Autonomous Vehicles and Avoiding the Trolley (Dilemma): Vehicle Perception, Classification, and the Challenges of Framing Decision Ethics

Martin Cunneen, Martin Mullins, Finbarr Murphy, Darren Shannon, Irini Furxhi & Cian Ryan

To cite this article: Martin Cunneen, Martin Mullins, Finbarr Murphy, Darren Shannon, Irini Furxhi & Cian Ryan (2019): Autonomous Vehicles and Avoiding the Trolley (Dilemma): Vehicle Perception, Classification, and the Challenges of Framing Decision Ethics, Cybernetics and Systems, DOI: 10.1080/01969722.2019.1660541

To link to this article: https://doi.org/10.1080/01969722.2019.1660541

© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC

Published online: 17 Sep 2019.

Article views: 785

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=ucbs20
Autonomous Vehicles and Avoiding the Trolley (Dilemma): Vehicle Perception, Classification, and the Challenges of Framing Decision Ethics

Martin Cunneen, Martin Mullins, Finbarr Murphy, Darren Shannon, Irini Furxhi, and Cian Ryan
Department of Accounting and Finance, University of Limerick, Limerick, Ireland

ABSTRACT
This article aims to introduce a degree of technological and ethical realism to the framing of autonomous vehicle perception and decisionality. The objective is to move the socioethical dialog surrounding autonomous vehicle decisionality from the dominance of “trolley framings” to more pressing ethical issues. The article argues that more realistic ethical framings of autonomous vehicle technologies should focus on the matters of HMI, machine perception, classification, and data privacy, which are some distance from the decisionality framing premise of the MIT Moral Machine experiment. To support this claim the article appeals to state-of-the-art technologies and emerging technologies concerning autonomous vehicle perception and decisionality, as a means to inform and frame ethical contexts. This is further supported by considering a context specific ethical framing for each time phase we anticipate regarding emerging autonomous vehicle technology.

KEYWORDS
Autonomous vehicles; AI; artificial intelligence; classification; computer vision; machine perception; trolley dilemma; trolley problem

Introduction
The development of autonomous vehicles offers many societal benefits. However, in order to avail of such benefits, a number of challenges must be overcome. Autonomous vehicle technology presents a new paradigm in sociotechnological relations and inhere numerous challenges in terms of both the technical complexities confronting engineers and the governance issues confronting governments, policy-writers and regulators. One consistent difficulty concerns how autonomous vehicle functionality, and in particular, how machine decisionality is communicated to other actors who often make important welfare decisions regarding the use of the technology. It is of critical importance to accurately frame the functionality and limitations of socially embedded technologies that concern human life and...
welfare. Our own human and linguistic biases lead us to misunderstand innovative technologies as, by default, we anthropomorphize machine perception and decision capacity. Therefore, we maintain we must avoid framings that further perpetuate our misunderstanding.¹ While, the article does not focus on confronting specific examples, such as MIT’s *Moral Machines* experiment, it defends the need to consider the more immediate, realistic, and important ethical questions that machine perception and decision-making present. We contend that the success of AVs depends, in part, upon the transparency and explainability of the technology. In the case of AVs, these center on nontechnical actors’ comprehension of the ability of AV to make complex driving decisions. This is an understandable response given that society and users are being asked to put their trust in AV technologies to provide them with a safe and efficient transportation.

“If you ask me whether autonomous vehicles will become commonplace, my unequivocal answer is yes, there’s no question about it. The technology is almost there, the world is almost there, there’s an economic motive for getting there, and drivers will slowly start to get used to the idea that you can get rid of the boring task of driving.” (Mobileye 2019)

“Artificial Intelligence Will Not Replace Us Behind the Wheel of a Car” (Floridi and Rodella 2019)

The above quotations present two contrasting world views regarding AVs which correspond to the two most consistent ethical framings of AVs. The first concerns the claim that AVs will provide a safer and superior driving decisionality which amounts to an AV safety argument (Cunneen et al. 2019b). This quote represents the most popular positive framing regarding the societal impact of AVs. This framing by Mobileye’s founder, Professor Amnon Shashua, is an important example of the commercial framings of AVs and the claim of AV decisional superiority. The second framing rehearses the claim that AVs will be significantly limited in terms of driving decisionality and will be unable to achieve full autonomous driving (Cunneen et al. 2019a, 2019b). This view is endorsed by Professor Luciano Floridi, a prominent AI researcher; in the above quote he is dismissing the possibility of superior AV decisionality. Floridi’s statement not only dismisses the claim of Shashua, but also underscores the social confusion surrounding AVs. As Shashua points out, there is an expectation that AVs will soon be a common occurrence on the human road network as billions are invested, and countless hours of driverless video from many different research groups abound our news feeds. However, Floridi is partly

¹The article brings together contributions from active researchers on state-of-the-art AV research projects. We have experienced the peak of the 2016 hype cycle concerning the expectation to achieve full automation within 5 years and the recent move away from this date. Accordingly, there is a renewed focus on using ADAS technologies to augment the driving phenomenon.
right; at the moment it is clear that state-of-the-art AI will not replace humans “behind the wheel of car” for some time. Yet, if Floridi is correct in his prediction, why is so much attention given to trolley dilemma questions regarding AVs, and in particular, the framing of AVs capacity to make complex human value based and moral decisions? The question of the societal and ethical impact of AVs is increasingly framed in terms of the safety arguments of decisionality superiority and “trolley-dilemma” considerations. These interrelated framings are largely based on an inaccurate understanding of the state-of-the-art and emerging AV technologies. In the near future, AVs are neither going to solve or dramatically reduce the 1.3 million global road fatalities nor have a level of machine perception or the ability to classify between doctors and criminals.

The article argues that AV literature and narratives are dominated by inaccurate framings regarding AV decisionality, machine perception, classification, and decisional ethics. AI is already replacing humans in modes of driving akin to closed envelopes of driving; from lane assist, park assist, and summon, to emergency-braking technologies. Such ADAS technologies create envelopes of operation which are stepping stones to larger envelopes of operation. Given time artificial driving intelligence will replace human driving intelligence. The salient point is that, as with all innovative AI framing, there is a significant distinction between the timeline of hype and the timeline of AI/technological realism. While both timelines may ostensibly align at certain moments, for the most part, they are actually contrasting positions. Accordingly, to remove some of the confusion from the AV narrative, we must, where possible, hold to technological reality regarding the decision capacity of AV innovation. With this in mind the article constructs an AV technology timeline as a means of creating a parallel contextualization regarding the more applicable or tangible ethical challenges that AV technology will present. This timeline offers a more pragmatic means of not merely understanding AV technological development, but also a means of beginning to frame and anticipate the ethical challenges and concerns.

This article makes the case that narratives around future ethical questions should be more directly embedded in technological developments. The extant literature demonstrates a tendency to address the ethical questions associated with automated vehicles whilst overlooking current “state-of-the-art” technologies; a critical lacuna which creates much confusion in attempts to engage with the ethical issues involved. All of which is not to suggest that such articles as The Moral Machine, in which MIT scholars canvas the ethical preferences of citizens around the world in respect of who should or should not be saved in the event of an UTA, are empty exercises (Awad et al. 2018). On the contrary, they provide an important
starting-point in heuristic debates on the potential deployment of AI in AVs.

However, the framing of these experiments in defining the ethical space regarding AVs plays more to populist appetites for headline and clickbait. Nonetheless, significant implications for societal acceptance of this new technology remain. For instance, nuanced debates within the scientific literature on the moral agency of future vehicles which have leached into the mass media inevitably impact both risk perception and societal acceptance of this technology. The same is true of vehicle manufacturing debates regarding the extent to which future AVs will prioritize the safety of car occupants. It is hardly surprising then that hystericized news coverage epitomized in headlines such as “automated driverless cars would run over a CHILD rather than swerve and risk injuring the passengers inside”\(^2\) has a negative effect on public reception of this new technology, which gives rise to a probable outcome of dread risk\(^3\) around such scientific advances.

It is generally held that a broad diversity of classes and/or ethical dilemmas will emerge as AVs develop. These will align with the increasing AV sophistication of the sensors and computational powers to process the data. Much will also depend of the communication infrastructure advances over the next 20 years, which are expected to include vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (T2I) communication. Key here is the irony that much of the technology currently sensationalized in lurid headlines about robot cars making life and death decisions, or obscure utilitarian algorithms attaching value to human beings, is some considerable way off. In response, this article proposes three distinct phases of vehicle automation development, differentiated by the capability of the vehicle to categorize objects and the granularity of the data available. To illustrate our thinking, and taking the example of how a machine might categorize humans, we posit that: 1) early stages of technology development afford AV ability to differentiate between objects such as trees, road-signs, and people; 2) in later technological iterations, AV ability will be further refined to distinguish between adults and children; and 3) cumulative AV advances will ultimately result in the introduction of facial recognition. As such, it is evident that only when the vehicle can access highly granular data, do many of the issues raised in the MIT study come to the fore. Moreover, as the technology advances there will doubtless be concurrent debates in civil society as to the desirability of AV ownership, with perhaps even insurers gaining access to this kind of data.


\(^3\)See Slovic et al. (1985)
Any taxonomic classification of AV technology must inevitably be partial, and within a highly dynamic field such as assisted and automated driving, potentially iterative. Nevertheless, the pragmatics of attempting to devise a system of governance around AVs and their future development certainly underscores the need for certain waypoints along the trajectory of technological developments. Reports on governance regimes for AVs across Europe and beyond (Brown et al. 2018; Metz 2018), and indeed, stakeholder responses to such reports, reaffirms the wisdom of a stepped response (Johnson 2017; KPMG 2018; Golias et al. 2019; Talebian et al. 2019). Several contingent issues arise along the trajectory of technology development, not least of which are the necessary liability regimes (Schellekens 2015), and the ethical underpinning of such. The typology of technology and the ethical problems which obtain to each stage of development offer a route to a more structured response to managing the roll-out of this emerging technology. It is held that juxtaposing the capabilities of the technology across time with the commensurate ethical problems is an essential first step on the path to the effective governance of AVs. This article therefore builds on the outputs of the EU-funded research projects, Cloud LSVA and Vi-das⁴, which seek to address governance, ethics, and the data-flows issues implicit in the roll-out of AVs.

In particular, the article engages with case studies developed in these projects as a means to rehearse the ethical issues which arise across the various stages of the AV development.

**Methodology**

As illustrated in Figure 1, in order to distill the expansive research arena, three time domains corresponding to the three tiers of increasing technical ability have been hypothesized: namely, current state-of-the-art; future scenario S1; and future scenario S2. While the two latter stages might be considered viable in 10 and 20 years respectively, unexpected breakthroughs

---

⁴See: https://cloud-lsva.eu/ & http://vi-das.eu/
and hurdles in development may alter the trajectory. As such, the current or anticipated technical capabilities and attendant limitations of each domain are elucidated. Building on this, we present an ethical analysis of each scenario in terms of the key areas of societal concern regarding the impact of AVs. In order to fully avail of the many societal benefits intrinsic to AV technologies, society will need to support their commercial development. Given the crucial understanding that AV technologies are commercial products which afford numerous societal benefits (Bagloee et al. 2016; Litman 2017) and attendant commercial opportunities, questions exploring how far AVs should be supported by fast-tracking policies and the allocation of public funds to support research must also be addressed.

For the moment, funding and policy concessions are the main instruments supporting AV research. However, as AVs exemplify radical and disruptive technologies which require informed and accurate governance regimes, both current concerns around the use of personal and societal data, and future concerns in terms of vehicle agency, provoke a rash of potential dilemmas. Indeed, there is a risk that should AV issues be overlooked by appropriate governance regimes, the potential benefits may be lost. To this end, each of the three scenarios illustrated below highlight the societal/ethical aspects of AVs which require further investigation:

**Current State-of-the-Art (SoA)**

While perception is perhaps the most critical component to facilitating autonomous driving, it simultaneously presents the most complex impediment. Perception enables an AV to continually monitor the surrounding environment, identify safe drivable locations, detect and classify objects, and calculate their location, velocity, and future predicted states (Pendleton et al. 2017). Environmental perception is achieved by acquiring and converting raw data from an on-board suite of vehicle sensors into scene understanding. Such sensors include cameras, stereovision, infrared, radar, ultrasonic (Van Brummelen et al. 2018; Rosique et al. 2019), and light detection and radar (LiDAR). This section therefore delineates how AVs assimilate the environment through vision and LiDAR-based sensing. Vision-based systems are typically used in the detection and classification of road features, other road-users, and objects (Pendleton et al. 2017).

Based on cameras and stereovision, vision-based perception is widely implemented using deep-learning capabilities. Semantic segmentation typically uses fully convolutional networks (FCN) to classify each pixel in an image. Region-based convolutional neural network (R-CNN), Fast R-CNN, and Faster R-CNN are effective and less costly computational implementations of CNN which also detect and classify objects (Girshick 2015; Ren...
et al. 2015). While cameras are obviously integral to object classification, they remain vulnerable to adverse weather (precipitation, fog, etc.), road quality (poor lane-markings, dirt, etc.), and lighting (shadows, reflections, etc.) conditions (Van Brummelen et al. 2018). To combat this, data sources are combined through data fusion techniques to minimize individual shortcomings and maximize overall AV perception capabilities.

LIDAR significantly enhances AV perception capabilities (Schwarz 2010), as LIDAR sensors rotate at high speeds, emitting laser beams to create a sparse 3D point cloud, with each data point signifying a reflection from an objects surface (Rosique et al. 2019). LIDAR is central for object detection and distance estimation but is limited with regard to object recognition. That said, numerous studies have measured the efficacy LIDAR-based object detection and classification algorithms. Yoshioka et al. (2017) found it to identify pedestrians, cyclists, and cars with approximately 90% accuracy.

This concurs with Melotti, Asvadi, and Premebida (2018) use of CNNs on LIDAR data for pedestrian classification, which also attained approximately 90% accuracy. Maturana and Scherer (2015) also apply a 3D CNN, called VoxNet, to LIDAR 3D point cloud data to classify objects with high accuracy. In the same way as cameras, LIDAR is vulnerable to adverse weather conditions including rain and snow. For instance, LIDAR beams which reflect off snowflakes or rainwater, can lead to the AV detection of “phantom obstacles” (Van Brummelen et al. 2018). While it is evident that Machine Learning (ML)-based algorithms have considerably advanced perception capabilities, a number of limitations and concerns persist. ML algorithms such as classifiers hinge on the extensiveness of the training dataset, training approaches, and a decisionality process which is often inexplicable to humans (Koopman and Wagner 2017). Banerjee et al. (2018) observe that mainly in terms of perception, ML were the leading cause of disengagements (roughly 44%) across all manufacturers. That figure includes the detection of traffic lights, lane-markings, and so on. As ML algorithms are statistical by nature they are inherently vulnerable to “unknown unknowns” and false-positive and false-negative situations (Koopman and Wagner 2017).

An important next step involves predicting potential future states of all objects. This is wholly dependent on the accuracy of vehicular environmental perception. Vehicle motion prediction is obtained through physics-maneuver or interaction aware-based motion models (Lefèvre, Vasquez, and Laugier 2014). Such prediction are used to evaluate the situational risk based on the probability of a collision or other risk indicators such as time-to-collision (TTC), time-to-react (TTR), and so on (Lefèvre, Vasquez, and Laugier 2014). As pedestrian motion prediction presents more of a
challenge, intent-based prediction models have been proposed which are based on physical motion data and semantic features (Gu et al. 2016; Habibi, Jaipuria, and How 2018). Yet, even with accurate scene understanding, the intrinsic variability and uncertainty surrounding human driving renders prediction and planning all the more difficult (Fridman et al. 2017).

At present, there are a number of open questions relating to scene understanding, localization, control, trajectory optimization, and higher-level planning decisions for the deployment of fully automated vehicles, even within restricted operational spaces (Fridman et al. 2017). Perception and scene understanding are particularly challenging since they hinge on the improvement of perception accuracy and minimization of sensor limitations (Van Brummelen et al. 2018). State-of-the-art (SoA) scene understanding is largely achieved through learning-based algorithms. These systems are reliant on large-scale manually annotated datasets such as KITTI (Geiger et al. 2013) and Cityscapes (Cordts et al., 2016).

While, for example, the Cityscapes dataset comprises 30 annotated visual classes (Cordts et al. 2016), AVs only focus on the salient features required to enable safe autonomous driving, including drivable regions, cars, trucks, pedestrians, cyclists, and so on. Approaches which learn with limited human guidance however, result in a number of problems. Systems can misrepresent objects outside the training set or defined as a class, leading to potentially dangerous (re)actions. In an unconstrained, real-world driving environment, the tolerable margin of error is small and the number of scenarios unbounded (Fridman et al. 2017). Contemporary systems are not sufficiently proficient to completely remove human input, and until the current challenges have been addressed, a human supervisor remains necessary (Fridman et al. 2017). As such, it follows that a human needs to recognize when the system is failing and be prepared to assume control. However, the transfer of driving control in these circumstances merely compounds the underlying problems.

**SoA: Ethical Issues**

It is clear that current AV technological limitations present a number of risks. While these are partially mitigated by human driver supervision, and are expected to be so for some time, the situation is far from ideal. In reality, the switch between machine/human and vice-versa, represent cognitive workloads which intensify safety concerns. The current state of technology, which rests on shared responsibility between the driver and machine, has generated issues around liability and responsibility. The ethical dilemmas which reside alongside such legalistically gray areas include the migration
of agency from the driver to the vehicle and vice-versa (McDuff 2018). This issue, not only has the potential to reduce moral questions, but also to disrupt the notion of virtue ethics around what is means to be a safe driver.

Goodall interrogates the measurement of AVs in terms of safety (Goodall 2018). One possible metric is to accept the artificially constructed reality of AVs as AV perception. In so doing, the perception deficiencies that give rise to decisional/operational gaps which require human driving intelligence to intercede may be better identified and investigated. But in order to tackle operational gaps through this approach, AV perception must significantly improve. One key issue relates to the inability of AV perception to evaluate scene contexts (Sivaraman and Trivedi 2013). The current SOA AV technologies described above are, for the present, focused on improving how the road environment is represented as a topographical map upon which the AV can assess, plan, and maneuver waypoints (Zhu et al. 2017). As such, the technological capability of human engineers to combine technologies which support an accurate representation of the environment for AI to analyze, classify, and path plan defines the limitation of AVs. Until AV perception reaches a level of proficiency which matches the driving ability of humans without also incorporating significant societal, ethical, and legal tensions, AVs will remain dependent upon human intelligence to fill the perception gaps:

Video detection, machine vision, laser scanning, and all other ways a vehicle can sense the environment will occasionally miss an object in one of three ways: detection (Is it there?), classification (What is it?), and prediction (What will it do next?). (Goodall 2018)

As Noah Goodall argues, the functionality of an AV is dependent upon the accuracy of three data dependent activities: namely, detection of objects in the environment; classification of the objects; and prediction of changing object values, such as speed, movement, and direction. This is similarly emphasized as the basis of robotics as “sense-plan-act” (Bagloee et al. 2016), and as the ability to perceive the environment and take “responsive action” in real-time (Zhu et al. 2017). To more accurately engage with and understand the phenomenon of AV ethics, proper attention should focus not on trolley type questions, but rather on weighing the challenges relating to the design, capacity, limitations, and societal impacts of AV perception/computer vision. The capacity of AVs to perceive the environment and apply intelligence in order to comprehend the objects populating the environment is not only intrinsic to accurate prediction and path-planning; it is also an ethically loaded dimension of AV technology. Any engineering/programming design which instructs a machine to classify objects and human beings differently has intrinsic ethical connotations. Even the ascription of
safety avoidance metrics which privilege dogs above cats is an example of ethical values. The demonstrable point is that if AV perception is to move toward operational independence from human input/supervision, AV perception must incorporate object classes which are augmented by a layer of additional values. In fact, the ability of a machine to learn object classes from human annotated data sets is proven to incorporate existing societal/cultural biases (McDuff 2018). Accordingly, as the culmination of many of the most complex components supporting AV functionality, AV perception is the most challenging aspect to AV research (Bagloee et al. 2016; Fagnant and Kockelman 2015). AV perception is perhaps also the most ethically challenging aspect of AV technologies since it calculates routes according to predetermined or acquired object values. As such, the actions supported by vehicle perception are particularly vulnerable to ethical challenges.

The SOA has now arrived at a point where AV perception consists of 100+ classes of objects. While these are mostly focused on road infrastructure such as signage, stop signs, and traffic lights, they also include variable classes relating to other road-users and vehicle types, such as bicycle, motorcycle, bus, truck, car, and train. It therefore follows that the design of AV perception presents both epistemological and ethical questions. These, in turn, are united by the necessity to grasp the ontology of AV perception as a functional map of relations between designers, law, ethics, and risk. At least three key aspects which define the ethical framing of AV decisions therefore prevail. The first relates to how both the technology and human designers build upon the capacity to design AV perception in such a way for it to function safely. This means that the design team must create a sufficiently robust system of perception. However, as AV perception is determined by the technological capacity to accurately fuse sensor data, this creates a relatively limited representation of the road phenomenon.

Future Scenario 1: (S1) Technology

Facial recognition (FR) technology has matured sufficiently to facilitate deployment in handheld applications ranging from mobile devices to security and surveillance systems. To do so, FR utilizes digital biometric or/and nonbiometric data and advanced machine learning pattern recognition algorithms such as Deep-Learning (Atallah et al. 2018) to provide an extensive review of face recognition comprising age estimation methods, databases, and algorithms, and including overall performance (accuracy rate)

5This figure was confirmed by lead researchers taking part in the VI-Das H2020 project relating to autonomous vehicle perception.

6See: both http://cocodataset.org/ and http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=2d datasets. Both were utilized in training the classification algorithms for VI-DAS vehicle perception.
and limitations. While the drawbacks of such techniques rest on the fact that facial appearance may change with time (age), availability of data (images), and/or processing speed, data collation and computer processing trends are expected to overcome these limitations in the near future. Moreover, body shape recognition corresponding to fitness attributes (Awad et al. 2018) may also be classified.

Sooklal, Hosein, and Teelucksingh (2016) review of body shape classification techniques determines that it this remains a challenging task, that little work has been undertaken in this area, and suggests low-cost alternatives to be used to monitor body shape.

Awad et al. (2018) investigates how individuals prioritize various attributes of a population involved in a car accident and discriminate between species (human versus pets), gender, age, fitness, criminals (pedestrians who cross illegally), and social status. Such attributes could arguably be harnessed to determine AVs reactions to an accident situation. This section therefore presents a brief overview of the limitations of extant technology in respect of variable recognition. Although not considered in the aforementioned article, Human Activity recognition is relevant to AV interaction within dynamic surroundings and the rapidly evolving technologies of robot-human interactions and video surveillance. Social status can be estimated by other data when income and education information is not available. Since deep analysis of fashion images based on clothes feature databases can identify clothing which yield estimates of the socio-economic status of a person, Liu et al. (2016) introduce a large-scale clothes dataset with annotations to enable the development of recognition algorithms and demonstrate the database comparison of different deep networks algorithms. The authors also underscore the current limitations of existing datasets annotation which challenges real-world applications. To this end, a recent review of the recognition of clothes patterns has been compiled by Naik, Shinde, and Thite (2017).

Activity recognition is mainly developed for video surveillance and human-robot interaction. An extensive review by Zhang et al. (2017) on human activity recognition highlights and compares the advances in state-of-the-art techniques and classification methods. The main challenges reported therein are real-time and portability limitations due to constrained computing power. Integration of datasets and architectures (video analysis) will be the next challenge. Ali, Moftah, and Youssif (2018) provide state-of-the-art depth maps-based image representations, feature extraction processes, and classification procedures. As with the aforementioned article, the challenges relating to variation inside one class and distinctions among the activities of various classes are also discussed. Awad et al. (2018) include a number of general moral indicators in their survey, such as gender, sex,
age, socioeconomic status, and so on. Near future technological advances will enable AVs to classify pedestrians according to these attributes using only pattern recognition based on images and video and preprogramed classifiers which bypass personal data protection issues.

**Future Scenario 1: Ethical Considerations**

In light of the potential gains in safety, mobility, and to the environment, AVs are frequently hailed as socially beneficial and ethically justifiable technology. Yet subsumed beneath this optimistic appraisal, the reality is that moving to AV and an intelligent transportation infrastructure comes at a price. Connectivity and the smart networking of the environment and surrounding objects can provide AVs with the necessary data to not only support better edge classification but to also fill in the perception gaps (Sivaraman and Trivedi 2013; Maddox, Sweatman, and Sayer 2015). However, with increasing connectivity comes increasing risks relating to cybersecurity and privacy which generate numerous societal, ethical, and legal tensions. While it is clear that AVs will form an intrinsic part of the social connectivity paradigm and thereby benefit from the available data provided by “V2V, V2I, I2I, V2P” (Maddox, Sweatman, and Sayer 2015), it is also evident that the benefits of AVs are often rehearsed as a justification for supporting increasing connectivity. Such ubiquitous connectivity presents a change in social identity and order and may facilitate a subtle form of surveillance. While the grouping of numerous data sources from the in-vehicle, exterior, and/or personal device provision of multiple data points could bolster greater contextual understanding to AV perception, such information stocks may also leech to third parties. This is not just a challenge to human rights frameworks regarding data ownership and privacy, but also reinforces the risk of escalating reliance on machine decisionality. This is most evident in the largely unregulated domain of the commercial use of facial recognition technologies by companies such as Sensetime, who offer paying clients a profiling service of private citizens.

AVs may well present an opportune mobile data collection technology, but it also carries a clear risk of significant societal intrusion and ethical impact. As with other core businesses, such as retail and services, the data commodification of users and acquired environmental data can offer a more profitable data stream than the actual core service. Data commodification will undoubtedly be part of the future car industry and connected mobility services. Maddox, Sweatman, and Sayer (2015) describe the

---

7See: Ann Arbour Connected Vehicle Test Environment for examples of the technologies under research: http://www.aacvte.org/

8See: https://www.sensetime.com/
datafication of transportation as the “‘Big’ Transportation Data” in the same way Big Data analytics has created a new market of data commodification. Thus, the dependency of AVs functionality upon data will also empower them to become one of the most important data analytics devices, and ripe for commercialization of data (Sivaraman and Trivedi 2013). The 2,000+ classes level of AV innovation will necessitate distinctions between types of pedestrians and basic risk/behavior profiling. It is possible that the AV perception will continue to focus on object avoidance, but at this stage it could also incorporate risk values relating to young and old, children with dogs, dogs with adults, and so on. However, the fact remains that the second phase of classification will be defined by the increasing need for individuated anonymity in order to respond to the overarching impact of unavoidable machine bias. Leaving aside questions around partiality then, axiomatic governance questions obtain to the potential for machines to make decisions which can have profound downstream impacts on human beings. Where errors do occur, they may stem from many layers of weighted bias relating to training which was derived from human annotated data sets which contained not only human bias, and social/cultural bias, but also stressed a reinforcement of the bias through the training process itself. The challenges are most acute in terms of algorithmic bias and attempts to understand and respond to them. The second ethical context of AV perception will therefore concern the need to understand the inherent bias in AV classification Algorithms\(^9\)

**Future Scenario 2: (S2) Technology**

Most autonomous vehicles in practice currently use Deep Neural Networks (DNNs) to discern the complexities of their environment. These DNNs are typically based on Convolutional Neural Networks (CNNs) and use the “long short-term memory” (LSTM) variant of Recurrent Neural Networks (RNNs) to process inputs and generate outputs. However, DNNs are just as fallible as traditional software and can demonstrate incorrect or unexpected behavior which can potentially lead to critical collisions. Given the brisk adaption of DNNs, validation testing of these models has been somewhat delimited (Koopman and Wagner 2016). In response to the rapid uptake of DNNs, Tian et al. (2018) have configured a systematic technique of ‘stress-testing’ recent advancements in machine learning models. This research adds realistic modifications to the input images, such as adverse environmental conditions. By supplementing the images in this way, Tian et al. (2018) maximize the number of neurons being used in a multi-layer

---

perceptron to classify objects. The validation method results in more robust models and the detection of thousands of erroneous behavioral decisions under realistic driving conditions.

However, validation of the DNN models is heavily dependent on the data collected through training or manual allocation. Since Tian’s scenarios are simply manipulations of the data available to the researcher, an inherent limitation exists in the current “stress-test” research. Hoffmann and Payton (2018) may offer a partial solution to the issues posed by AML. They propose a dynamic model selection process which enables the vehicle to choose from a set of simultaneously computed predictive models in order to recover from unexpected scenarios or critical events. By finding the predictive model that most closely corresponds to vehicle’s current state and surroundings, the on-board control system will automatically adopt the new predictive model to generate future control outputs for the vehicle. While previous applications of this system have seen positive results under human motor control (Haruno, Wolpert, and Kawato 2001), they have yet to be deployed in autonomous vehicles which typically use a single model to calculate decisions.

The rapid rise of DNNs used to classify and avoid potential critical events has matured in a short period, in tandem with the significant increase in computing power (Dutta 2018). The systematic ‘stress-testing’ technique proposed by Tian et al. (2018) is intended as a vigorous challenge to ML techniques; thereby driving improvements, robustness, and optimizations to enhance effectiveness. The current implementation of DNNs, which typically comprise a combination of Convolutional Neural Networks (CNNs) and the “Long Short-Term Memory” (LSTM) variant of Recurrent Neural Networks is being developed to encompass contextual output as well as image classification. Miikkulainen et al. (2019) propose a DNN which abstracts ambiguous images and infers situational context from an input image. The natural extension of this system would be a feedback loop of upcoming scenarios to other components of the autonomous vehicle, including the passengers. This has the advantage of offering transparency to a statistical technique which has become synonymous with the prototypical “black box” system, and could be the gateway for informing on, and/or adhering to ethical responsibilities.

A further resource under development is the application of Semantic Intention and Motion Predictions (SIMPs) to autonomous vehicles (Hu, Zhan, and Tomizuka 2018). By utilizing a probabilistic framework based on DNNs, it is possible for approximations of the intentions, final locations, and corresponding time information for surrounding vehicles and obstructions, to be made. This research lays the foundations for an introduction of Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I)
communications, and is further extrapolated by Kim and Liu (2016). By introducing a sudden obstacle seen only by a lead vehicle which communicates this information to following vehicles, the system proposed potentially enables the following vehicle to avoid the obstacle in a timely manner. By introducing V2V communication to avoid an unseen object, the avoidance rate increases from 38 to 96% in simulated scenarios. Cooperative measures in V2V technologies is also studied by During and Lemmer (2016), who conclude that a balance of altruism and egoism is necessary to attain a rational level of cooperative maneuvers to avoid safety-critical situations.

Recent advancements in machine learning algorithms remain imperfect. Although the introduction of V2V and V2I system will ultimately improve traffic flow, overall road, and environmental safety (Kim and Liu 2016; Bento et al. 2019), for the moment, these research methods rely on simulated scenario analyses which, in practice, could suffer from communication inaccuracies. In addition, privacy and connectivity concerns dominate the V2V and V2I problem space, in that multitudes of sensors will be used in tandem with machine learning techniques to navigate autonomously (Karnouskos and Kerschbaum 2018). Given the wealth of data collected on a continual basis within a hyperconnected domain, privacy issues are inevitable in the context of the nature of data being collected and the processing and transfer of unencrypted information. If current trends persist, such privacy and classification issues will be overcome and further levels of complexity added to arrive at a contextually cognizant information processing and transfer system. However, edge cases and complex scenarios will also likewise continue, resulting in problematic challenges for the classifier. Thus, while emergent and established models are adept at responding to complex scenarios, these models will continue to be imperfect. It is arguably probabilities, not complete confidence, which will ultimately underpin the model determinations (Hu, Zhan, and Tomizuka 2018; Miikkulainen et al. 2019).

Future Scenario 2: Ethical Considerations

As connected and intelligence technologies become increasingly embedded within society, the vast volume of data generated will feed a further wave of AV innovation. It is highly probable that the perception challenges associated with AV will remain a significant obstacle to full unsupervised autonomy. Nonetheless, this perception gap of classification may be overcome in part by utilizing secondary supporting technologies built around a connected AV phenomenon. Accordingly, the second phase of AV innovation will be predicated on a radically different social environment than that which we now experience. The increasing connectivity of society, road
infrastructure, buildings, and the geotagging of vehicles, people, and even pets, will generate a digitally connected road environment with an abundance of data relating to object identity, risk, and prediction values. As such, AV perception will be informed by the classification of objects and by communication directly with surrounding objects, to determine identities and access information specific to the AV and environment. To be fully operational, the AV intelligence system will require support in terms of policy. Aligned to this, are the ethical challenges obtaining to the data reliance on the emerging AV, societal, and commercial connectivity. Far richer AV perception will be possible if the necessary governance regimes are instantiated to support data use in this context. This may consist of several thousand object classes, incorporated into the system via a diversity of intelligence, and fused together to provide multiple layers of classification. The utilization of such connectivity, real-time transport management systems, behavioral analytics, and profiling of surrounding people will provide data to AVs which will mitigate many risks. In effect, this means that objects will not be classified as isolated values, but rather will become part of a more complex classification phenomenon consisting of the relational values between objects.

Far wider ethical implications relate to another phase of fully autonomous artificial driving intelligence. This context is not one of personal vehicle ownership, but rather foregrounds a radical new paradigm of the transportation services on offer by commercial actors. The current business models of Waymo, Uber, and others, suggest a decreasing demand in global individuated car ownership and point to a future of autonomous mobility. This model is based on configuring AVs as ride-hailing transport solutions. Full autonomy will remove the human input from the loop in three important ways. Firstly, ML will increasingly rely on unsupervised learning techniques and construct larger data sets which will be increasingly annotated by further ML algorithms. In this way the human in the loop will become less significant in the software development aspect of machine decisionality. Secondly, the human driver will no longer be required to compensate for the perception gaps in the context of AVs, as classification and perception will be exponentially improved and supported by a smart environment and transport system. Such intelligence will be increasingly supported by ubiquitous connectivity and cloud analytics. Thirdly, the growth of social dependency upon commercial transportation will generate risk relating to the authority of regulation and governance to function in an independent and non-coercive manner.

10See: “Alphabet’s Waymo begins charging passengers for self-driving cars” https://www.ft.com/content/7980e98e-d8b6-11e8-a854-33d6f82e62f8
The pressing, and perhaps more fundamental question here, is whether society will be prepared to offset the downsides of AVs, including the relinquishment of embedded rights to privacy and ownership and intrusive profiling and monitoring, in order to support innovation that may well offer lifesaving benefits. Zuboff (2015) delivers a critique in this regard which demonstrates the manifold risks of data capture on the part of large corporate actors. The changing social contract implicit in AVs is effectively achieved via a commercial product. It is evident that this is a product which may threaten the autonomy of society and individuals and potentially increase social dependency on commercial actors. As Fagnant and Kockelman (2015) suggest, personal transport could become safer, but only at the cost of personal privacy. Going forward, the data dependency of technologies such as AVs will necessarily entail a negotiation around the erosion of established human rights in order to realize their much-vaunted safety and utility based benefits. While vast global corporations such as Google and Uber may provide the infrastructure and support many benefits, the ethical questions regarding the third phase rehearse Floridi and Rodella’s (2018) insights into the issue of digital governance and the type of society we want to live in.

**Conclusion**

Anderson and Anderson (2007) elucidate the need for advanced technology to weigh ethical considerations when performing automated tasks. As such, they emphasize the futility of machine learning-based ethical judgments in the absence of understanding the underlying principles which justify any decision. Discussions of the philosophical and societal implications of AI applications have tended to take place at a distance from mathematical and computational approaches (Frank et al. 2019). The gap between these schools of enquiry not only does a disservice to both disciplines, but more importantly, has the potential to adversely color societal assimilations of the new technologies. Since vehicles are synonymous with social mobility and economic activity, and are thoroughly embedded in contemporary culture, a coherent discourse on a mutual apprehension of the technical capabilities and challenges of AVs represents a pivotal starting point. This article presents the current state-of-the-art and two future technical milestones in order to frame current and future ethical considerations. In doing so, we allow for a temporal appraisal of when, and in what form, ethical questions will need to be answered by AV technology. With just 62 companies registered to test AV technologies with a driver present\(^{11}\), current SoA

\(^{11}\)https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/permit
may at best be described as the ‘test phase’ of autonomous driving. As many AV engineers concede, the human driver will remain an essential part of the driving loop and transport phenomenon for some time to come (Fridman et al. 2017). At present SOA is more focused on acting as hybrid human/machine driving phenomenon wherein the technological benefits are supported as a guardian angel (Maurer et al. 2018). This relationship will continue until there is sufficient driving data available to support a more complex algorithmic driving decisional capacity which can cater to the wide array of unknown variables which define the human driving phenomenon. In Future Scenario 1, Facial Recognition matures and broadens to enable AVs to discern human activities, societal status, and even attitudes. We delineated the technical challenges to these accomplishments and posed questions relating to algorithmic biases and data privacy. In our Future Scenario 2, we examined algorithmic limitations in edge cases and demonstrated how increasingly robust techniques are being developed to overcome the concurrent issues of a more connected environment and reduced data privacy without regulatory intervention. While it is clearly worthwhile to engage in anticipatory research to prepare for emerging technologies, this can present risks to research when it distracts from more realistic and pressing impacts of state-of-the-art and emerging technologies. Therefore, we argue that some aspects of AV research, and indeed social perception, has been erroneously focused on pseudo-ethical questions and problems, as is evident in the promise of AVs as key to dramatically reducing global road deaths and the framing of MIT Moral Machine experiment. While both offer insights to far-off future possibilities, they have nonetheless fueled a misplaced and confused picture of the AV timeline in relation to ethical framings.

Acknowledgments

All authors are members of the Vision Inspired Driver Assistance Systems (VI-DAS) H2020 research consortia.

References


