Security and ethics in driverless vehicles

Cédric Prieels

March 9, 2020

1 Introduction

The society we live in is perpetual evolution and one of the latest major breakthrough of the last decades was the development of artificial intelligence techniques. These techniques were developed and introduced in order to cope with an ever increasing flow of data coming from all the parts of today's society with the objective of making our lives a bit easier, allowing for example the development of automation techniques, in particular the creation of self-driving cars that will be introduced in this work. There is nowadays little doubt that such autonomous vehicles will be a fundamental part of tomorrow's society. Indeed, a recent study done by Morgan Stanley [1] showed that autonomous cars are expected to first of all generate huge economic benefits, estimated to around 1.3 trillion of dollars each year just in the US, by reducing the number of traffic accidents, saving fuel, reduce congestion and by the global productivity increase induced by such a way of moving people. It is not only about money though, as the reduction of the accidents due to the development of such driverless cars is also expected to save lives, while car-to-car communication techniques can also help saving fuel and therefore reach our target towards a reduction of the global warming over the next few decades.

However, despite of all the advantages the development of such technology can have, it also raises a lot of questions, mostly political and ethical. Indeed, even though the driverless cars are expected to strongly reduce the estimated 1.35 million lives lost each year due to traffic accidents (WHO 2018 [2]), many challenges and questions remain, and need to be addressed over the next few years, before the emergence of this kind of vehicles. Such challenges include first of all several **security challenges**: how can we make sure that driverless cars are safe enough and cannot be hacked in any way? How can we protect the user data that will be generated, in such a way to protect his privacy? There are also quite a lot of **political challenges**: what happens when such a car will be involved in an accident? Who will be responsible and should be held accountable? Finally, and this will be the main focus of this work, a lot of **ethical challenges** and dilemmas are also raised by such a new technology: if the accident cannot be avoided, then the car software will have to take highly important decisions in a matter of a few seconds, such as deciding whether to save the occupant of the car or an eventual pedestrian crossing the street. A computer might be required to decide the worth of a human life in the future, and this obviously raises a lot of legitimate important questions, as we will now briefly discuss.

2 Issues and dilemma raised

2.1 Privacy issues

Even though not life threatening, the issue related to the privacy of the data generated by such vehicles has to be taken into account. This issue is already important today, even though we do not have access yet to fully autonomous cars: vehicles such as Tesla's cars already collect lots of data from their surroundings in order to improve its Autopilot software, mostly relying on machine learning techniques and which therefore globally benefits from more training data. It has actually been estimated that each autonomous car in the US will generate approximately 300Tb of (sensitive) data each year [3]. This data collected

by the GPS, cameras and sensors on the car and usually sent to the cloud can obviously threaten the privacy of the driver and, even though sending such sensitive data to centralized servers is not mandatory, most of the users are not even aware of such data collecting. All this generated data then needs to be completely anonymized using advanced techniques similar as the ones studied during this class, to make sure not to violate the privacy of the drivers. However, the debate on these kind of privacy issues is not new, especially in this post Cambridge Analytica scandal period, and several laws have already been put in place, especially in the EU (such as the GDPR on data protection), protecting the customers against eventual abuses at this level. Additionally, a car manufacturer alliance already published in 2014 seven "Privacy principles" guidelines governing the collection and the use of the driver related collection of data from self-driving vehicles, supposed to help protecting the customers [4].

2.2 Security issues

As any computational technology, the software of autonomous vehicles is susceptible of having flaws or being hacked, putting in danger the occupant or pedestrians around the car, especially given the fact the such vehicles heavily rely on the software developed and need to be in constant communication with the outside world. We need to consider such possibility now, to secure in the best way possible such cars before a complete fleet of driverless cars start circulating all around the world. These security issues can mostly grouped into two different categories, as we will now develop:

- **Software flaws.** The first security threat in this domain concerns eventual flaws in the software of self driving cars, especially at the beginning, when they are the weakest. The first case of a pedestrian fatality was registered in this context in 2018, caused by a self-driving Uber test car [5]. Even though this pedestrian was detected 6 seconds in advance, the software only decided to activate the breaks 1.3 seconds before the impact, which was too late. The accident was attributed to a sensor failure and a distraction of the actual person checking the behaviour of the car, and raised many questions about the safety of the current fleet of preliminary self-driving cars deployed.
- **Hacking.** In order to make the traffic a bit more fluid, some kind of car-to-car communication system will have to be put in place, based on the GPS information from each car and sent to the cloud in real time, making it vulnerable to an attack. Encrypting and securing the full data sent to the cloud is of course an option but this usually takes time, therefore delaying the transmission of the data, which can be problematic in some specific conditions. But the cloud is not the only way an hacker could harm a driverless car. Indeed, a recent study showed that road signs can be made completely invisible to current self-driving cars, or that their meaning can be altered using a very simple algorithm, giving the vehicle wrong information about its surroundings and leading to inappropriate reactions [6]. Finally, the fact that each car constructor typically uses a different programming language and environment for its software makes matter even worse, since the integration of different small systems may usually introduce weaknesses in the global application. Penetration and systematic testing of the different systems is therefore extremely important, in order to detect any eventual weakness or vulnerabilities, even though a perfect secure system does not exist, the risk has to be minimized as much as possible.

As we have just seen, security is extremely important when designing the software of autonomous vehicles and even though car constructors are usually not experts in cybersecurity, this will have to change in the future, especially since people are going to put their life in the hand of a software, which therefore needs to be extremely reliable. However, this is not the only challenge to overcome when designing such cars: ethical dilemma are also a great deal, as we will now see, and are probably even harder to address in this context, because in this case, there is no obvious "right" and "wrong".

2.3 Ethical issues

First of all, we can consider that the first cars of this kind will probably be quite expansive, and therefore reserved to richer people first, who will be the first to benefit from the increased safety and productivity introduced by such cars. Is it fair to reserve such an advantage in terms of safety to those people, or is it better to wait until such cars can be build in a cheap way to put them on the market?

Autonomous vehicles will probably heavily rely on machine learning techniques, which needs to be trained, therefore raising first of all many questions. Which proportion of the training of such methods should be dedicated to regular driving scenarios, and which proportion should be dedicated to the few cases where an accident is going to happen? Even though accident scenarios are quite rare, should the training be focused towards these particular cases to make sure to react properly when they happen, or should the training be instead focused on the regular driving, to try and avoid having an accident at all cost? This decision involves judgment and is typically subjective, making it an important ethical question to study. Additionally, should this training be more focused against scenarios happening in richer regions of the world, where the car is more susceptible to be used and bought, or towards every possible small differences between the countries? Avoiding any bias when defining the training dataset in extremely important, but not easy to achieve in reality.

Let's consider an additional ethical dilemma related to the training of such models, for the few cases where an accident is due to happen, which can be related to trolley problem, introduced in 1967 by the philosopher Phillipa Foot and which asks the philosophical question of whether people would be ready to decide of the death of one individual in order to save many others. This problem raises the following ethical question: what should a driverless car do in case where an accident cannot be avoided? Should it try to save the driver life at all cost, even if it means that an eventual pedestrian might be run over, or do the exact opposite? Should the car always try to save the greatest amount of people, or should it also consider the age of the different people involved in the crash, to try and save younger children, for example? In case a crash cannot be avoided, should a motorcyclist doing the right thing and wearing a helmet should be targeted by the car over someone not wearing such protection, because he has higher odds to survive the crash? These questions, and so many others, are extremely hard to respond to but need to be addressed since they will drive the software of future cars. In order to address such questions, the Moral Machine tool recently developed by the MIT [7] which collected more than 40 millions ethical choices made by users coming from 233 countries, in order to guide the development of future softwares based on the opinion of a large number of people, according to a study published in Nature [8]. The results obtained by this study are quite obvious but still interesting, with an obvious preference from people to spare humans over animals, to try and save the highest number of individuals and to spare the young over the elderly (even this last distinction depend quite a lot on the region, and is actually going the other way around in countries such as Japon).

There are different ways of programming softwares in order to deal with such ethical questions. The first way to proceed, even though not easy to implement programmatically, is to make the decision the will result in the greatest good, for the greatest amount of people. This would follow the Utilitarian principles introduced by the philosopher Jeremy Bentham. In this case, the fact that different regions of the world have different opinion about what is the "right" thing to do is of course an issue. Should a car then behave differently depending on the country where it is used or built? Another way to think about this issue would be to follow Immanual Kant's idea, according to which the car is supposed to keep its driver safe at all cost, since this is actually what it was designed to do: this is the so-called moral duty principle. The question remains on whether it is the car constructor or the customer itself that should be responsible of choosing between these two options. Several countries attempted to design preliminary guidelines

helping to solve such questions, such as Germany [9], stating that "the protection of individuals takes precedence over all other utilitarian considerations", but also that "in the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical, or mental constitution) is strictly prohibited", which goes in the same direction to the kind of discrimination strictly prohibited by the Institute of Electrical and Electronics Engineers (IEEE).

Apart from all these programming related questions, other ethical questions are worth studying at this point. First of all, the job of many different people actually depend on them driving (taxi, bus drivers, among many others). The development of artificial intelligence and automation in general, even though it might save a lot of money and increases productivity, of course results in many people losing their job. A 2017 study [10] investigated on the impact automation might have on the number of available jobs until 2030. This impact obviously depends on the country, wage rates and income levels, but globally concluded that the number of available jobs in 2030 should however stay more or less constant, even due to automation, but that workers will probably need to shift to other areas, where the employment is supposed to grow over the next few years. The question is to know whether the eventual loss of jobs in certain domains might be a reasonable cost to pay when we consider all the advantages previously quoted and introduced by autonomous cars.

Apart from this, the political and liability question is extremely important as well. In case of accident, who should be responsible? The driver or the car manufacturer? Should the laws regulating autonomous driving be decided country by country, or in a more global way? Should private companies such as Google have a word in the making of laws surroundings this kind of activities?

2.4 Conclusions

Society will probably benefit from the introduction of autonomous vehicles for the reasons previously quoted. However, such a change in the society comes from several challenges that need to be solved. First of all, concerning the security: before the introduction of such driverless cars in the streets, we need to make sure they are secure enough (since people lives will actually depend on this) and that they cannot be hacked, even though this a strong assumption to make, since no system is completely secure, especially when it involves the exchange of information with the outside world. Ethical questions are also extremely important in this context. The introduction of autonomous vehicles means that people will lose their jobs and might have to change their area of activity in the future. The machine learning models involved in the decision making can be quite easily hacked and their training has to be done in such a way that it allows to prevent accidents, but to also be able to react in case of accident, according to ethical principles that need to be decided before hand. At the end of the day, much work still needs to be done, in order to answer the most important question of all: will the benefits of developing such a new technology outweighs the inherent drawbacks?

3 Bibliography

[1] C. Weiss, S. Gaenzle and M. Romer, "How automakers can survive the self-driving era", as seen in February 2020
https://www.es.kearney.com/automotive/article?/a/how-automakers-can-survive-the-self-driving-era

[2] World Health Organization, "Global status report on road safety 2018", as seen in February 2020
https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/

[3] A. Chaturvedi, "Implications of data privacy once autonomous vehicles hit the roads", as seen in February 2020
https://www.geospatialworld.net/blogs/implications-of-data-privacy-once-autonomous-vehicles-hit-

[4] Autoalliance, "Privacy Principles for Vehicle Technologies and Services", as seen in February 2020
https://autoalliance.org/connected-cars/automotive-privacy

[5] National Transportation Safety Board, "Preliminary report highway HWY18MH010", as seen in February 2020
https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY18MH010-prelim.pdf

[6] K. Eykholt et all., "Robust Physical-World Attacks on Deep Learning Visual Classification" [arXiv: 1707.08945]
https://arxiv.org/pdf/1707.08945.pdf

[7] MIT, "Moral Machine", as seen in February 2020
http://moralmachine.mit.edu/

[8] E. Awad et all, "The Moral Machine experiment", Nature volume 563, pp. 59–64 (2018)
https://www.nature.com/articles/s41586-018-0637-6

[9] C. Lutge, "The German Ethics Code for Automated and Connected Driving", DOI: 10.1007/s13347-017-0284-0
https://www.researchgate.net/publication/320011270_The_German_Ethics_Code_for_Automated_and_Connected_Driving

[10] McKinsey & Company, "Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages", as seen in February 2020
https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-