Light in the Blackbox AI

Explainable AI looks into the "brain" of artificial intelligence and can explain how logarithms make their decisions. An important step, because the new General Data Protection Regulation requires traceability.

COPY —— Ingrid Kirsch

The Explainable AI research field has a new engine: the European Data Protection Regulation. Because AI systems require transparency. A demand that is still not easy to fulfill today. Why? To answer this, take a look back at Google's I/O Developer Conference in early May this year. The highlight: Google Duplex – an AI that independently arranges a hairdressing appointment on the phone. Spontaneous pauses in speaking, some interspersed "hmm's" – and already the computer voice could not be distinguished from that of a human being. The reaction? Cheers to the Google experts in the audience. Otherwise? The reaction was rather mixed. The reason: Google Duplex just sounds too real.

Because: Is it really okay if a software calls me and I think it's a human? "Clearly no," says the European General Data Protection Regulation (DSGVO), which forces companies to be transparent in terms of artificial intelligence. As soon as automated decisions affect people, they must be understandable and explainable. Companies are obligated to disclose the AI aspects of their services, products, analyses, and processes.

AI DECISIONS MUST BE TRACEABLE

However, the demand for transparency is usually more difficult to meet. What exactly happens during machine learning is often hidden in a black box. Even the programmers are in the dark when it comes to answering the question of how the AI makes its decisions. Which is why, for example, Microsoft Research's Kate Crawford calls for key public institutions in the areas of criminal justice, health, welfare, and education to stop using algorithms. Too many AI programs, according to the expert, have discriminatory tendencies or erroneous assumptions, it was discovered. Machines decide with high consistency, but also consistently inappropriately with unsuitable programming.

AI is relevant in more and more areas of life. Its importance will continue to grow. It can do many things: make medical diagnoses, buy or sell stocks for us, check our credit history, analyze whole business reports, or select job applicants. Software evaluates us according to certain mathematical criteria using so-called "scoring" methods. Therefore, the GDPR prescribes the "right of explanation" for the protection of every single person. This means: If an affected person submits an application, institutions or companies must be able to reasonably explain an AI decision or risk assessment.

MACHINE LEARNING REVEALS CASES OF FRAUD

It becomes difficult at this point. "The legality of decisions can only be examined by those who know and understand the underlying data, sequence of action, and weighting of the decision criteria," writes legal scientist Mario Martini in JuristenZeitung (JZ). Scientists around the world are working on this explanation. Their research field: explainable artificial intelligence. Or sexier: XAI. Explainable artificial intelligence or explainable machine learning want to look into the electronic brain. For example, the consulting firm PricewaterhouseCoopers (PwC) places XAI on the list of the ten most important technology trends in the field of artificial intelligence.

However, the literally enlightening view into the black box is difficult because neural networks have a very complex structure. Decisions are the result of the interaction of thousands of artificial neurons. These are arranged in tens to hundreds of interconnected levels – with their diverse interconnections, the neural networks of the human brain are modeled. Scientists are now also using the virtual dissecting knife in Berlin: The research group Machine Learning at the Fraunhofer Heinrich Hertz Institute (HHI) has developed a method called Layer-wise Relevance Propagation (LRP). Research Director Wojciech Samek and his team first published their explainable AI method in 2015 and already presented their XAI method at CeBIT.

LRP traces back the decision process of a neural network: The researchers record which groups of artificial neurons are activated and where – and what decisions they make. They then determine how much an individual decision has influenced the result.

EXPLAINABLE ARTIFICIAL INTELLIGENCE: THE PATH TO THE SOLUTION MATTERS

This type of transparent path, a kind of documentation, plays into the hands of the GDPR, because, as in the past, the solution and not just the results count in math lessons. Developing machine learning techniques that produce more predictable models should strengthen confidence in AI technology in the long term. PwC understands that many companies make use of explainable AI before embarking on algorithmic applications on a broader basis. The GDPR could even make explainable AI mandatory for state authorities.

And until then? Companies like Telekom are reviewing AI decisions through a review process. Employees constantly check whether the AI has decided on behest of the company and the person affected. If this is not the case, they can take corrective action at any time. “We should provide the algorithms with a sort of AI governance and prevent artificial intelligence from breaking out of ethical and moral guidelines,” recommends Claus-Dieter Ulmer, Group Representative for Data Protection at Deutsche Telekom AG. Under this condition, there is a lot of potential in AI. In their strategy paper “Künstliche Intelligenz als Innovationsbeschleuniger im Unternehmen” (“Artificial intelligence as an innovation accelerator in the company”), the PwC experts are sure that AI will develop into a decisive competitive advantage in the future, which decide the success and failure of every company.

“We should provide the algorithms with a sort of AI governance”

DR. CLAUS-DIETER ULMER,
Group Representative for Data Protection at Deutsche Telekom AG

bestpractice@t-systems.com
www.t-systems.com/gdpr
www.t-systems.com/perspective/iot-security

Photos: iStockphoto (2)