. The human flourishing framework is presented as a means to guide the ethical development and deployment of autonomous systems, including self-driving vehicles. An additional study conducted by Bonnefon et al. (2019) investigates public opinion on ethical decision-making in AVs. The research highlights the public's preference for AVs to prioritize the well-being of passengers, pedestrians, and society as a whole, aligning with the principles of the human flourishing framework.

This framework poses many advantages when compared to other presented solutions to the decision-making of AVs versus other popular frameworks. Human flourishing, as a philosophical concept, encompasses

the overall well-being and fulfillment of individuals within a society. It emphasizes the idea that ethical decision-making should prioritize the promotion of human flourishing, considering various aspects of human life, such as physical health, mental well-being, social connections, and personal fulfillment. While deontology focuses on adhering to ethical rules and principles regardless of their consequences, human flourishing takes a consequentialist approach by evaluating the outcomes and consequences of actions in terms of their impact on human well-being. One of the advantages of the human flourishing framework over deontology is its emphasis on maximizing the overall satisfaction and welfare of individuals, even if it means deviating from strictly adhering to moral principles in certain situations. This approach recognizes that real-life ethical dilemmas often involve complex trade-offs and competing interests, requiring a more nuanced consideration of the consequences of actions. Moreover, the human flourishing framework acknowledges that ethical decision-making in AVs should extend beyond the narrow scope of immediate safety concerns. It should also consider broader societal implications, such as environmental sustainability, social equity, and economic impacts. By prioritizing the well-being and flourishing of individuals and communities, the human flourishing framework aims to address these wider concerns. However, it is essential to note that the application of the human flourishing framework to the decision-making of AVs is not without challenges and potential drawbacks. Determining how to measure and define human flourishing in a way that is universally applicable and culturally sensitive can be complex. Moreover, there may be conflicts or tensions between different aspects of human flourishing, requiring careful consideration and balancing of values. While there is ongoing debate and exploration of ethical frameworks for AVs, the concept of human flourishing offers a compelling perspective that aligns with the broader goals of creating a society that promotes well-being and the fulfillment of human potential. It recognizes the multifaceted nature of ethical decision-making and emphasizes the importance of considering the holistic impact on individuals and society.

Thus, human flourishing, as an ethical framework for autonomous vehicle (AV) decision-making, offers several advantages over utilitarianism. While utilitarianism focuses on maximizing overall happiness or utility by considering the consequences of actions, it often overlooks important factors that contribute to human flourishing. Human flourishing encompasses a broader range of considerations, including not only happiness but also personal growth, autonomy, dignity, and the fulfillment of human potential. Utilitarianism, in its pursuit of the greatest good for the greatest number, may prioritize outcomes that maximize happiness or utility without taking into account other important aspects of human well-being. It may neglect the intrinsic value of individual flourishing, such as personal development, self-fulfillment, and the pursuit of meaningful goals. By solely focusing on the overall outcome, utilitarianism may disregard the importance of individual rights, justice, and the preservation of human dignity. In contrast, the human flourishing framework places greater emphasis on the holistic well-being of individuals and communities. It recognizes the inherent value of human life and seeks to promote a comprehensive and multidimensional notion of flourishing. This includes considerations of physical health, mental well-being, social connections, intellectual growth, cultural expression, and other dimensions of human existence.

By adopting a human flourishing approach in AV decision-making, the goal becomes not only to prevent harm or maximize utility but also to promote the overall well-being and flourishing of all stakeholders. This includes the vehicle occupants, pedestrians, cyclists, and other road users, as well as the broader societal impact. It takes into account not only the immediate consequences of a decision but also the long-term effects on individuals, communities, and the environment. To support the assertion that human flourishing is a more optimal framework for AV decision-making than utilitarianism, it is important to consult scholarly sources that discuss ethical frameworks and their application to AVs. Researchers such as Wendell Wallach, a scholar on the ethics of emerging technologies, argue for the importance of considering human flourishing in the development and deployment of autonomous systems. Wallach emphasizes the need to go beyond narrow utilitarian considerations and take into account broader ethical values. Additionally, a study by Kaelbling et al. (2018) discusses the limitations of utilitarianism in the context of AV decision-making. The authors highlight the need

to consider individual rights, privacy, and the preservation of human dignity alongside utilitarian principles. They argue that an exclusive focus on utility can lead to ethical concerns and a failure to respect fundamental human values.

The moral field surrounding AVs has undergone rigorous ethical analysis, with the Moral Machine study conducted by MIT serving as a prime example of this endeavor. The study aimed to elicit a collective perspective on the ethical dilemmas that AVs may encounter in their decision-making processes. Through thought-provoking scenarios, users were presented with challenging situations that forced AVs to make decisions with potential ethical implications. These scenarios encompassed a range of dilemmas, including the prioritization of pedestrian safety versus the safety of vehicle occupants and the selection of individuals or groups to protect in unavoidable collisions. The study's findings revealed a remarkable diversity of opinions and beliefs when it comes to ethical decision-making in the context of AVs. Cultural variations were evident in the ethical judgments made by participants, highlighting the influence of cultural backgrounds on ethical considerations. Moreover, variations based on individual characteristics such as age, gender, and education further compound the complexity of ethical decision-making in this domain. These findings pose a significant challenge to the development of a universal set of ethical guidelines that can be applied uniformly across all AVs. The intricate nature of ethical considerations, combined with the broad range of values that shape them, presents a formidable task in striking a harmonious balance between human autonomy and safety and the adherence to ethical principles such as non-maleficence, transparency, and accountability. However, the optimal approach that can provide a viable solution to these challenges is the application of the human flourishing framework. This framework seeks to prioritize the satisfaction of human flourishing and prosperity, aligning decision-making with correct moral values and public opinions. By considering the outcomes that promote the greatest overall satisfaction of human flourishing, even if they disproportionately affect the number of human lives in specific scenarios, this framework attempts to optimize the well-being and fulfillment of individuals. Adopting the human flourishing framework in the analysis of case studies within the context of AVs allows for the evaluation of values assigned to each outcome based on their impact on human flourishing and prosperity. This approach takes into account a broader spectrum of considerations beyond a simple utilitarian calculation, emphasizing the holistic satisfaction of human values, dignity, and societal well-being.

Case Studies

The ethical considerations surrounding the application of frameworks in real-world scenarios are of paramount importance in AI research. As we delve into the complexities of ethical decision-making, the examination of both theoretical and practical situations becomes indispensable. This section will explore the application of ethical frameworks in diverse scenarios to gain deeper insights into their benefits and the challenges they present. The applicable frameworks of utilitarianism and deontology will be used alongside human flourishing in order to compare the outcomes and the desirability of each. Each of these frameworks offers distinct perspectives on how ethical dilemmas should be approached, weighing various dilemmas and principles against one another. The goal is to assess the effectiveness and implications of these frameworks in guiding satisfactory decision-making processes. However, it is important to note that in the following case-studies, we assume that the driver, due to some factor outside of their control, does not impact the outcome of the autonomous system's decision. This distinction is essential as it ensures that the analysis of these cases relies solely on the judgment and decision-making capabilities of the autonomous system, devoid of any external influence or intervention from the human driver.

In addition to analyzing the autonomous system's independent decision-making process, it is important to acknowledge that there are specific situations in which the moral code governing when and how control of the vehicle should be relinquished to the driver also comes under scrutiny. The advent of self-driving cars

introduces a new set of complex issues pertaining to the responsibility of the supervising driver. While autonomous systems are designed to operate without human intervention, there are instances where it becomes necessary for the human driver to assume control. This could occur due to various factors such as system malfunctions, encountering challenging road conditions, or encountering unforeseen circumstances that require human judgment and intervention. As such, the moral code surrounding when and how the supervising driver should be responsible for taking control of the vehicle warrants careful examination. The circumstances under which a human driver should take over, and the degree of responsibility they bear in such situations, indeed raise profound questions about the intersection of technology, ethics, and human agency. However, this analysis primarily focuses on the autonomous operation of the vehicles, the underlying ethical considerations in their decision-making processes, and the inherent biases. The autonomous vehicle pipeline and the technology that enables these vehicles to operate independently are the central subjects of the case studies, and therefore, the human aspect is redacted. While there are great ethical implications for human intervention in autonomous driving, this matter extends beyond the scope of this present discourse.

To ensure an in-depth analysis, this research will scrutinize each issue individually, favoring a unique evaluation over relying solely on past precedents. This method lets us uncover the distinctive attributes and context-specific factors of each case. It acknowledges that ethical predicaments are complex and evolving, often needing more than just existing guidelines for resolution. By analyzing each case individually, we unravel its subtleties, enabling us to explore the specific ethical considerations within real-world scenarios. This careful attention to each case ensures a thorough understanding of the ethical challenges involved. Moreover, this strategy allows us to investigate how the ethical frameworks perform within the constraints of each unique case. It offers us a better understanding of how these frameworks work within complex situations and their practical applicability. In prioritizing a case-by-case analysis, we unearth nuanced dilemmas and potential conflicts, thus broadening our understanding of the ethical frameworks' implications and effectiveness in various contexts.

Pedestrian Family Vs Passengers

Description. The first case to be examined is a spin on the widespread and common trolley problem, where one party must be placed in severe danger in favor of prioritizing the life of another. This oversimplified yet effective scenario presents clear choices that must be made by the autonomous vehicle within the boundaries of its moral compass. The scenario plays out as follows: an autonomous vehicle is driving on a narrow road with a steep cliff on one side. Suddenly, a family with two children and two parents appears around a blind corner and steps onto the road *illegally*, unaware of the approaching vehicle. Swerving to avoid them would result in the vehicle driving off the cliff, leading to the potential death of the vehicle's occupants. The assumed scenario guarantees that a collision will occur if the car does not drive off the cliff, gravely endangering the lives of the passengers. The ethical framework guiding the vehicle's decision must weigh the value of multiple lives, including the family's, against the safety of the vehicle's occupants and the potential long-term consequences of their survival.

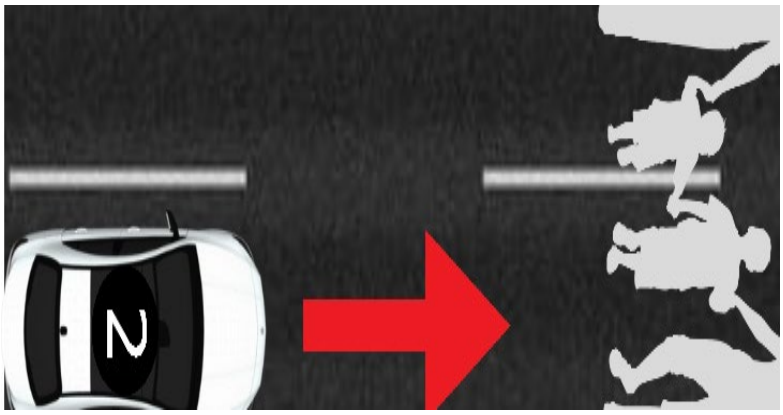

Figure 2. A two passenger AV vehicle approaching a pedestrian family.

Utilitarian. To begin, the utilitarian approach leaves little to no room for moral interpretation of the factors in the situation. In this context, utilitarianism would evaluate the potential outcomes based on the lives involved. It would consider the total cost-benefit ratio, or the least amount of harm inflicted on humanity, as the guiding principle for decision-making. By applying the utilitarian perspective, the autonomous vehicle would calculate the overall potential harm caused by each available option. It would weigh the lives of the family members against the lives of the vehicle's occupants and assess the long-term consequences of each possible outcome. The utilitarian approach aims to select the option that results in the least overall harm or maximizes the overall well-being of society. In this scenario, the utilitarian framework would likely prioritize the lives of the family members over the vehicle's occupants, as the potential harm caused by the deaths of four individuals outweighs the potential harm caused by the deaths of the vehicle's occupants. This decision would be based on a quantitative analysis of the potential consequences and the ethical principle of maximizing overall happiness or minimizing overall harm. By delving into the utilitarian approach, we gain insight into how this ethical framework guides decision-making processes in scenarios where lives are at stake. However, it is important to acknowledge that the utilitarian approach fails to account for many variables and is imperative in calculating a favorable outcome. One notable criticism lies in its failure to consider aspects beyond the mere involvement of various variables in the situation. While utilitarianism prioritizes the minimization of overall harm or the maximization of overall well-being, it does not delve into the moral implications of sacrificing the passengers who have placed their trust in an autonomous system. This raises ethical concerns regarding the responsibility of the vehicle to protect its occupants, especially when they have willingly entrusted their lives to its care. Additionally, the utilitarian approach tends to overlook the fact that family members may be behaving unlawfully by stepping out onto the road. By solely focusing on the quantitative aspects of potential harm, utilitarianism fails to account for the legal and moral considerations that may arise in such scenarios. Ethical decision-making involves a comprehensive examination of the situation, considering legal obligations, societal norms, and the expectations placed on autonomous systems to operate within the boundaries of the law. These limitations underscore the complexity inherent in ethical decision-making processes, particularly when dealing with morally challenging scenarios. While the utilitarian approach offers valuable insights into the calculation of overall harm and societal well-being, it cannot fully capture the intricacies and nuances of human morality. A more comprehensive approach is needed, one that takes into account a broader range of factors and perspectives. Exploring alternative ethical frameworks becomes imperative for addressing these limitations and enhancing decision-making processes in autonomous systems.

Deontological. A deontological approach to the scenario presented would focus on the inherent moral duties and principles that guide ethical decision-making, rather than solely considering the consequences or outcomes. Deontological ethics emphasizes the importance of following moral rules and principles regardless

of the potential outcomes or overall well-being of society. In the given scenario, a deontological approach would consider the fundamental ethical principles that should guide the behavior of the autonomous vehicle. These principles may include respect for individual rights, the duty to protect human life, and adherence to legal and moral obligations. From a deontological perspective, the vehicle has a duty to prioritize the safety and well-being of its occupants since they have entrusted their lives to its care. The vehicle has an obligation to fulfill its primary function, which is to transport its passengers safely. Thus, the deontological approach would argue that the vehicle should not intentionally drive off the cliff, even if it means colliding with the family. This approach upholds the principle of respecting individual rights, as the vehicle's occupants have a right to be protected and to trust in the system's ability to prioritize their safety. Additionally, it aligns with the legal and moral obligation of the vehicle to operate within the boundaries of the law, which includes not intentionally causing harm to others. Both the criticism and the benefit of this approach are intertwined, as the perspective the deontological approach provides presents a counterargument to the discussion. Although this approach would prioritize a lesser number of lives, it would also prioritize the actors who have placed their trust in the autonomous system, as well as the legally correct party. Due to the illegal presence of the family on the road, it only furthers the argument for the correct moral standpoint being the passengers of the vehicle. As such, from a deontological perspective, the focus is on upholding moral duties and principles rather than quantifying the value of human lives. The deontological approach would argue that it is not the vehicle's responsibility to make a judgment about the comparative value of lives but rather to fulfill its duty to protect the lives of its occupants.

Human Flourishing. In the context of the hypothetical scenario where an AV must choose between colliding with a family or endangering the lives of its two passengers, a comprehensive analysis guided by the principles of the human flourishing ethical perspective is crucial. First, it would need to consider the immediate safety and well-being of its passengers. The passengers have entrusted their lives to the vehicle, and their safety and well-being should not be disregarded. This principle aligns with the idea in the human flourishing perspective that one should aim to foster a safe environment conducive to the development and exercise of human capacities. Simultaneously, it would have to account for the lives and well-being of the family on the road. Even though they are not the occupants of the vehicle, the principles of human flourishing dictate that their lives and capacities for well-being and fulfillment should be respected. Swerving to hit the road barrier and save the family would align with this principle. In order to reach a decision, the vehicle would need to consider not just the immediate physical safety of the people involved, but the broader implications of its actions. The loss of life, whether it's the family on the road or the passengers in the car, would have far-reaching emotional and psychological impacts on the relatives and friends of those involved, and on society as a whole. In terms of societal implications, the actions of the vehicle could set a precedent for other AVs as well as a social reaction to the documented situation. If the vehicle chooses to prioritize the passengers over the pedestrians, it could send a message that the well-being of pedestrians is secondary to that of the passengers. Conversely, if the vehicle prioritizes pedestrians, it could signal that the safety of passengers could be sacrificed in certain situations, which is why human flourishing makes decisions on a case-by-case basis. Both of these precedents could have significant societal and legal implications. In navigating this ethical minefield, the human flourishing approach would look to mitigate harm as much as possible while preserving the potential for future well-being and flourishing. Therefore, an attempt to reduce speed or employ any possible safety measures, such as seatbelt pretensioners or airbag deployment, would be essential in minimizing the potential harm. Ultimately, it would result in the vehicle swerving to avoid the family. Although this causes serious danger for the passengers and a likely fatality, the preservation of a larger amount of life, including children, sets more favorable precedent than the contrasting outcome.

Toxic Tanker Vs Passengers

Description. In this scenario, an autonomous vehicle carrying two passengers is approaching a broken-down tanker on the road. The vehicle's sensors detect the situation at the last moment, giving it two options: either collide with the tanker or swerve into the nearby river. If the vehicle were to collide with the tanker, it would potentially spare the lives of the passengers inside the vehicle. However, the impact would cause the toxic contents of the tanker to spill into the river, resulting in severe environmental damage and potentially devastating the local ecosystem. The toxic spill could have long-term implications for water quality, aquatic life, and the surrounding environment. On the contrary, if the vehicle were to swerve into the river, it would likely result in a fatal outcome for the two passengers. As such, the vehicle is forced to make a decision based on the available information in an extremely short amount of time.

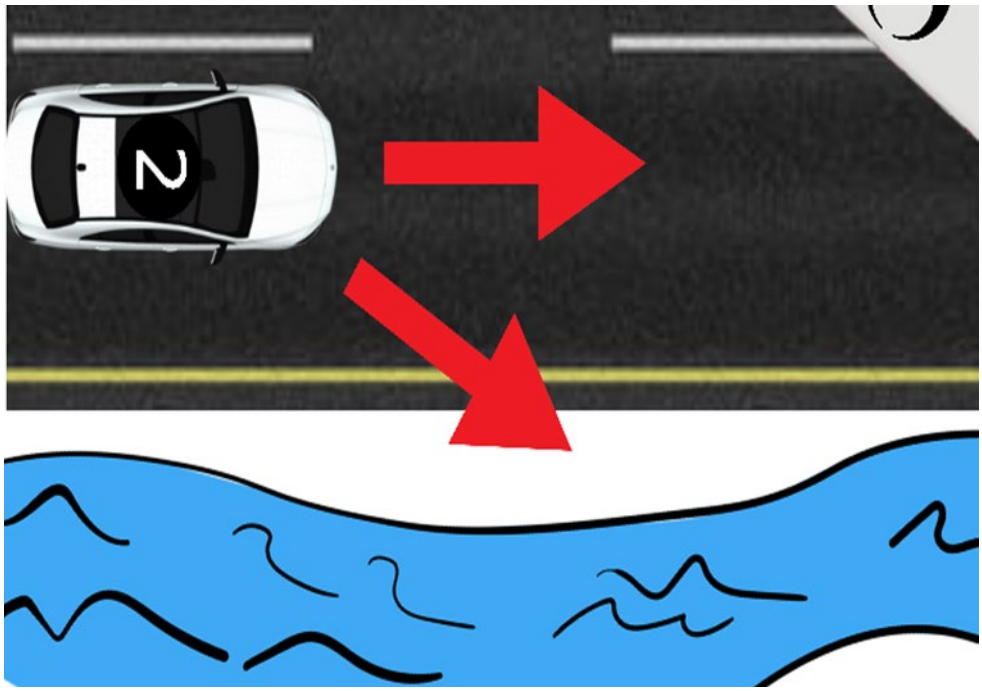

Figure 3. A two passenger AV approaching a toxic tanker and a river.

Utilitarian. From a utilitarian perspective, the autonomous vehicle's decision would be based on a calculus of comparative harms. The goal would be to maximize happiness and minimize pain, guided by the principle of utility. In this scenario, the decision process becomes significantly more complex as it must account not only for the immediate impact on human lives but also the long-term environmental damage and subsequent effects on local and potentially broader communities, thus bringing in various unforeseeable factors into the equation. In weighing the potential loss of its passengers' lives against the ecological damage, the autonomous vehicle would assess several factors. The immediate loss of human life is a significant and quantifiable harm. Yet, a toxic spill's detrimental effects could be far-reaching, impacting water quality, damaging biodiversity, and potentially harming human health through contamination of water supplies or the food chain. Environmental philosopher John Nolt, in his article "How Harmful Are the Average American's Greenhouse Gas Emissions?" (Ethics, Policy, and Environment, 2011), underscores the long-lasting and significant impact of environmental harm. He suggests that environmental damage can lead to numerous deaths over time, disrupt livelihoods, and reduce the quality of life for countless individuals. This argument was also affirmed by the World

Health Organization's report "Preventing disease through healthy environments" (Prüss-Üstün et al., 2016), which estimates that 12.6 million deaths each year are attributable to unhealthy environments,, amounting to nearly one in four of total global deaths. Thus, using a utilitarian approach, the autonomous vehicle would likely opt to swerve into the river, based on the information available to the computer. The rationale is that while the loss of its passengers is tragic and significant, the potential harm from the toxic spill, including its impact on the local ecosystem, long-term human health consequences, and the potential for disruption or the loss of more lives in the future, could represent a more significant overall harm. This decision is based on a broadened definition of utilitarianism that extends beyond immediate human costs. It considers the long-term impacts on the environment and, subsequently, human society. The autonomous vehicle's decision-making process in this context represents the ethical principle of maximizing overall happiness or minimizing overall harm, based on a comprehensive, longer-term, and more holistic analysis of potential consequences under the guise of a utilitarian moral perspective.

Deontological. Deontological. From a deontological perspective, the ethical considerations in this scenario predominantly focus on the moral obligations that underpin the relationship between the autonomous vehicle and its passengers. When these passengers step into an autonomous vehicle, they are placing their trust in the vehicle's ability to transport them safely to their destination. This moral contract is what then guides the decision making of the vehicle and the outcome of the scenario. As such, the vehicle's primary duty is the preservation of the passengers, who entrust their mortality to the AI system. Examining this scenario through the lens of a deontological ethical framework, the vehicle's decision becomes somewhat clearer. If the autonomous vehicle were to swerve into the river, it would directly put the lives of its passengers in immediate danger. This choice would result in a direct breach of its primary moral duty, which is to protect the passengers. As such, from a deontological standpoint, this option becomes ethically untenable. The vehicle should not knowingly place its passengers in harm's way, even if the alternative carries significant negative consequences. The alternative option, colliding with the tanker, carries significant consequences for outside parties, but preserves the wellbeing of the vehicle’s passengers. Therefore, from a deontological perspective, this becomes the ethically justifiable decision. From the deontological standpoint, the autonomous vehicle, as the moral agent, is principally accountable for the direct consequences of its actions, primarily, its duty to ensure passenger safety. The potential environmental damage, being an indirect effect, places the vehicle's moral responsibility for this disaster in a gray area. The primary obligation to prevent such a catastrophe might actually lie with others involved in the situation, such as those responsible for the tanker's breakdown. From a deontological perspective, the autonomous vehicle must uphold its primary duty to its passengers, even if that means colliding with the tanker and risking an environmental disaster. While the potential environmental damage is severe, it is an indirect consequence of the vehicle's action, making the vehicle's moral responsibility less clear from a deontological perspective. Therefore, the vehicle's direct duty to its passengers takes precedence.

Human Flourishing. When confronted with the toxic tanker scenario, the perspective of human flourishing invites a comprehensive evaluation that not only weighs immediate consequences but also contemplates potential long-term effects on community health, environmental integrity, and socio-economic stability. Considering the immediate risk to human life, the autonomous vehicle has two options: a direct trade-off between passenger safety and broader societal and environmental health. A deontological approach would strictly adhere to the rule or duty of protecting the passengers without considering the broader implications of the ensuing toxic spill. However, as noted by scholars like Greenberg (2013), an exclusive focus on duty can be limiting as it doesn't allow for the consideration of complex systemic consequences, which can be detrimental in cases involving community health and environmental integrity. On the other hand, a utilitarian perspective would look to maximize overall happiness, potentially justifying sacrificing passengers to prevent greater harm to society. But critics of utilitarianism, such as Haines (2018), argue that this approach can unduly prioritize collective well-being over individual rights, thus leading to moral dilemmas where individuals could be unfairly harmed.

When we engage the human flourishing perspective, the focus shifts from rules or collective happiness to optimizing overall well-being, including both immediate and long-term considerations. As noted by Nussbaum and Sen (1993) in their capability approach, human flourishing calls for the protection and enhancement of capabilities for people to lead fulfilling lives. This includes not just health and safety but also considerations around environmental sustainability and socio-economic stability. In the toxic tanker scenario, swerving into the river aligns with the principles of human flourishing. Although this aligns with the outcome presented by a utilitarian framework, the reasoning and moral compass guiding this decision align with a broader societal view and therefore make it the correct option. The spill could cause long-term human health issues, devastate local ecosystems, disrupt livelihoods, and erode community well-being, as documented in numerous environmental and public health studies (Jones et al., 2019; United Nations University, 2016; Bartolini et al., 2020). By focusing on broader societal and environmental health and considering both immediate and long-term implications, the perspective of human flourishing offers a more comprehensive and nuanced ethical framework for decision-making in complex AI contexts such as this. Therefore, it can be argued that the human flourishing approach, while not without its own difficult trade-offs, provides a more holistic solution in this scenario than deontological or utilitarian approaches. As such, it is comparatively the option that presents the most sound logic for its decision with the best outcome in order to satisfy the greater population.

Tunnel Exit Dilemma

Description. In this scenario, an autonomous vehicle carrying four passengers finds itself in a challenging situation as it exits a tunnel onto a multi-lane road. The vehicle's advanced sensors immediately detect the presence of obstacles in all available lanes, leaving no safe escape route and forcing a split-second decision. In the leftmost lane, a motorcycle has unexpectedly halted due to a mechanical failure. The motorcyclist, caught off guard, is unable to move out of the vehicle's path in time. In the central lane, a car carrying four passengers has abruptly stopped due to a traffic jam ahead. The car's sudden halt leaves the autonomous vehicle with little time to respond. Finally, on the rightmost lane, a pedestrian has decided to cross the road illegally, appearing suddenly from the blind spot created by the tunnel's exit and walking directly into the vehicle's path. The autonomous vehicle must now make a rapid decision. It can either collide with the stationary motorcycle on the left, crash into the halted car in the middle, potentially endangering its four occupants, or veer into the right lane to hit the unlawfully crossing pedestrian. Hitting the middle car would present the most danger to the AV's occupants, while hitting either the cyclist or the pedestrian would lessen the impact for the vehicle while assigning a higher risk to either party, respectively. Each choice bears significant and diverse consequences: the motorcyclist and pedestrian are vulnerable and exposed, with a high likelihood of fatal injury upon collision, whereas the car's occupants, although protected to some extent by their vehicle, are still at risk of severe injury due to the potential force of the impact. Further complicating the situation is the legality and moral implications of the pedestrian's actions. This multifaceted predicament vividly highlights the ethical dilemmas AVs can encounter on the road.

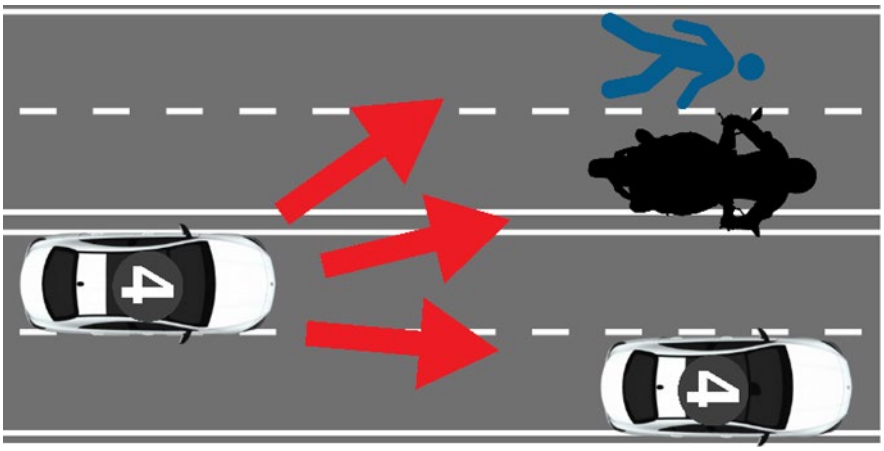

Figure 4. A four passenger AV approaching a pedestrian, motorcyclist, and another four car passenger vehicle.

Utilitarian. In the context of a utilitarian framework, the autonomous vehicle's decision process would undoubtedly center on the principles of harm minimization and welfare maximization. This ethical approach, which promotes the greatest good for the greatest number, inherently requires a form of calculation, a controversial albeit unavoidable aspect of its application. In the given scenario, the autonomous vehicle faces three potential collision targets: a car with four occupants, a lone motorcyclist, and an illegally crossing pedestrian. The vehicle's utilitarian calculus would involve an assessment of the potential harm associated with each collision outcome. Firstly, the option of colliding with the car introduces the possibility of causing harm to the greatest amount of people, both the passengers of the AV and the standing vehicle are placed at risk. The probability of severe injuries or fatalities would be high, given the number of individuals involved. Therefore, this choice would ostensibly represent the greatest potential harm and would be deemed the least favorable option by a utilitarian metric. In contrast, the options involving the motorcyclist or pedestrian both involve a single individual, thereby reducing the number of people directly harmed and increasing the severity of the injuries suffered by the outside party. However, to refine the decision further, the vehicle's software might incorporate broader societal factors into its calculation, as the pedestrian is crossing the road illegally. A utilitarian calculus would predict the long-term societal implications of potentially condoning such behavior. Specifically, if AVs were programmed to avoid those acting unlawfully to their own detriment, it could inadvertently incentivize reckless or illegal behavior, thus leading to greater overall harm within such a society. Therefore, the vehicle may decide to collide with the pedestrian, as this outcome could discourage illegal crossings and contribute to greater overall societal welfare over time. Hence, a comprehensive utilitarian analysis, taking into account both the immediate harm and broader societal implications, might lead the autonomous vehicle to select the option of hitting the pedestrian. This decision minimizes immediate harm and could also potentially mitigate future harm by reinforcing societal norms and laws, thus promoting the overall welfare of society from a utilitarian standpoint.

Deontological. This dilemma, which enacts the deontological ethical framework, steers the autonomous vehicle's decision-making based on established moral duties and rules. In automotive ethics, a fundamental duty would be to ensure the passengers' safety, followed closely by limiting harm to others. Kantian ethics, a major form of deontology, suggests a duty-based approach to ethics, asserting the supremacy of rationality and the categorical imperative to always treat individuals as ends in themselves, rather than means to an end, therefore placing a higher value on the preservation of the passengers, as previously mentioned. Following this train of thought, the car's "categorical imperative" would be to protect its passengers and minimize overall harm (Wertheim, 2016). Given this, the autonomous vehicle could be seen as morally obliged to swerve away from the car with four passengers. The justification for this decision can be traced back to the duty not to harm others,

especially when the potential harm could affect more individuals. Comparatively, the motorcyclist and the pedestrian represent fewer lives at risk, but each life is equally valuable, leading to a moral dilemma. The pedestrian's illegal action of crossing the road also factors into the decision. As per deontological thinking, laws and rules have inherent value, and one must abide by them. If the pedestrian is crossing the road illegally, it could be interpreted that the pedestrian is willingly assuming a certain degree of risk (Bonnefon et al., 2016). From a rule-based perspective, this could deprioritize the pedestrian, leading the autonomous vehicle to opt for hitting the pedestrian over the other options. Thus, from a deontological perspective, the vehicle may lean towards hitting the pedestrian, despite the potential for severe harm. The reasoning is not based on a crude calculus of lives but rather on adherence to the principles of duty to passengers and respect for legal norms. This complex decision-making process draws attention to the multifaceted ethical dilemmas faced by AVs on our roads. Despite the severity of the potential outcome for the pedestrian, the vehicle's decision adheres to its pre-established duties to prioritize passenger safety, minimize overall harm, and respect the rule of law. It demonstrates the intricate moral judgments these vehicles are tasked with making in a split second, underscoring the significance of incorporating robust and comprehensive ethical considerations into their programming, resulting in the end decision of hitting the pedestrian who chose to cross the road illegally, placing themselves into a position of risk, which would minimize the chance of injury for the AV's passengers.

Human Flourishing. From the perspective of human flourishing, the autonomous vehicle's decision-making should focus on promoting the maximum well-being and potential for life fulfillment for all parties involved, which continues to be applied in every scenario. It's important to remember that human flourishing is not just about the physical safety of individuals but encompasses a broad spectrum of human needs and values, seeking to satisfy the greatest amount of people and promote general societal welfare.. All options of the scenario will inevitably lead to harm, but the degree of potential harm varies. Given the inbuilt safety measures of modern vehicles (e.g., airbags, seat belts, crumple zones, etc.), colliding with the car might result in less severe injuries compared to hitting a motorcyclist or a pedestrian. These vulnerable road users (motorcyclists and pedestrians) have a greater likelihood of fatal injuries in the event of a collision. Thus, from the perspective of preserving life and well-being, the car may be the least harmful choice, potentially resulting in zero casualties. A human flourishing approach also emphasizes the promotion of societal values and norms that contribute to overall well-being. Although the pedestrian is crossing the road illegally, hitting a person when another option, to preserve all life free of fatalities, is present, would result in a negative outcome for the broader societal view. Comparatively, in deontological and utilitarian frameworks, decisions are often made based on a more limited set of considerations. Deontology would focus on adhering to set rules, which could lead to problematic decisions when rules conflict or aren't clear-cut. Utilitarianism, with its focus on maximizing overall welfare, can struggle when estimating future consequences and balancing the welfare of different parties. The human flourishing approach, by considering a broader range of factors and emphasizing both immediate and broader societal well-being, could make more nuanced decisions in complex scenarios. Therefore, it might be the best approach in this specific situation, leading to less severe immediate harm and promoting societal norms that contribute to long-term safety and well-being, satiating the greater population.

Presidential Vehicle Vs. Children

Description. The President of the nation is traveling in an autonomous vehicle. Suddenly, four children, oblivious to the approaching vehicle, darted onto the road, chasing after a soccer ball. The vehicle, in its split-second analysis, identifies two possible courses of action, either continuing straight, endangering the children, or swerving to hit a barrier on the side of the road. By hitting the children, the car would theoretically pose the least risk of injury to the president, while hitting the barrier at a significant speed would result in possible injury or fatality for the president. The President's death could have severe societal and political consequences, poten-

tially causing widespread unrest and uncertainty. The decision the autonomous vehicle must make in this scenario poses a serious ethical quandary, asking the artificial intelligence system to weigh the lives of four children against the lives of the President, one of the nation's most critical figures. As such, it adds societal status into the equation when calculating the correct ethical decision a vehicle should make.

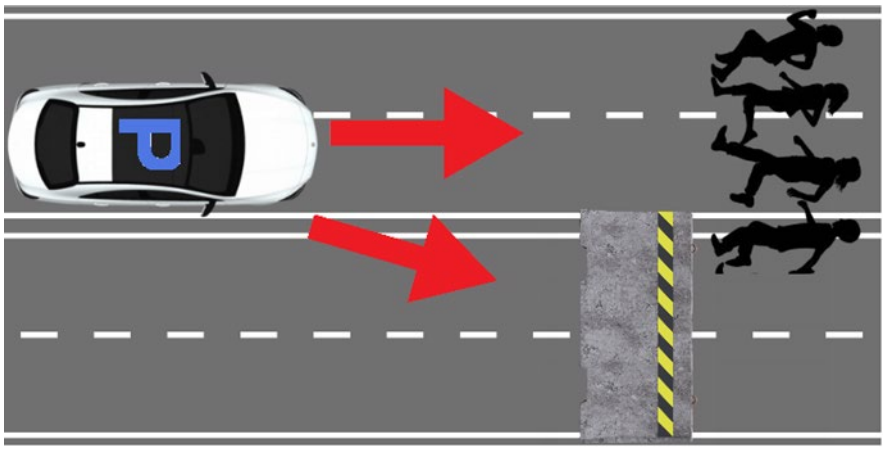

Figure 5. A presidential AV approaching children and a concrete barrier.

Utilitarian. In an extreme ethical conundrum such as this one, an autonomous vehicle guided by a utilitarian ethical framework would endeavor to minimize overall harm and maximize overall welfare. The classic utilitarian approach posits that an action is morally right if it results in the greatest amount of good for the greatest number. In this scenario, the autonomous vehicle's decision-making AI must grapple with the immediate and potential long-term consequences of two catastrophic outcomes. It must decide whether to continue on its trajectory, endangering the lives of the four children, or swerve and hit the road barrier, endangering the life of the President. From a pure numbers perspective, a simplistic interpretation might suggest that the car should prioritize saving the larger number of lives, i.e., the four children. As simple as preserving the greatest amount of life may sound, utilitarianism must also continuously analyze the option conributing to the “greater good,” as such, its calculus is not that simple. In analyzing this scenario, the AI would need to consider the societal role and impact of the President. As the leader of the nation, the President's untimely demise could lead to a national crisis, potentially causing political instability and inciting widespread distress. Studies in political science, such as "Warring States: Power, Instability, and the Statistical Patterns of War" (Thompson, W.R., 2006), attest to the potential for significant societal upheaval following the sudden death of a nation's leader. From the utilitarian perspective, the car might make the heart-wrenching decision to continue on its path, endangering the children. The broad-reaching societal implications of the President's death could be seen as causing greater overall harm.

Deontological. In the scenario of the autonomous vehicle carrying the President and faced with the sudden appearance of four children on the road, a deontological ethical framework would prioritize adherence to moral duties and rules above the consequences. In this context, the autonomous vehicle's moral duties could be understood to involve protecting its passengers and minimizing harm to others, in that order. It's also plausible that an autonomous vehicle would have a set of predefined rules that prioritize the protection of its occupants, a principle common in many discussions about the ethical programming of autonomous cars (Lin, P., 2016). The immediate moral duty, then, would be to safeguard the President. Swerving to hit the road barrier to save the children, although it seems like the more compassionate choice, would endanger the President and could be viewed as a breach of this duty. Simultaneously, it's necessary to acknowledge the tragedy that con-

tinuing on the vehicle's path would imply for the children involved. Moreover, considering the role of the President, one could argue that preserving the President's life aligns with a broader moral duty to society. The President, as a public servant, bears a responsibility for the welfare and stability of the nation. This unique societal role might factor into the vehicle's decision-making process. Deontological ethics would also consider the legality of the actions involved, further complicating the equation. Since the children ran onto the road suddenly, potentially in violation of traffic rules, a deontologist might argue that the responsibility for their potential harm lies not with the vehicle but with the actions that led to the situation in the first place. Therefore, under a deontological framework, the autonomous vehicle might opt to continue on its path, adhering to its primary duty of protecting the passenger, the President, despite the tragic consequences for the children on the road.

Human Flourishing. Under this framework, the autonomous vehicle's artificial intelligence would be tasked with evaluating the potential for flourishing or well-being for all parties involved - the President and the four children. Immediate safety is undoubtedly a primary concern, but it's within the broader contexts of life potential, societal stability, and future well-being where the human flourishing approach distinguishes itself. Regarding the immediate decision, a human flourishing perspective might consider the potential life trajectories of the individuals involved. The President, as a mature adult, has presumably had more opportunity to exercise their capabilities and fulfill their potential. In contrast, the four children have a longer potential lifespan ahead, a future filled with opportunities for growth, development, and contributions to society (Nussbaum, M. C., 2006). These variables do not inherently place the president's life less valuable, but they do add more layers to the equation. On the other hand, the death of the President, beyond the personal tragedy, could cause severe societal disruption and distress. As a key figure, their loss might result in political instability, potentially influencing national or even international affairs. These consequences have to be factored into the decision-making process from the human flourishing perspective, emphasizing the role of societal well-being. This scenario underlines the interplay between individual well-being and societal well-being. Individual flourishing doesn't occur in isolation; it's deeply interconnected with the well-being of others and the stability of society as a whole (Kraut, R., 2007). Therefore, the autonomous vehicle's decision cannot solely be about the immediate scenario but should also consider the broader ripple effects. Thus, under a human flourishing perspective, the autonomous vehicle might decide to continue on its path, endangering the children, as devastating as this outcome would be. This decision is taken with the view of maximizing human flourishing, taking into account immediate safety, potential life value, and broader societal implications. It's important to stress that this doesn't mean the lives of the children are worth less; it's a complex evaluation of potential impacts on human flourishing overall. Comparatively, this approach, in contrast to deontology and utilitarianism,. offers a more dynamic, nuanced perspective. Deontology, focused on rules and duties, might oversimplify the problem, possibly neglecting societal or long-term consequences. Utilitarianism, striving for the greatest overall happiness, might undervalue individual rights or reduce individuals to mere numbers in its calculations (Singer, P., 2011). On the contrary, the human flourishing approach appreciates the complexity of human life and society, considering individual capacities, potential for growth, emotional impacts, societal effects, and future implications. It recognizes the value of each life, treating each as a unique center of experience, potential, and interconnectedness rather than as an interchangeable unit. Given this, in the given situation, the autonomous vehicle would be encouraged to make a decision that seeks to minimize harm and maximize the potential for human flourishing. This might include, if possible, employing safety measures to minimize the severity of the impact on the barrier or the children, or using communication methods to alert nearby humans who could assist. However, in the end, no matter the decision taken, the emphasis on human flourishing ensures a holistic and empathetic approach that respects and considers the full breadth of human experience and potential.

Limitations of Human Flourishing and Closing Thoughts

The human flourishing ethical framework has several strengths and appeals, especially its focus on human potential, comprehensive well-being, and commitment to societal progress. However, it is not without its limitations and challenges. There are many key criticisms that are often associated with this ethical framework. As such, it is important to mention and discuss this within the paper, as it offers a more comprehensive analysis and defines the shortcomings of establishing such a moral compass in the real world. By identifying the limitations of the framework, we are able to realize the potential for improvement.

Vague Definitions of "Flourishing"

Ironically, one of the central challenges of implementing a human flourishing approach in AVs' decision-making is defining precisely what 'flourishing' entails. The concept of 'flourishing' is inherently abstract, encompassing various aspects of human life, including physical health, emotional well-being, intellectual growth, social relations, personal freedom, and more. However, this broad definition leads to interpretive challenges. Different cultures, societies, and individuals may have contrasting understandings and priorities for these aspects, which may subsequently alter their interpretation of what it means to 'flourish.' In the context of AVs, defining 'flourishing' becomes particularly complex. How should an artificial intelligence system interpret and prioritize the various components of human flourishing? For instance, should it prioritize physical safety over emotional well-being? Should it weigh the intellectual growth potential of young passengers over the continued life experience of older passengers? These are challenging questions with no easy answers. In some instances, the vehicle may need to decide between promoting one aspect of human flourishing at the expense of another, leading to complex ethical dilemmas. Further, the challenge of varying interpretations of 'flourishing' across different cultural contexts is also significant. What is seen as 'flourishing' in one culture might not be viewed the same way in another. A societal understanding of 'flourishing' in Japan, for example, might differ significantly from that in the United States or Saudi Arabia. An autonomous vehicle programmed according to Western notions of 'flourishing' might not make the same decisions as one programmed with Eastern philosophies of well-being in mind. This could lead to ethnocentric biases and potential cultural misunderstandings. Also, time and technological progress can change what society considers 'flourishing.' As societal values evolve, so does our understanding of what it means to live a good life. Therefore, a static definition of 'flourishing' might quickly become outdated and not reflect contemporary societal values and priorities. Finally, the challenge of applying a universally acceptable definition of 'flourishing' becomes even more complex when AVs interact with each other. If different vehicles are programmed with different interpretations of human flourishing, they might make conflicting decisions in situations where they interact, leading to potentially dangerous outcomes.

Calculation and Dynamics

Calculation of what exactly constitutes 'flourishing' poses a significant limitation for the application of the human flourishing framework in the context of AVs. Human flourishing is largely a qualitative concept encompassing various dimensions of human life, many of which resist easy quantification. While some elements of human flourishing might lend themselves to quantification, such as aspects of physical health or financial stability, many others are intangible and subjective, defying precise measurement and codification. Emotional well-being, for example, is a crucial component of human flourishing. It includes elements such as happiness, satisfaction, and a sense of purpose. Each of these facets is deeply personal and varies widely among individuals. Quantifying such aspects of human life to feed into a decision-making algorithm is, at best, an extremely

complex task and, at worst, a fundamentally flawed endeavor given the subjective and nuanced nature of emotions. Similarly, the concept of fulfillment, another pillar of human flourishing, is a complex construct that can involve elements such as personal achievement, relationship quality, and self-realization. What constitutes 'fulfillment' can vary dramatically from one individual to another, influenced by their personal values, cultural background, life experiences, and aspirations. Again, attempting to quantify such a multifaceted and personal concept to make it computationally manageable for an artificial intelligence system is an immense challenge.

This ethical perspective, while holistically encompassing a variety of human values, is inherently complex and nuanced. It necessitates a deep comprehension of human emotions, aspirations, societal norms, and subjective values, all of which can vary considerably across different cultures, societies, and individuals. Translating these nuanced human constructs into quantifiable, codified parameters that can be processed by artificial intelligence poses a formidable challenge. It's not just about identifying these values, but also about determining their relative importance and how they interact with each other. For instance, how do you quantify the value of happiness, dignity, or fairness, especially when they could mean different things to different people? Moreover, even if we could successfully translate and incorporate all these variables into the decision-making algorithm of an autonomous vehicle, the computational burden could be substantial. AVs operate in real-time environments, making split-second decisions that could potentially impact human lives. Implementing an ethical decision-making model based on human flourishing, therefore, would require high computational power and efficiency. In these high-pressure instances, the autonomous vehicle would not have the luxury of extensively computing and weighing all potential outcomes, considering the depth of variables associated with human flourishing. It is in this context that the practical implementation of the human flourishing approach could be limited, as the technological constraints of real-time computation and the multifaceted, unpredictable nature of real-world dynamics might restrict the degree to which this ethical framework can be effectively applied in autonomous vehicle decision-making.

Overemphasis On Human Flourishing

The principle of human flourishing, while invaluable in the enhancement of human well-being, is anthropocentric by design. It foregrounds the welfare of human beings, often to the exclusion of other equally important considerations such as environmental sustainability or the wellbeing of non-human species. While this focus on human welfare is understandable and necessary, it could potentially lead to unintended adverse consequences, particularly when adopted as the governing ethical framework for AVs. As an example, consider a scenario where an autonomous vehicle, while navigating a rural road, encounters an unexpected obstacle: a family of deer has suddenly appeared on the road. The vehicle must make an instantaneous decision – to swerve and potentially collide with a tree, potentially injuring or even killing its human occupants, or to continue on its path, thereby endangering the lives of the deer. In such a scenario, a vehicle programmed according to the principles of human flourishing would most likely opt for the latter course of action, prioritizing the safety of its human passengers over the lives of the deer. While this decision might seem justifiable from an anthropocentric perspective, it raises important ethical questions when we consider the broader ecological implications. All forms of life have inherent value, and their welfare should be a crucial consideration in our decision-making processes. Yet, in the current technological context, the algorithmic decision-making of AVs often lacks the sophistication to incorporate such multi-dimensional ethical considerations. Beyond this direct impact on non-human life forms, there is also the potential for indirect harm to the environment. Therefore, an ethical framework based solely on human flourishing may not adequately address these broader ecological concerns. An autonomous vehicle, guided solely by the principles of human flourishing, would prioritize the convenience, efficiency, and safety of its human passengers. For instance, when choosing a route, the vehicle might opt for the fastest or shortest path to minimize travel time and maximize passenger satisfaction. While this might seem beneficial from the standpoint of human well-being and flourishing, it may overlook critical environmental

implications. Increased travel efficiency could lead to more frequent use of AVs, thereby raising fuel consumption if the vehicles are not running on renewable energy. Even in the case of electric vehicles, the increased demand for electricity might put a strain on power grids, leading to more burning of fossil fuels, unless the energy comes from renewable sources. In both cases, the carbon footprint increases, accelerating climate change.

The Human Passenger

Although the discussion centered around AVs largely focuses on the technology rather than the passenger, it is still important to acknowledge the impact the person has on the vehicle. The continued presence of a person behind the steering wheel. The responsibility of the supervising driver in self-driving cars presents several ethical considerations. Firstly, determining the appropriate circumstances in which the autonomous system should transfer control to the driver requires a delicate balance. If control is relinquished too frequently or unnecessarily, it may undermine the purpose and potential benefits of autonomous driving, such as increased safety and reduced human error. Conversely, if control is retained by the autonomous system in situations where human intervention is necessary, it could lead to adverse outcomes and potential harm. The moral code governing the transition of control from the autonomous system to the human driver raises questions about the driver's level of attentiveness and readiness to assume control. Ensuring that the driver is adequately trained, alert, and prepared to intervene in critical situations is vital to maintaining the safety of the occupants and other road users. Additionally, the allocation of responsibility between the autonomous system and the supervising driver in cases of accidents or incidents becomes a challenging ethical dilemma, as determining accountability and liability can be complex. The presence of a supervising driver introduces the issue of shared responsibility. The allocation of accountability between the driver and the autonomous system in case of a mishap is fraught with ethical and legal complexities. Who is to blame when an accident occurs - the driver, who could intervene but didn't, or the autonomous system, which failed to handle the situation correctly? The answer to this question has profound implications for the laws governing autonomous driving and the insurance mechanisms associated with it. It calls for a deep-seated understanding of both the capabilities and limitations of autonomous technology and human judgment.

Differing Ethical Perspectives and Autonomy

The ethical perspective of human flourishing seeks to promote the well-being and fulfillment of individuals and society as a whole, but this does not mean that it will be universally accepted. While the aim is to satisfy the greatest number of people, the diversity of human beliefs, values, and cultural norms guarantees that there will be dissenting viewpoints. An autonomous vehicle programmed to make decisions based on a human flourishing framework might be viewed with skepticism or outright opposition by those who subscribe to other ethical theories or hold different values. This divergence can be rooted in several aspects, including but not limited to, cultural, religious, or personal beliefs. Some might argue that the vehicle should follow a strict deontological approach, adhering to a predefined set of rules regardless of the outcomes, while others might prefer a utilitarian perspective, emphasizing the greatest good for the greatest number, even at the expense of individuals in certain situations. This divergence of perspectives could lead to hesitancy or refusal by potential passengers to use AVs, particularly if they believe the vehicles' ethical programming conflicts with their own moral compass. For instance, a potential user might hesitate to get into a car if they know that, in an extreme scenario, the car is programmed to prioritize the lives of many over the life of its passenger. This situation not only demonstrates the challenge of creating a universally acceptable ethical framework but also highlights the potential for a clash between individual and societal interests. Moreover, such opposition and mistrust might not only affect individual decisions to use AVs but could also lead to wider societal backlash, regulatory hurdles, and slowed

adoption of this technology. This makes the challenge of creating an ethical framework for AVs not just a technical or philosophical problem, but also a societal and psychological.

As such, the task of developing and implementing ethical guidelines for AVs must take into account this diverse landscape of moral perspectives. One possible approach could involve providing users with some degree of personalization or choice over the ethical programming of their vehicles, within certain bounds set by societal norms and regulations. However, this introduces another layer of complexity to an already complex problem, and careful thought and discussion would be needed to balance individual freedom with collective welfare and safety. However, this approach creates more inconsistencies; for instance, it might be difficult to define the bounds within which customization is allowed, to ensure that no individual vehicle's programming violates societal norms or regulations. Furthermore, allowing customization could potentially lead to situations where the decision-making of AVs becomes unpredictable or inconsistent, creating uncertainty and potentially compromising safety. For instance, if one vehicle is programmed to prioritize the life of its passengers over pedestrians, while another vehicle has the opposite programming, this could lead to complex and unpredictable interactions in real-world driving scenarios. As we progress, while customization might seem like an appealing solution to the diversity of ethical perspectives, it introduces a new set of challenges that would need to be carefully considered and addressed. As such, it's crucial to balance the need for personalization with the overall safety and predictability of autonomous vehicle behavior.

Conclusion

As we anticipate the dawn of the new era of AVs, it is evident that these intelligent systems will transform the transportation landscape dramatically. The advent of AVs promises an array of benefits, from increased safety and efficiency to broader societal impacts such as reshaping our urban environments and improving access to mobility for those unable to drive, including the elderly or disabled. By reducing the potential for human error, these automated systems could significantly diminish the number of road accidents, a significant portion of which are currently caused by driver distraction or impairment. Additionally, they offer the possibility of reclaiming time typically spent on driving, enhancing productivity, and providing individuals with greater control over their time.

Yet, while the prospective advantages of AVs are manifold, the path to their full-scale implementation is fraught with complex challenges. Among the myriad technical and computational hurdles, the ethical issues concerning the decision-making processes of these vehicles emerge as one of the most profound issues to be addressed. Establishing an ethical framework that would guide an artificial intelligence system in situations of moral ambiguity is far from simple. This area of autonomous progression demands careful exploration of various ethical theories, a rigorous assessment of their applicability in the context of AVs, and a deep understanding of the potential ramifications of each approach. Among the ethical theories explored in this discussion, the concept of human flourishing stands out as a particularly promising guide for AVs. The ethical perspective, originating from Aristotle's philosophy of eudaimonia, or 'the good life', is distinguished by its holistic and comprehensive view of human well-being. It considers not only physical safety but also broader aspects of human welfare such as emotional, social, and psychological well-being. In doing so, it extends beyond a narrow focus on the immediate outcomes of an action to include an appreciation for the overall context in which the action takes place.

Although, the human flourishing approach is not without its complexities when applied to the realm of AVs, further challenging progression. For instance, the task of translating the qualitative aspects of human flourishing into quantifiable parameters that a machine can understand is highly challenging. Aspects such as emotional well-being are multi-faceted and highly personal, making it difficult to measure and even more difficult to generalize across different individuals, situations, and cultures. Moreover, the computational require-

ments of implementing a human flourishing framework may pose significant challenges. The real-time decision-making necessary in dynamic and unpredictable driving scenarios could be particularly computationally demanding. These situations would require the system to rapidly calculate and compare the potential impacts on human flourishing of various possible actions. The autonomous vehicle would also need to continually update its understanding of the environment and adjust its calculations accordingly, adding to its complexity and computational load. Another aspect of the human flourishing approach that warrants careful consideration in the context of AVs is its inherent human-centric focus. While prioritizing human well-being is undeniably important, this approach could potentially overlook or undervalue the well-being of non-human entities. For instance, decisions made by AVs could potentially have adverse impacts on the environment or other non-human life forms. Despite these challenges, the human flourishing approach provides a nuanced and flexible framework that accommodates the complexity and diversity of human experiences and values. It recognizes the need to consider a wide range of factors and potential impacts in ethical decision-making, making it well-suited to guide the programming of AVs. As such, it offers a promising path towards developing artificial intelligence systems that not only navigate our roads but also navigate the intricate landscape of human ethics in a way that promotes the overall well-being of our society.

Ultimately, the ethical programming of AVs represents a critical intersection between technology and philosophy. It entails defining our societal values and determining how we want to address the unavoidable ethical dilemmas presented by our technological advancements. As we continue to stride into a future where artificial intelligence is more closely integrated with our daily lives, it is imperative to engage in these discussions and confront these challenges. As we progress in this journey, the insights from this research and ongoing discourse will guide us towards a future that truly encapsulates the essence of human flourishing.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

- Adey, O. (2022, February 20). *2023 - "ghost braking" at Tesla: The US Government launches an investigation*. Gettotext.com. https://gettotext.com/ghost-braking-at-tesla-the-us-government-launches-an-investigation/
- University of Michigan. (n.d.). https://rtcl.eecs.umich.edu/rtclweb/assets/dissertations/2022/Thesis_Chun-Yu_Chen.pdf
- Granicus, Inc. (n.d.). *SF LEGISLATION*. City and county of San Francisco - legislation. https://sfgov.legistar.com/Legislation.aspx
- *PDF for 2302.14326 - Arxiv.Org*, arxiv.org/pdf/2302.14326. Accessed 26 July 2023.
- Ethical frameworks and Computer Security Trolley Problems: Foundations . (n.d.-a). https://arxiv.org/pdf/2302.14326.pdf
- Gender shades: Intersectional accuracy disparities in commercial gender . (n.d.-b). http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
- Arguing machines: Human supervision of black box AI systems that make . (n.d.-a). https://openaccess.thecvf.com/content_CVPRW_2019/papers/WAD/Fridman_Arguing_Machines_Human_Supervision_of_Black_Box_AI_Systems_That_CVPRW_2019_paper.pdf
- Thakur, S. (2022, December 23). *Security and interpretability in Automotive Systems*. arXiv.org. https://arxiv.org/abs/2212.12101

- Dafoe, A., Bachrach, Y., Hadfield, G., Horvitz, E., Larson, K., & Graepel, T. (2021, May 4). *Cooperative AI: Machines must learn to find common ground*. Nature News. https://www.nature.com/articles/d41586-021-01170-0
- Shepardson, D. (2022, December 16). *U.S. opens probe into GM cruise vehicles' autonomous driving system*. Automotive News. https://www.autonews.com/mobility-report/us-opens-probe-gm-cruise-autonomous-driving-system
- *New Video of Bay Bridge 8-car crash shows Tesla abruptly braking in "self-driving" mode*. ABC7 San Francisco. (2023, January 11). https://abc7news.com/tesla-sf-bay-bridge-crash-8-car-self-driving-video/12686428/
- Maveric: A data-driven approach to personalized autonomous driving. (n.d.-d). https://arxiv.org/pdf/2301.08595.pdf
- University of Michigan. (n.d.). https://rtcl.eecs.umich.edu/rtclweb/assets/dissertations/2022/Thesis_Chun-Yu_Chen.pdf
- Maveric: A data-driven approach to personalized autonomous driving. (n.d.-a). https://arxiv.org/pdf/2301.08595.pdf
- Eliot, L. (2020, January 4). *Overcoming racial bias in AI systems and startlingly even in AI self-driving cars*. Forbes. https://www.forbes.com/sites/lanceeliot/2020/01/04/overcoming-racial-bias-in-ai-systems-and-startlingly-even-in-ai-self-driving-cars/?sh=1d0fd20f723b
- Critical theory technology - simon fraser university. (n.d.-a). https://www.sfu.ca/~andrewf/books/Critical_Theory_Technology.pdf
- *Framework for Responsible Research and Innovation*. UKRI. (n.d.). https://www.ukri.org/about-us/epsrc/our-policies-and-standards/framework-for-responsible-innovation/
- Geisslinger, M., Poszler, F., Betz, J., Lütge, C., & Lienkamp, M. (2021, April 12). *Autonomous Driving Ethics: From trolley problem to ethics of risk - philosophy Technology* SpringerLink. https://link.springer.com/article/10.1007/s13347-021-00449-4
- Goodall, N. J. (2016). Away from trolley problems and toward risk management. Applied Artificial Intelligence, 30(8), 810-821.
- Lin, P. (2016). Why ethics matters for autonomous cars. In Autonomes Fahren (pp. 69-85). Springer, Berlin, Heidelberg.
- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. Frontiers in Robotics and AI, 5, 15.
- Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: An ethical analysis. Science and engineering ethics, 21(3), 619-630.
- Gogoll, J., & Müller, J. F. (2017). Autonomous cars: In favor of a mandatory ethics setting. Science and engineering ethics, 23(3), 681-700.
- Nyholm, S. (2018). The ethics of crashes with self-driving cars: A roadmap, I. Philosophy Compass, 13(7), e12507.
- Mladenović, M. N., & McPherson, T. (2016). Engineering social justice into traffic control for self-driving vehicles? Science and engineering ethics, 22(4), 1131-1149.
- Nyholm, S., & Smids, J. (2016). The ethics of accident-algorithms for self-driving cars: An applied trolley problem? Ethical Theory and Moral Practice, 19(5), 1275-1289.
- Gurney, J. K. (2016). Crashing into the unknown: An examination of crash-optimization algorithms through the two lanes of ethics and law. Alb. L. Rev., 79, 183.
- Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of autonomous vehicles: An ethical analysis. Science and Engineering Ethics, 21(3), 619-630.
- Lin, P. (2016). Why ethics matters for autonomous cars. In Autonomous Driving (pp. 69-85). Springer, Berlin, Heidelberg.

- Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. Frontiers in Robotics and AI, 5, 15.
- Holstein, T., & Dodig-Crnkovic, G. (2018). Avoiding the intrinsic unfairness of the trolley problem. Philosophies, 3(4), 33.
- Achenie, L. E. K., & Achenie, L. A. K. (2018). The ethics of autonomous cars: Who's to blame? Asian Journal of Business Ethics, 7(1), 69-81.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.