Autonomous systems such as self-driving cars take over important decisions for humans. But what if it's a matter of life or death? Is it at all possible to establish norms and programme systems to deal with such events?

Moral Dilemma

Autonomous systems capable of independent decision making are about to take over our world. But on what basis? What underlying ethical rules are they following? It is this human-machine interface that drives the research of philosopher Mark Siebel and neuropsychologist Jochem Rieger

We live in a world in which decision-making is being increasingly surrendered to technical systems. When you park your car nowadays, your assistant system guides you into the space. When you embark on a car journey, you allow yourself be led by your navigation system, trusting that it will find the right route. In the operating room robots assist surgeons, achieving a degree of accuracy to a tenth of a millimetre. Technological development – ridiculed for decades as science fiction – has long been firmly anchored in the present.

New questions impact society as a whole

For some the development of these cyber-physical systems represents a huge opportunity while others fear that automation will soon take over entire professional fields. Who needs lorry drivers when the lorries can drive themselves? Or surgeons when robots can operate much more precisely?

Whether we like it or not, the advance of technology gives rise to questions that impact society as a whole. If machines are increasingly able to make decisions for us and are learning how to react to unforeseen events, does this not pose the threat of society losing control? Is it possible to ever truly understand why a machine has made

a decision? And to what extent do we want to surrender decision-making to technical systems? Researchers from different disciplines at the University of Oldenburg are looking for answers to these questions. "Because one thing is clear: if research and politics do not address the issue, industry will ultimately dictate the development," Prof. Dr. Jochem Rieger firmly believes. And if this happens, the cognitive neuropsychologist who researches human-machine cooperation doubts that society will be able to reconstruct and understand how and why technical systems take certain decisions. His research focus is the "perception, information processing-cognition-action" cycle. "It is through this cycle that humans interact with their environment. We want to find out what exactly goes on in the brain here. Moreover, in this process decisions are a prerequisite for interaction with the world." His team measures brain activity to try to predict mental states: whether a person behind the wheel is exhausted, stressed or cognitively overloaded. "If we are able to measure this status and make predictions, we can provide the cyber-physical system with additional information about the human and whether he or she will be ‚amenable' to reaching a joint decision," Rieger explains. It might also be possible to determine which of the five senses the machine should best engage in order to communicate effectively with its partner at a given moment: "If the auditory channel is occupied because the person is in a conversation, the system would try to make contact via the visual channel," the scientist says. For human-cyber-physical-system cooperation to function, researchers must find a way to integrate humans and machines in goal tracking and actions in a way that they will compensate for each other's weaknesses and combine their strengths. "Potentially people will then be able to solve problems that are too complex for them at the moment," Rieger says. Someone creating a production plan, for example, gains a better overview of the production machines through interaction with a technical system, thus improving the working and production processes and saving resources. "This is an abstract example that illustrates how human-machine systems can expand the capacities and objectives within our reach as humans. "Society should stop being so pessimistic and seize the opportunity to make the most of human-machine cooperation," Rieger believes.

This sounds plausible but would require broad social acceptance, more precise system specifications and that ethical standards are taken also into account. "To put it simply, someone has to tell the system in advance which decisions are good or bad in different situations," Prof. Dr. Mark Siebel from the Institute of Philosophy explains. This is the area where Rieger and Siebel's interests intersect and where they can profit from each other's knowledge. "We neuroscientists are mostly interested in the technical side. We take a system, put it into action and ask the humans who interact with the system which actions they like and which ones they don't. But we won't be able to extract any ethical guidelines from the results. That's where the philosophers come in, because they are experts in ethical reflection," Rieger explains. "We, in turn, profit from the data provided by the neuroscientists," Siebel adds.

Jochem Rieger: "Society should stop being so pessimistic and seize the opportunities."

Will algorithms decide over life and death?

Siebel is mostly interested in the ethical standpoint here: what constitutes morally good and bad decisions? "Autonomous systems and how they learn to make good decisions are a perfect case in point," Siebel says. However, whether something is morally good or bad in a specific situation is often a matter for debate. "It's something even philosophers argue over," Siebel points out. In this context the "trolley problem" is frequently cited, a classic thought experiment revolving around the question of whether or not to pull a lever to divert a tram from one track where it would kill a group of people to another track where it would kill only one person. This situation raises an old philosophical question involving a moral dilemma. The person who has to make the decision is placed in an inextricable double bind. Is it acceptable to divert the tram onto another track where it would kill one person rather than killing several? This is the sort of dilemma an autonomous vehicle could find itself facing. How is an algorithm supposed to decide whether the car it is steering should kill an 80-year-old in order to save the lives of five children?

If you follow the reasoning of utilitarianism, one of the three main schools in ethics, the "net saving" of five lives would justify the ending of one. "This is difficult because it involves weighing up one life against another. On the other hand, what alternatives do you have? The decision can hardly be left to a random number generator," Siebel says. No wonder then that the "Ethics Commission on Automated and Connected Driving" set up by the German Transport Minister Alexander Dobrindt came to the conclusion that the "computer colleague" behind the wheel cannot solve difficult moral dilemmas. "Genuine dilemmatic decisions, such as a decision between one human life and another, depend on the actual specific situation, incorporating ‚unpredictable' behaviour by parties affected. They therefore cannot be clearly standardized, nor can they be programmed to be ethically incontestable." The explanation shows how automation forces people to confront ethical grey areas that have been left to philosophers and chance in the past. After all, is any driver in an extreme situation really able to reach a rational and ethical decision within a millisecond? Humans remain unpredictable here.

Mark Siebel: "Someone has to tell the system in advance which decisions are good or bad in different situations."

Culturally adaptive systems needed

Another factor that is relevant for decision-making in future autonomous systems is the cultural background of their field of application. "Cultural adaptation is something we should not forget," Siebel says. He explains that although in Germany priority is given to a child's wellbeing, other cultures are more protective of old people, cows are holy, or goats are valued over dogs – all details that must be taken into account when programming. Rieger goes one step further: "Autonomous driving systems like the ones we are discussing here in Germany would simply not function in a city like New Delhi!" Very few norms govern traffic there, and the few that exist differ considerably from our own due to cultural differences. Any autonomous vehicle equipped with "western European algorithms" would have huge problems there. So what kind of algorithm would be capable of making the right decisions there?

Decision-making processes must be transparent

Both academics agree that autonomous systems must function in a way that ensures that their decision-making processes are transparent. "We know from legal practice that the motivation for an action always plays a role when evaluating the consequences of that action," Rieger explains. If someone dies, it makes a difference whether the person was killed intentionally, whether it was planned or committed with malice aforethought, or whether it happened due to unfortunate circumstances. "The underlying reason plays a major role for the evaluation, so it must also play a role in evaluating the consequences of the actions of autonomous systems," Rieger states. But he gives one more reason why the decision-making process must be factored in: In test situations a self-learning machine delivers results conforming with the decisions that people have made. "But when you look inside the system you may find that the reasons it gives for its actions are completely unexpected, for instance the motives may be racist. But this is not apparent in the result." Will it be possible to trace an algorithm's motives in real situations? Rieger and Siebel are agreed: "This is a question for our colleagues from machine learning." At any rate the example shows that the aim is not to achieve the optimal decision, but one that is well founded.

For Rieger and Siebel alike, autonomous systems offer a multitude of opportunities for overcoming societal and economic challenges. They want to accompany and give direction to this process with their research. "We are a long way off being able to provide answers, first of all we have to ask the right questions." (kl)