Problems with the prospective connected autonomous vehicles regulation : Finding a fair balance versus the instinct for self-preservation

Mamak, Kamil

2022-11

Mamak , K & Glanc , J 2022 , ' Problems with the prospective connected autonomous vehicles regulation : Finding a fair balance versus the instinct for self-preservation ' , Technology in Society , vol. 71 , 102127 . https://doi.org/10.1016/j.techsoc.2022.102127

http://hdl.handle.net/10138/350498
https://doi.org/10.1016/j.techsoc.2022.102127

Technology in Society 71 (2022) 102127

Contents lists available at ScienceDirect

Technology in Society

journal homepage: www.elsevier.com/locate/techsoc

Problems with the prospective connected autonomous vehicles regulation: Finding a fair balance versus the instinct for self-preservation

Kamil Mamak [a,b,*], Jadwiga Glanc [1]

[a] University of Helsinki, RADAR: Robophilosophy, AI Ethics and Datafication Research Group, Finland
[b] Jagiellonian University, Faculty of Law and Administration, Department of Criminal Law, Poland

ARTICLE INFO

Keywords:
Crash algorithms
Ethical dilemmas
Responsibility
Autonomous vehicles

ABSTRACT

There would seem to be a potential regulatory problem with the crash algorithms for connected and autonomous vehicles that (normally) "kick in" should a crash be inevitable. Although the general regulatory considerations which have been put forward tend to seek a "fair balance" that would protect various users of the roads, a configuration which shielded other parties at the cost of the car user could be seen as posing an existential threat to the user by encroaching on his or her right to self-preservation. "Hacking the system" (i.e. modifying the configuration in order to obtain a more favourable outcome for the user) could therefore be understood as acting on one's instinct for self-preservation and – though illegal – could in certain situations turn out to be an action that is not punishable in law. The present article argues that in certain real post-crash situations, a person who has modified the code for his or her own benefit could be exonerated on the basis of existing legal provisions and thus go unpunished. This could create unforeseen flaws in the connected autonomous vehicles regulatory system.

1. Introduction

Although the advent of connected and autonomous vehicles (CAVs) has given rise to a rather fierce discussion on their regulation, certain sets of guidelines and rules on CAV traffic have gradually seen the light of day. The trait which they all have in common is that they well-nigh unanimously exclude the idea of a "selfish" algorithm setting that would prioritize the safety of the driver of the CAV over that of other road users. Although it comes as no surprise that the proposed rules should attempt to find a "fair balance" for all users of the roads, this would seem to leave a certain regulatory shortcoming or "loophole" that we would like to address in the present article. Indeed, this quite peculiar "loophole"— which in all probability would hardly ever be used— could eventually prove to be the "Achilles' heel" of all the CAV regulations that are currently being considered, as it lays bare a potential gap between legal practice and legal theory (in this case the "perfectly" regulated AI reality that many dream of) — a gap that would create a state of tension between the CAV user's instinct for self-preservation and a regulatory scheme that could be seen as posing a definite danger to that same user.

The key issue addressed here is the fact that by enacting CAV rules which attempt to find a fair balance between the interests of all users of the roads, the CAV user is being exposed to a permanent threat to his life that is based solely on the existence of crash algorithms and which therefore creates a strong incentive to find a solution to the problem by modifying the crash algorithm in favour of the CAV user. In any subsequent court proceedings, the CAV user might well be exonerated and thus evade criminal liability, in which case his or her crime would prove to be "unpunishable".

The present article does not in any way question the importance of creating rules for CAVs, nor does it propose to offer any "all-encompassing" solution for the afore-mentioned problem which is currently

* Corresponding author. University of Helsinki, RADAR: Robophilosophy, AI Ethics and Datafication Research Group, Finland.
E-mail addresses: kamil.mamak@helsinki.fi, kamil.mamak@uj.edu.pl (K. Mamak).
[1] Independent Scholar

https://doi.org/10.1016/j.techsoc.2022.102127
Received 21 June 2022; Received in revised form 14 September 2022; Accepted 15 September 2022
Available online 18 September 2022

posed by CAV regulation. Its aim is merely to pinpoint a certain legal problem which could arise and which could potentially undermine the system that is at present being built.[2] In particular, our arguments are based on the considerations and proposed rules that have been put forward by the European Commission Report of 2020 (because of their importance and the all-embracing character of the subject matter). In this article we shall look at certain regulatory considerations and ideas that have been proposed (**I**) and confront them with the possibility of "hacking" one's own car (together with the legal consequences thereof) (**II**), as well as with the possible social consequences of such an occurrence (**III**). In the article, we confront the possibility of regulatory intervention with the limits of rulemaking, as well as with the philosophical and legal basis for such limits.

2. The dream of the prefect CAV regulation

The concept of creating autonomous and connected cars has always had a regulatory layer attached to it (cf. [1,2]. From the very beginning, it was apparent that before CAVs could be allowed on public roads, some existing rules would need to be modified and/or new ones put in place.

This issue was (and still is) one of great consequence and one that is highly controversial, if only because – at least in part – it is a matter of life and death. Road traffic will most probably never be completely safe, owing to various unpredictable external contingencies, while the "Moral Machine" experiment has made it even more apparent that in certain situations a choice would have to be made as to who is going to be the "casualty" [3].

From a legal perspective, however, these decisions would have to conform to certain pre-defined ethical rules and be anchored in law and thus not based on arbitrary algorithms, as no one (outside the limited legal powers of the State) should be able to decide on the life or death of an individual citizen. This would also be necessary for the predictability of traffic.

Connected and autonomous vehicles (CAVs) are a general name for various kinds of the more or less autonomous car (cf. [4]; 12). The question of "autonomy" is also a question about the machine's judgement (cf. [5,6]. Much has been said on this subject and of late it has been one of those which are most widely discussed (cf. [3,5,7–15].

One particularly problematic issue has been the specific nature of decisions taken by autonomous vehicles, especially those decisions which lead to collisions, as these are based on certain algorithmic settings. They are therefore not the consequence of carelessness or ignorance, but the result of meticulously "thought-out" scenarios (cf. [7]; 74). It has been noted that while the driver may not have time for deliberation, an autonomous car might be able to react in an "optimal way" and make a more informed decision, taking into account many variables that are not available to or that cannot be processed by a human driver [7]; 74). When thinking about a non-avoidable accident involving a human driver, we understand that certain circumstances were beyond the driver's control, thus forcing him to fall back on instinct. Such a situation is also understandable from the point of view of the public, as we cannot make the driver fully responsible for the outcome. In the case of automated vehicles, however, we are dealing with a scenario that is planned and executed and that has usually been chosen from among many others. What the problem boils down to, therefore, is: who should be allowed to die (and why)?

Here we are getting into very murky territory, where decisions are made as to whose life is more "valuable". Given the variety of moral perspectives on this issue, the question of algorithm settings would at first sight seem to be far from straightforward. This is reflected in the sheer variety of opinions and standpoints. This issue was laid bare by the Moral Machine experiment, which showed that preferences and standpoints vary [3]. Thus the question of how to regulate the situation in real life became even more important. One of the authors of the original Moral Machine Experiment, for example, has been reported to be leaning towards the idea of manufacturers being allowed to take "cultural differences" into account while designing CAVs [16].[3] At the initial stage of development, certain car makers signalled a preference for constructing a vehicle that would prioritize the safety of the driver over that of other road users [12]. One journalist suggested that we should ultimately be able to choose between car models that would share our own moral point of view [17]. We can therefore see that ideas for algorithm settings abound and differ widely, being based as they are on a diversity of moral judgments.

When considering legal solutions, however, this concentration on various ethical choices is not very helpful. As has been rightly noted, ethical dilemmas dominate the discussions and obfuscate other issues [18]. CAVs are already with us and a response – a legal one, at least – is urgently needed. The ethical problem – to which there seems to be no "right answer" and which is therefore unsolvable [18] – can to a certain extent be addressed by the basic principles of legal systems – in particular rules of a constitutional kind (or something similar), e.g. the EU Treaties and the EU Charter of Fundamental Rights [19]. Although ethical reasoning is normally the source of legal solutions, paradoxically in this case existing legislation could help us navigate the dark waters of dilemmas connected with CAVs. Indeed, when we look at various regulatory propositions, they are based on certain well-grounded legal principles. We might even say that we are re-cycling the ethical considerations on which these principles are based.

It has been stressed that AI should be lawful, meaning that it is required to comply with laws and regulations (AI High Level Expert Group 2018, 2). This would also apply to CAVs (on different grades of autonomy) and their operation. The issue of including CAVs in legal frames in order to permit their operation has been discussed in various contexts: there have been very valuable publications discussing the legal dilemmas in the context of certain existing legal systems (e.g. Ref. [20], as well as more universal considerations, bringing certain general rules into the equation (see in particular: [21]. Also, the last few years have seen publications giving recommendations on how to include CAVs in normal road traffic and how to devise special legal regimes based on ethical recommendations issued in particular at a State or multi-State level [4,22,23]. As the CAV problem that we shall be facing is a universal one, all of these publications are of importance because they shed light on the likely ethical perspective and the possible regulatory response.

Discussions on life and death situations in the case of CAVs are a very particular issue for several reasons, some of which will be mentioned here. As far as the dilemma situations are concerned, it has been pointed out that "genuine dilemma situations" would probably be very rare [23]; 11) and that the present ethical dilemma is probably not the most pressing or problematic issue in the CAV regulations [4]; 17). To some extent it was also argued that – owing to current technical developments and the state of the art, etc. – they would in any case be unlikely to occur [23]. Having said that, it has also been recognized that the general public would be paying close attention to possible negative occurrences,

[2] During the process of its writing, this article has received many comments from lawyers and non-lawyers alike, for which we are very grateful. Interestingly, this "feedback" has revealed not only a great discrepancy between the understanding of the core purpose of the regulatory process and its possible "effectiveness", but — in many cases — a different understanding of the boundaries of law-making.Arguably, we believe, it could be attributed to the subject matter and the not-so-uncommon idea that we are dealing with something that is "magical", being the ultimate problem solver (AI). Some lawyers have tended to believe that in the case of AI we are dealing with something that is different from imperfect "human" reality and which therefore requires new rules. Some philosophers have shown a lack of understanding of the boundaries of law-making and also of the possibility of a legal excuse.

[3] Referring to Azim Shariff.

which alone could have a great influence on the acceptability and adoption of CAV technology [8]; 88). It would probably seem intuitive to agree with Lin, who says that some accidents would be unavoidable and would thus necessitate crash optimization, by which he means the necessity of making a choice "between two evils" in order to somehow minimize any harm done [7]; 72). Also, even a marginal number of potential casualties – perhaps even the death of one person – would suffice to make the question and the answers regulatorily relevant for CAVs to operate in a particular system without any vagueness or blind spots.

It has also been rightly noted by Bonnefon, Shariff and Rahwan in one of their contributions that although the classical trolley dilemma may well be over-sketched and that the cars may not be making decisions between the "outright sacrificing of the lives of some to preserve those of others", they will nevertheless be making decisions about who is to be put "at marginally more risk of being sacrificed", which on a large scale would result in a shifting of the risk and in a larger perspective could be detrimental to certain groups (e.g. pedestrians), having similar consequences to the classically framed trolley dilemma (the "statistical" trolley dilemma – [24]. Krügel and Uhl – the authors of a recent study which included a more realistic scenario of CAV accidents with an element of risk (uncertainty), including stochastic trolley problems – point out that the public expects a well-balanced reflection on the potential risks and harm as a whole and not only their consideration when accidents are imminent [25]. It has been pointed out that the standard trolley problems would seem to have retained their relevance for ethical questions about CAVs [25]; 9). It has been said that as a society, we will need to decide what we view as being a fair distribution of risks for road users [24]. This is also what the new regulations somehow attempt to address – at least in a preliminary fashion and on a more abstract level.

3. The idea of the "selfish" car as a public enemy

The last few years have brought not only ethical and legal discussions on the issue, but also the first regulatory and ethical recommendations and their explanations. From these endeavours, a slowly evolving framework for the regulation of CAVs has emerged and is notable for its attempt to strike a balance between the interests of various road users and also for its tendency to exclude – at least to a certain extent – the idea of the "selfish car" from those solutions which at present are available.

Having said this, we would like to analyse the particular tension that exists between the interests of the CAV car user and that of other road users and its influence on potential regulation and factual outcomes. The CAV car user is potentially in a more privileged position than most other road users, as – particularly in the case of a crash – he or she is protected by the car itself. This imbalance is and has been noticeable in the case of conventional (non-CAV-based) traffic, where cyclists or pedestrians are highly vulnerable. The envisioned system of CAV operations would ideally correct this imbalance, as has been provided for in Recommendation 5., which speaks of "redressing inequalities in vulnerability among road users" [4]; 7). This added protection would somehow have to be implemented in regulatory considerations and in the algorithms themselves.

To make this possible, however – and thus have a level playing field – we would have to combat the idea of the "egoistic car" (or "selfish car"). One of the important points of the ethical discussion has been the opposition between "selfish algorithms" and "utilitarian algorithms" [26]. Generally speaking, selfish algorithms would give the greatest priority to the occupant of the CAV (i.e. the car user), which would lead to the phenomenon of crash optimalization as described by Lin, who has pointed out that the choice (or target) of the crash could also depend on the particular interest that has been encoded – to protect the car's occupants or to protect other road users. If the "attitude" is to protect the occupant of the car, the CAV will choose what or who to crash into – and so will probably choose "easy" (i.e. lighter) "targets" such as a motorcycle instead of a car or a child instead of an adult [7]; 72). This would be equivalent to targeting [7]; 72).

This state of affairs would hardly be acceptable, especially when trying to improve the situation by introducing a balance of protection. To achieve a balance and add protection for vulnerable road users, the selfish algorithms would somehow have to be limited or curbed, which would constitute an encroachment on the "vital" interest of the CAV user.

Moving further on to the issue of when the car would have to decide whether to kill someone outside the car (especially uninvolved pedestrians) in order to save the life of the driver and his or her passengers, the answer should be in the negative (at least in theory). On a more abstract and theoretical plane, we must remember that any person who uses a car is knowingly putting himself/herself and other members of the public at risk (cf. [7]; 80). By allowing vehicles on the roads, Society as a whole is agreeing to a certain level of risk (and to a certain number of inevitable car accidents) for the sake of convenience and the needs of commuting. This is, of course, a necessity of our times and the idea that cars should simply be banned altogether is obviously a non-starter. However, the idea that preferential treatment should be given to those who are the root cause of the risk would seem to be going against the basic principles of fairness. An exception might be a situation in which, for example, we were dealing with a pedestrian who was acting against the rules, e.g. crossing the street in an irregular way or running onto a motorway. In such cases, the car should of course do all it can to avoid or minimize a collision, but it should not be expected to save the life of a "rogue road user" at the cost of that of the driver. The general rule of a German commission has stated that "those parties involved in the generation of mobility risks must not sacrifice non-involved parties" [22]; cf. [23]. It has been noted that such a principle would mean that the CAV's algorithm cannot make the idea of unconditionally saving the life of the driver a general rule [27]; 553).

The limiting or "curbing" of the "selfish car" concept with its driver prioritization could be variously tailored, with different scopes of protection being given to different parties. However, we might consider the example of a situation in which a CAV avoiding a crash swerves onto the pavement and kills uninvolved passers-by. Such a situation would be considered unacceptable by the State and would greatly detract from the general popularity of CAVs. Some curbing of "selfish algorithms" would therefore seem to be imperative.

It is also worth mentioning that in order to somehow avoid or at least overcome (in a more morally-neutral manner) the tricky question of preferences and algorithm settings, the idea of "randomization" has also been considered, whereby in the eventuality of a radical choice having to be made, such as the one under discussion, the machine would randomise its decision [28] To a certain extent, this would eliminate the burden of arriving at a morally grounded choice by basing the decision on randomness. A recent study on moral preferences and randomness has shown that randomization preferences were displayed in a stronger manner in situations with actual ethical dilemmas [29]. This could therefore seem to be a tempting scenario should we be facing the most radical choices – or at least it could be presented as being "the best of the completely bad solutions". This, however – though technically feasible on an individual level – could potentially lead to negative consequences (if, for example, two CAVs were to be involved in the collision, two random choices could end up worsening the possible outcome – not only systems, but also traffic normally benefits from the predictability of behaviour). It is also unclear if and exactly how that would appeal to car users [29] and – as far as our "hacking scenario" is concerned – whether it would make the alteration of the system less likely.

These are some of the most important points in the discussion and all of them display a certain tendency to establish a "fair" balance – in particular between the protection given to the user of the car and the protection given to other parties. In other words, they exclude the one-sided privilege of the CAV owner. However, one might ask what the latter himself/herself might think of such a noble approach. In

particular, the car owner may be tempted to tamper with the underlying algorithms in order to prioritize himself/herself in the event of a crash – be it in the case of algorithms prioritising other parties, or in the case of algorithms set to randomness.

4. Modyfing CAV settings for one's own benefit and criminal liability

At this juncture it is important to clarify that this paper is not about who should be responsible for the "regular" crashes of an autonomous car. This is a separate question that needs to be answered (cf. [30–32]. Here we are talking about a distortion of the system – i.e. about an atypical situation in which someone is changing something that has been set. This can be illustrated by the following example: let us imagine that there is an electricity company that is providing electricity to an individual customer. In general, the company should take care of the system's maintenance and if a problem arises or some damage is done, we can check to see whether the company has followed all its safety procedures. If, however, someone gains illegal access to the power grid and causes damage to another user, it is he who will be held legally responsible. In the same way, the autonomous car user who hacks the system could be held responsible for his actions.

In this article the word "hacking" is understood somewhat loosely in order to describe changes in the crash algorithms of CAVs that have been made by the owners. The actual hacking of autonomous cars is a separate problem that could arise together with the advent of autonomous and connected cars [33].

We make two assumptions. Firstly, that there will be legally binding guidelines covering the ethical dilemmas of crash algorithms. These guidelines cannot be based on the prioritization of cars, i.e. they reject the idea of the "selfish car", as this would be contradictory to the values described in the Ethics of Connected and Automated Vehicles Report (justice, solidarity and dignity) [4].

The second assumption is that the car user will generally not be responsible for the accident. Briefly put, responsibility assumes that there is an element of knowledge and control (see e.g. Ref. [34], which here would exclude the responsibility of the car user. "Generally", because we could imagine behaviours of the car owner which could be treated as being criminal, for example not keeping the car in good condition, not uncovering the sensors before starting the car, not changing the tyres or simply throwing something out of the window and causing a collision. The owner could, of course, be held criminally responsible for such behaviour. Traffic involving autonomous cars does not mean that road crimes will be a thing of the past.

Let us imagine that there is a collision in which a pedestrian dies under the wheels of a CAV. There will be an investigation. If the authorities find that the crash algorithm has been altered, then the person who changed it – and who therefore (partially, at least) determined the outcome (Recommendation 19. of the Ethics of Connected and Automated Vehicles Report) – may well be held responsible for the death of the pedestrian.

The problem here, however, is that although such behaviour may be unlawful, this does not necessarily mean that the car owner will be criminally liable, as – for a crime to be committed – there is a requirement that guilt be ascribed to the perpetrator, this reflecting the classic rule *nullum crimen sine culpa*. For example, the attribution of guilt is excluded when the perpetrator is mentally ill or is under age. In the case of tampering with algorithms, we could speak of a state of necessity such as self-defense [21]. The person who changes the algorithm in order to prioritize himself does so in order to protect his own life. The existence of a state of necessity means that the perpetrator is excused. He cannot be held culpable and therefore cannot have committed a crime. Thus he evades punishment.

Although the construct of legal excuse may be structured differently in various legal systems, it has a common basis: Legal excuse – especially self-defense – is one of the basic concepts of criminal law, which is based on the instinct for self-preservation. Ashworth notes that the right to life and physical security is considered to be "the most basic claim of every human being." [35]; 282). In his treatise entitled *Leviathan* (1651), Thomas Hobbes claims that "If a man by the terror of present death, be compelled to do a fact against the law, he is totally excused; because no law can oblige a man to abandon his own preservation." [36]; 199–200). It has been pointed out that the existence of the instinct for self-preservation is a barrier that cannot be transgressed by law [37]. On the more general level, therefore, even the very creation of a law that poses a permanent danger to the user is questionable.

Moreover, the legal excuse that could be invoked is not merely some general moral consideration that could be balanced against other moral considerations for the purpose of making the system "bullet-proof", as there are existing legal provisions on which such a "defense" could be based. Although different systems have different prerequisites (e.g. considering whether the danger was "imminent" or not), if it came to invoking someone's legal responsibility and using this excuse, this would mean that the danger of the accident had materialized.

Taking all this into consideration, even if a balanced CAV regulation were to be put in place — as indeed it is planned to be — it would not be an all-encompassing solution.

To conclude, although hacking one's own car in order to prioritize the person in the car might be seen as being unlawful, it may well be impossible to attribute blame to the perpetrator because he or she will be able to excuse himself or herself on the ground that he or she was merely following the dictates of the instinct for self-preservation.

5. Societal risks and ways in which they could be mitigated

This section will indicate potential risks to the desired values which should be embedded in the organization of traffic comprising automated and connected cars (see, for example, the Ethics of Connected and Automated Vehicles Report [4]). Identified risks result from the modification of car algorithms in order to prioritize the people in the vehicle. Such behaviour could be treated as a legal excuse that excludes the imposition of blame on the perpetrator. Although such an action could be treated as being illegal in the light of criminal law, the perpetrator could not be punished, as he could not be held criminally responsible. Such a state of affairs could have serious implications, as it might undermine the desired balance between the values which we want to be at the core of CAV traffic control.

The scale of the risks depends on the scale of the modifications. If changes are marginal, the problem will also be marginal. If the scale of the modification is broader, the risks will be higher. However, we should also be concerned even if there is a marginal scale of modification, as the impact on Society does not solely depend on the scale of the changes. Even individual cases of modification could have a huge social impact, irrespective of the factual scale of the problem. Initially at least, the public's attention will be focused on every anomaly connected with the new traffic system, especially if deaths are involved. The worldwide press gave wide coverage to the first death connected with the operation of an autonomous vehicle, even though at that same time thousands of people were dying in "regular" car accidents all around the world. After that experience, we should carefully assess every potential risk associated with the functioning of the system, as any irregularities could cause a public outcry against these new technologies (see e.g. Ref. [38].

Below we present the potential spheres in which the problem of attributing blame could militate against the values on which system should be built.

5.1. The risk of looking for a scapegoat for unpunished harm

In the [4] report, there is a claim that the lack of culpability for the behaviour of CAVs that cause harm could lead to people looking for a scapegoat (Recommendation 19, [4]. The report even uses the term "culpability gap" [4]; 61), which has connection with the similar term

"responsibility gap" that is often associated with autonomous systems (cf. [39–42]; see also [43]. Here there is no problem of ascribing responsibility, for we do know who changes the algorithms and we do know the person's intention. The problem is that it will not be possible to ascribe blame and, accordingly – given the rule *nullum crimen sine culpa* – this means that no crime has been committed. This in turn could lead to a "retributive gap" [44] because there will be no one to punish. Danaher draws attention to the risk of scapegoating. The risk indicated in the report materializes in the problem discussed. The person who is responsible for altering the algorithm is excused, despite the fact that such an action could be unlawful. Thus there may be a death that cannot be punished in criminal law. The perpetrator could even publicly admit that he had changed the algorithm, yet still be acquitted.

5.2. The excessive privileging of people in cars in relation to other road users

Creating a CAV system based on the values presented in the Report assumes that no category of participator is privileged at the expense of others. The report clearly states that "[…] no category of road user (e.g. pedestrians, cyclists, motorbike users, vehicle passengers) should end up being more at risk of harm from CAVs than they would be against this same benchmark." [4]; 25). All human beings share the instinct for self-preservation, but – on the roads, at least – human beings in cars are in a more favourable situation than other road users. Cyclists and pedestrians hardly have a chance in confrontations with people in cars. The far-reaching consequences of using legal excuses to protect users of cars mean that they are overprivileged when they are on the road. This in turn would seem to contradict the value of justice, as exposure to harm is not distributed fairly. We are already seeing a trend to "reclaim the streets" for pedestrians [45,46]. It is doubtful that Society at large would accept the fact that CAVs are overprivileged.

5.3. The risk of rising social inequalities

Another far-reaching consequence could be the deepening of social inequalities. In certain places around the globe – and especially in developing countries – people in cars are usually wealthier than pedestrians or cyclists (cf. [47]. The fact that people in cars are overprivileged by the system might send a message to Society saying that the lives of car users are worth more than those of pedestrians or cyclists. Such a state of affairs would run counter to the principle of dignity. According to the report, "Dignity is the basis of the equality of all human beings and forms the normative point of reference that grounds human rights." [4]; 21). The CAV system should not give the impression that the lives of poorer citizens are worth less than those of the rich.

Thinking about how we could mitigate the above-mentioned risks, we are forced to admit that there is not much that can be done. However, we would like to draw attention to four possible new ideas.

5.4. Abandon the force of the legal excuse

Could and should we change the law regarding the legal excuse? This is technically feasible. The law is a social technology (cf. [48] and human beings could change it, since it was they who created it in the first place. Here, however, it is the word "should" that is problematic. Based as it is on the instinct for self-preservation, the legal excuse is at the core of criminal law and would therefore seem to be untouchable: any attempt to remove it from the law would simply be unacceptable, as this could lead to the creation of laws that were unethical. In a life-threatening situation, everyone should have the right to self-defence. Any violation of the law on this point would merely create a new set of problems.

5.5. Creating new crime

Could we create a crime that will "indirectly" forbid tampering with algorithms? For example, "Whoever impacts the functioning of the autonomous car is subject to a penalty". This is also technically possible and such a crime could be a proxy crime (see e.g. Ref. [49]. Leaving aside the issue as to whether such a crime would meet the criteria for criminalization [50] – the harm principle, for example (cf. [51–53] – the main problem will remain. As with every crime, we will have to examine the motive of the person's action, the elements of the crime and then attribute guilt. The judge will assess the motives of the crime and motives based on the instinct for self-preservation could be assessed as before, i.e. as an excuse that excludes the attribution of guilt.

5.6. The infrastructure of the roads

We are at the stage of thinking about how to design traffic that includes CAVs. We should think not only about how cars will be designed, but also about all other elements of road infrastructure. The building of the system should be based on how the participants actually behave and not on how we would like them to behave. This is also recommended directly in the report: "[…] CAV technologies should be designed to reflect the road users' psychological capabilities and motivations […]" [4]; 18). If the instinct for self-preservation is unavoidable and we cannot force car users to participate equally in the distribution of harm, we should at least limit the spheres where there could be mutual interactions. What should this mean in practice? The number of areas in which collisions are possible should, for example, be reduced to a minimum and there should be a warning that an autonomous car is approaching. Admittedly, this is not very much. There will still be places where mutual interactions are unavoidable, but designing a new road in a city centre would be an entirely different enterprise if the CAV issue was not problematic.

5.7. Resistance to modification

The most promising measure depends on a design that would restrict the possibility of tampering with the system that governs crash algorithms. The success of this measure depends on technical possibilities. This part of the software should be designed with the awareness that changes could hinder the success of the whole project, making it impossible for other participants and Society as a whole to accept CAVs in the public sphere. At the designing stage, therefore, an effort should be made to make it impossible to change the algorithms and there should also be a monitoring system recording any changes made after the cars have been sold. Potential changes should be undone by updates, while security loopholes that allow car owners to make alterations should also be eliminated.

6. Conclusions

To conclude, although hacking one's own car in order to prioritize the person in the car might be seen as being unlawful, it may well be impossible to attribute blame to the perpetrator because he or she will always be able to excuse himself or herself on the ground that he or she was merely following the dictates of the instinct for self-preservation. This is something that cannot be ignored when considering how to organize the system and is an issue that should be the subject of further debate. We hope that our article will trigger new investigations. We believe that the use of the qualitative and quantitative approach – for example, taking into consideration the interviews carried out with several lawyers and other specialists – could be of particular value (see other studies, cf. [54–59].

Ethics approval and consent to participate

Not applicable.

Consent for publication

We consent to publish this paper.

Availability of data and material

Not Applicable.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article uses results of a research project financed by the Academy of Finland: 333873.

Authors' contributions

The authors' contribution is equal. All parts were written jointly by both co-authors.

Declaration of competing interest

We declare not to have competing interests concerning the submitted manuscript.

Data availability

No data was used for the research described in the article.

References

[1] Thomas A. Hemphill, Autonomous vehicles: U.S. Regulatory policy challenges, Technol. Soc. 61 (May) (2020), 101232, https://doi.org/10.1016/j.techsoc.2020.101232.
[2] Shane Epting, Ethical requirements for transport systems with automated buses, Technol. Soc. 64 (February) (2021), 101506, https://doi.org/10.1016/j.techsoc.2020.101506.
[3] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, Iyad Rahwan, The moral machine experiment, Nature 563 (7729) (2018) 59–64, https://doi.org/10.1038/s41586-018-0637-6.
[4] European Commission, European Commission, Directorate-General for Research and Innovation. Ethics of connected and automated vehicles : recommendations on road safety, privacy, fairness, explainability and responsibility. Publications Office, 2020. https://data.europa.eu/doi/10.2777/035239.
[5] Martin Cunneen, Martin Mullins, Finbarr Murphy, Autonomous vehicles and embedded artificial intelligence: the challenges of framing machine driving decisions, Appl. Artif. Intell. 33 (8) (2019) 706–731, https://doi.org/10.1080/08839514.2019.1600301.
[6] Debbie Hopkins, Tim Schwanen, Talking about automated vehicles: what do levels of automation do? Technol. Soc. 64 (February) (2021), 101488 https://doi.org/10.1016/j.techsoc.2020.101488.
[7] Patrick Lin, Why ethics matters for autonomous cars, in: Markus Maurer, J. Christian Gerdes, Barbara Lenz, Winner Hermann (Eds.), Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte, vols. 69–85, Springer, Berlin, Heidelberg, 2015, https://doi.org/10.1007/978-3-662-45854-9_4.
[8] J. Christian Gerdes, Sarah M. Thornton, Implementable ethics for autonomous vehicles, in: Markus Maurer, J. Christian Gerdes, Barbara Lenz, Winner Hermann (Eds.), Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte, vols. 87–102, Springer, Berlin, Heidelberg, 2015, https://doi.org/10.1007/978-3-662-45854-9_5.
[9] Filippo Santoni De Sio, Ethics and Self-Driving Cars: A White Paper on Responsible Innovation in Automated Driving Systems, 2016. https://repository.tudelft.nl/islandora/object/uuid%3A851eb5fb-0271-47df-9ab4-b9edb75b58e1.
[10] Coca-Vila, Ivo, Self-driving cars in dilemmatic situations: an approach based on the theory of justification in criminal law, Criminal Law and Philosophy 12 (1) (2018).
[11] Jeffrey Gurney, Crashing into the Unknown: an Examination of Crash-Optimization Algorithms through the Two Lanes of Ethics and Law, Social Science Research Network, Rochester, NY, 2016. SSRN Scholarly Paper ID 2622125, https://papers.ssrn.com/abstract=2622125.
[12] Azim Shariff, Iyad Rahwan, Jean-François Bonnefon, Whose life should your car save? The New York Times (2016). November 3, 2016, sec. Opinion, https://www.nytimes.com/2016/11/06/opinion/sunday/whose-life-should-your-car-save.html.
[13] Jean-François Bonnefon, Azim Shariff, Iyad Rahwan, The social dilemma of autonomous vehicles, Science 352 (6293) (2016) 1573–1576, https://doi.org/10.1126/science.aaf2654.
[14] Filippo Santoni De Sio, Killing by autonomous vehicles and the legal doctrine of necessity, Ethical Theory & Moral Pract. 20 (2) (2017) 411–429, https://doi.org/10.1007/s10677-017-9780-7.
[15] Darius-Aurel Frank, Polymeros Chrysochou, Panagiotis Mitkidis, Dan Ariely, Human decision-making biases in the moral dilemmas of autonomous vehicles, Sci. Rep. 9 (1) (2019) 1–19, https://doi.org/10.1038/s41598-019-49411-7.
[16] Caroline Lester, A study on driverless-car ethics offers a troubling look into our values, January 24 (2019) 2019. https://www.newyorker.com/science/elements/a-study-on-driverless-car-ethics-offers-a-troubling-look-into-our-values.
[17] Leetaru, Kalev n.d. "Could buying A driverless car in the future mean selecting its ethical values too?" Forbes. Accessed 1 21, 2020. https://www.forbes.com/sites/kalevleetaru/2019/06/25/could-buying-a-driverless-car-in-the-future-mean-selecting-its-ethical-values-too/.
[18] Tobias Holstein, The misconception of ethical dilemmas in self-driving cars, Proceedings 1 (3) (2017) 174, https://doi.org/10.3390/IS4SI-2017-04026.
[19] European Union, Ethics of Connected and Automated Vehicles : Recommendations on Road Safety, Privacy, Fairness, Explainability and Responsibility." Website, Publications Office of the European, Publications Office of the European Union, 2020, 9 17, 2020, http://op.europa.eu/en/publication-detail/-/publication/89624e2c-f98c-11ea-b44f-01aa75ed71a1/language-en.
[20] Armin Engländer, Das Selbstfahrende Kraftfahrzeug Und die Bewältigung Dilemmatischer situationen, Zeitschrift Für Internationale Strafrechtsdogmatik (9) (2016).
[21] Ivó Coca-Vila, Self-driving cars in dilemmatic situations: an approach based on the theory of justification in criminal law, Criminal Law and Philosophy 12 (1) (2018) 59–82, https://doi.org/10.1007/s11572-017-9411-3.
[22] The Ethics Commission on Automated, Connected Driving, Report of the Ethics Commission Automated and Connected Driving, 2017.
[23] The Task Force on Ethical Aspects of Connected and Automated Driving, Report of the Task Force on Ethical Aspects of Connected and Automated Driving, 2018.
[24] Jean-François Bonnefon, Azim Shariff, Iyad Rahwan, The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars [point of view], Proc. IEEE 107 (3) (2019) 502–504, https://doi.org/10.1109/JPROC.2019.2897447.
[25] Sebastian Krügel, Matthias Uhl, Autonomous vehicles and moral judgments under risk, Transport. Res. Pol. Pract. 155 (January) (2022) 1–10, https://doi.org/10.1016/j.tra.2021.10.016.
[26] Peng Liu, Jinting Liu, Selfish or utilitarian automated vehicles? Deontological evaluation and public acceptance, Int. J. Hum. Comput. Interact. (2021) 1–12, https://doi.org/10.1080/10447318.2021.1876357.
[27] Christoph Luetge, The German ethics code for automated and connected driving, Phil. Technol. 30 (4) (2017) 547–558, https://doi.org/10.1007/s13347-017-0284-0.
[28] Patrick Lin, The Robot Car of Tomorrow May Just Be Programmed to Hit You, 2014. Wired, 2014, https://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/.
[29] Anja Bodenschatz, Matthias Uhl, Gari Walkowitz, Autonomous systems in ethical dilemmas: attitudes toward randomization, Comput. Human Behav. Rep. 4 (August) (2021), 100145, https://doi.org/10.1016/j.chbr.2021.100145.
[30] Sabine Gless, Emily Silverman, Thomas Weigend, If robots cause harm, who is to blame? Self-driving cars and criminal liability, New Criminal Law Rev. 19 (3) (2016).
[31] Kamil Mamak, Odpowiedzialność karna za wypadek drogowy z udziałem samochodu bez kierowcy, Paragraf Na Drodze, 2015 no. 4.
[32] Kamil Mamak, *Rewolucja Cyfrowa a Prawo Karne*. Kraków, Krakowski Instytut Prawa Karnego Fundacja, 2019.
[33] Maurice Schellekens, Car hacking: navigating the regulatory landscape, Comput. Law Secur. Rep. 32 (2) (2016) 307–315, https://doi.org/10.1016/j.clsr.2015.12.019.
[34] Mark Coeckelbergh, AI Ethics, The MIT Press, Cambridge, MA, 2020.
[35] A.J. Ashworth, Self-defence and the right to life, Camb. Law J. 34 (2) (1975) 282–307.
[36] Thomas Hobbes, in: J.C.A. Gaskin (Ed.), Leviathan, Reissue edition, OUP Oxford, 1998.
[37] Jerzy Śliwowski, Prawo Karne, 1979. Warszawa.
[38] Leila Ouchchy, Coin Allen, Veljko Dubljević, AI in the Headlines: the Portrayal of the ethical issues of artificial intelligence in the media, AI Soc. 35 (4) (2020) 927–936, https://doi.org/10.1007/s00146-020-00965-5.
[39] Andreas Matthias, The responsibility gap: ascribing responsibility for the actions of learning automata, Ethics Inf. Technol. 6 (3) (2004) 175–183, https://doi.org/10.1007/s10676-004-3422-1.
[40] David J. Gunkel, Mind the gap: responsible robotics and the problem of responsibility, Ethics Inf. Technol. 22 (4) (2020) 307–320, https://doi.org/10.1007/s10676-017-9428-2.
[41] Peter Remmers, Would moral machines close the responsibility gap? in: Birgit Beck, Michael Kühler (Eds.), Technology, Anthropology, and Dimensions of Responsibility Techno:Phil – Aktuelle Herausforderungen Der Technikphilosophie. Stuttgart: J.B. Metzler, 2020, pp. 133–145, https://doi.org/10.1007/978-3-476-04896-7_10.

[42] Zoë Porter, Ibrahim Habli, Helen Monkhouse, John Bragg, The moral responsibility gap and the increasing autonomy of systems, in: Barbara Gallina, Amund Skavhaug, Erwin Schoitsch, Friedemann Bitsch (Eds.), Computer Safety, Reliability, and Security, Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018, pp. 487–493, https://doi.org/10.1007/978-3-319-99229-7_43.
[43] Daniel W. Tigard, There Is No Techno-Responsibility Gap, Philosophy & Technology, 2020, https://doi.org/10.1007/s13347-020-00414-7. July.
[44] John Danaher, Robots, law and the retribution gap, Ethics Inf. Technol. 18 (4) (2016) 299–309, https://doi.org/10.1007/s10676-016-9403-3.
[45] "Reclaiming City Streets for People : Chaos or Quality of Life?", Publications Office of the European Union, 2004. http://op.europa.eu/en/publication-detail/-/publication/94a8a003-be86-467a-9a85-63a5d52bf7ae.
[46] Casado Pérez, Vanessa, Social Science Research Network, Rochester, NY, 2020. Reclaiming the Streets." SSRN Scholarly Paper ID 3747436, https://papers.ssrn.com/abstract=3747436.
[47] Robert Cervero, Linking Urban transport and land Use in developing countries, J. Transport Land Use 6 (1) (2013) 7–24.
[48] Joshua A.T. Fairfield, Runaway Technology: Can Law Keep up? Cambridge University Press, Cambridge, 2021 https://doi.org/10.1017/9781108545839.
[49] Piotr Bystranowski, Retributivism, consequentialism, and the risk of punishing the innocent: the troublesome case of proxy crimes, Diametros 53 (October) (2017) 26–49, https://doi.org/10.13153/diam.53.0.1099.
[50] Thomas Søbirk Petersen, Why Criminalize?: New Perspectives on Normative Principles of Criminalization, first ed., Cham: Springer, 2019, 2020 Edition.
[51] Hamish Stewart, The limits of the harm principle, Criminal Law and Philosophy 4 (1) (2010) 17–35, https://doi.org/10.1007/s11572-009-9082-9.
[52] Dennis J. Baker, Constitutionalizing the harm principle, Crim. Justice Ethics 27 (2) (2008) 3–28, https://doi.org/10.1080/0731129X.2008.9992238.
[53] Arthur Ripstein, Beyond the harm principle, Philos. Publ. Aff. 34 (3) (2006) 215–245.
[54] Stephen Rice, Scott R. Winter, Do gender and age affect willingness to ride in driverless vehicles: if so, then why? Technol. Soc. 58 (August) (2019), 101145 https://doi.org/10.1016/j.techsoc.2019.101145.
[55] Andrea Sestino, Alessandro M. Peluso, Cesare Amatulli, Gianluigi Guido, Let me drive you! The effect of change seeking and behavioral control in the artificial intelligence-based self-driving cars, Technol. Soc. 70 (August) (2022), 102017, https://doi.org/10.1016/j.techsoc.2022.102017.
[56] Yi-Ching Lee, Momen Ali, Jennifer LaFreniere, Attributions of social interactions: driving among self-driving vs. Conventional vehicles, Technol. Soc. 66 (August) (2021), 101631, https://doi.org/10.1016/j.techsoc.2021.101631.
[57] Brent Smith, Personality facets and ethics positions as directives for self-driving vehicles, Technol. Soc. 57 (May) (2019) 115–124, https://doi.org/10.1016/j.techsoc.2018.12.006.
[58] Taşkın Dirsehan, Ceren Can, Examination of trust and sustainability concerns in autonomous vehicle adoption, Technol. Soc. 63 (2020), 101361, https://doi.org/10.1016/j.techsoc.2020.101361. November.
[59] Diana Escandon-Barbosa, Jairo Salas-Paramo, Ana Isabel Meneses Franco, Carlos Giraldo Gonzalez, Adoption of new technologies in developing countries: the case of autonomous car between Vietnam and Colombia, Technol. Soc. 66 (August) (2021), 101674, https://doi.org/10.1016/j.techsoc.2021.101674.