On the Problem of Making Autonomous Vehicles Conform to Traffic Law

Henry Prakken[1,2]

[1] Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands,
h.prakken@uu.nl
[2] Faculty of Law, University of Groningen, Groningen, The Netherlands

Abstract. Autonomous vehicles are one of the most spectacular recent developments of Artificial Intelligence. Among the problems that still need to be solved before they can fully autonomously participate in traffic is the one of making their behaviour conform to the traffic laws. This paper discusses this problem by way of a case study of Dutch traffic law. First it is discussed to what extent Dutch traffic law exhibits features that are traditionally said to pose challenges for AI & Law models, such as exceptions, rule conflicts, open texture and vagueness, rule change, and the need for commonsense knowledge. Then three approaches to the design of law-conforming AV are evaluated in light of the challenges posed by Dutch traffic law, which includes an assessment of the usefulness of AI & Law models of nonmonotonic reasoning, argumentation and case-based reasoning.

Autonomous vehicles are one of the most spectacular recent developments of Artificial Intelligence. Among the problems that need to be solved is the one of making the behaviour of autonomous vehicles (AV) conform to the traffic laws. Solutions to this problem may well profit from computational models of legal reasoning but so far the field of AI & Law has hardly addressed this issue. The present paper[3] puts this topic on the AI & Law research agenda by way of a case study of Dutch traffic law and its implications for the design of fully autonomous self-driving cars. In particular, the challenges are discussed that Dutch traffic law poses for AV and how existing AI & Law techniques are relevant for dealing with these challenges. In the literature on AV design there have to the best of my knowledge so far not been any systematic studies of the problem of making AV conform to traffic law through its design. Therefore, the present study, while still a conceptual one, fills an important gap in the literature.

The problem of making AV conform to traffic law is a special case of the more general problem of making intelligent autonomous systems conform to the relevant laws. The present study of the problem for autonomous vehicles may therefore also have relevance for the study of the more general problem. Computer systems are increasingly being employed in practice with some degree of autonomy. Their behaviour is not fully specified by the programmer but is the result of the implementation of more general cognitive or physical abilities. Such artificially intelligent systems can do things which, when done by humans, are regulated by law. Apart from self-driving cars, some examples are care robots that help sick or elderly people and whose actions can damage

[3] The full paper has been published in *Artificial Intelligence and Law* 25 (2017): 341-363.

property or the health of the person (spilling coffee over an iPad, failing to administer medication on time), intelligent fridges that can order food or drinks when the supplies run out and thus have to conform to contract law, and financial trading programs that have to conform to financial regulations.

When such autonomous systems are being used, legal rules cannot any more be regarded as regulating human behaviour, since it is not the humans but the machines who act. This raises the problem of how the autonomous systems can be designed in such a way that their behaviour complies with the law. Note that this question needs to be asked irrespectively of the legal question whether machines can be assigned responsibility in a legal sense. Even if a human remains legally responsible or liable for the actions of the machine, the human faces the problem of ensuring that the machine behaves in such a way that the responsible human complies with the law. Currently, this kind of problem is mainly studied under the heading of 'machine ethics'. While this may be the appropriate field for studying the related problem of making intelligent autonomous systems behave ethically responsibly, the problem of making them conform to the law arguably belongs to AI & Law.

For AV designers the paper will hopefully create a deeper awareness of the challenges that traffic law poses for AV design. To start with, an account is given of the similarities and differences between the tasks traditionally modelled in AI & Law and the new task of autonomous driving on public roads. Then three different approaches to the design of law-conforming AV, are discussed, namely, the *regimentation* approach (designing the system in a way that guarantees that the system will not exhibit unwanted behaviour), the *reasoning* approach (giving the system the ability to reason about the lawfulness of its own behaviour) and the *training* approach (letting the system acquire the ability to behave legally correctly implicitly by training). The latter is the one currently usually applied in AV design.

Next, the need is identified for formal specification of traffic law as a component of AV design in all three approaches, even in the training approach. As regards the applicability of knowledge representation and reasoning techniques previously developed in AI & Law, the conclusion is that logic-based techniques for rule-based (monotonic or nonmonotonic) reasoning are largely suitable for representing the logical and hierarchical structure of the regulations but cannot deal with the interpretation problems arising from open texture, vagueness and the need for new exceptions. On the other hand, existing argumentation-based techniques for dealing with open texture, vagueness and the need for new exceptions are largely inapplicable, partly since the adversarial setting presumed by these techniques is lacking in the AV problem and partly because the existing case law is too sparse and inconclusive. A more promising approach is to develop standards and guidelines for implementing law-conforming behaviours in a collaborative effort between the government, industry and possibly insurance companies. However, this does not avoid the need for formal representation of the relevant traffic regulations. Finally, the present paper provides reasons for the relevant research communities to focus less on moral algorithms for decision making in emergency situations and more on careful and anticipatory driving behaviour (while not sacrificing traffic efficiency).