The world's first ethical guidelines for driverless cars

September 4 2017, by David Tuffley

Credit: AI-generated image (disclaimer)

Humans before animals and property. No discrimination as to who should survive. Safeguards against malicious hacking.

These are just some of the world-first ethical rules being implemented in Germany regarding how autonomous vehicles are to be programmed.

The federal transport minister Alexander Dobrindt presented a [report on automated driving](#) to Germany's cabinet last month. The [report](#) is the work of an Ethics Commission on Automated Driving, an expert panel of scientists and legal experts.

The [report](#) notes the technological advances being made to increase automation in cars to make them safer and reduce accidents, but it adds:

"Nevertheless, at the level of what is technologically possible today […] it will not be possible to prevent accidents completely. This makes it essential that decisions be taken when programming the software of conditionally and highly automated driving systems."

The report lists 20 guidelines for the motor industry to consider in the development of any automated driving systems. The minister says that cabinet has adopted the guidelines, making it the first government in the world to do so.

The report allows German car makers to maintain their technological lead, setting a strong [example](#) for the rest of the world to follow.

Automated driving is safer

The moral foundation of the report is simple – since self-driving vehicles will cause [fewer human deaths and injuries](#), there is a moral imperative to use such systems since governments have a duty of care for their citizens.

So what are some of the situations the report considers?

If an accident cannot be avoided, the report say human safety must take precedence over animals and property. The software must try to avoid a collision altogether, but if that's not possible, it should take the action

that does least harm to people.

The report also recognises that some decisions could be too morally ambiguous for the software to resolve.

In these cases, the ultimate decision and responsibility, at least for now, must be with the human sitting in the driver's seat, as control is swiftly transferred to them. If they fail to act, the vehicle simply tries to stop. In the near future, as capability improves, vehicles might well become fully autonomous.

It's acknowledged that no system is perfect. If harmful outcomes cannot be reduced to zero, at least it will be below the current human level.

If a collision is unavoidable, the report say systems must aim for harm minimisation. There must be no discrimination on the basis of age, gender, race, physical attributes or anything else of any potential accident victim.

All humans are considered equal for the purposes of harm minimisation.

This makes the famed Trolley Dilemma irrelevant in as much as the software is not allowed to prioritise an individual's relative worth.

Who is in control?

The report mentions the possibility of fully autonomous systems, but recognises that the technology is not yet capable of solving tricky "dilemma situations" in which the vehicle has to decide between the lesser of two evils. As the technology becomes sufficiently mature, full autonomy will be possible.

According to the report, at all times it must be known who is driving –

human or computer. Perhaps by means of scanning their license, everyone who drives a vehicle must first be validated as being legally qualified to drive that class of vehicle.

The vehicle should have an aviation-style "Black Box" that continuously records events, including who or what is in control at any given time.

In the event of an accident involving an autonomous vehicle, an investigation should be carried out by an independent federal agency to determine liability.

The driver of a vehicle retains their rights over the personal information collected from that vehicle. Use of this data by third parties must be with the owner's informed consent and with no harm resulting.

The threat of malicious hacking any autonomous driving system must be mitigated by effective safeguards. Software should be designed with a level of security that makes malicious hacking exceedingly unlikely.

As for the set-up of a vehicle's controls, they must remain ergonomically optimal for human use, as they are in a conventional car.

The [vehicle](#) can react autonomously in an emergency, but the human may take over in morally ambiguous situations. The controls should be designed to smoothly and quickly make the transfer.

The report says the public must be made aware of the principles upon which autonomous vehicles operate, including the rationale behind any of those principles. This should be incorporated into school curriculums so that people understand both the how and why of [autonomous vehicles](#).

A good start, but a work-in-progress

The guidelines will be reviewed after two years of use. Doubtless there will be fine tuning in the light of experience, in this the first of many reviews in the years and decades to come.

The guidelines are solidly reasoned and comprehensive enough to provide a legal basis for German car-makers to move forward with their plans.

Since other countries appear to have taken a wait and see position on such legislation, they may well decide to follow Germany's example, and not for the first time. This would be no bad thing, since a piecemeal approach from one country to the next would be in no-one's interests.

This article was originally published on The Conversation. Read the original article.

Provided by The Conversation