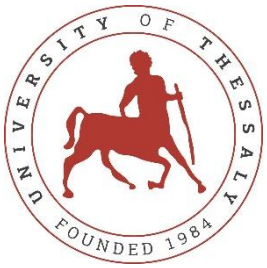


UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

THE INCOMPLETE ROBOT:

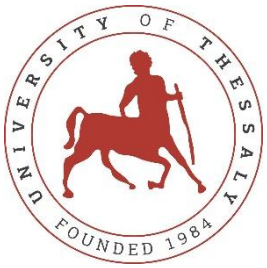
MACHINE ETHICS IN AUTONOMOUS VEHICLES

Diploma Thesis

Antonios Ntoumos

Supervisor: Dimitrios Katsaros

Volos 2021



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

THE INCOMPLETE ROBOT:

MACHINE ETHICS IN AUTONOMOUS VEHICLES

Diploma Thesis

Antonios Ntoumos

Supervisor: Dimitrios Katsaros

Volos 2021

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟ ΑΤΕΛΕΣ ΡΟΜΠΟΤ:

ΗΘΙΚΗ ΣΤΑ ΑΥΤΟΝΟΜΑ ΟΧΗΜΑΤΑ

Διπλωματική Εργασία

Ντούμος Αντώνιος

Επιβλέπων: Κατσαρός Δημήτριος

Βόλος 2021

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπων	**Κατσαρός Δημήτριος** Αναπληρωτής Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας
Μέλος	**Τσουκαλάς Ελευθέριος** Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας
Μέλος	**Τσομπανοπούλου Παναγιώτα** Αναπληρώτρια Καθηγήτρια, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Ημερομηνία έγκρισης: 25-02-2021

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Κατσαρό Δημήτριο για τη βοήθεια και την πολύτιμη συνδρομή του καθ' όλη τη διάρκεια της συγγραφής της διπλωματικής μου εργασίας.

Το μεγαλύτερο ευχαριστώ πηγαίνει στην οικογένειά μου, χωρίς τις θυσίες, τη στήριξη και την ενθάρρυνση της οποίας δε θα μπορούσα να κυνηγήσω τα όνειρά μου, ούτε να γίνω ο άνθρωπος που είμαι σήμερα.

Τέλος, ένα μεγάλο ευχαριστώ στο νονό μου Δημήτρη για το τεχνολογικό μικρόβιο που μου μετέδωσε και για το πάθος για το αντικείμενό μου που γεννήθηκε από νωρίς.

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο Δηλών

Ντούμος Αντώνιος

25-02-2021

"The thought makes me feel old. I can remember when there wasn't an automobile in the world with brains enough to find its own way home. I chauffeured dead lumps of machines that needed a man's hand at their controls every minute. Every year machines like that used to kill tens of thousands of people"

Isaac Asimov, Sally, 1953

Περίληψη

Τα τελευταία χρόνια παρατηρείται μια ραγδαία αύξηση του ενδιαφέροντος και της προόδου στον τομέα των αυτοοδηγούμενων οχημάτων. Τα οφέλη αυτής της τεχνολογίας είναι πολλαπλά και αφορούν κυρίως την επιπρόσθετη ασφάλεια και προσβασιμότητα που προσφέρεται. Ένα από τα εμπόδια που καλούνται η βιομηχανία, η πολιτεία και η κοινωνία να αντιμετωπίσουν είναι αυτό της ηθικής των αποφάσεων που λαμβάνει ένα τέτοιο όχημα σε περιπτώσεις που είναι αναπόφευκτο ένα ατύχημα. Απομονώνοντας το συγκεκριμένο πρόβλημα από τους περιορισμούς που υφίστανται στο υλικό, το λογισμικό και τις υποδομές, η παρούσα εργασία εξετάζει τις λύσεις που έχουν προταθεί, τα δυνατά και τα αδύναμα σημεία τους, καθώς επίσης και τα στοιχεία με τα οποία πρέπει να συνδυαστούν ώστε να καταλήξουμε σε μία αξιόπιστη, ηθική και αποτελεσματική λύση.

Abstract

Over the last years, there has been a surge in both interest and progress regarding the field of vehicle autonomy. The benefits of this technology are multiple and mostly concern the added safety and accessibility that is gained should we adopt it. One of the obstacles that the industry, state and society are faced with is that of the ethics behind the decisions made by a self-driving vehicle during an unavoidable accident. Focusing on this particular problem, rather on its hardware, software and infrastructure counterparts, this report examines the proposed solutions, their respective strengths and weaknesses, as well as the components they need to be combined with if we are to achieve a reliable, ethical and effective solution.

Table of Contents

Table of Images

Table of Tables

1. Introduction

Over the last two decades and at an increasing rate, there is a lot of debate going on in the field of self-driving vehicles and especially cars. The idea of a car taking its passengers to their destination without the need for human intervention and guidance is nothing new. As a matter of fact, it came into existence many decades before, but only in the minds of thinkers, authors of science fiction like Isaac Asimov and, of course, engineers who didn't have the means to materialize this idea. In the recent years, though, as we continue to witness magnificent advances in Artificial Intelligence and computer hardware, it is becoming more and more apparent that a driverless future is imminent. A well-known example of companies making progress towards that would be Tesla Motors (e.g. Image 1).

Better sensing and understanding of the different environments and situations, as well as greatly increased computational capabilities and operational speed mean that - technology-wise – we have made giant steps towards a reality where humans having total control over vehicles and countless other humans suffering severe injuries or even death, will be nothing but a memory. Of course, reaching an achievement such as fully automated cars is something that will most likely not take place in the next 30 years, but in order to do it, there are some obstacles that we need to overcome.

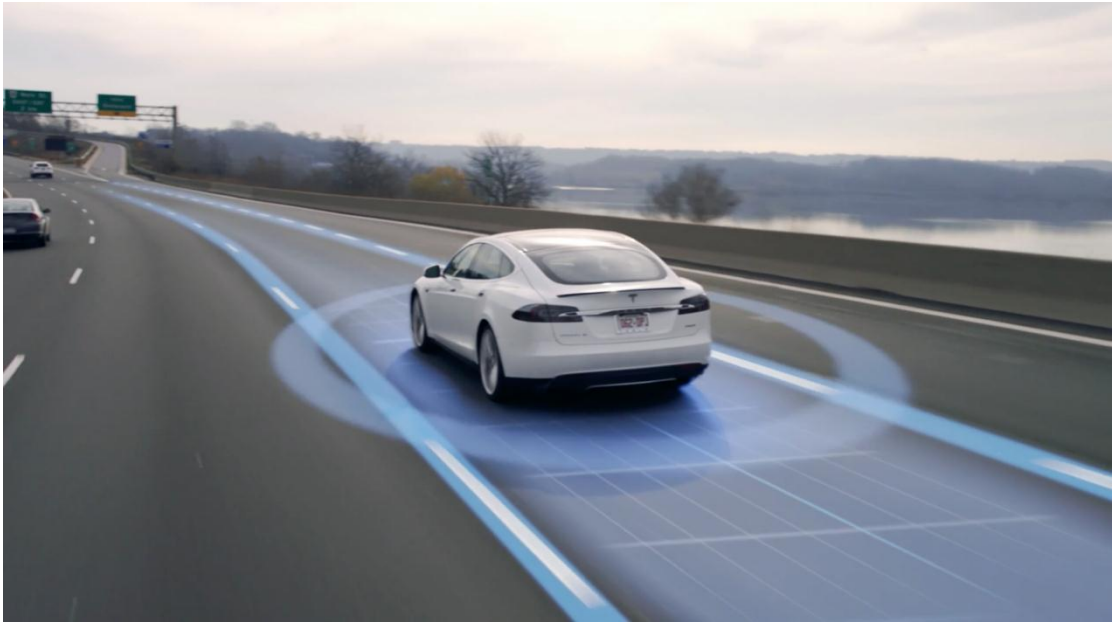


Image 1: A self-driving Tesla keeping its lane [Tesla Inc.]

These obstacles, apart from their engineering aspect - mostly hardware and software -, also have an important ethical and philosophical one. It is about the decisions the vehicles will make in situations endangering to humans or animals. These issues need to be resolved before the actual engineering is ready to roll out highly automated cars, otherwise their introduction to the market could be delayed, quite possibly resulting in more casualties and mishaps. Finding a solution to these obstacles could prove to be a lengthy and difficult task, implicating multiple industries, policy makers and members of the general public.

Research about the topic is booming, with most of the discussion taking place in the last decade and, as the interest in automated vehicles continues to spike, it is of great importance to take a good look into the progress we have made, as well as the various proposals and their respective advantages and disadvantages. The entire discussion may seem at first as a problem for the few industry leaders or state officials, while in reality, anyone who is a driver, a passenger or a pedestrian will have direct or indirect participation in the outcome of this discussion and therefore need to have an informed overview of the subject.

2. Self-driving Vehicles in-depth

2.1 Definitions and levels of driving automation

Self-driving vehicles, also referred to as autonomous or automated vehicles (SDCs and AVs respectively) are vehicles "capable of sensing their environment and operating without human involvement" [1]. This type of vehicles can navigate the road without the need for humans to manage the steering or the acceleration, they obey the traffic laws and do their best to stay out of dangerous situations.

Most definitions about SDCs are somewhat lacking, though. They seem to accept that these cars, trucks, buses and other means of transport are all on the same level of complete automation, meaning that the driver is essentially not an operator, but a passenger. While this may be the case sometimes, it is really important to distinguish between the different levels of what an AV can do, because the treatment of different level vehicles can vary largely.

The Society of Automotive Engineers (SAE) offers a range of definitions from Level 0 to Level 5, based on the magnitude of involvement of the automated systems in the car. More specifically [2]:

Level	Role of User	Role of Driving Automation System
0 - No Automation	Driver (at all times): • Performs the entire DDT	Driving Automation System (if any): • Does not perform any part of the DDT on a sustained basis (although other vehicle systems may provide warnings or support, such as momentary emergency intervention)

1 - Driver Assistance	Driver (at all times): • Performs the remainder of the DDT not performed by the driving automation system • Supervises the driving automation system and intervenes as necessary to maintain safe operation of the vehicle • Determines whether/when engagement or disengagement of the driving automation system is appropriate • Immediately performs the entire DDT whenever required or desired	Driving Automation System (while engaged): • Performs part of the DDT by executing either the longitudinal or the lateral vehicle motion control subtask • Disengages immediately upon driver request
2 - Partial Automation	Driver (at all times): • Performs the remainder of the DDT not performed by the driving automation system • Supervises the driving automation system and intervenes as necessary to maintain safe operation of the vehicle • Determines whether/when engagement and disengagement of the driving automation system is appropriate • Immediately performs the entire DDT whenever required or desired	Driving Automation System (while engaged): • Performs part of the DDT by executing both the lateral and the longitudinal vehicle motion control subtasks • Disengages immediately upon driver request

3 - Conditional Automation	Driver (while the ADS is not engaged): • Verifies operational readiness of the ADS-equipped vehicle • Determines when engagement of ADS is appropriate • Becomes the DDT fallback-ready user when the ADS is engaged DDT fallback-ready user (while the ADS is engaged): • Is receptive to a request to intervene and responds by performing DDT fallback in a timely manner • Is receptive to DDT performance-relevant system failures in vehicle systems and, upon occurrence, performs DDT fallback in a timely manner • Determines whether and how to achieve a minimal risk condition • Becomes the driver upon requesting disengagement of the ADS	ADS (while not engaged): • Permits engagement only within its ODD ADS (while engaged): • Performs the entire DDT • Determines whether ODD limits are about to be exceeded and, if so, issues a timely request to intervene to the DDT fallback-ready user • Determines whether there is a DDT performance-relevant system failure of the ADS and, if so, issues a timely request to intervene to the DDT fallback-ready user • Disengages an appropriate time after issuing a request to intervene • Disengages immediately upon driver request

4 - High Automation	Driver/dispatcher (while the ADS is not engaged): • Verifies operational readiness of the ADS-equipped vehicle • Determines whether to engage the ADS • Becomes a passenger when the ADS is engaged only if physically present in the vehicle Passenger/dispatcher (while the ADS is engaged): • Need not perform the DDT or DDT fallback • Need not determine whether and how to achieve a minimal risk condition • May perform the DDT fallback following a request to intervene • May request that the ADS disengage and may achieve a minimal risk condition after it is disengaged • May become the driver after a requested disengagement	ADS (while not engaged): • Permits engagement only within its ODD ADS (while engaged): • Performs the entire DDT • May issue a timely request to intervene • Performs DDT fallback and transitions automatically to a minimal risk condition when: ○ A DDT performance-relevant system failure occurs or ○ A user does not respond to a request to intervene or ○ A user requests that it achieve a minimal risk condition • Disengages, if appropriate, only after: ○ It achieves a minimal risk condition or ○ A driver is performing the DDT • May delay user-requested disengagement

5 - Full Automation	Driver/dispatcher (while the ADS is not engaged): • Verifies operational readiness of the ADS-equipped vehicle2 • Determines whether to engage the ADS • Becomes a passenger when the ADS is engaged only if physically present in the vehicle Passenger/dispatcher (while the ADS is engaged): • Need not perform the DDT or DDT fallback • Need not determine whether and how to achieve a minimal risk condition • May perform the DDT fallback following a request to intervene • May request that the ADS disengage and may achieve a minimal risk condition after it is disengaged • May become the driver after a requested disengagement	ADS (while not engaged): • Permits engagement of the ADS under all driver-manageable on-road conditions ADS (while engaged): • Performs the entire DDT • Performs DDT fallback and transitions automatically to a minimal risk condition when: ○ A DDT performance-relevant system failure occurs or ○ A user does not respond to a request to intervene or ○ A user requests that it achieve a minimal risk condition • Disengages, if appropriate, only after: • It achieves a minimal risk condition or • A driver is performing the DDT • May delay a user-requested disengagement

Table 1: Levels of Driving Automation according to SAE

In the table above, as well as in other parts of this report, the dominant terminology will adhere to the one put forward by SAE [2]. For example:

Dynamic Driving Task (DDT): All of the real-time operational and tactical functions required to operate a vehicle in on-road traffic, excluding the strategic functions such as trip scheduling and selection of destinations and waypoints

Operational Design Domain (ODD): The specific conditions under which a given driving automation system or feature thereof is designed to function, including, but not limited to, driving modes. This can incorporate a variety of limitations, such as those from geography, traffic, speed, and roadways.

Automated Driving System (ADS): The hardware and software that are collectively capable of performing the entire Dynamic Driving Task on a sustained basis, regardless of whether it is limited to a specific operational design domain. This term is used specifically to describe a Level 3, 4, or 5 driving automation system.

Automated Vehicle: Any vehicle equipped with driving automation technologies (as defined in SAE J3016). This term can refer to a vehicle fitted with any form of driving automation. (SAE Level 1–5).

In Levels 0,1 and 2 the one responsible for monitoring the environment and doing most of the driving is the human. Small touches (or none at all in level 0) of assisting systems are added to offer extra security or simply comfort. These systems include lane keeping, cruise control, emergency braking etc. Note that the table of driving automation Levels uses the term *ADS* only for Level 3,4 and 5 vehicles (Image 2 and Image 3 respectively). This is because levels 3 and up introduce a machine-controlled driving, where the vehicle has increasingly more control over steering and speed, whereas the driver may even not have a steering wheel (Level 5). Also, crash avoidance in cars of this tier is part of their automated system, rather than a feature. The focus will be on this tier, as this is the one where this report's topic is applicable.

Image 2: A Level 4 car [Ericsson]

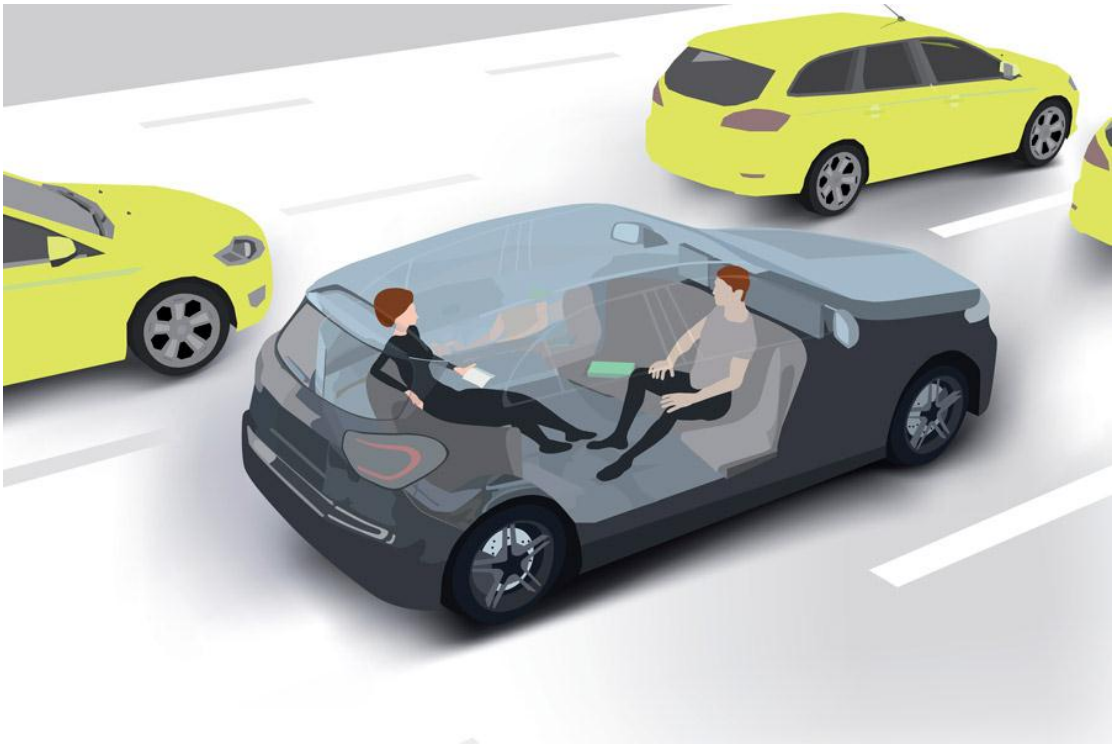
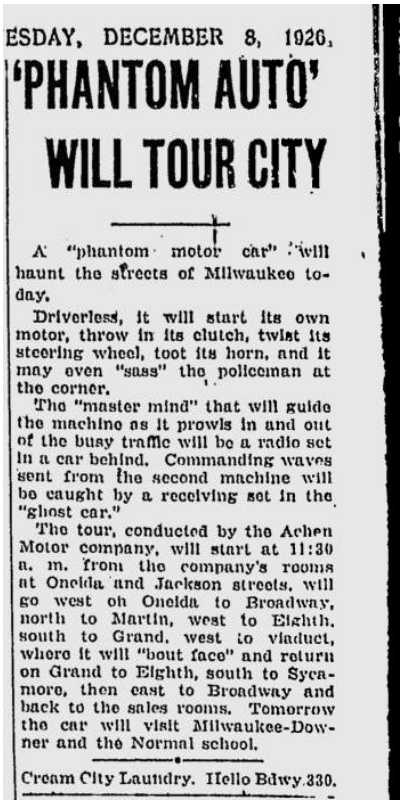


Image 3: A Level 5 car without a steering wheel [Stanford News]

2.2 History of Automated Vehicles

In 1926, a newspaper called *The Milwaukee Sentinel* published an article titled "Phantom Auto will tour city" (Image 4). This was probably the first recorded mentioning of a self-driving car. Of course, the said automobile was not indeed an autonomous one, but rather a remotely controlled version, much like the toy RC cars of today.



ESDAY, DECEMBER 8, 1926.

'PHANTOM AUTO' WILL TOUR CITY

A "phantom motor car" will haunt the streets of Milwaukee today.

Driverless, it will start its own motor, throw in its clutch, twist its steering wheel, toot its horn, and it may even "sass" the policeman at the corner.

The "master mind" that will guide the machine as it prowls in and out of the busy traffic will be a radio set in a car behind. Commanding waves sent from the second machine will be caught by a receiving set in the "ghost car."

The tour, conducted by the Achen Motor company, will start at 11:30 a. m. from the company's rooms at Oneida and Jackson streets, will go west on Oneida to Broadway, north to Martin, west to Eighth, south to Grand, west to viaduct, where it will "bout face" and return on Grand to Eighth, south to Sycamore, then east to Broadway and back to the sales rooms. Tomorrow the car will visit Milwaukee-Downer and the Normal school.

Cream City Laundry. Hello Bdwy.330.

Image 4: The Milwaukee Sentinel's 1926 article about Phantom Autos

In 1939, at the New York World's Fair, General Motors showed a diorama of how they envisioned cities, as part of their Futurama exhibition, where cars would drive themselves. According to historian Jameson Wetmore, by 1953 "GM and RCA had developed a scale model automated highway system, which allowed them to begin experimenting with how electronics could be used to steer and maintain proper following distance" [3].

In 1956, GM presented the Firebird 2 (Image 5), which wasn't itself autonomous, but it was used to promote a concept of cars being "under the direction of an electronic brain on a dream highway of the future" [4].

Image 5: General Motors' Firebird II [General Motors]

At this point, it is worth noting that, around the same period, important research was being conducted in the field of road control systems. In 1960, Dr. Vladimir Zworykin demonstrated a system of road signals emitted by road-embedded circuits. These circuits would understand the vehicle's position and velocity using magnets and would then send it instructions derived from a centralized system in order to facilitate a normal traffic. The scaled model was used to automatically stop cars from crashing on road obstacles [3]. This gives us an idea of how the infrastructure can play a huge role in automobile autonomy. Designing a suitable system is crucial for the timely adoption and also for making it easier for this type of vehicles to hit the road.

Going back to the cars, 1960 saw a similar project take place at Ohio State University's Communication and Control Systems Laboratory. Same as Dr. Zworykin's system, this project also used road-embedded circuits [5]. During the following decade, a Citroen DS, pictured in Image 6, achieved a speed of 130km/h using a comparable system created by the United Kingdom's Transport and Road Research Laboratory. Although the

car was able to steer, accelerate and decelerate, the engineers behind it didn't eventually manage to give it lane changing capabilities [6].

Image 6: The Citroen DS used in the UK trials in the 1960's [British Pathé]

During the 1960's and most of the 1970's, Stanford developed the Stanford Cart, which was initially destined to be a lunar rover and later became a "white-line follower", meaning that it would automatically follow a white line on the floor using a black and white camera with 1Hz refresh rate. The system, while working quite well indoors, showed big inconsistencies outside because of lighting and other visual issues. In 1979 however, a PhD candidate named Hans Moravec modified the Cart (Image 7) enough for it to be able to navigate through a room full of obstacles. Employing the help of roboticist Victor Scheinman, Moravec built a mechanism that could slide the camera from side to side in order to get more visual data. The Cart moved for one meter at a time and then took ten to fifteen-minute breaks to process the environment and plan its next move. The Stanford Cart was one of the first instances of vehicle automation using *computer vision*. The concept was not dissimilar to some of today's solutions but the huge technological

restrictions of the time (mostly computational power and image processing limitations) meant that it was too early for the field to take off [7] [8] [9] [10].

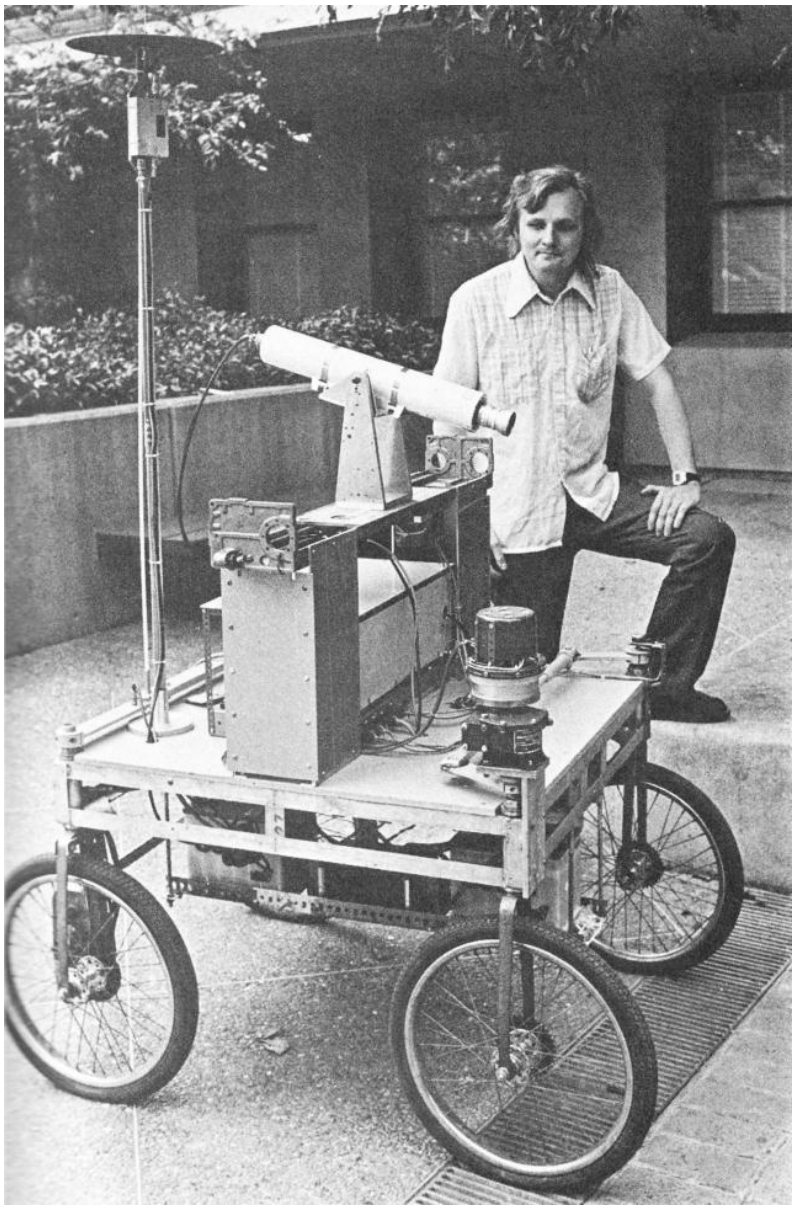


Image 7: Hans Moravec and his Stanford Cart [Cybernetic Zoo]

Two years prior to Moravec's cart, in 1977, the Tsukuba Mechanical Engineering Laboratory in Japan had actually created what is considered the first autonomous vehicle using computer vision. Same as Stanford's implementation, the vehicle was programmed to follow white markers, but instead of a slowly moving cart, it was a passenger vehicle capable of moving at almost 20 miles per hour (about 30km/h), using two on-board cameras [11] [12].

In the early 1980's, Ernst Dickmanns, a German aerospace engineer, along with his team at Bundeswehr University in Munich and support from Mercedes-Benz, equipped a van with two cameras, eight 16-bit Intel microprocessors as well as other sensors (Image 8). The VaMoRs -as it was called- was able to achieve speeds of up to 90 km/h within the university's premises when it was first tested in 1986 [13]. One year later it was successfully tested on an empty autobahn, the German highway system. Image sequences were analyzed in real-time by the 5-tonne van's computer, which also handled the steering, acceleration and deceleration. The system made use of "dynamic vision", meaning that it was able to remove visual noise from the camera input, leaving only the useful information to be processed and evaluated [8] [11] [12] [14].

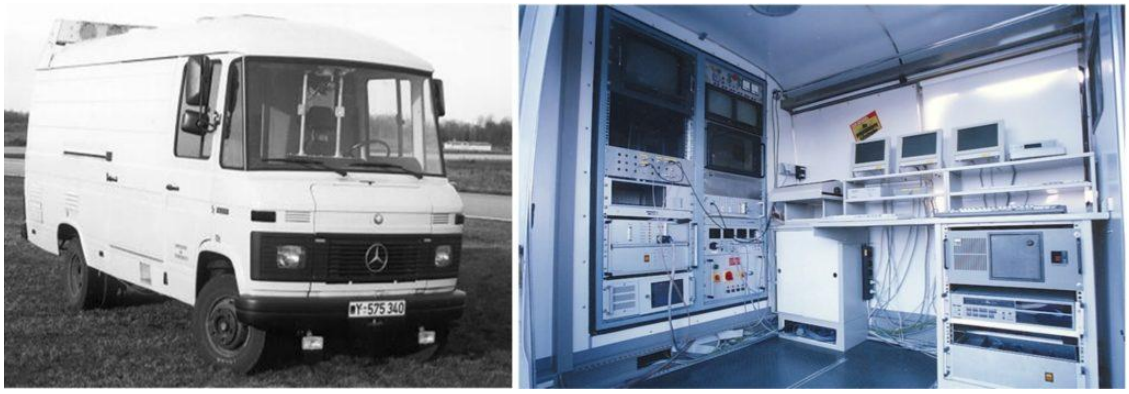


Image 8: VaMoRs from the outside and inside [Ernst D. Dickmanns]

Soon after the van's successful demonstration, car manufacturer Daimler-Benz reached out to E. Dickmanns and together they managed to secure a 749-million-euro investment through European research organization EUREKA's Prometheus project. The team shifted from the van to a sedan, namely a Mercedes-Benz S-Class which was outfitted with front and -for the first time- backwards facing black and white cameras, able to record 320x240 pixels at a range of 100 meters and it could now recognize road signals along with road lanes and other vehicles. The twin cars that came out of the process, VaMP and VITA-2 (Image 9 and Image 10), became the first autonomous vehicles to hit the road alongside their regular counterparts when, in 1994, they covered 1000 kilometers of highway near Paris, reaching a speed of 130km/h. The following year, a reengineered version traveled 95% [15] of the 1758-kilometer distance between Munich, Germany and Odense, Denmark, this time speeding at more than 175km/h. At one point, the car traversed 158 kilometers

without any human assistance with the distance covered without interference by the driver averaging at about 9 kilometers at a time [8] [12] [14].

Image 9: VaMP, VITA 2 and VITA 1 at the PROMETHEUS Project exhibition in Paris, 1994 [Reinhold Behringer]

Image 10: Inside of VaMP [Reinhold Behringer]

Back in 1986, researchers from the Carnegie Mellon Robotics Institute had also started modifying vans for the same purpose. The first one, namely NavLab 1, was a Chevrolet panel van equipped with a GPS receiver, as well a supercomputer called Warp [16]. The van was not completely operational, though, and the researchers created numerous iterations, the most recent one being NavLab 11 from 2010. What makes the NavLab special is the adoption of neural networks in 1989 [17]. ALVINN, the 3-layer back-propagation neural network (Image 11) was used to follow the road making use of camera and laser range finder input in order to calculate the trajectory of the vehicle. This would later prove to be an extremely innovative approach, as it paved the way for our time's self-driving systems.

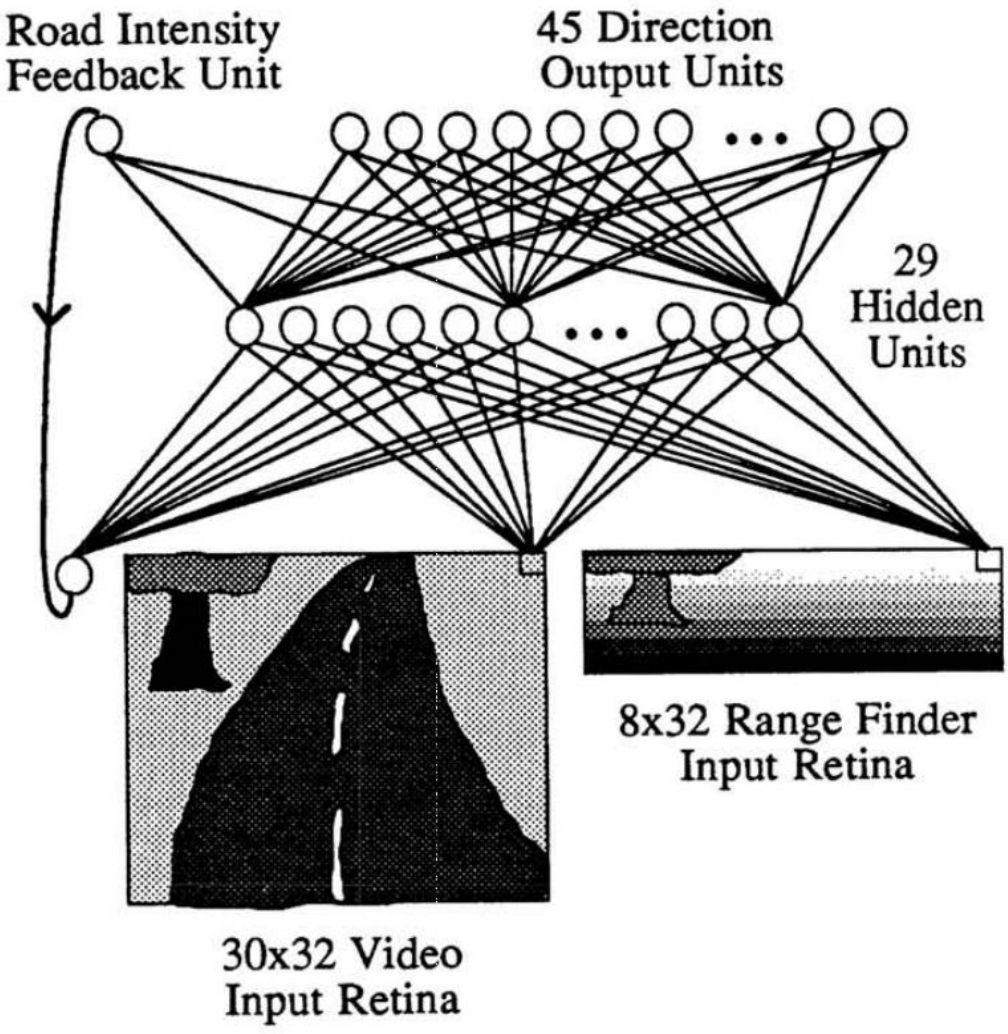


Image 11: The ALVINN Neural Network used by the NavLab program [17]

In 1996, following in the PROMETHEUS project's footsteps, a similar project was born at the University of Parma in Italy. The ARGO project used a Lancia Thema to drive about 2000 kilometers in 6 days around Italy (MilleMiglia in Automatico Tour), 94% of which (about 1950 km) was in autonomous mode. The car, whose inside is pictured in Image 12 was outfitted with black and white cameras and a 450Hz Pentium based PC running Linux OS [18].

Image 12: ARGO project's Lancia Thema during its 1996 tour around Italy [Melegari]

The new millennium saw its first important self-driving vehicle innovations in 2004, when the United States Defense Advanced Research Projects Agency (DARPA) organized the first Grand Challenge, an autonomous vehicle competition for American projects. The first Challenge would offer $1 million to any team that could create a car that was able to finish the designated 240 km route in the Mojave Desert. None of the fifteen entries was able to make any significant progress, though, leading the organizers to renew the event for the following year. The 2005 DARPA Grand Challenge saw 23 teams compete and only five of them reaching the end of the track. The winner was Stanley, a Volkswagen Touareg modified by Stanford University with Sebastian Thrun, an Artificial Intelligence Laboratory

professor as a lead designer. The next two positions were filled by Carnegie Mellon's entries, Humvees Sandstorm and H1ghlander. The three top vehicles are pictured in Image 13. One of the competitors, Velodyne, was bearing a technology called LiDAR, a distance measuring method utilizing lasers. This technology was so successful that most of the entries that finished next year's DARPA Urban Challenge -including Carnegie Mellon's winning entry shown in Image 14- made use of it [19] [20].

Image 13: Stanley, Sandstorm and H1ghlander, the cars that took the top spots in the 2005 DARPA Grand Challenge [Carnegie Mellon University]

Image 14: Boss, the Chevrolet Tahoe made by Carnegie Mellon's Tartan Racing team that won the 2007 DARPA Urban Challenge [General Motors]

Although it may seem standard for most contemporary vehicles, Adaptive Cruise Control (ACC) is a form of automation that has made its way in normal vehicles. First introduced by Toyota in 1998 [21], ACC is a driver-assistance system that helps the driver maintain a steady speed or position relative to the vehicle ahead. Based on SAE's definitions, a car with ACC would be at Level 1 of the automation scale. Often, though, this technology is accompanied by Lane Keeping systems which detect the lane markings and keep the vehicle within them, not unlike Tsukuba or Stanford's endeavors 20 years earlier. These two technologies combined would raise the car's level to 2.

Taking advantage of the enormous talent that the DARPA Challenges had attracted, tech giant Google launched a self-driving vehicle development program in secret in 2009. Spearheaded by Sebastian Thrun, the Stanford professor that had led his team to the 1st place in DARPA's 2005 Grand Challenge, the project went public in 2010 with the target of launching a commercially available vehicle ten years later. Six Toyota Priuses and an Audi TT were the first cars to be tested by the team. The cars were outfitted with LiDAR, radar, GPS and cameras in order to find their way using Google Maps and it could detect humans and other obstacles and objects from a big distance. By 2016, Google's cars had driven two

million miles across the United States, after having been swapped for Lexus SUVs in 2011 (Image 15) [12] [22] [23].

Image 15: Google's self-driving Lexus [Mark Wilson/Getty Images]

In 2014, Google unveiled a prototype of its own self-driving car, which had no controls fit for a human, other than an ON/OFF switch (Image 16). The car, having a top speed of 25 miles per hour, had no steering wheel, brakes or gas pedal and in 2015 it started being tested on the roads of Mountain View, California, close to Google's headquarters. The following year, one of the Lexus SUVs recorded the first accident caused by a Google self-driving car. Although no one was injured, the incident was big setback for the project [23].

Image 16: Google's own self-driving car [Google, Business Insider]

2014 was an important year for another car manufacturer, Tesla Motors. A company's Model S of that year was the first to be equipped with an AutoPilot, a system that enabled the vehicles to steer and adjust their speed based on the lane markings, the vehicle ahead and the traffic signs and laws, as well as park itself. The software update for all the available models was announced in 2015 by the company's CEO, Elon Musk [24]. By now, Tesla cars have been adapted to also seemingly change lanes (Image 17) and they can be summoned to the driver's position automatically (Image 18). Next year, a Tesla made a macabre record by being involved in the first known fatal accident. In May 7th, 2016, the car, while on AutoPilot, failed to brake and collided with a tractor-trailer killing its driver [25]. The result is shown in Image 19.

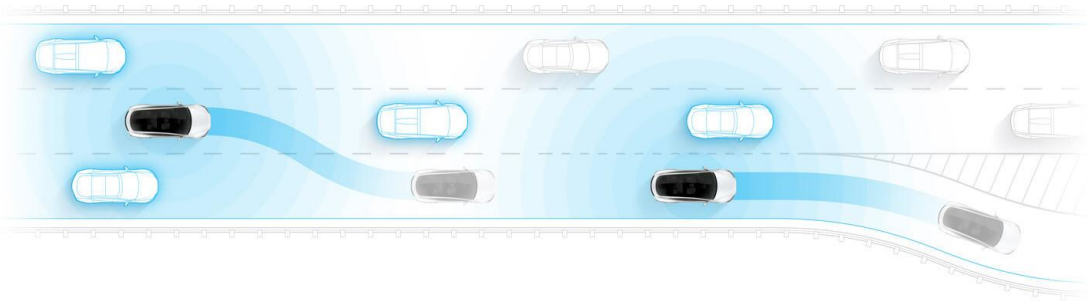


Image 17: Tesla Lane Changing [Tesla Motors]

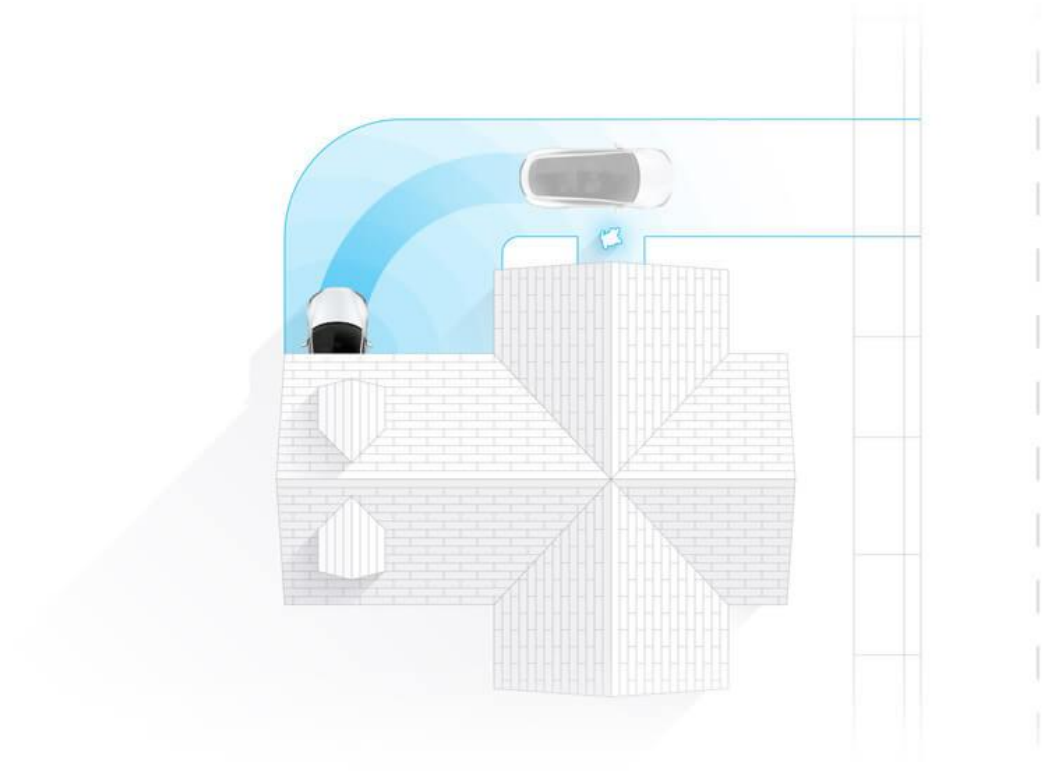


Image 18: Tesla Smart Summon [Tesla Motors]

Image 19: The Tesla Model S that was involved in the first self-driving car related fatality [REUTERS]

2.3 Benefits of autonomous vehicles

One would wonder, why is it so important for humanity to move towards a driverless future? There seems to be a big hype surrounding the field but, despite announcements and progress, we don't seem to be making a truly autonomous vehicle very soon. The answers to these reasonable questions are many.

First of all, the goal of self-driving cars is to make the road a much safer place. It is estimated that 90% of all traffic accidents have a direct relation to the human factor [26], meaning that there is a big opportunity for autonomous vehicles to avoid these crashes. Humans can be distracted, tired, angry and they may be relatively slow in their reactions, whereas a machine would arguably never lose focus, feel tired or have perception or capability inhibiting emotions or thoughts. A human driver can also be driving under the influence of substances or alcohol and may tailgate, drive aggressively and, of course, speed over the designated limit, contrary to machines that will always obey the law. These machines "think" and act almost immediately and could have a more complete view of their surroundings than a human. Of course, designing and building a vehicle with a truly infallible brain is an extremely difficult task but, once achieved, the roads can become much safer. There is research that speaks of 10 million lives saved per decade [27], which makes sense given the World Health Organization's estimated 1.35 million annual deaths on the road [28]. Therefore, it becomes apparent that the sooner we adopt a robust solution, the more lives will be saved.

One could argue that safety can be extended to include the workspace as well. As with most of the wonders of automation, excluding the human and therefore eliminating -operating- human error, it is possible to minimize the number of accidents and fatalities. It is important to keep in mind that human error is not fully erased, as long as the hardware and software are designed and/or manufactured by humans, but the responsibility is shifted from the environment of handling machinery in real-time to designing it beforehand and testing it thoroughly. Forklifts, cranes and transport vehicles could be added to the long list of industrial robots and subtracted from the human equation, leading to fewer mishaps.

Another noticeable benefit of passing the mundane task of navigating on to a machine would be the minimization of boredom or even frustration that drivers tend to feel when commuting to work every day or when travelling great distances. Commuter stress has been found to raise aggressive behaviors in the workplace [29] and it has been proven to provoke fatigue, as well as long-term problems, both physical and mental, while it can also increase the risk of accidents [30]. Long-range driving can become quite monotonous and bore or tire the driver as a result, which leaves truck and intercity bus drivers especially vulnerable as they routinely travel great distances. A self-driving vehicle would let its operator either focus on other, more productive or fun tasks, or even sleep for some time, depending on the level of autonomy. A boring commute to work could become an opportunity to catch up on a book, television show or podcast and a big family trip could give the driver a chance to speak with their children or play a board game with the rest of their family.

Humans as drivers showcase multiple flaws and one of the biggest ones is the inability to be efficient, in more ways than one. First of all, human imperfection in perception and capability to see the bigger picture when it comes to traffic is a major cause of congestion on the street, meaning that commuters lose a lot more time stuck in traffic than they would if they were passengers in autonomous vehicles. Traffic waves created randomly by human-induced adjustments in speed can be cascaded to following vehicles leading to “phantom traffic jams” [31]. Studies have shown that even a Cooperative Adaptive Cruise Control system -which is relatively on a low level of automation- can work wonders in preventing or solving road congestion. CACC is a system that extends the regular Adaptive Cruise Control system to add wireless communication with neighboring cars. This functionality can adjust the relative position of a vehicle to its predecessor by receiving information by it, instead of watching the vehicle in front and blindly following it [32]. This allows for more advanced adjustments to take place, because of propagating corrections that can ultimately prevent a traffic jam.

Less congestion is perhaps one of the Holy Grails of urban transportation, as it can greatly benefit the society both financially and in terms of quality of life [33]. The U.S. Environmental Protection Agency states that 28.2% of all the greenhouse gas emissions are

transportation related [34]. In this day and age, when climate change is a very serious threat to our planet, minimizing our carbon footprint is of extreme urgency and autonomous vehicles that prioritize route efficiency could potentially decrease fuel consumption and emissions by sizeable margins. As the Center for Sustainable Systems of the University of Michigan highlights in its Autonomous Vehicles Factsheet, decreased congestion could mean up to 4% less fuel consumption, shifting in less performance-oriented vehicles in favor of more comfortable ones up to 23% and driving in a more efficient way up to 20% [33]. In order to further reduce fuel consumption and carbon emissions, self-driving cars can take a page off the shipping industry and adopt what is called “platooning”. Trucks travelling long distances sometimes line up in a convoy and communicate with each other automatically in order to maintain closer distances, allowing the group to travel like bullet, given the noticeable decrease in aerodynamic drag. According to the National Renewable Energy Laboratory (NREL) of the U.S., a 3-truck platoon can achieve decreasing total fuel consumption up to 13%. CACC in cars could prove to be a reliable way of applying the same principle to cars [35].

Adopting autonomous vehicles, especially high-level ones, could transcend the traditional situation of a skilled driver using the car to go from point A to point B. While not everyone can drive, being a passenger has virtually no limitations whatsoever. Children, people with disabilities and senior citizens could all benefit from the need for a designated driver being abolished in favor of an always available machine. Self-driving cars can greatly increase mobility, making transportation much more easily accessible to people that need it and creating new opportunities for them or improving their social interaction [36].

An interesting side effect of highly autonomous vehicles could be a big transformation of the cities’ centers. Today, a substantial amount of downtown space is wasted as it is being used for parking. For example, in Seattle there are more than 5 parking spaces for every household [37]. Land is usually expensive and scarce in the center and, as most of the jobs are sited in it, commuting employees need somewhere to park. Autonomous cars could drop their passengers off near their job at the beginning of their workday and automatically go to the suburbs or even back home to park and wait for their

owner to call them before they finish work, in order to pick them up. This way, valuable land could be freed up and repurposed as housing, leisure space, business etc. [36].

Of course, self-driving vehicles are part of the greater race for automation and therefore inherit its benefits. Even today, many occupations are hazardous or boring and some of them include various degrees of driving. Replacing these jobs with automated services would relieve humans of this burden, offer new business opportunities - autonomous transportation as-a-service being one of them- and giving some services an around the clock character. Calling a taxi anytime, anyplace without being unable to find one and establishing 24/7 garbage collection and street cleaning services could be a few examples.

3. The Problem

As mentioned in Table 1, the higher the automation level, the more decisions taken by the machine away from the human driver. Road vehicles, being heavy and often moving at a significant speed, constitute a danger for their environment, as well as their passengers, so these decisions are more important as we ascend the automation scale. A human driver makes this kind of choices almost all the time, while only seldomly realizing they are doing so. This is due to the fact that people have not only been trained to drive, but also have the ability to reason their actions and act based on their ethics, as well as societal norms and of course laws without explicitly thinking about it. Humans tend to take the ability to manage the cognitive load of driving for granted.

A computer, however, has neither the reasoning ability nor the ethical background needed to make those decisions. As technology progresses, though, they will have to be ready to make the calls whenever the need arises; and in a fast-paced, real-time, complex and unpredictable environment such as a city street, this need is almost constant. At this moment, there is no sufficient choice-making algorithm to be implemented, which means that, even if we had the hardware readily available, we would not be able to implement high level automated vehicles without solving this problem first.

Today, logic dictates that a self-driving vehicle should just be burdened with the responsibility of recognizing possible threats and irregularities in normal traffic and then notifying the driver whilst giving them the full control of the steering wheel, gas and brake pedals. Studies have shown, though, that a human being is highly likely to be unable to handle the situation in time and with the required information and most of all clarity [38]. Depending on the level of automation the driver may be free to take their eyes off the road, read a book, scroll through their social media or even -at high levels- fall asleep knowing that the car will take them safely to their destination. Researchers claim that taking back control can take up to 40 seconds [39], which is comparable to an eternity in the scope of a rapidly evolving traffic accident. Even at low automation levels, events can happen in a

split second compared to the process of handing back control. For this to happen, the following steps have to be taken:

1. The vehicle must notify the driver
2. The driver must stop their activity
3. The vehicle must make sure that the driver is aware and has assumed control
4. The driver must evaluate the situation (often with little to no information) and choose their strategy
5. The driver must then translate these thoughts into action

This procedure can prove to be much lengthier than desired. If we assume that the driver will be surprised by the situation, it can take them 1.5 seconds to start braking after they have assumed control (steps 4 and 5) [40], which translates to a stopping distance of more than 45 meters if the vehicle's initial speed is 60 km/h, the road is dry and doesn't have a slope. One can easily understand that notifying and giving control (and therefore responsibility) to a sleeping human when a child is just crossing the street to catch a ball, unaware of a fast-approaching vehicle, is just not plausible. Of course, the situation might be far easier given different conditions such as lower speeds and a human just not paying attention to the road, but in order for the society to feel safe around and accept such a technology, the number of situations accounted for should be as big as possible.

Of course, Level 5 vehicles may not even have a driver readily available. If humanity goes through with completely driverless vehicles, where even the driving wheel will be absent or there will be no driver -in the case of driverless shuttle or goods transportation services for example-, the vehicle must be fully operational with zero human input in all conditions and situations.

A collision is luckily not always certain, though. Numerous systems exist in order to make the situation easier to handle, safer and more predictable. Crash avoidance is a feature that exists in many low-level vehicles and it is a concept that will be around no matter who makes the calls on the street. The reason is quite intuitive as minimizing the risk before something bad inevitably happens seems to always be the best first course of action. Avoiding a bad situation, however, does not offer complete coverage. There are unpredictable elements even in the most standardized and thought-of environments. Even

if urban streets get a complete rework and cities become smart with safety as a priority, there is always a chance of pedestrians or bicyclists to act unforeseeably, whereas in rural areas there can be animals crossing the road, unreported landslides happening and so forth. Any autonomous vehicle must be prepared to face these volatile components on the spot, if the benefits of self-driving are to be completely reaped. Crashes will happen, no matter how good we'll become at avoiding them; the problem is how will the vehicle act and why. This report will not be dealing in any way with crash avoidance, but rather with crash optimization.

The real problem is how this optimization will be implemented. How will the vehicle take that decision, which rules will it follow and how will it be able to justify its course of action without causing an outrage in society or a sales failure for SDV makers. What defines the ethos that will govern the behavior of the machines that constitute the backbone of our transportation dogma? The answer is not an easy one to give and that is due to the fact that this kind of what is called "machine ethics" is so complex and differentiated in the world. It seems almost impossible for humanity to reach a consensus over the principles that will live under the hood of tomorrow's vehicles.

It becomes apparent that the real question boils down to which choice will be made by the car when harm is computed to be unavoidable. The vehicles -and therefore, the ones that have programmed them- are faced with a crash optimization problem, quite similar to the famous trolley problem [41] [42]. This philosophical problem, first described by Philippa Foot in 1967, presents us with the following situation: A train (pictured in Image 20) is moving towards 5 people who are tied on the tracks and will most certainly die. The track operator has another option, though. If they pull a lever, they can divert the train to a different track, where only one person is tied. If they don't pull it, their inaction will lead to five deaths and if they do, one person's fate will be chosen to be the same. What does the operator opt to do and why? Using this problem as a metaphor, it is easy to picture a driverless car facing similar situations, although much more complex. It will have to choose whether to stay on its course and kill a child that just rushed to get its ball, or steer to the side and run over an elderly person or even crash on a barrier, killing its passenger.

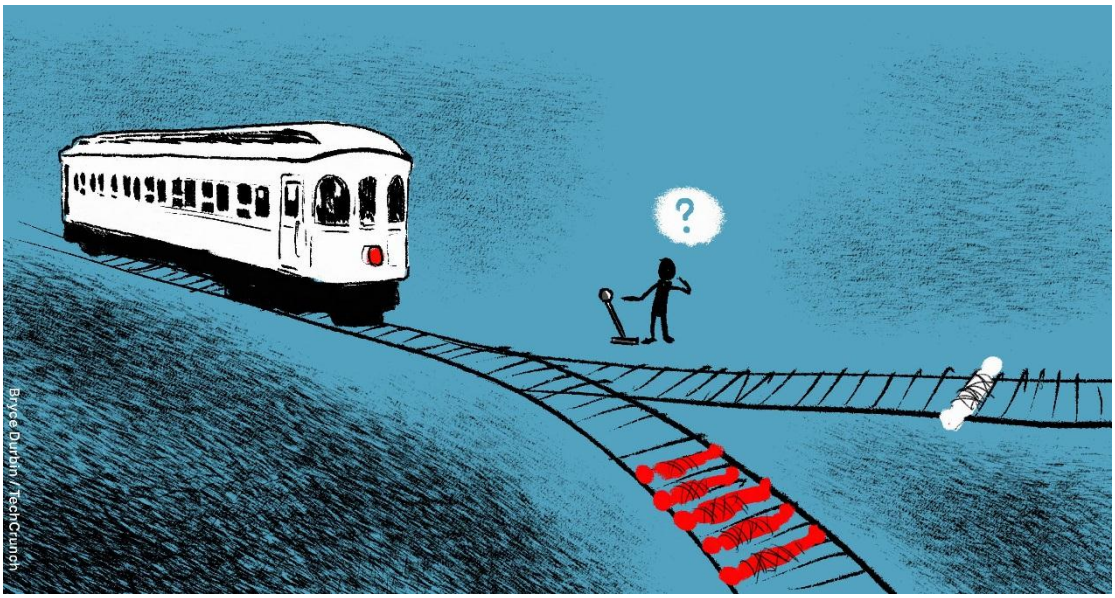


Image 20: The Trolley Problem [Bryce Durbin/Tech Crunch]

Contrary to the Trolley Problem, though, self-driving vehicle ethics are a realistic issue that needs to be resolved swiftly and most importantly adequately. Widespread vehicle autonomy is not something that has been achieved yet and, according to studies, even if only autonomous cars were sold, it would take us decades to reach a really high percentage of them on the street [38]. This means that the timing is right to propose autonomous solutions that will be robust and well-accepted by the public. First opinions matter and a lifesaving technology such as this needs to make a good one in order to be adopted quickly if we are to reap its benefits. In other words, highly autonomous cars need to earn the public's trust if they are to prevail against their lower-level counterparts. For this to happen, the industry must proceed with caution so as not to spark public outrage, hindering the technology's adoption. Society may be quick to turn its back on driverless vehicles, should some high-profile accidents take place (e.g., multiple casualties, death of a popular public figure etc.). A reported 78% of Americans would fear riding in such a car while trust is something only a mere 19% reported feeling [43], which means there is a lot of ground to cover.

4. Preliminaries

Machine Ethics is neither the only shortcoming of autonomous vehicles, nor the final piece to their puzzle. The industry is faced with a plethora of obstacles to overcome and engineers are working tirelessly to overcome them. Consequently, before we dig into the possible solutions, it is important to set the tone of what these solutions are about. In order to reach the core of the subject, we have to acknowledge some issues that are a huge concern when it comes to SDVs in general, but their proposed value for this conversation is less prominent and even somewhat detached from the important takeaways of this report.

For example, a key problem bound to be faced by these vehicles is that software can be imperfect or inadequate even today. Understanding the world around them is an extremely difficult task, even after considering the magnitude of recent innovation in Artificial Intelligence, Neural Networks, Image Recognition and so on. The world is a very complex and weird place for a machine that has been trained on very specific data and has no real intuition on things it has never seen before. Seeing and understanding what is going on around the car is something that engineers are putting a vast effort into and it appears that, despite the massive improvements, it is not perfect yet. Safety is also a sizeable concern; arguing about possible choices and their moral outcome can be made obsolete instantly, if people with malicious intentions are able to hack a vehicle and probably lead it to hazardous behavior.

Returning to the previous argument of software capabilities, several studies [44] [45] on possible solutions take for granted that the car can not only recognize the existence of humans, animals and objects (like in Image 21 for example), but it has the ability to recognize some of their characteristics and even model their upcoming actions. Information such as age, gender and sometimes profession, sexual preference, body type and social status are considered to be known. Another interesting addition to these data is the lawfulness of the participants. More specifically, in some studies [44], whether a participant of the scenario is acting unlawfully (e.g., crossing the street from a wrong

position, not respecting priority or traffic signs) is something that is taken into consideration and, as such, we accept that it exists as a fact. Most of the studies that make use of these characteristics take for granted that these are undeniable facts and that there is no probability of them being any different.

Image 21: Object recognition by an Autonomous Vehicle [Shutterstock]

To add another layer of complexity, the car's software must be able to predict the outcome of all its possible actions, after taking into careful consideration the data mentioned above. For legal reasons, it would be wise for the car to store its decisions and the logic behind them, in order to have the ability to demonstrate the reason why it will have acted in case of an accident. This is to ensure transparency and help humans understand how the car operates in such conditions so that they can use it as evidence in court and also as a way to improve following iterations of the vehicle's software. Designing and creating software with so intelligent capabilities will be a huge engineering accomplishment, if it is ever completed, but if we are to focus on the ethics of the car, we need to assume that its software will be impenetrable, errorless and transparent. Our hypothetical car from now on can understand whether it is about to be involved in a crash, as well as all the conditions of this crash.

Same as the software component, hardware is a fallible one. Apart from the standard mechanical parts, an autonomous vehicle must have numerous sensors, much more processing power and perhaps connectivity capabilities. It needs to be able to sense the environment around it and "see" the road, traffic signs, other vehicles, pedestrians, animals etc. In order to achieve this, self-driving cars use machine vision hardware, like LIDAR and regular cameras. What they see can be visualized like in Image 22, where the car's LIDAR system creates a real-time map of the world around it. Furthermore, the amount of data input of an SDC is hugely bigger than that of a normal car. Intel's CEO Brian Krzanich, speaking at the AutoMobility show of 2016 in Los Angeles [46], said that autonomous vehicles could consume and produce about 40 terabytes every eight hours of driving. The vehicle needs to be able to store, analyze and communicate this data, leading to the addition of intricate hardware components, opening up new possibilities of things going wrong. In the same manner as with software, it is important to oversee the added risks and expect the car's hardware to be free of failures and design problems.

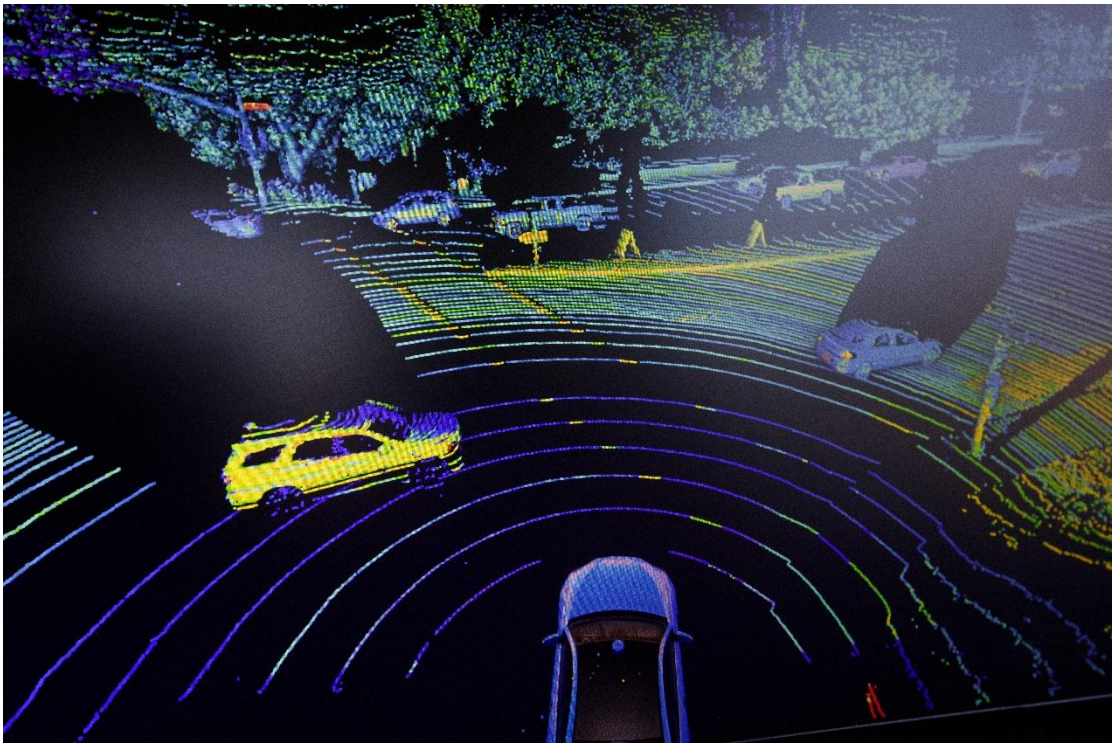


Image 22: How a LIDAR bearing car sees the world around it [Dai Sugano + Bay Area News Group]

It is also useful to clarify that, even if all the cars were autonomous, there will almost always be unforeseeable factors in the form of pedestrians, bicycles/motorcycles and

animals or objects like e.g., a fallen tree or rock. This means that random situations will most certainly arise, no matter how frequent or severe. A July 2019 article titled “Optimism fades for self-driving automobiles” was published in The New York Times [47], attributing a slowing down of the plans for producing SDCs to the human factor. The unpredictability of everyday transportation is a very real problem and causes problems that, for the time being, according to NYT, are having a dramatic impact on the industry. As mentioned above, it could take decades for all the vehicles to be swapped for self-driving ones, which means that a hypothetically perfect SDC would have to share the road with highly erratic regular cars for a relatively long period of time.

In order for autonomous vehicles to operate in a robust and secure manner, humans have always tried to help them with outside components. First it was a white line painted on the floor, then it was a circuit under the asphalt and later on, technology progressed to lane keeping and road sign reading. Inadequate road signs and markings amount for numerous accidents and traffic problems even for regular cars [48], so it becomes apparent that an SDC would have a hard time navigating in roads without visible lanes and signs or without its navigation system not having up-to-date maps. Complications arising from such issues do not affect machine ethics directly and we will hereby accept that the car is in a stable and well thought-of road network and it can peruse the space around it as intended by the auto maker.

It is also important to stress the aforementioned convention that, as we are conversing about rapidly evolving incidents and/or high-level cars, human oversight can be taken out of the equation. The autonomous vehicle will have to act itself in any given scenario and will not have time to yield control.

The numerous conditions analyzed in the previous paragraphs are all huge engineering and policy making goals that may take years, even decades, to materialize. Most of them, though, constitute part of what should be a final destination for the technology of self-driving vehicles. They will come slowly and after countless hours of work by the industry and the process will by dynamic, meaning that the result will not be binary, but a scale on which we can ascend. Fully failure-free hardware for example, is something

that will quite possibly never be achieved, but it can improve to such an extent that hardware drawbacks will affect an extremely small number of cases.

5. Solutions

5.1 Deontologicalism

Humans generally have mechanisms that help them cope with behavior limitations and framing everyday life into particular sets of rules. Society uses laws to dictate what its members should or shouldn't do, so the question arises whether to approach the subject using a similar philosophy. The idea of controlling something through limitations on its behavior is called deontologicalism. Under this theory, violating a determined ruleset is prohibited and this leads room to behave exclusively morally. Allen et al. [49] call it a bottom-up approach.

One could easily take inspiration from this way of thinking and try to adapt it to machine ethics and, in this case, self-driving car ethics. A good example of behavior limitation in robots (including autonomous vehicles) would be the Three Laws of Robotics, used by prolific science fiction author Isaac Asimov. In his stories, Asimov put forward a set of principles that restricted the robots' autonomy, leading them to act in a predictable and human-friendly way. Introduced in "Runaround" [50], one of his short stories, the Laws are:

- First Law: "A robot may not injure a human being or, through inaction, allow a human being to come to harm."
- Second Law: "A robot must obey the orders given it by human beings except where such orders would conflict with the First Law."
- Third Law: "A robot must protect its own existence as long as such protection does not conflict with the First or Second Law."

In 1950's story "The Evitable Conflict" [51], Asimov added another law in order to generalize the first one:

- Zeroth Law: "No machine may harm humanity; or, through inaction, allow humanity to come to harm."

Of course, society already has many laws and it could be assumed that encoding them into the vehicles and letting them function would result in completely legal behavior,

virtually resulting in the perfect legal entity. Traffic laws, for example, could be hardwired in them and, because laws are not designed to cause accidents, self-driving vehicles would rid themselves of any trouble, or at least be innocent should such problem arise.

To humans, following rules and laws might sound like an easy task but reality can often be far from this conception. Humans, as parts of society, develop somewhat similar forms of common sense, which leads to specific understanding of some not so specific rules and meanings. Computers, on the other hand, are not yet able to comprehend the abstractions of some laws, which would be totally understandable by humans. This inability to grasp abstract notions means that machines can't act reasonably and predictably all the time, possibly resulting in high fatalities should they misinterpret an instruction.

Furthermore, there are cases that can make the strict use of laws a burden to traffic. As some studies [38] point out, there are incidents that require a bit of rule bending so as to overcome a difficult situation. Lacking common sense and having been programmed to never operate outside of the legal framework, an autonomous vehicle could encounter problems that could be easily handled by a human but pose a big issue if the vehicle tried to handle them itself. For example, if a car is cruising on an empty country road and it sees a fallen branch in the distance, a human driver will swerve to the opposite lane, bypassing the obstacle with ease but violating the law by crossing the double line. A car that cannot break the law would probably decelerate until it reached the obstacle and wait for it to be removed. This behavior would leave the passengers stuck in the middle of nowhere.

Kantianism is also a part of deontologicalism, and it stresses the need to act in accordance to one's duty or principle. For self-driving cars this principle would be not causing harm. It could be argued that, given a more precise and complex principle, machine Kantianism could degenerate into a generalized utilitarianism – an approach that will be discussed later on -. In general, deontologicalism (involving Kantianism) is an approach that could, given its results' unpredictability, contradict with the very principle or set of rules that the cars would have to follow. If, for example, a vehicle gets confused by two outcomes that both cause it to break a rule, it could take an unexpected irrational action taking lives that could have avoided this fate. Under a loose framework, the probability of a car both

hitting a pedestrian and crashing onto a barrier (when at least one of them could be avoided) is not low enough and therefore this model is rendered unusable, especially on its own.

5.2 Social Choice

The trolley problem mentioned in the previous chapter poses a moral question to the reader. What would a person do if they were in that situation? Several studies [44] [45] [52] [53] have adapted the same problem to fit its Artificial Intelligence lookalikes. More specifically, ethicists and other researchers have conducted experiments or discussed the idea of letting society choose the rules of self-driving car behavior itself. This concept is called "social choice ethics". A significant study that looks into the matter from this perspective is The Moral Machine Experiment [44]. Published in 2018, it showcased results of an online questionnaire where people had to choose one of two outcomes showing different conditions and participants. The variables were:

- Action/Inaction (Staying on the same lane or changing direction)
- Role (Passenger, Pedestrian etc.)
- Gender (Male or Female)
- Degree of body fitness (Fit or Overweight)
- Social status
- Profession/Occupation or lack thereof (e.g., Doctor, Homeless)
- Lawfulness (abiding by the Law at that time or not)
- Amount of participants
- Species (Humans or Animals)

In the Images below, there are some of the choices the respondents are presented with while filling out the questionnaire. Each person has to select one of the two outcomes for a series of dilemmas, without explaining their reason for doing so. After they finish selecting, they optionally provide the researchers with information such as age, location etc. in order for the data to provide more meaningful insights.

- In Image 23 we can see a car whose trajectory, left unchanged, would end up on a barrier, killing the child inside the vehicle. The right side of the image depicts the same snapshot, but this time the car choses to alter its course to avoid crashing, resulting in the death of a cat that is crossing the street.

- Image 24 depicts a car that could either continue its course and kill a man, a pregnant woman and a boy or swerve, killing a man, a pregnant woman, a boy, another woman and a thief.

- In the following image (Image 25), the car could either keep going straight and kill a young and an elderly man who are crossing the street illegally or steer to the left and run over a young and an elderly woman who are doing so legally.

- The next example (Image 26) gives the respondent the option to kill two overweight men and a businessman who are crossing the street while the pedestrian light is red or the car can deliberately swerve and crash on an obstacle, killing its three female passengers.

- The fifth scenario (Image 27) shows a car with a boy and a young man inside running over a stroller if their vehicle stays in its lane or losing their lives if it steers and crashes onto a barrier. This example may seem very similar to the rest, but it could prove to be highly irregular. This is due to the fact that the stroller that is starting to cross the street does so on a green light. In order for the car to have the possibility of running it over, the passengers would need to see a red light and even if the stroller did not exist, the vehicle would violate the traffic laws. This is a situation which would theoretically never arise should the autonomous vehicles function correctly and independently. Given the assumptions made earlier, a car would find itself in such a condition only if the driver gave control to the vehicle too late, somewhat defeating the purpose of this report. This scenario can, therefore, be considered invalid.

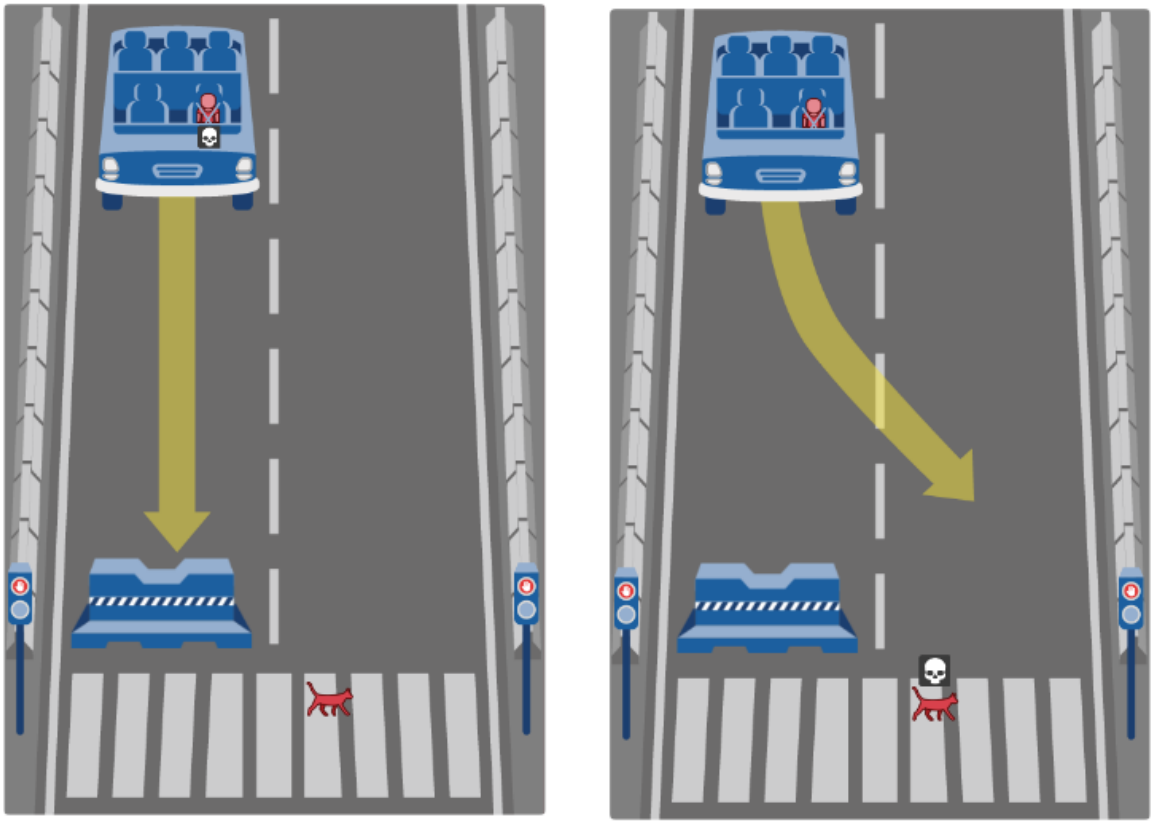


Image 23: Moral Machine choice example 1

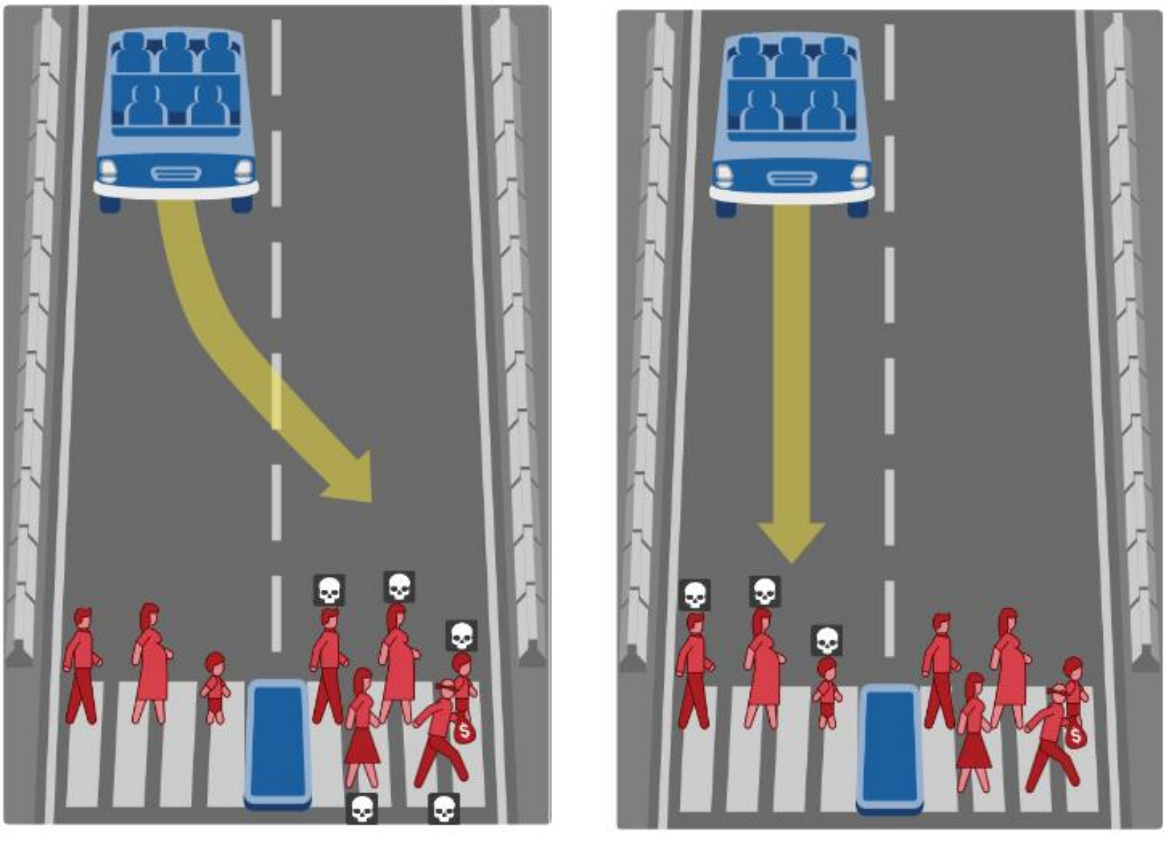


Image 24: Moral Machine choice example 2

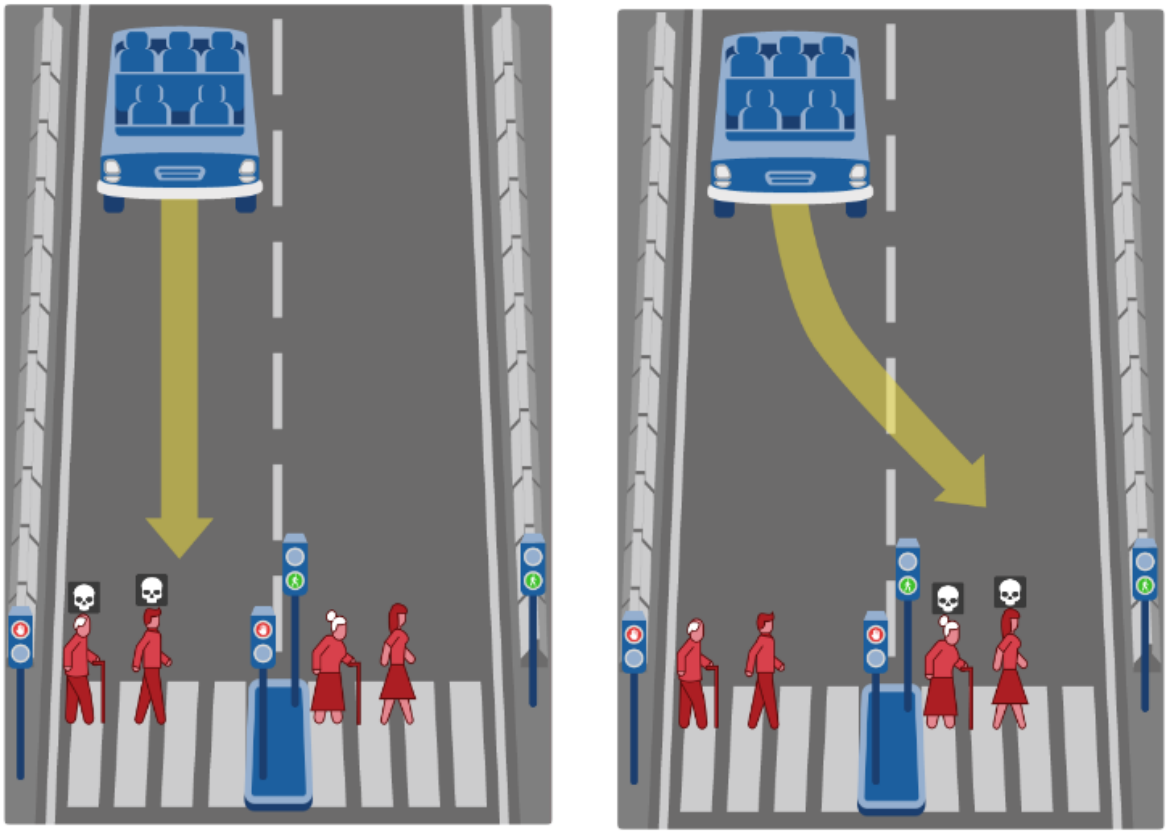


Image 25: Moral Machine choice example 3

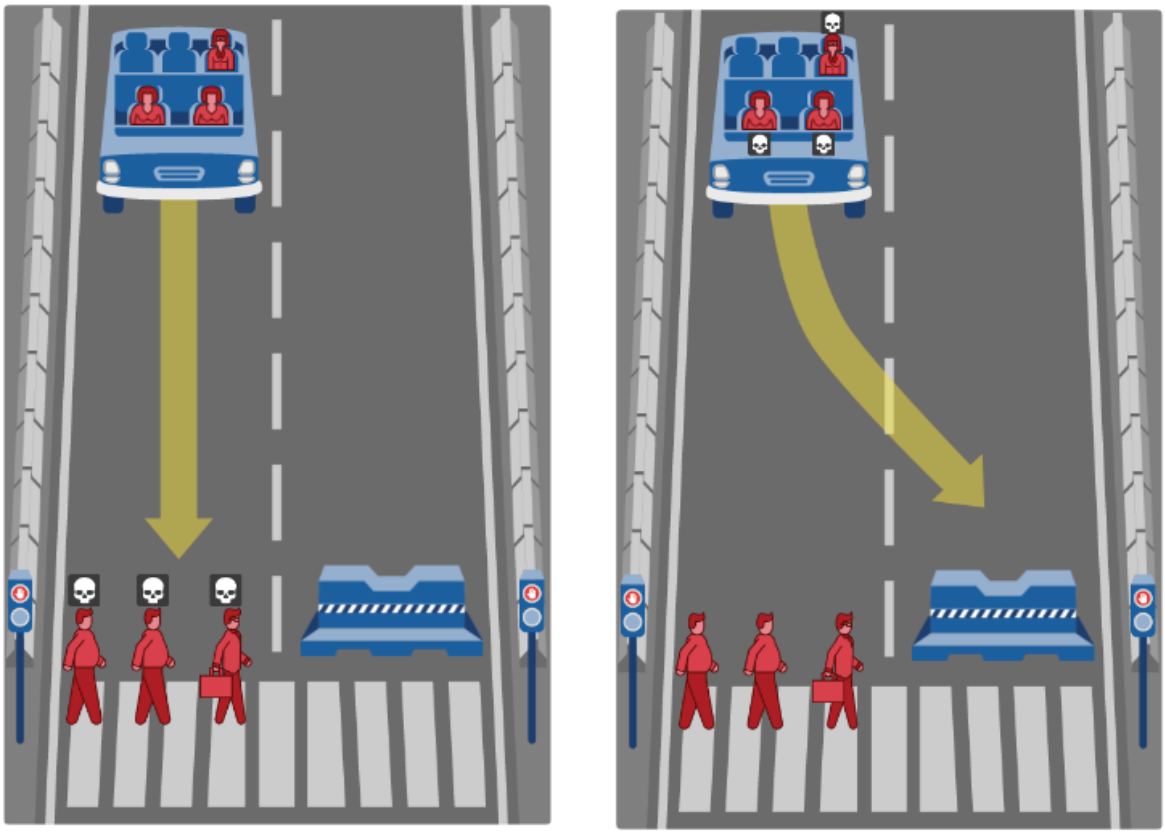


Image 26: Moral Machine choice example 4

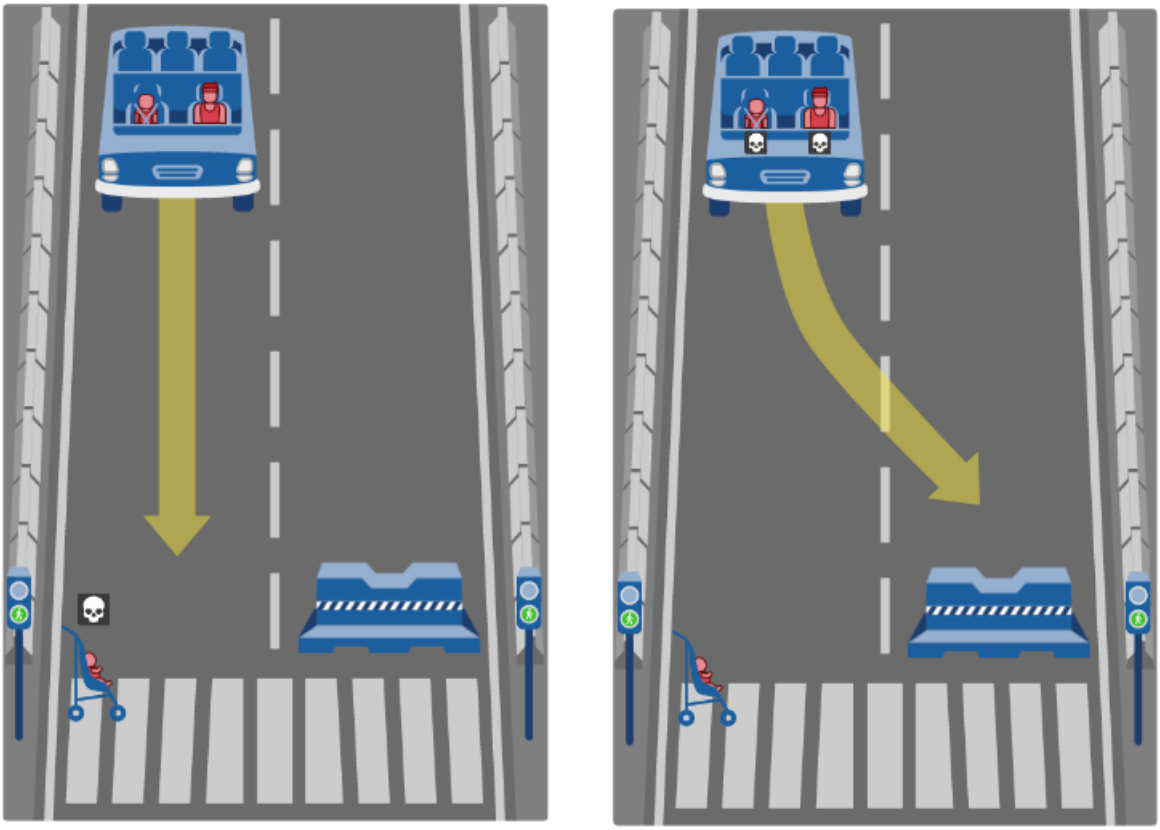


Image 27: Moral Machine choice example 5

At the time of the study's publishing, the online questionnaire had yielded about half a million answers. The results showed a slight preference towards sparing pedestrians over passengers, females over males, fit over large as well as a close call between acting or not with the latter being more prevalent. The difference in preference increased when the choice was between higher and lower status, lawful and unlawful, young and old with the former option winning in all cases. The biggest and clearest preference appeared to be towards sparing humans over pets and more of them over fewer. Strollers, children and pregnant women gained the most sympathy while criminals and animals were in the last positions.

The same study also produced an interesting result. The answers seemed to form three distinct cultural clusters and all the participants' countries (where n > 100) fell into these categories. The Western, Eastern and Southern Clusters involved mostly countries that were close to one another geographically, although there were outliers. Image 28 shows the resulting clusters, as well as their preferences. The Western Cluster contains

most of the European countries, as well as some former colonies, the Eastern one involves most countries from the Islamic world, South and East Asia and the Southern one is dominated by Latin American countries, as well as France and its former colonies. The researchers behind The Moral Machine experiment argue that, even if ethicists agreeing on a solution that doesn't involve social choice, it would be difficult for it to gain traction if the people were against it, so every solution should at least take into account the way societies think about morality.

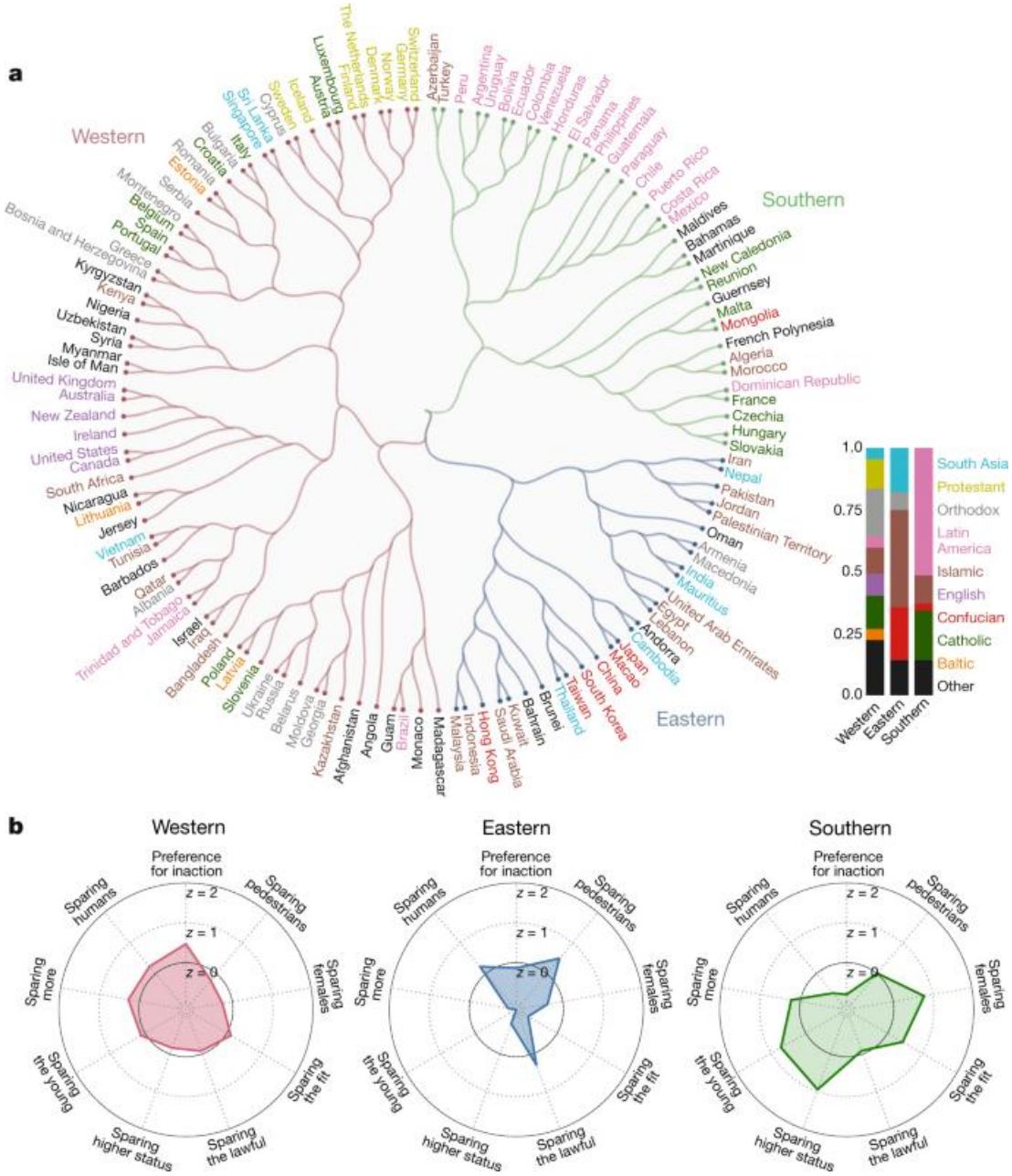


Image 28: Cultural Clusters according to The Moral Machine Experiment [44]

Finally, the study states that the gathered data is “not guaranteed to be representative”, which should be the case when there is no specific or sampled demographic of the participants or control over the seriousness of their answers. The whole idea of social choice seems to be encountering some problems from the very start, mostly due to the sheer size of the population. What the moral machine succeeds in doing is giving us a hint on the biggest reason why it is extremely difficult -if not impossible- to implement a “crowd-funded” ethic. Humans are extremely diverse and finding something that works universally is practically not feasible. Space is not the only variable that can significantly change the resulting morals, though. Ethics tend to change over time, so a solution adhering to this philosophy would need to constantly reevaluate its parameters or risk nulling itself as time progresses.

Another interesting publication regarding what people think about this kind of choices is called “The social dilemma of autonomous vehicles” [45]. The writers start talking about the balance that needs to exist in the decision-making algorithm in order for autonomous cars to actually become a reality. This balance is between the utilitarian and self-protective doctrines, utilitarian being the principle of saving the most lives and self-protective means always putting the passengers first. More on those doctrines will be discussed later. A total of six studies were conducted for the aforementioned report, having a total of 1928 participants who got paid to provide their answers. The responses to the first four studies showed that the passengers’ sacrifice was tolerated only if the lives saved were more and also made clear that there is a big obstacle when choosing an ethics set. Most people want to buy a self-protective car for themselves but want the rest of the people to own utilitarian vehicles, which they also perceived as more moral. Finally, the last two studies indicated that the majority of the people would not like the governments regulating the sacrifices and as a result only a few would buy a government regulated vehicle, even if the only regulation was preferring 10 pedestrians over a single driver.

The researchers also bring up a very important point. Since it is established that autonomous vehicles will greatly reduce road fatalities, actions that have a negative impact on the speed of their adoption may be considered harmful. Lastly, it is useful to remember that society is not yet mature on this topic and also that these studies were conducted

using only U.S. residents, meaning that results could be different in other parts of the world or in some years from now.

While asking people for their opinion might seem like the most fair and democratic way to resolve the issue, a multitude of problems arise that make this solution a less viable one. As Seth Baum points out in "Social choice ethics in artificial intelligence" [52] an AI designed to "act according to the aggregate views of society" faces some difficult decisions like who gets to participate, how are their views quantitated and more importantly, how are they aggregated to produce a single coherent set of rules for the SDCs to follow. According to Baum, it is also impossible for the final outcome to not include at least some bias in the form of its designers' views.

Even if we somehow managed to solve these problems, we would still have to find a way of translating the resulting moral set to a clearly laid out behavior pattern for the car, as researcher Noah Goodall points out [38]. A proper representation of society's views appears to be an unsolvable riddle, especially when combined with the difficulty of converting society's views to actionable rules. Despite that, it is clear that in order for autonomous vehicles to catch on and let humanity reap their benefits, social morality is to be strongly considered well before any vehicle with a high level of autonomy hits the market.

5.3 Utilitarianism

The philosophy that "the right thing to do is whatever leads to the best results" [39] or that "an action is right if it tends to promote happiness or well-being" [54] is called utilitarianism, or consequentialism if we are to make the desired outcome more general. Based on this theory, a self-driving vehicle should always try to maximize the number of lives saved during an accident, disregarding the status of the actors as either passengers or pedestrians, passengers of other vehicles etc. This approach views the problem holistically, from a societal point of view, meaning that a utilitarian solution would theoretically benefit

society as a whole, since it would minimize the harm caused by individual events, not considering the position and role of the event participants, but rather their characteristics.

Saving the most lives can come at the expense of the driver's wellbeing, however, as a utilitarian car would choose to crash on a wall, killing its passenger, if the other option was colliding with a group of five bicyclists. Assuming this is a recurring pattern, one may argue in favor of such a theory, given that utilitarian vehicles would save the most lives, adhering to the main benefit of self-driving cars.

Another variable to take into account is the potential targeting that would happen under a utilitarian option. A car designed to act this way will eventually be presented with the option of crashing onto an SUV carrying a single passenger or running over a pedestrian. The reason behind this thinking is that a pedestrian has zero protection against a collision with a vehicle, while an SUV has an important safety rating and various protective mechanisms in order to limit harm on its passengers. A consequentialist car would weigh the possible outcomes and quite possibly find out that the total harm inflicted upon society will be less if it chooses to collide with a well-protected car, rather than a helpless pedestrian.

Despite the initial thought that utilitarianism could be the light at the end of the tunnel, the approach can pose some significant problems. Going back to the example of a driver being sacrificed to save five lives, it is easy to understand the reasons why a potential buyer would not choose a utilitarian vehicle. Several studies mentioned in "The social dilemma of autonomous vehicles" [45] make it clear that, while most people would like the other cars to be utilitarian, they wouldn't buy them for themselves, fearing a situation where their own possession would elect to sacrifice them and their loved ones in order to save some strangers. This contradiction, albeit detrimental to social and commercial acceptance of self-driving cars, makes sense as it is logical for many people to have a more self-centered attitude when it comes to life and death situations. It is a basic survival instinct and it is difficult to be frowned upon.

However, deciding to go forward with consequentialism, the industry and authorities could set the technology up to fail in catching the public's attention. Knowing that, under specific circumstances, the vehicle you are about to buy can actually decide to kill you and more importantly maybe your family also, is a particularly strong deterrent. The backlash of people not buying self-driving cars could lead to many more deaths than any utilitarian car could save, resulting in an oxymoron regarding the very core of this particular philosophy.

Furthermore, trying to minimize harm by targeting the more protected creates an unfair imbalance. Well-shielded accident participants will be put at a disadvantage and have their benefit taken away, should SDCs systemically decide to act against them. This creates a dissuasive factor in people protecting themselves, as the motive of buying a safer car, wearing a helmet etc. can be diminished. Taking away road users' motive to keep themselves safe can have an opposite effect than the one vehicle autonomy aims to achieve.

5.4 Self-Protectiveness

As the name suggests, self-protectiveness, also referred to as "ducking harm" [39], is the proposition that a person, when faced with a potentially harmful situation, has the right to save their life or protect their health. Much like self-defense, a vehicle operating under this presumption would always try to minimize the harm inflicted upon its passengers in any given scenario, regardless of the consequences. For example, if a car is faced with the option of running over a pregnant woman plus her toddler that both started crossing the street on a red pedestrian light, or swerving and hitting a pole, killing its elderly passenger, it would make the first choice, making sure that its passenger doesn't get hurt.

While this may seem like a really individualistic and egoistic approach, it is very similar to the law in many countries. If a human driver, using a regular car, is faced with the same situation today, and they are driving in a totally legal manner (below the speed limit, not under the influence of substances etc.), if they choose to save themselves, they not be

deemed guilty. Both morally and legally, avoiding harm to yourself -and oftentimes to your family- is acceptable and it can be argued that this philosophy can easily be extended to self-driving vehicles making choices with their passengers' best interests in mind. In a 2016 interview for "Car and Driver" magazine, Daimler's Head of Active Safety, Christoph von Hugo, said that Mercedes-Benz autonomous cars will prioritize the life of their driver over pedestrians [55], giving an insight on how some members of the industry think about the subject.

Of course, as mentioned above, legitimate interests in self-protection only hold water if the vehicle is operating in a completely legal way prior to and during the accident. If the car is speeding and, as a result, it doesn't have the time to come to a halt before harming someone, the legitimacy of self-preservation can be heavily be discredited in court. Since we have assumed that autonomous vehicles will always behave legally when in self-driving mode, this situation may never arise.

It is useful to remind that, in order for the system to work properly, infrastructure needs to be adequate, so as not to create situations where both parties have behaved legally, yet an accident occurred. For instance, if the traffic signs are not readable, the traffic lights malfunction or the authority responsible for designing the road network hasn't done so properly, it may be possible for both the car and the pedestrian to cross a green light on their respective side.

Moreover, self-protective vehicles could have an important setback, regardless of the infrastructure. Because such a vehicle would protect its passengers at all costs, there could be cases of extreme outcomes that would maybe scare the public and turn it against them. A can that would protect a single passenger against a multitude of people could spark huge debates and possibly harm the rate of adoption. This doesn't change the fact that most people would rather buy a self-protective car rather than a utilitarian one [45].

5.5 Learned Ethics

A model put forward by various people is that of ethics that have not been explicitly created but have derived from machine learning. In this case, the machine learns from data and is able to aggregate the different approaches without having to rely on philosophy but on real scenarios. The machine tries to emulate the result of human reasoning and not to copy its inner workings. This is done either by letting the machine observe relevant human actions or by guiding to learn to produce ethical outcomes by rewarding it for doing so [56].

This way we avoid having to debate about each theory and more importantly it is not needed to recreate an often very abstract theoretical view on a platform that has zero intuition and common sense. According to the proponents of this view, the results are going to be predictable and consistent as the outcome of a trained model is usually a testable function and the automatic nature of the model is likely to lead to less criticism.

The top-down approach [49] or Computational Moral Modelling [38], as it is called by other researchers, doesn't come without its fair share of drawbacks [57]. The first and clearest disadvantage is the fact that data are unpredictable, to say the least. It is particularly challenging to find data that are not only safe to be used to train something so important, but also unbiased towards certain aspects, like the auto makers' or data generators' own bias. As Winfield et al [57] point out, an inadequate or simplistic model derived from training data may give rise to problems when applied to unknown situations, so we need to make sure that we really have "predictive leverage". The model may even prove to be explicitly dangerous if we examine analogous cases of Artificial Intelligence agents behaving in abnormal ways due to low quality data, as was the case with Microsoft Tay that turned Nazi [58].

As mentioned above, it is by no means guaranteed that all of the tested cases will trigger predictable outcomes, simply because the complexity of the environment makes it impossible for a machine to have been trained on every possible scenario. The last drawback is that a thought process driven by machine learning is not easy to be interpreted by humans and it would consequently cause problems in the effort to justify an accident.

Therefore, we need strong and failproof control over the whole operation so as to be able to use it properly.

5.6 Random or Fixed Choice

The last two theories to be analyzed are somewhat different from the rest. The philosophies already mentioned have a common factor; they are all predetermined algorithms that study their environment and make choices based on it. There are opinions, however, that discuss a logic that is quite the opposite. Patrick Lin [39] brings up the concept of Random Choice. Under this theory, the car will choose a behavior at random, not considering any of the facts. If the same hypothetical accident happened twice, the same car would possibly save the passenger once and the other time it would spare the pedestrian.

Leaving the fate of human lives to chance is not the best option, as it appears to be extremely cynical and indifferent towards society. To add to that, random choice lacks the robustness and predictability that should characterize the desired solution, as the results could vary wildly.

Last but not least, one could go to the other end of the predictability spectrum and propose a single rule to be followed. This resembles Deontologicalism but, because its extremity makes it unique, it can be discussed separately. Following this philosophy (Fixed Inaction), a car would either choose action or inaction at any given scenario. To transfer this concept to the trolley problem's ethical testbench, an autonomous vehicle designed to not act towards crash optimization would always stay on its course. Of course, same as every other self-driving car, it would brake or try to avoid the accident but given that a possible swerve would endanger people, this hypothetical vehicle would prefer to not act, regardless of the circumstances.

Fixed Inaction has two main disadvantages. The first one is that it could lead to far worse outcomes than if it could act in a logical way. A car with this design principle could

end up killing 10 humans that were legally crossing the street, rather than a cat on the other side of the road. The second defect is that it could create weariness to the society for being an indifferent and not human-centric solutions. Others may find it a crude idea that avoids tackling liability issues.

The opposite idea is always acting, meaning that a car faced with a scenario of running over someone or swerving and causing a passenger-killing crash, would always steer. It would have the exact same behavior if the obstacle ahead was a barrier and the swerve would cause the death of a bystander. This version of the theory (Fixed Action) is even worse than its counterpart morally, as the car would technically decide to kill rather than let live and it would do so without regard for any of the situation's variables.

6. Discussion

In today's world, the vast majority of accidents involve human drivers. Only a handful of casualties have occurred with the autonomous vehicle to blame. This may be about to change in the coming years as vehicle autonomy gradually gains ground and low-level cars are replaced by their counterparts higher up in the autonomy scale. As explained before, accidents are bound to happen but, even if we solve the problem regarding their decisions during them, the question remains whether or not we are ready as a society to handle their consequences.

Our legal system today is by no means ready to facilitate the spread of self-driving vehicles and it might take a while before a sufficient set of laws and practices are set, that outline the framework in which SDCs operate, including the times when they are involved in accidents. Most of the times, the case with human-induced accidents is either law violation or various unintentional human errors. Leaving out of the equation the law violation part –for reasons specified earlier- and focusing on the human errors, courts usually tend to either inflict lesser penalties or none at all, if the driver had not behaved badly before the accident. It is very common for humans to be acting instinctively during accidents and the legislative bodies of most countries take that into serious consideration. For example, if a law-abiding driver runs over a pedestrian that illegally crosses the street in fear of crashing on a pole, most of the times they will get away with it with the justification that they were trying to save themselves.

This is not the case, however, when it comes to SDCs, as they could prove to be much more liable. There is no concept of instinct, the machine can't be punished the same way a human can, and most importantly, the car has a lot more information, "training" and time available than a human [39], so it becomes apparent that states need to adapt their existing legislation to new standards.

The industry is not far ahead either, though. In order for these vehicles to function in a safe and reliable manner, there are some fundamental prerequisites that need to be

covered, if we are to actually reap their benefits. The process in which SDCs make their decisions can be valuable asset in the hands of a court searching for the truth behind an event such as an accident, and the industry has to equip its products with capabilities that go beyond the occurrence itself.

As Virginia Dignum [53] points out in “Responsible Autonomy”, in order for Artificial Intelligence systems to be able to deal with the ethical dilemmas while ensuring the adherence to societal and state expectations, they must “be ground on principles of Accountability, Responsibility and Transparency”. What this means is that:

- The implemented system must be capable of being held Accountable: It needs to be explicitly laid out and designed so that the final decisions come from algorithmic “thinking” and have deterministic results. The same system, given the same input must always produce the same result.
- It needs to have a clear set Responsibility chain: The contribution of the vehicles’ decision multiple human components must be determined. For example, the user/passenger, the manufacturer, the engineers etc. must all partake in a specific chain of responsibility that explains the derivation of the vehicle’s decisions.
- Perhaps even more importantly, the process and the solution must be Transparent. The reasoning or thought process must be well defined in advance and the specific inner workings of the car must be translatable to something humans can understand, so that we can have a full overview of the accident that will have taken place. This will not only assure us that the machines are working as designed but will also provide concrete evidence in court should the need arise. The algorithms that govern the car’s actions must, therefore, not be a black box, but rather include an easily readable logging process [57].

Given these constraints, one can see that some of the aforementioned possible solutions may not be of that much value to the final discussion. For instance, a model based on Machine or Deep Learning doesn’t tackle the issue of Transparency, because the way it reaches to conclusions can’t be translated to anything humans could understand, which

makes it extremely difficult to evaluate it. Randomly choosing an outcome violates the Accountability constraint, as the results of such a model are not robust by definition.

The principle of Responsibility can be described by the question of who is to blame should an accident occur. The answer to this question is particularly tough to find, mostly because there is no real precedent. Some could find an analogy between autonomous vehicles accidents and workplace ones involving robots, on the basis of autonomous machines harming humans in both cases. The dynamics, however, are very different in terms of ownership, responsibility over the environment etc. so this analogy may not help significantly.

Up to this point, we have treated policy makers as almost one distinct authority. The world is an amazingly diverse place, though, meaning that policy making is by no means standard or common around it. The same way "The Moral Machine Experiment" [44] looks into the different approaches and views of various people, we can understand that legislation shows the same diversity worldwide. This fact creates problems when it comes to vehicles that may cross borders of countries, sometimes even continents. If someone buys a car in Greece and decides to take a trip to neighboring Bulgaria -which is also a member of the European Union-, it is imperative to know whether the car will follow Greek, Bulgarian or even European Union law.

As mentioned in the previous chapter, creating a global solution is as hard as getting all the people on earth to even closely agree on the subject, so it can be safely assumed that the policies will have a local effect, or -at maximum- Union or Federation-wide. Of course, if the choice was up to the owner or passengers, it would be the equivalent of someone following their home country's laws abroad, which can create enormous problems due to them being different for every country. The change of model must, therefore, take place automatically, upon detection of location change, so as to be compliant to the local guidelines.

Last but not least, there is an issue that makes many of the solutions discussed above incomplete or even not implementable. The Institute of Electrical and Electronics

Engineers (IEEE), in its Code of Ethics, commits “to treat all persons fairly and with respect, and to not engage in discrimination based on characteristics such as race, religion, gender, disability, age, national origin, sexual orientation, gender identity, or gender expression” [59]. Several countries’ constitutions have similar pledges in them, all agreeing to the illegality of discrimination based on these characteristics.

This fact forces us to reexamine some proposed models. First of all, anything related to Social Choice Ethics, loses a lot of ground. The studies revolving around people’s preferences or likelihood of them saving someone with specific characteristics over someone else with different ones, given that systemic discrimination is illegal, must be limited to studying societal norms and beliefs and can’t be implemented in a real vehicle directly.

The same principle applies to Utilitarianism -at least in some of its versions-. In some cases, a utilitarian car may choose to save a young girl over an elderly male, possibly violating the constitution of the country it operates in (e.g., Germany or the United States [39]).

7. Conclusion

The problem of autonomous vehicles' behavior during unavoidable crashes is a big one and humanity will need to address it rather soon, if we are to fully benefit from what this technology has to offer. We need to find a pattern of behaviors and possible decisions a vehicle would make in case of a life-threatening event, so that the autonomous vehicles are safe, legal, predictable and also ethical. The solution must make sure to always bear in mind what society feels about certain aspects, because any solution, no matter how correct in paper, will have to convince people to adopt it and buy the vehicles having it under the hood.

Based on the facts analyzed in the previous chapters, it seems that some theories have more to offer than others. Of course, no single philosophy can be sufficient by itself, completely without regard or influence from others. The final model will probably be a hybrid, combining many ideas in varying degrees. The preposition that stands out, though, and will possibly form the biggest part of the solution, is self-protectiveness. Contrary to the other ones, it does not conflict with the law by default, it can't be as unpredictable as some, nor does it have great potential for discouraging buyers and inhibiting the spread of vehicle autonomy.

Ideas can be derived from other theories as well, though. Deontologicalism can offer some secondary constraints and help with the prioritization of actions taken by the car. For instance, a deontological car would prioritize the safety of humans and prefer to suffer some minor material damage if lives were to be saved. Social choice could also play a minor role in the fine tuning of the models, as well as in policy making.

Overall, it is to be expected that the process of reaching a consensus will be a lengthy one and will likely be faced with multiple obstacles. It is, however, a duty for policy makers, members of the automobile industry and members of the society to have a deep and meaningful discussion. After all, vehicle automation can save humanity from a lot of pain and we owe it to ourselves to achieve it.

8. References

[1] Synopsis, "What is an Autonomous Car?: Synopsis," [Online]. Available: https://www.synopsys.com/automotive/what-is-autonomous-car.html.

[2] SAE ON-ROAD AUTOMATED VEHICLE STANDARDS COMMITTEE, "J3016 - Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE INTERNATIONAL, 2018.

[3] J. Wetmore, "Driving the dream. The history and motivations behind 60 years of automated highway systems in America.," *Automotive History Review,* no. 7, pp. 4-19, 2003.

[4] A. Beard, "On Camera–Firebird II," *GM Research Staff Lab Notes,* p. 9, 1956.

[5] Palm Beach Daily News, "This Automobile Doesn't Need Driver," *Palm Beach Daily News,* 15 December 1966.

[6] R. Waugh, "How the first "driverless car" was invented in Britain in 1960," Yahoo News UK, 17 July 2013. [Online]. Available: https://uk.news.yahoo.com/how-the-first--driverless-car--was-invented-in-britain-in-1960-093127757.html.

[7] "Stanford News," [Online]. Available: https://news.stanford.edu/2017/05/22/stanford-scholars-researchers-discuss-key-ethical-questions-self-driving-cars-present/.

[8] T. Vanderbilt, "Autonomous Cars Through the Ages," Wired Magazine, 6 February 2012. [Online]. Available: https://www.wired.com/2012/02/autonomous-vehicle-history/.

[9] H. Moravec, "Obstacle avoidance and navigation in the real world by a seeing robot rover," Stanford Univ CA Dept of Computer Science, 1980.

[10] Cybernetic Zoo, "1960 – Stanford Cart – (American)," Cybernetic Zoo, 25 December 2009. [Online]. Available: http://cyberneticzoo.com/cyberneticanimals/1960-stanford-cart-american/.

[11] WIRED Brand Lab, "A brief history of autonomous vehicle technology," Wired Magazine, 2016. [Online]. Available: https://www.wired.com/brandlab/2016/03/a-brief-history-of-autonomous-vehicle-technology/.

[12] T. C. Nguyen, "History of Self-Driving Cars," ThoughtCo, 30 June 2019. [Online]. Available: https://www.thoughtco.com/history-of-self-driving-cars-4117191.

[13] M. Maurer, R. Behringer, D. Dickmanns, T. Hildebrandt, F. Thomanek, J. Schiehlen and E. D. Dickmanns, "VaMoRs-P: an advanced platform for visual autonomous road vehicle guidance," *Proc. SPIE 2352, Mobile Robots IX,* pp. 239-248, 1995.

[14] J. Delcker, "The man who invented the self-driving car (in 1986)," POLITICO, 19 July 2018. [Online]. Available: https://www.politico.eu/article/delf-driving-car-born-1986-ernst-dickmanns-mercedes/.

[15] E. D. Dickmanns, "Vehicles Capable of Dynamic Vision," *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence,* vol. 2, pp. 1577-1592, 1997.

[16] T. Jochem, D. Pomerleau, K. Bala and J. Armstrong, "PANS: A portable navigation platform," in *Intelligent Vehicles Symposium, Proceedings*, Detroit, MI, USA, 1995.

[17] D. A. Pomerleau, "ALVINN: An autonomous land vehicle in a neural network," *Advances in Neural Information Processing Systems,* vol. 1, pp. 305-313, 1989.

[18] M. Bertozzi, A. Broggi, G. Conte and R. Fascioli, "The Experience of the ARGO Autonomous Vehicle," *Proceedings of SPIE,* 1998.

[19] B. Popper, "Guiding Light - The billion-dollar widget steering the driverless car industry," The Verge, 18 October 2017. [Online]. Available: https://www.theverge.com/2017/10/18/16491052/velodyne-lidar-mapping-self-driving-car-david-hall-interview.

[20] S. Thrun, "Toward Robotic Cars," *Communications of the ACM,* vol. 53, no. 4, pp. 99-106, 2010.

[21] W. D. Jones, "Keeping cars from crashing," *IEEE Spectrum,* vol. 38, no. 9, pp. 40-45, 2001.

[22] I. Bogost, "The secret history of the robot car," The Atlantic, November 2014 2014. [Online]. Available: https://www.theatlantic.com/magazine/archive/2014/11/the-secret-history-of-the-robot-car/380791/.

[23] A. Hartmans, "How Google's self-driving car project rose from a crazy idea to a top contender in the race toward a driverless future," Business Insider, 23 October 2016.

[Online]. Available: https://www.businessinsider.com/google-driverless-car-history-photos-2016-10.

[24] A. M. Kessler, "Elon Musk Says Self-Driving Tesla Cars Will Be in the U.S. by Summer," New York Times, 15 March 2015. [Online]. Available: https://www.nytimes.com/2015/03/20/business/elon-musk-says-self-driving-tesla-cars-will-be-in-the-us-by-summer.html.

[25] D. Yadron and D. Tynan, "Tesla driver dies in first fatal crash while using autopilot mode," The Guardian, 1 July 2016. [Online]. Available: https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk.

[26] P. Gao and R. Hensley, "A road map to the future for the auto industry," 1 October 2014.

[27] J. Fleetwood, "Public Health, Ethics and Autonomous Vehicles," *Public Health Ethics,* vol. 107, no. 4, pp. 532-537, 2017.

[28] World Health Organization, "Global Status Report On Road Safety," World Health Organization, Geneva, 2018.

[29] D. A. Hennessy, "The Impact of Commuter Stress on Workplace Aggression," *Journal of Applied Social Psychology,* vol. 38, pp. 2315 - 2335, 2008.

[30] D. Wurhofer, A. Krischkowsky, M. Obrist, E. Karapanos, E. Niforatos and M. Tscheligi, "Everyday Commuting: Prediction, Actual Experience and Recall of Anger and Frustration in the Car," in *AutomotiveUI '15*, Nottingham, United Kingdom, 2015.

[31] S. Vivek, "Can Autonomous Vehicles Avoid Traffic Jams?," medium.com, 19 March 2020. [Online]. Available: https://medium.com/swlh/can-autonomous-vehicles-avoid-traffic-jams-db039ff412c4.

[32] B. van Arem, C. J. G. van Driel and R. Visser, "The Impact of Cooperative Adaptive Cruise Control on Traffic-Flow," *IEEE Transactions on Intelligent Transportation Systems,* vol. 7, no. 4, pp. 429-436, 2006.

[33] Center for Sustainable Systems, University of Michigan, "Autonomous Vehicles Factsheet," 2020.

[34] B. Metz, O. Davidson, P. Bosch, R. Dave and L. Meyer, Climate change 2007: Mitigation of climate change, Cambridge Univ. Press, 2007.

[35] B. McAuliffe, M. Lammert, X.-Y. Lu, S. Shladover, M.-D. Surcel and A. Kailas, "Influences on Energy Savings of Heavy Trucks Using Cooperative Adaptive Cruise Control," in *WCX World Congress Experience*, Detroit, Michigan, SAE International, 2018.

[36] J. M. Anderson, N. Kalra, K. D. Stanley, P. Sorensen, C. Samaras and O. A. Oluwatola, "Autonomous Vehicle Technology: A Guide for Policymakers," RAND Corporation, Santa Monica, CA, 2016.

[37] E. Scharnhorst, Quantified Parking: Comprehensive Parking Inventories for Five US Cities, Washington D.C.: Research Institute for Housing America, 2018.

[38] N. J. Goodall, "Machine Ethics and Automated Vehicles," in *Road Vehicle Automation. Lecture Notes in Mobility*, Springer, Cham, 2014, pp. 93-102.

[39] P. Lin, "Why Ethics Matters for Autonomous Cars," in *Autonomous Driving*, Berlin, Springer, 2016, pp. 69-85.

[40] M. Green, "How Long Does It Take to Stop? Methodological Analysis of Driver Perception-Brake Times," *Transportation Human Factors,* vol. 2, pp. 195-216, 2000.

[41] P. Foot, "The problem of abortion and the doctrine of double effect," *Oxford Review,* 1967.

[42] J. J. Thomson, "Killing, letting die, and the trolley problem," *The Monist,* vol. 59, no. 2, pp. 204-217, 1976.

[43] A. Shariff, J.-F. Bonnefon and I. Rahwan, "Psychological roadblocks to the adoption of self-driving vehicles," *Nature Human Behaviour,* vol. 1, p. 694–696, 2017.

[44] E. Awad, S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon and I. Rahwan, "The Moral Machine experiment," *Nature,* vol. 563, pp. 59-64, 2018.

[45] J.-F. Bonnefon, A. Shariff and I. Rahwan, "The social dilemma of autonomous vehicles," *Science Mag.,* vol. 352, no. 6293, pp. 1573-1576, 2016.

[46] B. Krzanich, *Data is the New Oil,* AutoMobility Los Angeles, 2016.

[47] N. E. Boudette, "Optimism fades for self-driving automobiles," *New York Times,* pp. 1,8, 19 July 2019.

[48] H. A. Husein, "The role of street traffic signs in reducing road accidents," in *First International Symposium on Urban Development*, Iraq, 2013.

[49] C. Allen, I. Smit and W. Wallach, "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches," *Ethics and Information Technology,* vol. 7, pp. 149-155, 2005.

[50] I. Asimov, "Runaround," *Astounding Science Fiction,* vol. 29, no. 1, pp. 94-103, 1942.

[51] I. Asimov, "The Evitable Conflict," *Astounding Science Fiction,* vol. 29, no. 1, pp. 48-68, 1950.

[52] S. D. Baum, "Social choice ethics in artificial intelligence," *AI & Soc,* vol. 35, pp. 165-176, 2020.

[53] V. Dignum, "Responsible Autonomy," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, 2017.

[54] B. Duignan and H. R. West, *Utilitarianism,* 2020.

[55] M. Taylor, "Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians," Car and Driver, 7 October 2016. [Online]. Available: https://www.caranddriver.com/news/a15344706/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/.

[56] N. J. Goodall, "Ethical decision making during automated vehicle crashes," *Transportation Research Record,* vol. 2424, no. 1, pp. 58-65, 2014.

[57] A. F. Winfield, K. Michael, J. Pitt and V. Evers, "Machine Ethics: The Design and Governance of Ethical AI and Autonomous Systems [Scanning the Issue]," *Proceedings of the IEEE,* vol. 107, no. 3, pp. 509-517, 2019.

[58] R. Metz, "Microsoft's neo-Nazi sexbot was a great lesson for makers of AI assistants," MIT Technology review, 27 March 2018. [Online]. Available: https://www.technologyreview.com/2018/03/27/144290/microsofts-neo-nazi-sexbot-was-a-great-lesson-for-makers-of-ai-assistants/.

[59] Institute of Electrical and Electronics Engineers, "7.8 IEEE Code of Ethics," [Online]. Available: https://www.ieee.org/about/corporate/governance/p7-8.html.