Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

Final draft

STEPS TOWARDS REAL-WORLD ETHICS FOR SELF-DRIVING CARS. BEYOND THE TROLLEY PROBLEM

Tobias Holstein[1,2], Gordana Dodig-Crnknovic[1,3], Patrizio Pelliccione[3,4]

[1]*Mälardalen University, Sweden*
[2]*Hochschule Darmstadt, Germany*
[3]*Chalmers University of Technology | University of Gothenburg, Sweden*
[4]*University of L'Aquila, Italy*

ABSTRACT

Research on self-driving cars is transdisciplinary and its different aspects have attracted interest in general public debates as well as among specialists. To this day, ethical discourses are dominated by the Trolley Problem, a hypothetical ethical dilemma that is by construction unsolvable. It obfuscates much bigger real-world ethical challenges in the design, development, and operation of self-driving cars. We propose a systematic approach that connects processes, components, systems, and stakeholders to analyze the real-world ethical challenges for the ecology of socio-technological system of self-driving cars. We take a closer look at the regulative instruments, standards, design, and implementations of components, systems, and services and we present practical social and ethical challenges that must be met and that imply novel expectations for engineering in car industry.

Keywords: Self-Driving Cars, Autonomous Driving, Smart Cars, Future Transportation, Ethical Principles, Ethical Guidelines, Artificial Intelligence, Responsibility, Sustainability, Safety

INTRODUCTION

Future autonomous (self-driving, driverless, smart) cars are attracting big societal attention as causing the revolution of transport systems that is expected to affect society in profound ways {Formatting Citation}. There is a public debate all around the world about the possibility and desirability of the self-driving cars. The interest of the general public up to now was mostly focused on the machine decision-making. The discussion has been connected to the trolley problem, an idealized and unsolvable (human) decision-making conundrum.

In this chapter, we present ethical and social aspects of the emerging technology of self-driving cars addressed through an applied engineering ethical approach. Instead of the discussion of specific unsolvable hypothetical moral dilemma, we present ethical analysis focused on the study of ethics of complex real-world engineering problems from the systemic perspective of the socio-technological system.

Modern automated cars are steadily increasing their level of automation, from no automation to driver assistance, partial automation, conditional automation, high automation and continue towards full automation or autonomy, as defined by the Society of Automotive Engineers (SAE, 2016) and United

States National Highway Traffic Safety Administration (NHTSA) (National Highway Traffic Safety Administration (NHTSA), 2020). Concrete examples of self-driving cars are the Waymo car (Waymo, 2020) and Cruise (Cruise, 2020).

The intense industry development of increasingly automated cars is accompanied by the interest of many domains, such as engineering and computer science (Aydemir & Dalpiaz, 2018; Dennis, Fisher, Slavkovik, & Webster, 2016; Pelliccione et al., 2017), design and human-computer interaction (Eden, Nanchen, Ramseyer, & Evéquoz, 2017), cognitive science (Zhu & Tang, 2015), sociology (Bissell et al., 2020), behavioral science (Awad, Dsouza, Bonnefon, Shariff, & Rahwan, 2020), and ethics and law (Coca-Vila, 2017). Moreover, they increasingly attract the interest of decision-, policy-, and law-makers (Jobin, Ienca, & Vayena, 2019).

From the engineering and scientific perspectives, technical problems of this development are challenging, but they are successively being solved by an engineering approach. Automation might positively affect the system performance, safety and utility of cars (Favarò, Nader, Eurich, Tripp, & Varadaraju, 2017). Two recent studies that compare crash experiences of automated vs. conventional vehicles show that automated vehicles perform better (Blanco et al., 2016; Schoettle & Sivak, 2015). We might expect that higher levels of autonomy will further increase safety. The process goes via step-wise improved driving capabilities through machine-learning. New capabilities are added to highly automated cars after they have been thoroughly tested under human supervision.

The chapter is structured as follows. After the introduction, the state of the art of the ethical analysis for autonomous cars is presented, with the account of the problems introduced by the focus of the debate on the hypothetical unsolvable trolley problem. It is followed by the argument for the necessity of re-orienting the focus to the real-world ethics of practical importance. The methodology of the current work is outlined in the subsequent section. Identifying ethical challenges in the techno-sociological ecology of self-driving cars is divided into two sections, addressing technical aspects, and social aspects, respectively. Conclusions and future work close the chapter.

STATE OF THE ART OF THE ETHICAL ANALYSIS FOR AUTONOMOUS CARS

The autonomous cars ethics analysis has been introduced through the trolley problem, which has been dominating the debate since then, and has been discussed in a huge number of publications, including IEEE (Goodall, 2016), ACM (Frison, Wintersberger, & Riener, 2016; Kirkpatrick, 2015; McBride, 2016), Scientific American (Deamer, 2016; Greenemeier, 2016; Kuchinskas, 2013), Science (Bonnefon, Shariff, & Rahwan, 2016; Greene, 2016), other publication venues (Coca-Vila, 2017; Goodall, 2014; Goodman & Flaxman, 2017). The topic is also discussed in workshops (Alavi, Bahrami, Verma, & Lalanne, 2017; Riener et al., 2016) and other sources (Mooney, 2016; Shashkevich, 2017).

Some of the authors directly apply trolley problem scenarios, as (Islam & Rashid, 2018) who present a crash-optimization algorithm that takes number, age, and gender of people as input in order to assess the outcomes in case of inescapable accident. Noothigattu et al. use the collected data from the Moral Machine Experiment (MIT Moral Machine Lab, 2016) to implement a decision making algorithm (Noothigattu et al., 2018), while Kim et al. are introducing a computational model by learning and generalizing from moral judgments of humans (Kim et al., 2018). Addressing ethical issues, numerous publications suggest implementing moral principles into algorithms of self-driving cars (Dennis, Fisher, Slavkovik, & Webster, 2014; Dennis et al., 2016; Goodall, 2016).

Alternative approaches, oriented towards real-world engineering problems include (Aydemir & Dalpiaz, 2018) who suggest *an* analytical framework to help stakeholders analyse ethical issues and apply it

towards ethics-aware software engineering. Karnouskos (Karnouskos, 2018) investigates the impact of variety of ethical frameworks on self-driving car acceptance, with the conclusion that *there are currently many intertwined aspects that need to be carefully addressed in an interdisciplinary manner.*

For self-driving cars, we are experiencing the typical "policy-vacuum" problem (Moor, 1985) of computer ethics, which arises in situations for which we lack policies, e.g. we have no experience, no ethics, and no laws. The fact that in the context of self-driving cars everyone is focusing on the Trolley Problem as it would represent something central for self-driving cars has an effect of directing ethical deliberation in the wrong direction. This leads to focusing the public imagination on AI as a decision-maker about life or death of people. However, this idea is fundamentally misguided and it is setting the wrong emphasis on hypothetical unsolvable ethical dilemma instead of relevant challenges for real-world self-driving cars.

Hypothetic vs. Engineering Ethics Problems

Even though in the debate, the trolley problem has been given a very prominent role, there are critical approaches arguing that it *is an ill-suited benchmark for an automated algorithm,* see (Foot, 1967; Mirnig & Meschtscherjakov, 2019; Thomson, 1976). Johansson et al. (Johansson & Nilsson, 2016) add that *the self-driving car shall not be* unprepared in a way that the Trolley Problem suggests. Based on the early data about the road situation, assessments of internal state, it should adjust its own behavior appropriately.

In the following, we present further arguments for the inadequacy of the trolley problem for the analysis of ethical and social aspects of self-driving cars. Beard (Beard, 2019) points out the fact that autonomous cars are not deterministic in the way trolleys are, so that neither the trajectory is known with certainty, nor how the pedestrians will react. Additionally, the trolley problem assumes that self-driving car would have a system that will make a precise and reliable distinction, not only between humans and other kinds of obstacles on the road (which is already a problem), but even distinctions among people. However, there are already principles and laws that forbid to differentiate among humans based on attributes such as age, nation, wealth, social status, gender, etc., in regard to the *right to life*. Therefore, choosing which human one will kill is not an option (Holstein & Dodig-Crnkovic, 2018).

The assumption about the deterministic nature of all the involved processes made in trolley problem scenarios is fundamentally wrong. It means that all the objects have perfectly known positions from which only one perfectly calculable consequence will follow. In the real world with humans, we have a complex system and it is not possible to predict exactly in real-time. We are dealing with statistical phenomena under uncertainty and the only way we can handle those is by constant machine learning.

As a most prominent example, the Moral Machine online experiment (Awad et al., 2018; MIT Moral Machine Lab, 2016) asking people what they would do in different versions of the trolley problem, is about humans and it is not a suitable basis for the design of self-driving cars. Cars should not mimic people, as human drivers are the major cause of accidents. 94% of serious crashes are due to human errors (NHTSA's National Center for Statistics and Analysis, 2018). The sequel of the Moral Machine paper (Awad et al., 2020) presents a new attempt to explain and justify the idea of the Moral Machine experiment as a way to explore the cultural differences between preferences whom to kill - young or old, women or men, and so on. The authors mention that the German code of ethics does not allow making this kind of discrimination of humans (Luetge, 2017). Actually, discrimination is not only rejected by experts but also forbidden by European and UN laws on fundamental human rights. Moral Machine is an experiment about cultural differences in answers of people to the question what they believe they would do in certain traffic situations, but it is not instructive for the design of self-driving cars.

As prototypes of self-driving cars are increasingly participating in public traffic, it is important to investigate how self-driving cars are actually envisaged, designed, developed, and built, how real ethical challenges are addressed, and how decisions in all those stages are justified. Discussing this before self-driving cars are regularly introduced into the traffic, allows taking part in the setting of ethical approaches.

From the emerging normative work (Ethics Commission, 2017; Floridi et al., 2018; High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019; Spiekermann, 2015) with value-based approach, we extract a list of ethic-sensitive technical issues such as safety, security, privacy, trust, responsibility/accountability, quality assurance/auditing, sustainability (many of them connected to AI and machine-learning such as transparency, explainability, fairness, etc.), together with social challenges of disruptive technology with stakeholders interests, legislation, norms and standards, as continuation of our research in the field (Holstein & Dodig-Crnkovic, 2018; Holstein, Dodig-Crnkovic, & Pelliccione, 2018).

We present a systematic conceptual ethical model that connects the different stakeholders as responsible for ethical aspects in the development of fully autonomous self-driving cars. This is an iterative process that involves providers of laws and regulations, society (such as public acceptance), research (e.g., AI-, ML-, engineering- ethics) as well as the actual development (e.g., automotive/sensors/transport industry). It is necessary to establish knowledge exchange between stakeholders, to build the common ground for the solutions for future self-driving cars.

Focusing on the real-world ethical challenges that should currently be addressed is the first step before ethical aspects of self-driving cars can be meaningfully discussed from the point of view of societal, individual and professional/organizational stakeholders. It is important to base our conclusions, not on abstract thought experiments unrelated to autonomous cars, but concrete circumstances of their development and introduction. We should focus on factors we as stakeholders can influence in our different roles, via the design, development, engineering, and organizational solutions.

THE METHODOLOGY

As we study the ethics of emergent technology to identify challenges for the development of autonomous cars, we use a hybrid and interdisciplinary methodology. It builds on insights from ethics theory with the significance of “policy vacuums” that are being filled by adding the most important aspects of the emerging technology. We combine ethical theory literature with the technical characteristics of present-day automated cars and their anticipated developments. Our aim is to address the gaps in the understanding of ethics of automated cars as they develop towards autonomy. From the literature on value-based design and current guidelines (European Group on Ethics in Science and New Technologies (EGE), 2018; Floridi et al., 2018; Friedman & Kahn Jr., 2003; Friedman, Kahn, Borning, & Huldtgren, 2013; High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019; The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019), we extract the list of topics of relevance for real-world automated/self-driving cars that we present in a table form, for both technical and social ethical challenges.
We did not make a full systematic literature review, but we made extensive searches on the most used scientific repositories, like IEEE Explore and ACM Digital Library, as well as more precise searches in leading ethics journals such as Philosophy & Technology, Ethics and Information Technology, and conferences in the field.

Thus, concluding, our hybrid methodology is based on methods from ethics, the typical approach of humanities with logical argument starting from existing theories, applied to the case of self-driving cars, taking into account basic knowledge of the technology of automated cars today.

We discussed our findings with different stakeholders in several steps. First, we presented our findings in Sweden, at seminars at Chalmers University of Technology (interaction design aspects) and at Science festival in Gothenburg for the general public, where we collected reactions from those two groups of important stakeholders. Two more seminars were held in Norway, one at NTNU with focus on software engineering for the software engineering audience, and the more general ethical aspect views presented at The Big Challenges Science Festival workshop Code of Ethics in Trondheim.

Finally, we discussed our findings with four experts - two senior software engineering researchers, and two leading philosophers from different universities in Europe, as well as with a practitioner working in a company involved in the production of self-driving vehicles. We integrated all the experiences from the discussions with different stakeholders and expert advice into the present version of this chapter.

IDENTIFYING ETHICAL CHALLENGES IN THE TECHNO-SOCIOLOGICAL ECOLOGY OF SELF-DRIVING CARS

The concrete ethical challenges of the decision making must consider the current state of the art of technology and its development, also in comparison with traditional human-driven cars. Human drivers are far from perfect and there is a clear expectation that self-driving cars will eventually be much better than humans. But decision making is not the only aspect of autonomous cars that has ethical consequences. We see ethical challenges in the whole ecology of techno-social system. From the choices on the part of stakeholders of what we want self-driving cars to do for us (such as – will they be shared? – how will they fit into “smart” or “intelligent” cities?) to the design decisions, such as whether a certain technology is used because of its low price, even though the decision making would be substantially improved with more expensive technology. Even today sensors are an important issue, but in an envisaged completely autonomous cars they will be essential for the behavior of the vehicle. This also poses the question about hardware updates, and how long self-driving cars that do not fulfil the state-of-the-art safety technology are allowed to be used and how will this safety threshold be defined and measured. Since building and engineering of self-driving vehicles involve various stakeholders, such as designers, software/hardware engineers, salespeople, management, the general public, etc., we will also explore the questions of stakeholder involvement. One of the important questions is how will responsibility in the socio-technological system be distributed and assigned (Dignum, 2019) so to assure maximum benefit for the maximal number of stakeholders, at the same time respecting principles of fairness and justice?

ETHICAL ASPECTS OF THE TECHNICAL CHALLENGES

In the following, we will discuss ethical deliberations regarding specific requirements for autonomous vehicles, regarding technological aspects, including costs versus quality. The multifaceted and complex nature of reality emphasizes the importance to look broader and from the transdisciplinary, systemic point of view, for each of the requirements on expected properties of autonomous vehicles.

A decision-making process implemented in a self-driving car can be summarized as follows. It starts with a detection of the environment, identifying obstacles (nearby objects, including humans and animals), as well as the current context/situation of the car, using external systems such as GPS, maps, street signs, connection to other cars (V2V) or available infrastructure (V2X), or locally measurable information

(speed, direction, etc.). Engineering ethics is focused on the assurance of hardware and software involved in the function of the car, as well as maintenance of ethical standards of the socio-technological system.

Safety

Safety is the most fundamental requirement of autonomous cars. The central question is: how should a self-driving car be made safe and its safety reliably tested? What guidelines should be fulfilled to ensure that it is safe?

Leveson (Leveson, 2020) provides an excellent advice regarding safety of hardware-software systems, by presenting the list of major misconceptions regarding safety, such as common believes that: software itself can be unsafe; reliable systems are safe; the safety of components in a complex system is a useful concept and we can analyse the safety of software in isolation from the system design; software can be shown to be safe by testing, simulation, or standard formal verification. According to Leveson, creating the safety-related requirements is the most effective approach to dealing with the safety of computer-controlled systems. It is necessary to take a system-approach and include both controls implemented in software and hardware and the ones delegated to human controllers, organizations and social controls.

Currently, there are several standards, such as the ISO 26262 (ISO, 2011), that specify the safety standard for road vehicles. For self-driving cars standards, such as the ISO/PAS 21448 (ISO, 2019), which is also known as Safety of the Intended Functionality (SOTIF), are under development.

As an argument for safety of their cars, Waymo stated that since its start a decade ago, they have *more than 10 million miles on the road, 7 billion miles in simulation* testing their cars under diverse circumstances (Hersman, 2019). But is this enough to certify their software? Kalra and Paddock calculated the "*way to safety"* concluding that *hundreds of millions of miles and sometimes hundreds of billions of miles* are necessary (Kalra & Paddock, 2016).

The source code of autonomous cars and its components are typically proprietary and not publicly available. The legislators have chosen, instead of controlling the software, to focus on the behavior of a vehicle that is being tested, based on the "*Proven in use*" Argument. It directly connects to the necessity pointed out by Leveson (Leveson, 2020) of testing the whole software-hardware system.

As testing of present-day cars should demonstrate the compliance of their behavior with legislative norms, it is important to have detailed data about their behavior on the road. The DMV sets an example an provides collision and annual reports online, covering multiple manufacturers (Department of Motor Vehicles (State of California), 2020a). Faverò et al. present an in-depth analysis of accident reports (DMV data from 2014 to 2017) in (Favarò et al., 2017). Disengagements, accidents and reaction times based on data released in 2016 from the California trials are discussed in (Dixit, Chand, & Nair, 2016). As autonomous cars are learning from experience, constant improvements connected to continuous software and hardware development, present a new challenge.

Related to safety, when it comes to hardware and hardware-software systems, there have been discussions about the prices of different equipment such as laser radars compared to cameras or ultra-sonic sensors. Laser radars are very expensive but deliver high-quality data in diverse weather conditions. Ultra-sonic sensors or cameras are less accurate and sensitive under weather conditions like rain. A study of (Combs, Sandt, Clamann, & McDonald, 2019) compared different sensors and analyzed the number of pedestrians killed in the accidents which would not happen if autonomous cars were involved. The capacity of adequate detection of pedestrians is in the range from <30% to >90% of analyzed cases. They point out

that the price of the best sensor technologies *may be unrealistically expensive.* However, it is often the case that the prices of advanced technologies get to affordable levels quite quickly with its wide use.

One of interesting questions for the safety of autonomous cars is the possibility for the people to intervene if something goes wrong. In advanced driving assistance systems, the driver would take over if a critical situation could not be handled by the system. What would happen in a self-driving car? Will passengers be allowed to intervene? Under which conditions? Would the police have a possibility to intervene, and stop the car, when it behaves inadequately or dangerously? Learning from experience is the most important basis for the improvement of safety in self-driving cars. Tesla CEO Elon Musk envisages a vehicle self-driving capability that is ten times safer than manual, developed by massive fleet learning (Musk, 2016).

Both for security and for safety it makes the difference if the vehicle is connected to the infrastructure and other vehicles or completely disconnected. Connected vehicles might receive information from other systems that will enhance the understanding of the reality thus opening new and promising safety scenarios.

Security

For autonomous cars, security is of paramount importance, and software security is a fundamental requirement. As an indication of the development, we mention that in August 2017 UK's Department for Transport published the document “Key principles of vehicle cybersecurity for connected and automated vehicles” (Department for Transport (DfT) & Centre for the Protection of National Infrastructure (CPNI), 2017). Similar documents have been developed, such as Microsoft Security Development Lifecycle (SDL), SAFE Code best practices, OWASP Comprehensive, lightweight application security process (CLASP), and HMG Security policy framework, mentioned in (Department for Transport (DfT) & Centre for the Protection of National Infrastructure (CPNI), 2017). In order to break security of the existing advanced automated cars, there have been a number of attacks at car systems and sensors (e.g., LIDAR and GPS) that were used to manipulate the behavior of the car. Attacks might be inevitable in the real life, which actualizes the question: should there be a minimum-security threshold for a self-driving car to be used? This leads to another question: How secure must the systems and its connections be? In the case of accidents, in aircrafts “black boxes” are used after a crash to determine what happened. Should this also be a part of a self-driving car? Software updates can bring security issues. Should a self-driving car be allowed to drive, if it does not have the latest software version running? It is also important to regulate how are bugs in the new software are handled.

Security in relation to the connected driving brings new challenges as well as new possibilities. On one side, the most secure system is the one that is disconnected from the network. On the other side, it would be unethical not to deploy new software or a new version of the software in the car if there is evidence that the new update will fix important problems.
In order to enable massive fleet learning and to enable software updates, connectivity is needed.

Privacy and Personal Integrity

The more information taken into consideration for the decision making, the more it might interfere with the data and privacy protection. For example, a sensor that detects obstacles, such as human beings in front of the car is based on visual information. Even the use of a single sensor could invade privacy if the data is recorded/reported and/or distributed without the consent of the involved people. The general question that regulation should answer is: How much data is the car supposed to collect for the decision

making? Who will have access to those data? When will these data be destroyed? In Europe, this refers to the General Data Protection Regulation (GDPR) (European Union, 2016).

Following and applying legal frameworks to protect personal data, such as regulation (EU) 2016/679 of the European Parliament (European Union, 2016), also discussed in (Wachter, Mittelstadt, & Floridi, 2017), concerns questions of privacy and personal integrity. It includes the idea of using devices, such as mobile phones, that could send active signals to the surrounding environment which the car is connected to, in order to improve obstacle detection and awareness. People who do not carry such devices would not be be possible to detected that way, which should be taken into account. Questions related to privacy are how much data is used for evaluation, whether it is anonymous and whether it contains more data than "just" the position of a human? Can it be connected to other types of data like the phone number, the bank account, the credit cards, personal details, or health data? Those and similar questions are met by legislations such as Regulation (EU) 2016/679 of the European Parliament and of the Council (the General Data Protection Regulation) setting a legal framework to protect personal data (European Union, 2016), which is discussed in (Wachter et al., 2017).

Transparency

The transparency is of central importance for many of the previously introduced challenges as well as for social sustainability of the techno-social ecology. Without transparency, none of them could be analyzed, because the important information would be missing. According to (McBride, 2016), Transparency is a precondition for the possibility of ethical development of this technology. It is a multi-disciplinary challenge to ensure transparency while respecting e.g., copyright, corporate secrets, security concerns and many other related topics. How much should be disclosed, and disclosed to whom? The car development ecosystem includes many companies acting as suppliers that produce both software and hardware components. In what way and for whom should the entire ecosystem be transparent? Some initial formulations are already present in the current policy documents and initial legislative that will be discussed later on.

The problem of the development and sharing of knowledge on automated and connected driving, and the necessity of interoperability and common guidelines is addressed in the Declaration of Amsterdam (Ministry of Infrastructure & Environment – Netherlands, 2017) adopting a "learning by experience" approach.

Algorithmic decision-making and a "right to explanation" are part of the privacy and integrity complex covered by EU regulations expressing the right for a user to ask for an explanation of an algorithmic (machine) decision that was made about them (Goodman & Flaxman, 2017).

The Department of Motor Vehicles provides the law requirements (Department of Motor Vehicles (State of California), 2020a) *Under the testing regulations, manufacturers are required to provide DMV with a Report of Traffic Accident Involving an Autonomous Vehicle (form OL 316) within 10 business days of the incident*. The list of all incidents can be found in (Department of Motor Vehicles (State of California), 2020b).

Algorithmic Fairness

Algorithmic decision making is required to be fair, and not to discriminate on the grounds of race, gender, age, wealth, social status, etc. This requirement is related to transparency of decision making and expectation of explainability of the ground for decision making. The quality of recognition algorithms, i.e. their capability to detect human obstacles (Wilson, Hoffman, & Morgenstern, 2019) is central.

Reliability

Besides systems related to reliability of classical cars, such as starters, fuel injection, headlights, anti-lock braking systems (ABS), automatic transmission control, airbags, emission controls, or collision detection radar, autonomous vehicles will heavily rely on advanced driver assistance systems, Wi-Fi connectivity, and vehicle-to-vehicle communication (V2V). Emerging smart vehicles contain processing communication modules, such as parallel processor, Ethernet controller, cell modem, Wi-Fi controller, data storage as well as human-machine interfaces (HMI) displays and screens. Some of the basic questions are: what if sensors fail? What level of redundancy is necessary? Is there a threshold that determines when the car is no longer reliable, in terms of component fail? How reliable is the cell network? What if there is no mobile network available? A major issue with the connected cars is their vulnerability to hacking. The car must be able to deal with incorrect data, broken communications, including “denial of service” attacks. Many of issues above are safety-critical and it is of highest importance to develop a reliability-aware culture in product design and subsequent phases of the entire system lifecycle.

Environmental Sustainability

Environmental sustainability is expected to permeate all steps of the sociotechnical process from the system design and development, to operations and management of smart cars.

Intelligent Behavior

The main difference between classical cars and automated/autonomous ones is the intelligent behaviour. Thus, one of the approaches to the ethics of real-world autonomous cars is via artificial intelligence (AI) ethics. As the worldwide overview of AI ethical guidelines given in (Jobin et al., 2019), the following ethical principles are common for all 84 guidelines analysed, where the number in the parenthesis indicates how many guidelines contain this principle explicitly: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, sustainability, dignity and solidarity.

Quality

Detailed quality assurance (QA) program covering all relevant steps in the lifecycle of a car must be developed to ensure high-quality functionality that holds ethical challenges. A non-autonomous vehicle today has often more than 100 electronic control units and this makes it very complex (Pelliccione et al., 2017). For the self-driving functionality complexity will increase. As it is argued in (Sapienza, Dodig-Crnkovic, & Crnkovic, 2016), ethical aspects, today implicit, should be made visible as a part of the process of design and development; this requires developing ethics-aware decision making in all processes.

Transdisciplinarity - Systemic approach

Experiences from advanced modern technologies show vital importance of transdisciplinary collaboration between diverse involved parties, disciplines and stages in the design, development, implementation, testing, verification, maintenance, etc. Transdisciplinary collaboration contributes to the assurance of system-level properties in accordance with the values and ethical principles for both technological and social side of the process. The most important ethical aspects of technical challenges grouped by requirement are given in the following table.

Table 1. Summary of the technical challenges and approaches, grouped by requirement

Requirements	Challenges	Approaches
Safety	Hardware and software adequacy. Vulnerabilities of machine-learning algorithms. Trade-offs between safety and other factors (like economic). Possibility of intervention in self-driving cars (including for the Police). Systemic solutions to guarantee safety in organizations (regulations, authorities, safety culture).	Setting safety as the first priority. Learning from the history of automation. Learning from driving experience - perception and input interpretation processes. Specification of how a self-driving car will behave in cases when the car is not able to operate autonomously. Clarification of the role of the police. Regulations, guidelines, standards being developed as the technology develops.
Security	Minimal necessary security requirements for deployment of self-driving cars. Security in systems and connections. Deployment of software updates. Storing and using received and generated data in a secure way.	Technical solutions that will guarantee minimum security under all foreseeable circumstances. Anticipation and prevention of the worst-case scenarios regarding security breaches. Provide active security. Accessibility of all data, even in the case of accidents, has to be provided, so that it can be analysed to foster knowledge and to provide facts for next generation developments.
Privacy	Trade-offs between privacy and data collection/recording and storage/sharing.	Following/applying legal frameworks to protect personal data, such as GDPR.
Transparency	Information disclosure, what and to whom. Transparency of algorithmic decision making. Transparency in the techno-social ecosystem.	Assurance of transparency and insight into decision making. Active sharing of knowledge to ensure the interoperability of systems and services.
Algorithmic Fairness	Algorithmic decision making is required to be fair and not to discriminate on the grounds of race, gender, age, wealth, social status etc.	This requirement is related to transparency of decision making and expectation of explainability of the ground for decision making.
Reliability	Reliability of sensors and software and need for redundancy. Reliability of required networks and solution for the case when the network is unavailable.	Definition of different levels for reliability, such as diagnostics, vehicle input sensors, software, and external services, set the ground for reliability measures of the car as a system and its components. Standardized process required to shift from fail-safe to fail-operational architecture.
Environmental Sustainability	Environmental sustainability ethics refers to new ways of production, use, and recycling for autonomous vehicles.	Production, use, and disposal/recycling of technology rises sustainability issues (batteries, car sharing) that must be addressed.
Intelligent behavior control	Intelligent behavior may lead to unpredictable situations resulting from learning and autonomous decision making.	Development of self-explaining capability and other features ensuring desired behavior in intelligent software.
Transdisciplinarity - Systemic approach	Ethics in design, requirements engineering, software-hardware development, learning, legal and social aspects, software-hardware interplay.	Adoption of transdisciplinarity and system approaches is increasing and should be given even more prominent role.
Quality	Quality of components. Quality of decision making. Lifetime and maintenance. QA process. Adherence to ethical principles/guidelines.	Ethical deliberations included in the whole process starting with design and development. Ethics-aware decision making to ensure ethically justified decisions.

ETHICAL ASPECTS OF SOCIAL CHALLENGES

The emergence of self-driving brings social challenges as well. Autonomous cars will influence job markets, for example for taxi- or truck- drivers. The perception of cars will change, and cars might be seen as a service that users pay for and no longer as a goods that users buy. The idea of vehicles specialized for the specific use, e.g. off-road, city road, or long travels might become attractive. This might impact the business models of car manufacturers and their markets. This poses ethical problems: what strategy should be applied for people losing jobs because of the transition to self-driving cars? It is expected that the accident frequency will decrease rapidly, so car insurances may become less important. This may affect insurance companies in terms of jobs and business. There is a historical parallel with the process of industrialization and automatization, and there are experiences that may help anticipate and better plan for the process of transition.

Stakeholders Involvement

Stakeholders concerns must be taken into account in the development of emergent technology, which means involvement of professional groups as well as users, and general public. Stakeholders should share information and base their opinions on adequate information.
From the user's perspective, the possible choices given to the user are of interest. How for example will a route be planned? In an extreme scenario, self-driving cars might even avoid or reject some route. Would that be an interference with the freedom of choice, will passengers be informed about the reasons for such decisions? It is important to determine how much control the human should have, that will be taken into account when making design choices for a self-driving car. Here transparency of the system for the user is critical. The question that arises in this context is the number of choices users will have towards control, route planning, and other services of the self-driving car, which we have also discussed in focus of fairness in (Holstein & Dodig-Crnkovic, 2018). For example, Tesla provides different predefined settings for the autopilot (Tesla, 2018), which represent different behaviour regarding speed and lane change suggestions.

Non-maleficence

This requirement of doing no harm becomes especially important in the case of smart and autonomous cars. The first priority is not to harm people inside or outside the car. However, potential users will have different expectations, depending on who will own the cars - companies, social institutions, or individual users, as they all have different preferences. Among those preferences, besides protection of humans, environmental and social sustainability criteria can be expected to play central role so do no harm to the environment or society.

Beneficence

Beneficence is a stronger requirement than non-maleficence. Technology is expected to do good, such as a United Nations "AI for Good" platform ('AI for Good', 2020) or Microsoft "AI for good" initiative (AI for Earth, Health, Accessibility, Humanitarian Action and Cultural Heritage). Autonomous cars can actively contribute to sustainability goals and increase accessibility of transport.

Responsibility and Accountability

Responsibility refers to the role of people themselves and to the capability of AI systems to answer for one's decision and identify errors or unexpected results. Accountability is the need to explain and justify one's decisions and actions. The question how responsibility in the socio-technological system will be

distributed and assigned so to assure maximum benefit for the maximal number of stakeholders, at the same time respecting principles of fairness and justice is addressed by (Dignum, 2019). Regarding ethical aspects of responsibility, a lot can be learned from the existing Roboethics and the debate about responsibility in autonomous robots, e.g., (Dodig-Crnkovic & Persson, 2008). This is still an open problem even though important steps forward are being made by legislators, such as mentioned "Key principles of vehicle cyber security for connected and automated vehicles" (Department for Transport (DfT) & Centre for the Protection of National Infrastructure (CPNI), 2017).

Freedom and Autonomy

Freedom and personal autonomy are essential part of human rights. Every person has to autonomy and is free to make decisions. The Universal Declaration of Human Rights (1948) is the foundation of international law for all people. The Universal Declaration principles protecting human rights are incorporated in the laws of more than ninety countries globally. Human freedom and autonomy are now being defined in relation to intelligent decision-making machines, such as self-driving cars.

Social Sustainability

Social sustainability includes according to Wikipedia, such topics as: social equity, livability, health equity, community development, social capital, social support, human rights, labor rights, placemaking, social responsibility, social justice, cultural competence, community resilience, and human adaptation.

According to United Nations Global Compact, in the domain of business, *social sustainability is about identifying and managing business impacts, both positive and negative, on people*. Both the broader and the business-oriented view of social sustainability are relevant for the field of self-driving cars.

Social Trust

Trust is an issue that appears in various forms in autonomous cars e.g. in production (when assembled, trust is the requirement for both hardware and software components) as well as in use of the car. A human might define where the car has to go, but the self-driving car will make the decisions on how to get there, following the given rules and laws. However, the self-driving car might already distribute data like the target location to a number of external services in order to receive traffic information or navigation data, which are used in the calculation of the route. But how trustworthy are those data sources (e.g., GPS, map data, external devices, other vehicles), the sensors and other hardware and how can trust be implemented, when so many different systems are involved? Also, the opposite can pose a challenge, when untrustworthy systems are already in place and have to be identified.

Social trust may evolve by fulfilling multiple requirements, such as accountability, transparency, diversity, non-discrimination and fairness, as suggested by (High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019) in their guidelines for trustworthy AI.

Social Fairness

Fairness refers to absence of bias and treating people equally. Even if it is applicable to organisations and relationships between humans, in the context of autonomous cars fairness refers most often to the quality of decision-making algorithms, but also to the quality of recognition algorithms, and other components contributing to the decision (Holstein & Dodig-Crnkovic, 2018).

Dignity and Solidarity

In the overview of AI ethical guidelines world-wide, given in (Jobin et al., 2019), dignity and solidarity are one among ten most frequent principles of AI ethics in the contemporary guidelines. Applied to autonomous cars it can refer to respect for and solidarity with humans who are affected by negative consequences of the emergent technology such as unemployment.

Justice: Legislation, Standards, and Guidelines

Present-day regulatory instruments for transportation systems are based on the assumption of human-driven vehicles. As the development and introduction of increasingly automated and connected cars proceed, from no automation at all (level 0) towards full automation (level 5), legislation needs constant updates (Ethics Commission, 2017; *Ethics Commission on Automated Driving presents report: First guidelines in the world for self-driving computers*, 2017; National Highway Traffic Safety Administration (NHTSA), 2020; Pillath, 2016). It has been recognized that state regulatory instruments and the existing NHTSA authority for human-controlled vehicles will not be adequate for self-driving cars and NHTSA is constantly evaluating and updating its regulations in order to provide up-to-date guidelines, which meet the challenges of autonomous cars, while technologies advance (National Highway Traffic Safety Administration (NHTSA), 2020).

The Declaration of Amsterdam (Ministry of Infrastructure & Environment – Netherlands, 2017) addresses legislation frameworks, use of data, liability, exchange of knowledge and cross-border testing for the emerging technology. It prepares a European framework for the implementation of interoperable connected and automated vehicles (Ethics Commission, 2017). It also considers the roles of stakeholders. Terminology has been developed in order to facilitate communication between technology and politics domains, with the definition of levels of automation in vehicles (Pillath, 2016).

The question is thus how to ensure that self-driving cars will be built upon ethical guidelines, which will be adopted by society. The strategy is to rely on rigorously monitoring the behaviour of cars, while the details of implementation are within the responsibility of producers. That means among others that design and implementation of software should follow ethical guidelines. An example of ethical guidelines trying to think one step further is described in the book Ethical IT innovation (Spiekermann, 2015).

The approach based on “learning by experience” and “Proven in use” argument (Ministry of Infrastructure & Environment – Netherlands, 2017; Schäbe & Braband, 2015; *What is the ISO 26262 Functional Safety Standard?*, 2014) presupposes a functioning socio-technological assurance system that has strong coupling among legislation, guidelines, standards and use, and promptly adapts to lessons learned. Ethical analysis in (Dodig Crnkovic & Çürüklü, 2012; Johnsen, Dodig-Crnkovic, Lundqvist, Hanninen, & Pettersson, 2017; Thekkilakattil & Dodig-Crnkovic, 2015) addresses this problem of establishing and maintaining a functioning learning socio-technological system, while Johnsen et al. discuss why functional safety standards are not enough (Johnsen et al., 2017).
The most important ethical aspects of social challenges, grouped by requirement, are given in the following table.

Table 2. Summary of social challenges and approaches, grouped by requirement

Requirements	Challenges	Approaches
Non-maleficence	Technology not causing harm. Disruptive changes on the labor market. Change of related markets and business models (e.g., insurances, manufacturers).	Partly covered by technical solutions. Preparation of strategic solutions for people losing jobs. Learning from historic parallels to industrialization and automatization.
Stakeholders involvement	In this field different stakeholders are involved – from professionals designing, developing, maintaining cars, to their users, and general public.	Active involvement of stakeholders in the process of design and requirements specification as well as decisions of their use.
Beneficence	Values and priorities: Ensure that general public values will be embodied in the technology, with interests of minorities taken into account.	Initiatives as "AI for good" exemplify this expectation that new technology not only do not cause harm, but actively do good for its stakeholders.
Responsibility and Accountability	Assignment and distribution of responsibility and accountability are among central regulative mechanisms for the development of new technology. They should follow ethical principles.	The Accountability, Responsibility and Transparency (ART) principle based on a *Design for Values* approach includes human values and ethical principles in the design processes (Dignum, 2019).
Freedom and Autonomy	Freedom of choice hindered by the system (e.g. it may not allow to drive into a certain area)	The freedom of choice determined by regulations. Determination and communication of the amount of control a human has in context of the self-driving car
Social Sustainability	In the domain of business, social sustainability is about identifying and managing business impacts on people	Pursuing social equity, community development, social support, human rights, labour rights, social responsibility, social justice, etc.
Social Fairness	Ascertaining fairness of the socio-technological system.	Fairness of the decision-making. Related to transparency and explainability.
Dignity and Solidarity	This requirement refers to the entire socio-technological system.	Challenges come from the lack of common wholistic view.
Social Trust	Establishing trust between humans and highly automated vehicles as well as within the social system.	Further research on how to implement trust across multiple systems. Provision of trusted connections between components as well as external services
Justice: legislation, standards, norms, policies and guidelines	Keeping legislation up to date with current level of automated driving, and emergence of self-driving cars. Creating and defining global legislation frameworks. Including ethical guidelines in design and development processes	Legislative support and contribution to global frameworks. Ethics training for involved engineers. Establishment and maintenance of a functioning socio-technological system in addition to functional safety standards

CONCLUSIONS AND FUTURE WORK

Self-driving vehicles have been envisioned as the future of transportation systems and will be successively introduced into the transport systems globally (*Ethics Commission on Automated Driving presents report: First guidelines in the world for self-driving computers*, 2017; National Highway Traffic Safety Administration (NHTSA), 2020; Pillath, 2016). It is time to start an investigation into the manifold of ethical challenges surrounding self-driving and connected vehicles (Ethics Commission, 2017). As this new technology is being gradually allowed on public roads under controlled conditions, the focus is on the practical technological solutions and their social consequences, rather than on idealized unsolvable problems such as much discussed Trolley Problem. Unlike idealized thought experiments, the real-world engineering ethics deliberation involves characteristics of the whole techno-social system supporting new technology, with the emphasis on maximizing learning from experience, on machine-, individual-, and social-level (Charisi et al., 2017; Dodig Crnkovic & Çürüklü, 2012).

For the future a lot of systemic transdisciplinary work remains to be done. A system-level analysis is crucial, as pointed out by Leveson (Leveson, 2020) who argues that the analysis of software in isolation does not guarantee the safety of the whole software-hardware system. The decision-making process and its implementation, which are central to the behaviour of a car, might use unreliable, insecure or inadequate hardware technology, as we described in our earlier study (Holstein et al., 2018). Leveson (Leveson, 2011, 2020) argues that in present-day technology we can no longer separate engineering from human, social and organizational factors, but we have to take a systemic view. This is achieved through the following: collaboration, systems thinking, solving real problems guided by the needs of stakeholders, communication and cooperation, technology transfer from research to practice, design, development, and management of the entire system lifecycle within a socio-technical view.

It is necessary to develop more elaborated ethical principles, guidelines, and analyses, as well as regulatory and legal documents that involve all stakeholders. This affects all stages - from the existing to the new regulatory infrastructure to the requirements engineering, development, implementation, testing and verification and back to the regulatory structure in the iterative process of continuous improvement (Charisi et al., 2017; Greene, 2016; Mooney, 2016). It is also necessary to ensure the transparency of those processes so that independent evaluations become possible.

Finally, it is important to point out the total ecology of the socio-technological system, where ethics is ensured through education, constant information sharing and negotiation of priorities in the value system. In the development process, values and ethics come first, then follows legislation and standardization processes, which are constantly monitored in practice and validated with value- and ethical standards. That is why we emphasize the central role of real-life system-level ethics as a basis that will sustain and inform ethically sound emerging technology of autonomous cars.

ACKNOWLEDGEMENT

One of the authors (P.P.) acknowledges financial support from the Centre of EXcellence on Connected, Geo-Localized and Cybersecure Vehicle (EX-Emerge), funded by Italian Government under CIPE resolution n. 70/2017 (Aug. 7, 2017).

REFERENCES

AI for Good. (2020). Retrieved 8 March 2020, from https://en.wikipedia.org/w/index.php?title=AI_for_Good&oldid=933562584

Alavi, H. S., Bahrami, F., Verma, H., & Lalanne, D. (2017). Is Driverless Car Another Weiserian Mistake? In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems* (pp. 249–253). New York, NY, USA: ACM. https://doi.org/10.1145/3064857.3079155

Awad, E., Dsouza, S., Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2020). Crowdsourcing Moral Machines. *Commun. ACM*, *63*(3), 48–55. https://doi.org/10.1145/3339904

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., … Rahwan, I. (2018). The Moral Machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Aydemir, F. B., & Dalpiaz, F. (2018). A Roadmap for Ethics-Aware Software Engineering. In *2018 IEEE/ACM International Workshop on Software Fairness (FairWare)* (pp. 15–21). https://doi.org/10.23919/FAIRWARE.2018.8452915

Beard, S. (2019, September). The problem with the trolley problem. *Quartz*. Quartz. Retrieved from https://qz.com/1716107/the-problem-with-the-trolley-problem/

Bissell, D., Birtchnell, T., Elliott, A., & Hsu, E. L. (2020). Autonomous automobilities: The social impacts of driverless vehicles. *Current Sociology*, *68*(1), 116–134. https://doi.org/10.1177/0011392118816743

Blanco, M., Atwood, J., Russell, S., Trimble, T., McClafferty, J., & Perez, M. (2016). Automated Vehicle Crash Rate Comparison Using Naturalistic Data. *Vtti*. https://doi.org/10.13140/RG.2.1.2336.1048

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654

Charisi, V., Dennis, L. A., Fisher, M., Lieck, R., Matthias, A., Slavkovik, M., … Yampolskiy, R. (2017). Towards Moral Autonomous Systems. *CoRR*, *abs/1703.0*. Retrieved from http://arxiv.org/abs/1703.04741

Coca-Vila, I. (2017). Self-driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law. *Criminal Law and Philosophy*. https://doi.org/10.1007/s11572-017-9411-3

Combs, T. S., Sandt, L. S., Clamann, M. P., & McDonald, N. C. (2019). Automated Vehicles and Pedestrian Safety: Exploring the Promise and Limits of Pedestrian Detection. *American Journal of Preventive Medicine*, *56*(1), 1–7. https://doi.org/10.1016/j.amepre.2018.06.024

Cruise. (2020). Cruise - Create what's next. Retrieved 7 March 2020, from https://www.getcruise.com/technology

Deamer, K. (2016, July 1). What the First Driverless Car Fatality Means for Self-Driving Tech. Retrieved from https://www.scientificamerican.com/article/what-the-first-driverless-car-fatality-means-for-self-driving-tech/

Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2014). Ethical choice in unforeseen circumstances. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Vol. 8069 LNAI, pp. 433–445). https://doi.org/10.1007/978-3-662-43645-5_45

Dennis, L., Fisher, M., Slavkovik, M., & Webster, M. (2016). Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, *77*, 1–14. https://doi.org/10.1016/j.robot.2015.11.012

Department for Transport (DfT), & Centre for the Protection of National Infrastructure (CPNI). (2017). *The Key Principles of Cyber Security for Connected and Automated Vehicles*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/624302/cyber-security-connected-automated-vehicles-key-principles.pdf

Department of Motor Vehicles (State of California). (2020a). Testing of Autonomous Vehicles. Retrieved 7 March 2020, from https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing

Department of Motor Vehicles (State of California). (2020b, February 25). Report of Traffic Collision Involving an Autonomous Vehicle (OL 316). Retrieved 8 March 2020, from https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/autonomousveh_ol316

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Springer. https://doi.org/10.1007/978-3-030-30371-6

Dixit, V. V, Chand, S., & Nair, D. J. (2016). Autonomous Vehicles: Disengagements, Accidents and Reaction Times. *PLOS ONE*, *11*(12), 1–14. https://doi.org/10.1371/journal.pone.0168054

Dodig-Crnkovic, G., & Persson, D. (2008). Sharing Moral Responsibility with Robots: A Pragmatic Approach. In *Proceedings of the 2008 Conference on Tenth Scandinavian Conference on Artificial Intelligence: SCAI 2008* (pp. 165–168). Amsterdam, The Netherlands, The Netherlands: IOS Press. https://doi.org/10.3233/978-1-58603-867-0-165

Dodig Crnkovic, G., & Çürüklü, B. (2012). Robots: ethical by design. *Ethics and Information Technology*, *14*(1), 61–71. https://doi.org/10.1007/s10676-011-9278-2

Eden, G., Nanchen, B., Ramseyer, R., & Evéquoz, F. (2017). On the Road with an Autonomous Passenger Shuttle: Integration in Public Spaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1569–1576). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3027063.3053126

Ethics Commission. (2017). *Automated and Connected Driving*. Retrieved from https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

Ethics Commission on Automated Driving presents report: First guidelines in the world for self-driving computers. (2017). *online*. Retrieved from https://www.bmvi.de/SharedDocs/EN/PressRelease/2017/084-ethic-commission-report-automated-driving.html

European Group on Ethics in Science and New Technologies (EGE). (2018). *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*. European Commission. https://doi.org/10.2777/78651

European Union. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Da*. Retrieved from http://data.europa.eu/eli/reg/2016/679/oj

Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M., & Varadaraju, N. (2017). Examining accident reports involving autonomous vehicles in California. *PLOS ONE*, *12*(9), 1–20. https://doi.org/10.1371/journal.pone.0184952

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People---An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707.

Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review*, *5*.

Fraedrich, E., Kröger, L., Bahamonde-Birke, F. J., Frenzel, I., Liedtke, G., Trommer, S., … Heinrichs, D. (2017). *Automatisiertes Fahren im Personen- und Güterverkehr. Auswirkungen auf den Modal-Split, das Verkehrssystem und die Siedlungsstrukturen*. Retrieved from https://elib.dlr.de/117868/

Friedman, B., & Kahn Jr., P. H. (2003). The Human-computer Interaction Handbook. In J. A. Jacko & A. Sears (Eds.) (pp. 1177–1201). Hillsdale, NJ, USA: L. Erlbaum Associates Inc. Retrieved from http://dl.acm.org/citation.cfm?id=772072.772147

Friedman, B., Kahn, P. H., Borning, A., & Huldtgren, A. (2013). Value Sensitive Design and Information Systems. In N. Doorn, D. Schuurbiers, I. van de Poel, & M. E. Gorman (Eds.), *Early engagement and new technologies: Opening up the laboratory* (pp. 55–95). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7844-3_4

Frison, A.-K., Wintersberger, P., & Riener, A. (2016). First Person Trolley Problem: Evaluation of

Drivers' Ethical Decisions in a Driving Simulator. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 117–122). New York, NY, USA: ACM. https://doi.org/10.1145/3004323.3004336

Goodall, N. J. (2014). Vehicle automation and the duty to act. In *Proceedings of the 21st world congress on intelligent transport systems* (pp. 7–11).

Goodall, N. J. (2016). Can you program ethics into a self-driving car? *IEEE Spectrum*, *53*(6). https://doi.org/10.1109/MSPEC.2016.7473149

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a 'right to explanation'. *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Greene, J. D. (2016). Our driverless dilemma. *Science*, *352*(6293), 1514–1515. https://doi.org/10.1126/science.aaf9534

Greenemeier, L. (2016). Driverless Cars Will Face Moral Dilemmas. *Scientific American*. Retrieved from http://www.scientificamerican.com/article/driverless-cars-will-face-moral-dilemmas

Hersman, D. (2019, August). Safety at Waymo | Self-driving cars & other road users. Waymo. Retrieved from https://blog.waymo.com/2019/08/safety-at-waymo-self-driving-cars-other.html

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2019). *Ethics Guidelines for Trustworthy AI*. European Commission. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Holstein, T., & Dodig-Crnkovic, G. (2018). Avoiding the Intrinsic Unfairness of the Trolley Problem. In *Proceedings of the International Workshop on Software Fairness (FairWare '18)* (pp. 32–37). https://doi.org/10.1145/3194770.3194772

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2018). Ethical and Social Aspects of Self-Driving Cars. *ArXiv E-Prints*. Retrieved from http://arxiv.org/abs/1802.04103

Islam, M. A., & Rashid, S. I. (2018). Algorithm for Ethical Decision Making at Times of Accidents for Autonomous Vehicles. In *2018 4th International Conference on Electrical Engineering and Information Communication Technology (iCEEiCT)* (pp. 438–442). https://doi.org/10.1109/CEEICT.2018.8628155

ISO. (2011). ISO 26262 -- Road vehicles -- Functional safety. ISO, Geneva, Switzerland.

ISO. (2019). ISO/PAS 21448:2019 -- Road vehicles -- Safety of the intended functionality. ISO, Geneva, Switzerland. Retrieved from https://www.iso.org/standard/70939.html

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Johansson, R., & Nilsson, J. (2016). Disarming the Trolley Problem --Why Self-driving Cars do not Need to Choose Whom to Kill. In M. Roy (Ed.), *Workshop CARS 2016 - Critical Automotive applications : Robustness & Safety*. Göteborg, Sweden. Retrieved from https://hal.archives-ouvertes.fr/hal-01375606

Johnsen, A., Dodig-Crnkovic, G., Lundqvist, K., Hanninen, K., & Pettersson, P. (2017). Risk-based decision-making fallacies: Why present functional safety standards are not enough. In *Proceedings - 2017 IEEE International Conference on Software Architecture Workshops, ICSAW 2017: Side Track Proceedings*. https://doi.org/10.1109/ICSAW.2017.50

Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? *Transportation Research Part A: Policy and Practice*, *94*(Supplement C), 182–193. https://doi.org/https://doi.org/10.1016/j.tra.2016.09.010

Karnouskos, S. (2018). Self-Driving Car Acceptance and the Role of Ethics. *IEEE Transactions on Engineering Management*, 1–14. https://doi.org/10.1109/TEM.2018.2877307

Kim, R., Kleiman-Weiner, M., Abeliuk, A., Awad, E., Dsouza, S., Tenenbaum, J. B., & Rahwan, I. (2018). A Computational Model of Commonsense Moral Decision Making. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 197–203). New York, NY, USA: ACM. https://doi.org/10.1145/3278721.3278770

Kirkpatrick, K. (2015). The Moral Challenges of Driverless Cars. *Commun. ACM*, *58*(8), 19–20. https://doi.org/10.1145/2788477

Kuchinskas, S. (2013, April 13). Crash Course: Training the Brain of a Driverless Car. Retrieved from https://www.scientificamerican.com/article/autonomous-driverless-car-brain/

Leveson, N. (2011). *Engineering a safer world: Systems thinking applied to safety*. MIT press.

Leveson, N. (2020). Are You Sure Your Software Will Not Kill Anyone? *Commun. ACM*, *63*(2), 25–28. https://doi.org/10.1145/3376127

Luetge, C. (2017). The German Ethics Code for Automated and Connected Driving. *Philosophy & Technology*, *30*(4), 547–558. https://doi.org/10.1007/s13347-017-0284-0

McBride, N. (2016). The Ethics of Driverless Cars. *SIGCAS Comput. Soc.*, *45*(3), 179–184. https://doi.org/10.1145/2874239.2874265

Ministry of Infrastructure & Environment – Netherlands. (2017, May 18). On our way towards connected and automated driving in Europe. Retrieved from https://www.government.nl/binaries/government/documents/leaflets/2017/05/18/on-our-way-towards-connected-and-automated-driving-in-europe/On+our+way+towards+connected+and+automated+driving+in+Europe.pdf

Mirnig, A. G., & Meschtscherjakov, A. (2019). Trolled by the Trolley Problem: On What Matters for Ethical Decision Making in Automated Vehicles. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 509:1--509:10). New York, NY, USA: ACM. https://doi.org/10.1145/3290605.3300739

MIT Moral Machine Lab. (2016). Moral Machine. Retrieved 7 March 2020, from http://moralmachine.mit.edu/

Mooney, C. (2016, June 23). Save the driver or save the crowd? Scientists wonder how driverless cars will 'choose'. *Online*. Retrieved from https://www.washingtonpost.com/news/energy-environment/wp/2016/06/23/save-the-driver-or-save-the-crowd-scientists-wonder-how-driverless-cars-will-choose/

Moor, J. H. (1985). What is Computer Ethics?*. *Metaphilosophy*, *16*(4), 266–275.

Musk, E. (2016, July 20). Master Plan Part Deux. Retrieved from https://www.tesla.com/blog/master-plan-part-deux

National Highway Traffic Safety Administration (NHTSA). (2020). Automated Vehicles for Safety. Retrieved 8 March 2020, from https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety

NHTSA's National Center for Statistics and Analysis. (2018). *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey* (No. DOT HS 812 506). Retrieved from https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812506

Noothigattu, R., Gaikwad, S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P., & Procaccia, A. (2018). A Voting-Based System for Ethical Decision Making.

Pelliccione, P., Knauss, E., Heldal, R., Ågren, S. M., Mallozzi, P., Alminger, A., & Borgentun, D. (2017). Automotive Architecture Framework: The experience of Volvo Cars. *Journal of Systems Architecture*, *77*(Supplement C), 83–100.

Pillath, S. (2016). Briefing: Automated vehicles in the EU. *European Parliamentary Research Service (EPRS)*, (January), 12. Retrieved from http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/573902/EPRS_BRI(2016)573902_EN.pdf

Riener, A., Jeon, M. P., Alvarez, I., Pfleging, B., Mirnig, A., Tscheligi, M., & Chuang, L. (2016). 1st Workshop on Ethically Inspired User Interfaces for Automated Driving. In *Adjunct Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 217–220). New York, NY, USA: ACM. https://doi.org/10.1145/3004323.3005687

Ryan, M. (2019). The Future of Transportation: Ethical, Legal, Social and Economic Impacts of Self-driving Vehicles in the Year 2025. *Science and Engineering Ethics*. https://doi.org/10.1007/s11948-019-00130-2

SAE. (2016). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *Global Ground Vehicle Standards*, (J3016), 30.

https://doi.org/10.4271/J3016_201609
Sapienza, G., Dodig-Crnkovic, G., & Crnkovic, I. (2016). Inclusion of Ethical Aspects in Multi-criteria Decision Analysis. In *Proceedings - 2016 1st International Workshop on Decision Making in Software ARCHitecture, MARCH 2016*. https://doi.org/10.1109/MARCH.2016.8
Schäbe, H., & Braband, J. (2015). Basic requirements for proven-in-use arguments. *CoRR*. Retrieved from http://arxiv.org/abs/1511.01839
Schoettle, B., & Sivak, M. (2015, October). A Preliminary Analysis of Real -World Crashes Involving Self -Driving Vehicles. Retrieved from http://www.umich.edu/~umtriswt/PDF/UMTRI-2015-34.pdf
Shashkevich, A. (2017, May 22). Stanford professors discuss ethics involving driverless cars. *Standford News*. Retrieved from https://news.stanford.edu/2017/05/22/stanford-scholars-researchers-discuss-key-ethical-questions-self-driving-cars-present/
Spiekermann, S. (2015). *Ethical IT Innovation: A Value-Based System Design Approach*. Taylor & Francis.
Tesla. (2018, October 28). Introducing Navigate on Autopilot. Retrieved from https://www.tesla.com/blog/introducing-navigate-autopilot
The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*.
Thekkilakattil, A., & Dodig-Crnkovic, G. (2015). Ethics Aspects of Embedded and Cyber-Physical Systems. In *2015 IEEE 39th Annual Computer Software and Applications Conference* (Vol. 2, pp. 39–44). https://doi.org/10.1109/COMPSAC.2015.41
Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *Monist*. https://doi.org/10.5840/monist197659224
Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. *International Data Privacy Law*, *7*(2), 76–99. https://doi.org/10.1093/idpl/ipx005
Waymo. (2020). Technology - Waymo. Retrieved 6 March 2020, from https://waymo.com/tech/
What is the ISO 26262 Functional Safety Standard? (2014). Retrieved from http://www.ni.com/white-paper/13647/en/
Wilson, B., Hoffman, J., & Morgenstern, J. (2019). Predictive Inequity in Object Detection. *CoRR*, *abs/1902.1*. Retrieved from http://arxiv.org/abs/1902.11097
Zhu, X. L., & Tang, S. M. (2015). Autonomous vehicle: from a cognitive perspective. In *Multimedia, Communication and Computing Application: Proceedings of the 2014 International Conference on Multimedia, Communication and Computing Application (MCCA 2014), Xiamen, China, October 16-17, 2014* (p. 401).