

A NORMATIVE SYSTEMS PERSPECTIVE ON AUTONOMOUS SYSTEMS: THE FUTURE OF DRIVING

Wouter van Haaften / Tom van Engers / Robert Meijer

Ph.D. researcher, University of Amsterdam, Leibniz Center for Law
Nieuwe Achtergracht 166, 1018 WV, Amsterdam, NL
vanHaaften@uva.nl; <http://www.leibnizcenter.org>

Professor, University of Amsterdam, Leibniz Center for Law
Nieuwe Achtergracht 166, 1018 WV, Amsterdam, NL
vanengers@uva.nl; <http://www.leibnizcenter.org>

Professor, University of Amsterdam, Institute for Informatics / TNO
Science Park 904, Postbus 94323, 1090 GH Amsterdam, NL
R.J.Meijer@uva.nl

Keywords: *normative systems, autonomous systems, autonomous vehicles, legal framework*

Abstract: *The recent growth in autonomous systems brings challenges to our society. Autonomous vehicles in particular bring challenges to road authorities in playing their role in initial road allowance. Hence, in this article we will look further into the desirability and the possibilities of in-car artificial intelligence to support the automated vehicle in becoming a safe road user. Meaningful human control is required in order to rightly assign social responsibility for automated vehicles. Providing the automated vehicle with the proper law-based driving software should be based on that principle. Previous work on relevant normative systems could help to accelerate this research.*

1. Introduction

Autonomous systems such as autonomous driving vehicles, robots and softbots are no longer phenomena that only exist in our laboratories. In fact, such autonomous systems are hot. Newspapers are publishing articles on the development of autonomous systems on a weekly basis and the number of scientific conferences on the subject is obviously growing. Autonomous systems are becoming more and more applied in different fields like warfare, health care and financial trading, and penetrating our households; robotic vacuum cleaners and lawn mowers have become wide spread. All these particular fields raise a variety of ethical and legal questions on subjects like, can autonomous systems independently perform legal acts and can they be held responsible for those acts. Under most current legal systems this is not possible. These systems are, and probably will be, very much human-orientated, with human beings as subjects behind them. Humans are considered responsible for their actions and able to bear that responsibility due to the fact that they can be corrected in case of a failure. They are «punishable», so to speak, socially responsible.

Our normative framework addresses a clear need: To fit in any aspect of human society. The robotics industry for example is perfecting the behaviour of robots to be as human-like as possible. Yet, robots cannot take social responsibility. This element specifically plays a role when autonomous systems are being charged with activities that could be potentially harmful to the environment, like driving a vehicle. Nevertheless, car companies claim that autonomous driving will be a reality in the near future. Apparently, they are willing and able to take on the responsibility for autonomous driving. In 2017, a few of the largest car manufacturers (OEMs) worldwide have announced the introduction of self-driving cars. In the US, Tesla is already advertising an autopilot capable of self-driving under most highway-conditions. Also other OEMs have presented supporting services for the driver that might also be able to drive automatically in a limited amount of circumstances and time, the so-called Advanced Driver Assistance Systems (ADAS). In that case, the driver is still in control of

the vehicle, while in the case of a fully autonomous driving¹ vehicle a human driver is no longer required. In the transfer phase, the level of control may be distributed between human drivers and the vehicle controller who will take control and therefore responsibility in specific contexts, e.g. driving on a motorway. The burden of responsibility is dependent on having control over the vehicle and therefore lies either with a human driver or with a vehicle controller.

One of the first challenges to come up is whether the automated vehicle, besides trial and error testing, will be instructed with the basic traffic rules as laid down in the Geneva or Vienna Conventions. Human drivers are supposed to have knowledge of these rules, like driving on the right side of the road, even if that means driving on the left side in some cases, and stopping when a traffic lights goes red, after having turned yellow for some time. Nothing special to a human driver, but how will the automated vehicle know these rules and how to apply them? How does an automated vehicle comply with the traffic rules in order to make it a safe road-user?

In this legal perspective it becomes interesting to see what OEMs are actually doing to prepare their vehicles for an independent life as road user. We know they are driving millions of test miles with their vehicles, and rightly so, that is what all OEMs do before they launch a new model. But this time they not only test the car's hardware and some supporting software, but they also have to test the active driving behaviour of the vehicle as a road user in many different traffic situations. To support this, they use, for example, data on traffic accidents in order to «teach» the car how to avoid collisions. They are in fact training the car, i.e. its (deep) learning algorithm, by offering it a large training set of features of situations, trusting that this will enable the car to learn and eventually become a sufficiently adequate «driver». But will that be enough? What are OEMs doing to make sure that collisions are avoided? As far as the statistics on accidents are concerned: they may be useful in mixed traffic, along with human drivers, but how likely is it that automated cars will cause the same kind of accidents as human drivers? Automated vehicles are not easily distracted, are not checking their email, but they may have difficulty checking out situations on the road that human drivers overlook in a split second. And what if an accident happens? Wouldn't we want the car to be able to explain which decisions it made and for what reasons, if only to be able to determine responsibility for the accident? So it really seems worthwhile to look more in depth into the traffic education of the automated vehicle. The responsibility for the accident with an autonomous vehicle has at least two sides. The first one is the tort law liability. When the autonomous vehicle causes an accident while not obeying the traffic rules, the controller of the vehicle will be liable. It will have to pay for the damages. And if it happens again it will have to pay again. To avoid this situation, the OEM will have to explain how the accident could happen and how it will make sure that it doesn't happen again. To achieve that level of control the software of the vehicle must be legislation/rule based and moreover, be absolutely transparent in the development of its driving intelligence. Besides the tort liability, this will be the incentive for the controller to make sure its vehicle will function flawlessly, thus taking on its societal responsibility.

2. Autonomous driving car software

Modern vehicles are already largely software-driven on aspects like motor management, breaking system, suspension and steering. Many cars have ADAS-features like lane departure warning and control, traffic sign recognition as well as speed control and automatic distance control. The self-driving road user capabilities of the vehicle are obviously software-driven as well, but the possibilities for the driver to take back control over the car will be quite limited. As a consequence of the car taking over driving from humans, the vehicle does not only need a type approval and an admission for the vehicle's software, it will also need a drivers licence.

¹ SAE level 4 and 5.

The first part, just like for human drivers, starts with the theory, the basic traffic rules as represented in the various national legislations. Those traffic rules should be the basis for the self-driving software of the vehicle, meaning that these rules will have to be explicitly modelled thus allowing for rule-based driving software that can also explain the cars' decisions in certain circumstances, such as in the case of a traffic accident. Such a basic module should be in every self-driving vehicle permitted on the road. A test for compliance to the traffic regulations may even have to be generic, taken by the proper authorities rather than the OEMs, like it is the case in the human traffic theory exam.

The second part of the development of driving skills with regards to automated vehicles will be more challenging. It is hard to imagine that road authorities would allow automated vehicles on the road without serious insight in the driving skills of the vehicle. A half hour practical exam will give some insight in the capabilities of the vehicle, but it will probably not show the depth of the driving capabilities present, and the algorithms lying underneath. On top of that, just like the human driver, the self-driving vehicle may be learning every day on the road from the traffic situations it encounters. This learning process must be controlled. Moreover, it will bring up a more fundamental artificial intelligence question: What has the machine learned and what will the machine learn in the future, i.e. how will it interpret situations and transform them into future behaviour?² This is where the concept of «meaningful human control»³ comes in. How can human control of a vehicle that is self-driving and using machine learning algorithms be maintained?

3. A future scenario of autonomous driving vehicles

In Fig 1 we show a future scenario of autonomous driving vehicles. In this scenario we distinguish the minimally required two to three phases of autonomous driving, neglecting potential accidents and dispute settlements after such incident. In such a «happy flow scenario», the first phase is the admission test. In addition to regular tests that all cars are subjected to, the software of the autonomous car will be subjected to millions of generated traffic conditions to see if it has the right behaviour and proper implementation of rules. The second phase is the check before the car starts driving. Because of the fact that the software will be regularly updated each autonomous drive could start with a specific internal software admission test. This would include a check if the software and hardware of the car are admitted. If so, then the controller of the vehicle could release it for use on the road. The third phase is the actual driving of the vehicle, where sensors of the vehicle allow the vehicle's software to build an actual representation of the vehicle's environment and – crucially – its normative position towards all other participants in traffic that are or may be impacted by the vehicle's behaviour. In Fig 1. this is indicated by the normative model that is being constantly updated, or better instantiated, when the vehicle participates in traffic. The normative model is abstractly represented as a series of Hohfeldian cubes as these are the basis for our modelling. This paper is too short to explain all details of this normative modelling so we refer to DOESBURG/VAN ENGERS (2016) for more details on this modelling approach.

² DEI SIO 2016.

³ UNIDIR 2014, HOROWITZ/SCHARRE 2015.

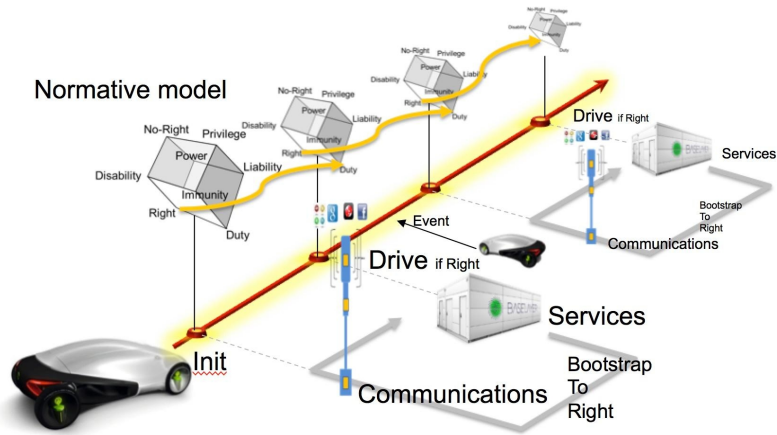


Fig. 1. Requirements check before driving and communication with other traffic participants during driving.

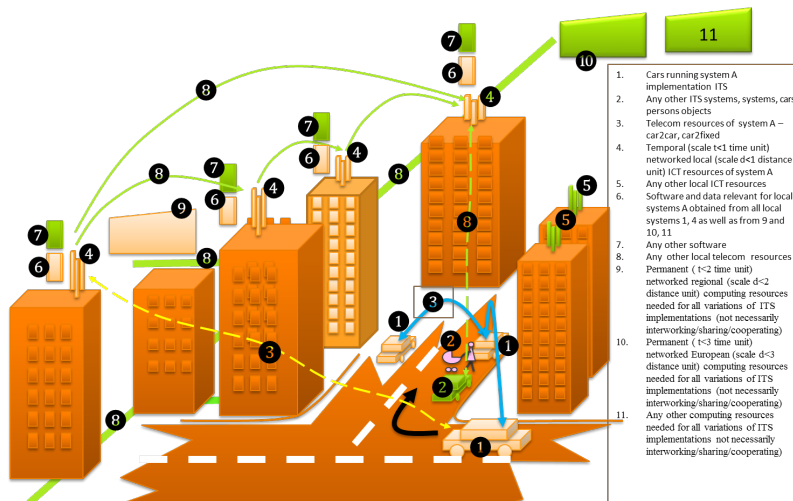


Fig. 2. ITS systems for cooperative driving, supported by communication infrastructure.

If we would look at this future scenario of autonomous driving vehicles from a legal perspective then we could conceptualize the autonomous driving vehicles as «instrumental agents» controlled by an OEM. The entire traffic situation would thus consist of multiple OEMs whose presence in the situation is manifested through the vehicles that operate as «instrumental agents». Consequently, there will be liability relations between passengers and OEMs, amongst different OEMs and between OEMs and other traffic participants etc. Also other parties, including the road authorities and road controller, may have responsibilities and consequent liabilities. In the ITS systems (see Fig. 2.) currently being developed for cooperative driving for example, various parties play a role in communicating information essential for safe driving to the vehicles.

One could think of various messages that are sent to those vehicles, such as the electronic version of traffic signs, traffic jams and roadblocks. The reliability of this information is important even if the actual situation

will always have to be compared with information received through the other sensors of the vehicle. Being able to produce evidence of what information was available at the moment a decision was made, next to the decision model it self, is essential for establishing responsibilities in case of accidents.

4. Allowing for meaningful human control

An important challenge that holds for all autonomous systems is how we can allow for meaningful human control. In the context of autonomous driving vehicles this can be phrased as: «How to build «fully autonomous» vehicles that do not carry the risk of getting out of meaningful human control?⁴». Here, a basic set of rules is required to ensure that the software in the vehicle, which may be using various forms of artificial intelligence, including sub-symbolic ones, will not take over beyond human comprehension. Incidents should always be imitable in order to be analysed in such a way that the results are transparent and accountable, in order for human judges to be able to subject them to our human values and norms as included in our legal systems. This means that two modelling actions should be performed:

1. Modelling traffic rules as represented in the national legislation into software
2. Modelling transparency and accountability in or on top of the deep learning module of the vehicle.

The result should be that the self-driving vehicle verifiably «knows» the rules and is learning from its practice in a transparent and accountable way.

5. A formal model of traffic regulation

Already in the early 90 's in the TRACS-project supported by the Dutch Association for Scientific Research into Traffic Safety (SVOW), a formal model of Dutch traffic rules was created and translated into software⁵. The aim of the project was to construct an intelligent teaching system for traffic law. However, since the SVOW was also concerned with testing new versions of the traffic law, presented to them by the Ministry of Transport and Public Works, looking for anomalies in these new versions of traffic law became the researchers' core interest. Although the project ran long before artificial intelligence would enter into the vehicle itself, one of the conclusions drawn could be very relevant to the implementation of traffic rule based software into the control module of the self-driving vehicle. By bringing two paragraphs of the law into the system it was uncovered that in the relatively simple test case two of the three positions of road users in a certain situation were misinterpreted by following the rules of the law.

While the work in the paper was almost forgotten, it has become extremely relevant in the context of today's autonomous driving vehicles for which the approach taken in the TRACS project provides us with a good basis of research on meaningful control of autonomous systems.

Research on meaningful control of autonomous vehicles is expected to go in at least four directions:

1. Testing the rules (law) on suitability for modelling and for application, in this case traffic regulations on autonomous driving vehicles.
2. Modelling the relevant rules into control software, in this case traffic regulations in control software of autonomous driving vehicles.
3. Testing the control software, in this case the control software of autonomous vehicles, in a virtual environment that covers all possible situations.
4. Monitoring the actual behaviour and intervene if necessary.

This research will cover the implementation of the legislation into rule-based software.

Modelling transparency and accountability will take a separate research approach.

⁴ Hearing on Self driving cars, March 2016.

⁵ BREUKER/DEN HAAN, 1991.

6. Actors and responsibilities

Along with the modelling and testing of the autonomous driving software, it also has to be determined who will be responsible. The current relation, when it comes to admitting new vehicles on the road, is between the OEM and the vehicle admission authority. This relation provides for an admission largely based on conventional vehicle technology. Since the introduction of software in vehicles it is extended with the corresponding steering software, as far as the safety-related features of the vehicle are concerned. In recent years another extension took place, namely to the emission performance of the vehicle. In the Volkswagen (VW) case it turned out that compliance with the admission regulation is not self-evident. The next step could well be the extension to the autonomous capabilities of the vehicle. Does the autonomous vehicle comply with the traffic regulations and how does it comply? Does it learn from experience on the road and how do we know it does not pick up bad habits specifically from fellow human drivers? This means that not only the starting position of an autonomous vehicle has to be established, but also its lifetime learning mechanisms.

Due to these developments the relation between OEMs and the Authorities intensifies and gains impact for both parties, and in fact for society as a whole. The vehicle has to be admitted to the road based on the criteria set out in the admission legislation. These criteria include build quality, road safety and driver/passenger/fellow road users protection⁶. Admittance is a public interest. It is in the interest of the public that cars are safe, that the emission is within the legal standards and that the software that is running many functions in the vehicle is well designed and tested. So, where we now have norms for the deceleration of a vehicle and the steering clearance and the CO_x and NO_x emissions, in the future we will also need norms for autonomous software.

In order to maintain those legal norms, the processes to which they relate should be sufficiently transparent. This transparency nowadays often lacks in autonomous functions although it is necessary for another legal requirement: accountability. The emission fraud was not noticed at the time the admission tests were performed, partly due to a lack of transparency regarding the software. Would the software have been transparent or even certified to the legal norms than the fraud would not have occurred, or at least would have been immediately noticed. When it comes to software for autonomous vehicles the same demand for transparency arises. Software is not an autonomous entity even if it steers an autonomous vehicle. Ultimately, someone has to be responsible for the consequences, as the emission case showed. If not, OEMs and software developers will not have much incentive to comply with the regulations and autonomous cars may become unpredictable black boxes.

7. Vehicle admission process

Looking at the admission process and the developments in that respect we start at the first phase, the late nineteenth century, when no admission rules had come into effect yet. At a certain point in time this situation was no longer acceptable and so the first admission rules were established, mainly to have some guarantee on the build quality and the breaking, steering and lightning of the vehicle, basic requirements that preserve product-, user- and road safety. The formally approved car was born. In Table 1 we give an overview of this development.

The formally approved car still exists today although it has developed into a high-tech machine, supported by loads of software while retaining its basic functions like breaking, steering and lights to facilitate the use in night hours. With a human driver at the wheel, the legal sky was open and sunny. But what if the car becomes autonomous? From the public perspective the «autonomous» part of the vehicle should be able to drive safely and up to certain standards. Specifically, during the transition period, that will cover many decades, it will be vital that the autonomous vehicle «knows» how to behave in human traffic. This means that the autonomous vehicle has to drive at normal speeds and be capable of maintaining this speed safely in various circumstances.

⁶ Regulations 167/2013/EU and 168/2013/EU.

Not like, for instance, in tests with public autonomous vehicles that will reduce speed whenever a potential danger may be within a «hundred-meter» range. The autonomous car should behave as human-like as possible in the mixed situation where both autonomous and human-driven cars will be taking part in traffic. But it has its limitations. Reading the body language of a pedestrian approaching a crossing, or quick and subtle eye contact and understanding will not be possible. Admittedly, the vehicle itself will need and have a lot of sensors and cameras to feed the autonomous driving unit with information. Also, it will have to be connected to other road users in order to be able to reduce safety margins, for instance while bouncing in and out of motorways. The autonomous vehicle will still lack the human capability of dealing with complex interrelated situations in split seconds. Nevertheless, the sensors, cameras and the telecommunication with the roadside and other vehicles should be able to compensate for the limitations in communication with fellow human road users as much as possible. This new set of tools to preserve road safety should be publicly established and put into legislation, just like the requirements concerning the admission of vehicles. In order to avoid competition law problems, the requirements should be appointed as functional as possible.

| | <i>Scheme responsibilities vehicles</i> | | | | 3 January 2018 | | Version 0.1 | |
|-------|---|--|--|---|---|-------------------------------------|--|---|
| Phase | Activity | Goal | Public actors | Private actors | Responsibilities | Responsible entity | Compliance tools | Result |
| A | None | Facilitate driving/innovation | Road authorities/police | OEMs, owners/drivers | Product-related, traffic-related | OEM, User | Civil Law, Penal Law | no control at public level |
| B | Admission vehicles | Road user protection, product/traffic safety | Politics/Legislator-Admission authorities/police-judge | OEMs | Product-related, traffic-related, use-related | OEM, Admission Authorities | Legislation, admission requirements | Public control over admitted vehicles at the moment of admission |
| C | Extended admission, periodical check on the maintenance of the vehicle | Road user protection, product/traffic safety, | Politics/Legislator-Admission authorities/inspection stations/police-judge | OEMs, owners/drivers, inspection stations | Product-related, traffic-related, use-related | Owner, Admission Authorities | Legislation, admission requirements, certification | Public control over admitted vehicle during life cycle |
| D | Extended admission, emissions and safety-related items like airbags etc., comfort apps, Advanced Driving Assistance Systems | Road user protection, product safety, active endorsement of traffic safety, emission reduction | Politics/Legislator-Admission authorities/inspection stations/police-judge | OEMs incl. software update providers, owners/drivers, inspection stations | Product-related, use-related, traffic-related, software-related | OEM, owner, Admission Authorities | Legislation, admission requirements, certification | Public control over admitted vehicle during life cycle and over ADAS software |
| E | Extended admission, emissions and safety-related items like airbags etc., self-driving abilities | Road user protection, product safety, active endorsement of traffic safety, emission reduction | Politics/Legislator-Admission authorities/inspection stations/police-judge | OEMs incl. software update providers, [owners], inspection stations | Product-related, use-related, traffic-related, software-related | OEM, [owner], Admission Authorities | Legislation, admission requirements, certification | Public control over admitted vehicle during life cycle, over autonomous driving software and the development of autonomous driving software |

Table 1. The development of regulated responsibilities of the various actors over time.

8. Conclusions

The previous analysis leads to the conclusion that the responsible entity for the actual driving software will bare most of the responsibility. To be able to live up to this responsibility this entity should have control over the autonomous vehicles status. The vehicle could be considered as an agent of the controlling entity, the controller. Although it is too early to make definite decisions on how the legal framework for autonomous vehicles should look like, it seems a workable solution that puts the liability where it can best be bared, thus allowing authorities, occupants and other road users to feel comfortable with autonomous driving vehicles in mixed traffic. In the years to come, the vehicle will only be partly autonomous, for instance on high ways. One of the issues to solve is the take-over of the responsibility from the vehicle, and its controller, to the human driver. The driving software should foresee a smooth and flawless take-over in order not to create uncertainty about the driving responsibility. We expect the driving software to contain several parts.

The first part should be the normative aspect based on the general traffic rules. All relevant traffic rules should be known to the vehicle software and should form the basis of both handling in traffic and of learning. This will not be easy and it could become such a task, that OEMs will not have it done individually, but would like to implement this software as a commodity, provided for by certified IT companies.

The second part of the driving software will be the heart of the autonomous vehicle. It will combine the traffic rules software with the data from the sensors and the connection with other vehicles and roadside stations. Additionally, the software will be collecting data from its experience as a road user. These data can be used to enhance the performance of the autonomous vehicle, thus facilitating the software to integrate all data into safe and compliant driving behaviour.

But what if something goes wrong and the autonomous car ends up in an accident? What will be the position of the controller of the vehicle and of the victims? Will we, as a society, want to know what happened as precisely as possible, in order to address the liability and to be able to learn from the mistakes that were made? We think and hope this will be the case. And that is where meaningful human control comes in. Software subject to this control will be able to reproduce the decision making in the last moments before the accident, because it will be transparent. All in order to achieve that the information coming from the EDR (Event Data Recorder) that registers the last short period before the crash, will lead to the accountability of the controller as a legal entity responsible for the autonomous vehicle and its road behaviour.

The focus in this paper was on autonomous vehicles, but autonomous systems are used in other areas too. These other autonomous systems are to be human controlled in some way as well and clarifying the ethical and normative rules is essential for enabling trust in those systems. Obviously, creating the legal frameworks and technologies to enforce the desired behaviour of those autonomous systems will require further inter-disciplinary research. Application domains such as autonomous vehicles, industrial robots, autonomous decision-making systems in public administration and the banking sector all require solutions that need such a normative system approach, which is essential to keep humans in control of the artefacts they create. Power to the people!

9. References

- J. BREUKER/N. DEN HAAN, Separating world and regulation knowledge: where is the logic, in: ICAIL '91 Proceedings of the 3rd international conference on Artificial Intelligence and Law, pp. 92–97, AOEM, 1991.
- UNITED NATIONS INSTITUTE FOR DISARMAMENT RESEARCH (UNIDIR), The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control might move the discussion forward, <http://www.unidir.ch/files/publications/pdfs/considering-how-meaningful-human-control-might-move-the-discussion-forward-en-615.pdf>, 2014.
- R. VAN DOESBURG/T.M. VAN ENGERS, A Formal Method for Interpretation of Sources of Norms, AI and Law Journal, 26-1, 2018 (to be published).
- R. VAN DOESBURG/T.M. VAN ENGERS, Perspectives on the Formal Representation of the Interpretation of Norms, Artificial Intelligence and Applications, 2016, pp. 183–186.
- R. VAN DOESBURG/T.M. VAN ENGERS, CALCULEMUS: Towards a Formal Language for the Interpretation of normative Systems, Artificial Intelligence for Justice Workshop, ECAI 2016.