

Ethics Discourses between Fat Man and Drunk Robot, or: Reflections on the Search for Good Ways of Dealing with Possible Damage from Automated Vehicles

Statement by Dipl.-Phys. Dr. Torsten Fleischer

on the basis of the given, concrete questions for the event

"Im Fokus: Die ethischen Herausforderungen des automatisierten Fahrens", TU Graz, 11 March 2019

Automated driving (AF) now appears to be largely controllable in the course of technical change and, in conjunction with social performance promises, also economically and politically attractive. Research, development and politics are currently concentrating primarily on technical, organisational and legal feasibility.

Both in technical discussions and in public debates, there is a great variety of positions and ideas regarding automation, vehicle and service concepts, the degree of technical maturity, entrepreneurial implementation strategies, market launch periods and regulatory conditions. For the subsequent discussions, it is also necessary (at least) to differentiate between these dimensions; this often does not take place sufficiently within science and very rarely outside science.

The innovation process is accompanied by comparatively early and extensive political, media and internal scientific discussions about "ethical challenges", "ethical prerequisites", "ethical guidelines", "ethical dilemmas", etc. of automated driving. An exact framing of "ethical" is often avoided.

One of the reasons for this may be that the "ethicisation of automated driving" may have been promoted from very different directions. Ethicization of technology in general is a process that has been noticeable since the 1970s, first of all in the field of biomedicine, which in turn is based on a series of more profound social developments. In a nutshell, these include

- a growing cultural and social diversity in modern societies, which goes hand in hand with an extensive pluralization and privatization of moral concepts and a requirement of tolerance in this respect (and which has in turn made collectively binding decisions on difficult moral issues more complicated),
- decision-making and responsibility problems of professional actors such as technology developers, entrepreneurial decision-makers or rule-setting institutions, which are increasingly breaking down as a result of technological change,
- as well as a growing public awareness of the immoral behaviour of scientists and engineers.

In the course of this, ethics has gained considerably in importance in a relatively short period of time, in several roles:

- as a discourse relevant to regulation in disputes about science and technology, not least in the expectation of "clarifying words" in social controversies;
- as a versatile instance for reflection, (legitimation), experimentation and consultation in the form of ethics councils and committees;
- as a group of "governance techniques", including the formalization and bureaucratization of weighing and decision-making processes, the containment of social protest and
- as an emerging sub-discipline of academic philosophy and important actor in interdisciplinary research contexts such as technology assessment (TA), Science and Technology Studies (STS) or Responsible Research and Innovation (RRI), it is subject to the practices and weaknesses of the existing scientific system;

- and as a central rhetorical reference in crises of acceptance of new technologies and in their early recognition (anticipation), prevention and avoidance.

The discussion about "ethical questions" of automated driving is by no means, as is often insinuated, solely about finding and setting rules, which are then translated by engineers and computer scientists into action programmes for technical artefacts (a kind of "ethical specification sheet", if that is at all possible). It is about a broader programme of shaping the transformation of the socio-technical systems "transport" or "mobility", about finding and deciding between social futures. To want to comprehensively investigate this complexity would go beyond the scope of the contribution - as probably also of the entire event. Some of its facets will be discussed on the basis of the following questions and theses:

- Will we (as a society) allow or tolerate people to be harmed by automated vehicles? If we expect complete freedom from damage ("Vision Zero", 1st law of Azimov), these vehicles will never be introduced or dysfunctional in everyday traffic (probably because they are practically inaccessible). Is it justified not to use potential gains in road safety through automation because of this strong demand?
- What and how (on the basis of which heuristics) should automated vehicles "decide" in situations in which all options for action identified by the system will or could lead to personal injury? Must there be a uniform set of rules for this? (Who sets this up? National legislators? International standardisation bodies???) Or would it be conceivable that every user "sets her" and every user "his" vehicle according to their own moral preferences (so to speak as a "moral double")? And thus remains responsible?
- In the event of a loss, how will any tensions between a situational decision and its results on the one hand and a subsequent moral and legal evaluation of these on the other be dealt with? Automation could also get into situations with damage consequences, in which the following analysis shows that other (potentially less momentous) options would have been available, but which automation did not recognize in time or did not use? Would we "excuse" this (as a person and as a society)? Could we learn from this? To what extent are citizens willing to accept limits to the performance of such decision-making systems?
- Should automated vehicles one day be able to move more safely in traffic than those with human drivers (actually, this is a minimum condition of their importability - but how exactly would one measure this ex ante?) - shouldn't people then actually be banned from driving a vehicle?

07/02/2019, Karlsruhe

Bibliographic data:



After studying physics, Torsten Fleischer opted for research and policy advice on processes of technical change and their interaction with social changes. Today he is head of the research area "Innovation Processes and Technology Consequences" at the Institute for Technology Assessment and Systems Analysis (ITAS) at KIT. Within the framework of several research projects located there, he is currently also concerned with the possibility conditions and consequences of automated driving.