Artificial intelligence – Future scenarios

Assignment

Artificial intelligence offers our society many opportunities, but it also presents us with many challenges. Three main trade-offs can be distinguished:

1. Responsibility versus trust
2. Security versus progress
3. Freedom versus dependence

Read through the following future scenario. Consider how and why the three trade-offs are relevant to this scenario by applying the key questions from the “Artificial intelligence – Social trade-offs” graphic to the scenario. The questions are listed below. Also think about the further consequences of your answers.
You should not make a personal judgment about the future scenario, but rather contemplate the ethical challenges of using AI in this area.

Future scenario

A criminal case is being tried in court: A married couple broke into a house and stole items worth 10,000.00 euros. The homeowner was slightly injured. The controversial aspect of this case is that for the first time, the judge is not a person, but an artificial intelligence system that will render judgment.
A developer team trained the AI system before its use. For this training, numerous files on criminal acts and the judgments from the past ten years were digitalized and entered into a central database. The AI system draws on this database to recognize patterns and develop algorithms that it will apply in future use. Thanks to the training, AI is capable, for example, of computing how likely it is that an offender will relapse. Taking into account these computations and other considerations, it determines what judgment is appropriate.

Solution

Fundamental considerations are provided below as an example for each trade-off with its key questions.

Responsibility versus trust

- Who bears the responsibility when a robot makes “wrong” or critical decisions?

The question arises as to who bears responsibility when AI makes an automated decision and a mistake occurs and whether mistakes will even be apparent. Consequently, a major challenge would be making the AI decision transparent. Only then could the judgment be contested in most cases because reasonable doubt in the decision would have to be established and ultimately proven.

- Are people abandoning responsibility?

This question is closely related to the previous one. If the AI judgment is regarded as unappealable and/or the AI decision-making process is not transparent, people transfer responsibility completely to a machine. In the case at hand, the question of whether a representative who could be made responsible for the AI decision should be named merits discussion. However, this then leads to the next question of who that person should be.

- In what areas does AI act more reliably than people?

Areas exist in which AI is more reliable than people can be, perhaps because AI can incorporate more data (for example, similar cases) than a human brain. Furthermore, AI consistently makes decisions according to the same patterns. Human emotions that distract from the pure facts would not play a role in the judgment rendered by an AI judge. According to studies, when rendering a decision, human judges are affected by hunger pangs, for example. We would therefore have to ensure that the AI developers have trained the AI system with extensive and diverse data and that the objectives match the actual fulfillment of those objectives.

- When and under what conditions do we start trusting in decisions made by machines?

Deep trust that AI works reliably and that the decisions made by AI systems are based on acceptable ethical views of society and applicable law is also required. To that end, we must be sure that the training data are diverse and flawless. Human presuppositions and prejudices that enter into the selected data and defined objectives must also be thought through as diligently as possible. Otherwise, systematic discrimination against people may occur that is not transparent or that leads to flawed results. This goal of impartial AI appears to be infeasible given the current state of knowledge. In the case at hand, for example, discriminating patterns could be embedded in the AI training data because the files on criminal acts and their judgments from the past ten years are assuredly not flawless or unprejudiced.

Security versus progress

- How secure are AI systems against hacker attacks?

Artificial intelligence intensifies the security risks associated with digitalization. It increases the masses of data tremendously. In our scenario, for example, large volumes of data would arise due to digitalization of the criminal acts and their judgments from the past ten years. In addition, AI requires comprehensive data on the cases in order to make a decision. Some of the data could be very personal or sensitive. In the example, it must also be kept in mind that AI would have to integrate data from witnesses. Moreover, a sophisticated security system would be needed to prevent third parties from hacking into the comprehensive data.

- How can AI systems be manipulated?

Manipulation is possible, for instance, if hackers modified the objectives and as a result AI could render judgment on a basis other than the basis actually defined. Publication of sensitive data without the consent of the affected parties, perhaps with the intent to harm the involved parties, is also possible.

- Can we foresee all consequences of this technology and prevent risks?

Precisely because AI would be rendering judgment for the first time, the consequences and risks would not be foreseeable. A particular risk would be that discriminating patterns would continue and even become systematized. Because the AI system would have been trained with digitalized files on criminal acts and their judgments from the past ten years, AI would continue the patterns learned from the training data – including any discrimination. The further social consequences also cannot be foreseen (see freedom versus dependence). A possible control mechanism that could minimize risks is not mentioned in the scenario.

- What opportunities does AI help find that never would have been possible without it?

Compared with humans, AI can incorporate much more data when making decisions and solving problems and it processes the data faster. This accelerates decisions, which would lead to significantly faster judgments in the given scenario. People also cannot analyze these vast amounts of data as systematically: AI can discover correlations in data that humans cannot identify. In the given scenario, these correlations could also be useful for preventing or solving crimes. AI could thus contribute to the prevention and solving of criminal acts.

Freedom versus dependence

- To what extent does the technology provide new freedoms?

If AI were utilized as the judge, this would save an enormous amount of time and resources for people, for the previously stated reasons. People could concentrate on other tasks that AI cannot perform. The technology could be transferred to other areas, such as to criminal defense. Thanks to its efficiency, AI would also be less expensive in this area. This means that more people could afford a defense. In addition, the victims of a crime would possibly not have to endure as many tedious and incriminating questions or spend as much time in the courtroom with the offenders.

- How does coexistence in society change as a result of AI?

Courts determine people's fate. Currently, this job is performed by people who distinguish themselves through a particular course of study and many years of experience. Professional judges are considered to be neutral and trustworthy. At the same time, we are aware that these judges have feelings; they do not only render judgments in a cool, calculating manner, but they are also empathetic, toward both the victims and the offenders. If an AI system were to take over the job of rendering judgments, this structure would change fundamentally, as would the overall trial: Trials involve active processing and negotiation, and the offender's guilt is not yet proven at the outset. The crime is analyzed through examination and questioning. By contrast, an AI system would collect data and very quickly come to a decision. People would no longer determine other people's fate, but rather judgments would be quickly computed by a machine; thus, the machine would determine the defendant's fate. As a result, the persons involved in the trial might feel that a judgment from an AI system is less fair, partly because AI cannot see itself in the victim's situation. As a result, it is possible that the very legitimacy and trustworthiness of the administration of justice as a whole would decrease.

- Are we surrendering our self-determination to machines?

As already described, the use of AI would change the nature of court rulings. The court's decision would no longer be the result of negotiation but rather the result of a computation. People's behavior in the trial (the judge's considerations; a defendant's confession and cooperative behavior) would have less influence on the judgment rendered. Rather, the judgment would appear predetermined by the data.

On the other hand, people develop and train AI. AI could therefore be viewed as a tool of those who build and use it. In this sense, it can be argued that self-determination would not be reduced by an AI judge, but only changed.

- To what extent are people becoming more dependent on technology due to AI?

Processes that are taken over by AI depend on smooth functioning. The more AI systems are used, the greater the dependency and also the greater the consequences if AI makes a mistake or is hacked. With an AI judge that would be responsible for all trials, the resulting dependency would be very high. In the example, this would have tremendous consequences since it deals with people's lives. In addition, it would generate great dependency on decisions made by a machine. If the machine were to make a mistake or fail, the consequences would be huge.