

# The conscience of artificial intelligence

**29/01/2020** Professor Christoph Lütge heads one of the world's first research institutes for the ethics of artificial intelligence. At TU Munich, interdisciplinary teams will examine the consequences of software decisions. A central topic will be autonomous driving.

What if I simply stopped doing anything? The question shot through the mind of Professor Christoph Lütge as he drove in a highly automated car on the A9 from Munich to Ingolstadt for the first time in 2016. The test route had been opened a year earlier and vehicles that can accelerate, brake, and steer independently are allowed to use it. When a warning signal sounds, the driver has ten seconds to reassume control of the vehicle. And if the driver does not? What criteria does the on-board computer use to decide how to proceed? How does it prioritize? Lütge couldn't stop thinking about these questions. He had come across a new, cutting-edge field of research.

The 49-year-old professor of business ethics at the Technical University of Munich (TUM) has been researching how competition promotes corporate social and ethical responsibility for the past nine years. Before his test drive on the A9, he had had only a casual acquaintance with artificial intelligence

(AI). Then he read studies, researched, talked to manufacturers. It quickly became clear to him that AI raises a number of ethical questions: who's liable if something goes wrong? How comprehensible are the decisions made by intelligent systems? The transparency of the AI algorithms is also still insufficient: it is still often impossible to understand the criteria on the basis of which they make their decisions—the AI becomes a black box. "We must face up to these challenges, whether AI is used for diagnosing medical findings, fighting crime, or driving cars," says Lütge. "In other words: we need to address the ethical issues surrounding artificial intelligence."

The idea that our life in the future will be determined by machines that know only logic but no ethics is an unsettling one for many people. In a survey conducted by the World Economic Forum (WEF) in 27 countries, 41 percent of a total of 20,000 respondents said they were concerned about the use of AI. 48 percent want stronger regulation of companies, and 40 percent more restrictions on the use of artificial intelligence by governments and authorities.

## Ethics of autonomous driving

Autonomous driving is a particularly topical and difficult field, because it very quickly moves into the realm of human lives being at stake. For example, what should the AI algorithm do if the brakes fail and the fully loaded vehicle can either collide with a concrete barrier or drive into a group of pedestrians? What priorities should the AI set in this case? Should it value the lives of occupants higher than those of passers-by? Should it prioritize avoiding child victims as opposed to older people? "These are typical dilemmas that are explored by social scientists," says Lütge. They also played an important role in the Ethics Commission on Automated Driving set up by the Federal Ministry of Transport and Digital Infrastructure, to which Lütge belonged. "We agreed that there should be no discrimination based on age, gender, or other criteria. That would be inconsistent with the Basic Law." However, programming that minimizes the number of personal injuries is permitted.

The question of liability is also a completely new one. If the vehicle was driven autonomously, the manufacturer would be liable—because then product liability applies. Otherwise the driver is liable. "However, in the future we will need a kind of flight recorder in the car," says Lütge. "It will indicate whether the autonomous driving functions were switched on at the time of the crash. This, of course, gives rise to questions concerning data protection." Despite all the challenges, the scientist is convinced that autonomous vehicles will make traffic safer. "They will be better than humans, because beyond not getting tired or losing focus, their sensors also perceive more of the environment. They can also react more appropriately: autonomous vehicles brake harder and evade obstacles more skilfully. Even in normal road traffic situations, they will ultimately outperform people."

## Lots of AI competence at the Munich location

So there are many exciting questions for Lütge to chew on at his new Institute for Ethics in Artificial Intelligence at the TU Munich. It is being financed with 6.5 million euros by Facebook. There were no

strings attached, he says. But of course Facebook is interested in the scientific results—it is, after all, one of the first research institutes to get started in this field. Munich was an obvious choice due to the TUM's reputation for AI expertise. Moreover, data protection is taken particularly seriously in Germany, and the population is generally comparatively critical of technological developments. At the new research institute, these skeptics will also find a hearing: "We want to bring together all the important players to jointly develop ethical guidelines for specific AI applications. The prerequisite for this is for representatives from the worlds of business, politics, and civil society to engage in dialog with each other," says Lütge.

For the research on ethics in AI, the scientist wants to form interdisciplinary teams to investigate the ethical salience of the new algorithms: "Technicians can program anything," says Lütge. "But when it comes to predicting the consequences of software decisions, you need the input of social scientists." That's why he wants to form interdisciplinary teams, with each tandem consisting of one employee from the technical sciences and one representative from the humanities, law, or social sciences.

In addition, Lütge is planning project teams whose members will come from different faculties or departments. They will examine concrete applications, such as the use of care robots, as well as the ethical questions that arise in this context.

Lütge is already convinced that AI will find its way into many areas of life—because it offers enormous added value, for example in traffic: "In a few years, autonomous vehicles with varying degrees of automation will be part of the traffic landscape," predicts the researcher. According to the rules set forth by the Ethics Committee, the prerequisite is fulfilled when the autonomous vehicles are at least as good as a human driver—for example in terms of assessing the traffic situation and their reactions. Personally, he's looking forward to it: "When I get into an autonomous vehicle, I always feel a certain uncertainty in the first few minutes—until it becomes clear that the car reliably accelerates, brakes, and steers. Then I can hand over the responsibility very quickly. I enjoy the situation, because then I have time to think." For example, about what would happen if autonomous vehicles cross borders: would the same ethical decision algorithms apply there? Or would you need an update at every border? Such questions will undoubtedly keep Lütge busy for a long time to come.

## Guidelines on ethics and AI

## EU High-Level Expert Group: Ethics Guidelines for Trustworthy AI (April 2019)

In setting forth ethics guidelines, a High-Level Expert Group on artificial intelligence aimed to create a framework for achieving trustworthy AI. The guidelines are intended as an aid for the potential implementation of principles in socio-technical systems. The framework addresses concerns and fears of members of the public and aims to serve as a basis for promoting the competitiveness of the EU

across the board.

## Ethical fundamentals/principles in the AI context

**1. Respect for human autonomy**

AI systems should not unjustifiably subordinate, coerce, or deceive humans. They should support humans in the creation of meaningful work.

**2. Prevention of harm**

AI systems should not cause harm. This entails the protection of mental and physical integrity. They must be technically robust to ensure that they are not open to malicious use.

**3. Fairness**

AI systems should promote equality of access to education and goods. Their use should not lead to people being impaired in their freedom of choice.

**4. Explicability**

Processes and decisions must remain transparent and understandable. Long-term trust in AI can only be achieved through open communication of its capabilities and uses.

## Ethics Commission: Automated and Connected Driving (report from 2017)

The interdisciplinary Ethics Commission convened by the German Federal Ministry of Transport and Digital Infrastructure developed guidelines for automated and connected driving.

## Automated and Connected Driving: Excerpt from the rules

**Rule 2**

The licensing of automated systems is not justifiable unless it promises to produce at least a diminution in harm compared with human driving, in other words a positive balance of risks.

**Rule 7**

In hazardous situations, the systems must be programmed to accept damage to animals or property in a

conflict if this means that personal injury can be prevented.

**Rule 9**

In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited.

**Rule 10**

In the case of automated and connected driving systems, the accountability that was previously the sole preserve of the individual shifts from the motorist to the manufacturers and operators of the technological systems and to the bodies responsible for taking infrastructure, policy, and legal decisions.

## About

Prof. Christoph Lütge is an expert in business and ethics. Since 2010 he has held the Peter Löscher Endowed Chair of Business Ethics at the TU Munich. Lütge is also a visiting researcher at Harvard University.

## Info

Text: Monika Weiner
Photos: Simon Koy

Text first published in the Porsche Engineering Magazine, Issue 2/2019

**Link Collection**

Link to this article
https://newsroom.porsche.com/en_AU/2020/technology/porsche-engineering-artificial-intelligence-professor-christoph-luetge-19808.html

External Links
https://www.porscheengineering.com/peg/en/