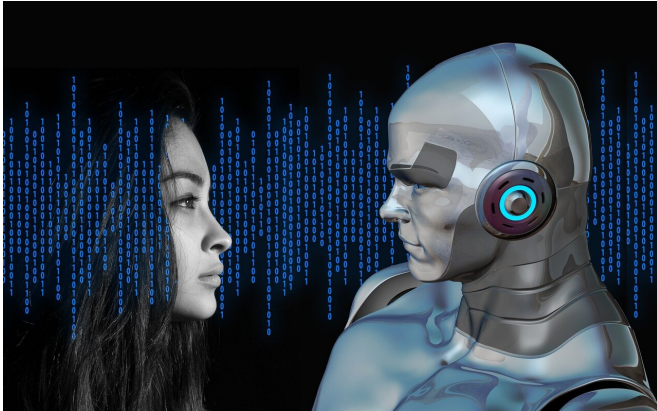


# Building trust in artificial intelligence

2 July 2019



Credit: CC0 Public Domain

From telecommunications to road traffic, from healthcare to the workplace—digital technology is now an intrinsic part of almost every area of life. Yet how can we ensure that developments in this field, especially those that rely on artificial intelligence (AI), meet all our ethical, legal and technological concerns? In a project led by the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS, and with the participation of Germany's Federal Office for Information Security (BSI), an interdisciplinary team of scientists from the Universities of Bonn and Cologne are drawing up an inspection catalog for the certification of AI applications. They have now published a white paper presenting the philosophical, ethical, legal and technological issues involved.

Artificial intelligence is changing our society, our economy and our everyday lives in fundamental ways. And in doing so, it is creating some exciting opportunities in how we live and work together. For example, it already helps doctors to better evaluate x-rays, which often leads to a more accurate diagnosis. It is the basis of chatbots that provide helpful answers to people looking for advice on, for example, insurance. And, before too long, it will be enabling cars to become more and more

autonomous. Current forecasts indicate that the number of AI applications is set to increase exponentially over the coming years. McKinsey, for example, projects additional global growth from AI of up to 13 billion U.S. dollars by 2030.

At the same time, it is clear that we need to ensure that our use of AI and the opportunities it brings remains in harmony with the views and values of our society. Acting under the aegis of Kompetenzplattform KI.NRW, an AI competence platform in the state of the North Rhine-Westphalia, an interdisciplinary team has come together to develop a [certification process](#) for AI applications to be carried out by accredited examiners. This will confirm compliance with a certified quality standard that, in turn, will enable technology companies to verifiably design AI applications that are technically reliable and ethically acceptable.

"The purpose of the certification is to help establish quality standards for AI made in Europe, to ensure a responsible approach to this technology and to promote fair competition between the various players," says Prof. Dr. Stefan Wrobel, director of Fraunhofer IAIS and professor of computer science at the University of Bonn.

## Focusing on the human aspect

Artificial intelligence has the potential to enlarge our capabilities and provide us with new knowledge. However, once we begin to base our decisions on machine learning that is either fully or partially automated, we will face a host of new challenges. The technical feasibility of such applications is a one consideration. First and foremost, however, we must resolve the basic philosophical, ethical and legal issues. To ensure that the needs of people are firmly embedded at the center of development of this technology, close dialog between the fields of information technology, philosophy and law is necessary.

The team of experts has now published a [white paper](#) in which they detail their interdisciplinary

approach to the certification process. For example, the use of AI—and that their rights and freedoms are they explain the ethical principles involved. "Anyone not curtailed in any way.

using AI should be able to act properly in accordance with their moral convictions, and nobody should be curtailed in their rights, freedom or autonomy," says Prof. Dr. Markus Gabriel, professor of philosophy at the University of Bonn. Legal questions have also been addressed. "For example, we need to determine how AI applications can be made to conform to the basic values and principles of a state governed by the rule of law and subject to the principles of freedom," explains Prof. Frauke Rostalski, professor of law at the University of Cologne.

### **Priorities for building trust in the use of AI**

This interdisciplinary approach has identified a number of ethical, legal and technological issues of relevance to the use of AI. These are all examined in the white paper. The criteria employed in a certification process should include fairness, transparency, autonomy, control, data protection, safety, security and reliability. Recommendations of the EU thus also serve as orientation for the KI.NRW certification project.

The certification process will revolve around questions such as: Does the AI application respect the laws and values of society? Does the user retain full and effective autonomy over the application? Does the application treat all participants in a fair manner? Does the application function and make decisions in a way that are transparent and comprehensible? Is the application reliable and robust? Is it secure against attacks, accidents and errors? Does the application protect the private realm and other sensitive information?

### **Building in checks and controls at the design stage**

According to the white paper, it must be determined during the initial design process whether the application is ethically and legally permissible—and if so, which checks and controls must be formulated to govern this process. One necessary criterion is to ensure that use of the application does not compromise anyone using it in their ability to make a moral decision—as if the option existed to decline

Transparency is another important criterion: the experts emphasize that information on correct use of the application should be readily available, and the results determined through the use of AI in the application must be fully interpretable, traceable and reproducible by the user. Conflicting interests, such as transparency and the nondisclosure of trade secrets, must be balanced against one another.

The plan is to publish an initial version of the inspection catalog by the beginning of 2020 and then begin with the certification of AI applications. This project also involves the participation of Germany's Federal Office for Information Security (BSI), which has extensive experience in the development of secure IT standards. This know-how will feed into the certification process. Finally, given that AI is constantly evolving, the inspection catalog itself will be a "living document" in need of continual updating.

**More information:** White paper: [www.iais.fraunhofer.de/ki-zertifizierung](http://www.iais.fraunhofer.de/ki-zertifizierung)

Competence platform KI.NRW: [www.ki.nrw](http://www.ki.nrw)

AI certification project: [www.iais.fraunhofer.de/ki-zertifizierung](http://www.iais.fraunhofer.de/ki-zertifizierung)

Provided by University of Bonn

APA citation: Building trust in artificial intelligence (2019, July 2) retrieved 17 November 2019 from <https://phys.org/news/2019-07-artificial-intelligence.html>

*This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.*