Lernende Systeme

GERMANY'S PLATFORM FOR ARTIFICIAL INTELLIGENCE

Certification of AI systems

White Paper by Jessica Heesen,
Jörn Müller-Quade, Stefan Wrobel et al.
Working Group IT Security, Privacy,
Legal and Ethical Framework
Working Group Technological Enablers
and Data Science

Executive Summary

The certification of AI systems can help to strengthen trust in technology and application. And thus increase AI application. The goal of a successful certification process should basically be to guarantee standards and at the same time avoid overregulation and enable innovation.

Led by two working groups IT Security, Privacy, Legal and Ethical Framework as well as Technological Enablers and Data Science, members of all working groups of the Plattform Lernende Systeme, together with other guest authors, have defined criteria that can be used as orientation for the certification of AI systems on the one hand and can support the decision on the necessity of certification on the other.

The experts also provide an overview of efficient infrastructural and organisational requirements for certification. This takes up, at the same time, the current state of the discussion and connects to the already published **discussion paper Certification of AI systems** (see Heesen et al. 2020a).

Existing certification initiatives and procedures for AI systems

There are already numerous national and international starting points for successful certification initiatives for AI systems. These include political initiatives such as the European Commission's White Paper on Artificial Intelligence, the German government's statement on the European Commission's White Paper, good practice examples from AI research and application, and other initiatives on technical solutions, standardisation and the testing and auditing of AI systems. These can form a first starting point for a successful certification of AI systems, for which there are hardly any valid and recognised norms and standards in Germany to date.

In which cases is certification of AI systems necessary?

Certification will not be necessary for every AI system against the backdrop of its respective application context. For a successful certification of AI systems, a distinction must be made between use cases in which certification is necessary and those in which there is no need for an independent third-party verification of conformity. The need for certification of AI systems can be derived from the AI system's criticality in a specific context of its application. First, this depends on the risk to human life and other legal goods and second on the range of possible human interventions in a certain context of application. The extent of criticality indicates possible need for regulation. Who is responsible for determining criticality depends on the context of use: If the government wants to introduce mandatory approval or certification for certain contexts of use, it is responsible for the criticality assessment as the regulating authority. In cases where certification is initiated on a voluntary basis, the criticality assessment can be carried out by the companies themselves.

Which objects and criteria should be used as a basis for the certification of AI systems?

Once the need for certification has been identified, the question arises as to how and according to which test criteria AI systems can be certified. In principle, the certification of AI systems should be based on or linked to general and industry-specific norms, standards, test procedures and legal regulations (ISO and DIN standards, applicable (european) law). Where necessary, gaps may need to be closed or adjustments to be made. A situation should not arise in which AI-specific standards and regulations compete with broader certifications and regulations.

Subject of certification

Different types of certification can be distinguished that are useful for the certification of AI systems. For the area of AI systems, the experts recommend either a **product certification** or a mixed form of **product and process certification**. Product and process certification differ in terms of objective and object of consideration, which is why some test criteria can be better queried or implemented as part of product certification and others within the framework of process certification.

Product certification is a neutral verification of compliance with guaranteed product properties at the physical product level. Product certification should start at an early stage (optimally already during the specification of the product). A characteristic of product certification is that often several procedures must be combined to verify the criteria. **Process certification** can be an alternative or supplement to product certification. It examines the quality of the manufacturing, development as well as the implementation process in general and the implementation of the AI solution in particular. It serves to reflect on the processes to be tested and can, under certain circumstances, also be carried out by the manufacturer or the operator itself. If a certified process is applied with tools specifically tailored to AI, process certification can provide important implications for questions of responsibility and liability. At the same time, well-executed processes can also prevent possible malfunctions and thus lead to better products.

Test criteria for certification

The test criteria which should be basically applied can be divided in terms of their binding nature within the framework of certification into **minimum criteria**, which must always be fulfilled and tested in the respective application context, and **additional criteria** that go beyond this, which can be tested and thus enable a type of „certification plus". These criteria are of great importance for a positive and value-oriented development of trustworthy AI and go beyond the minimum requirements, which „primarily" serve to prevent evident and immediate hazards.

Minimum criteria that must be verified as part of certification:

Minimum criteria

- Transparency, traceability, verifiability and accountability
- Functional security/safety/incl. product safety and reliability
- Avoidance of non-intended consequential effects (on other systems, people, and the environment)
- Equity in the sense of equality and non-discrimination
- Protection of privacy and personality
- Self-determination, including transparency about the use of the AI system and the role of the human being in the decision-making process

Additional criteria beyond this, which can be checked as part of a certification:

Additional criteria

- Open interfaces and system operability
- Human-centeredness and user-friendliness (usability) incl. participation, protection of the individual, sensible division of work, and conducive working conditions
- Sustainability
- Marking and limiting the systems functionality

Prerequisites for successful certification

AI systems are particularly dynamic, especially continuously self-learning (and self-advancing) systems. This touches on the question of when, how often and to what extent certification should be conducted. Certification should be carried out before the product or service is launched, for systems that continue to learn, certification should be repeated regularly. The level of detail and depth of testing for certification should also be based on an AI system's level of criticality in its application area – the higher the criticality is assessed in the context of application, the more extensive the level of detail and depth of testing for certification should be.

In addition to the assessment of criticality and the creation of a test catalogue, successful certification of AI systems also requires an effective organisational and technical infrastructure. For the conformity assessment of AI systems to succeed, technical requirements must be met regarding tools, software, and environments for testing. Organisational structures and processes in companies should also become an important, complementary component of the certifi-

cation of AI systems in the future. To react adequately to AI innovations the cooperation between certification bodies and research institutes is particularly important in order to account for a dynamic nature of the testing bodies.

Possible recommendations for action

In line with these considerations, concrete measures for establishing successful certification of AI systems can be derived that address different groups of actors:

Research could...

- investigate the details of certification procedures in more detail in interdisciplinary research networks to help develop testing tools for evaluating AI systems and to make general criteria such as „transparency" operational for business, users and technology development. On this basis, research can advise policy-makers, companies and civil society even more thoroughly on the opportunities, risks and consequences of the individual technologies and areas of application.
- develop interdisciplinary technological solutions and methods to ensure that AI systems are trustworthy.
- develop trustworthy AI methods together with companies (explainable AI (XAI)).
- make their state-of-the-art infrastructures available so that these can form starting points for initial certification projects.
- explore where traditional signal processing methods end and AI begins to enable more accurate law enforcement.
- participate in the development of a concept for training AI test engineers.

Companies could...

- support the formation of trust in AI systems by **voluntarily** elaborating and disclosing ethical and technical standards and by devoting more attention to the use of explainable AI. This provides a basis for the debate on certification and regulation of AI systems.
- participate in the creation of appropriate standards and identify corresponding needs.
- exchange industry-specific information on which aspects of AI systems are to be regarded as critical in their application context and how best practices could be established by manufacturers for such cases. Existing concepts and design guidelines can serve as a point of orientation for such an exchange. Furthermore, this exchange could provide a basis for companies to invest in trustworthy AI and develop corresponding business models.
- make their state-of-the-art infrastructures available so that these can form starting points for initial certification projects.
- offer employees special training courses as part of their in-house continuing education, the aim of which is to teach them how to deal with AI systems in a confident manner.

Civil society could...

- identify areas for which regulation is necessary from the perspective of consumers and citizens. In the same way, areas could be identified for which no regulation is necessary, and areas which could lend themselves to a conformity check by civil society organisations, for example via a quality seal.
- based on existing and future criteria and design guidelines as well as the legal framework, take on the role of a „watchdog" and thus press for compliance with the criteria, guidelines and rules in order to help shape the use of AI.

In addition, some aspects require a **social discourse** involving all relevant stakeholders from business, science and civil society. These concerns, among other things:

- the definition of criticality levels. This requires a discussion on acceptable risks and on the fair distribution of the benefits arising from AI applications. Furthermore, education about the function and mode of action and possible applications of AI is necessary to arrive at a realistic assessment of its potential.
- the need for certification of AI systems: This applies above all the question of the extent to which further norms and standards are needed beyond the already existing security and transparency standards of technical (industrial) systems.
- the way we want to live, learn and work with AI in the future: The goal is to develop AI systems that are implemented in a way that enhances rather than curtails human competencies. Such a broad discourse of AI represents the basis on which evaluation and testing criteria for AI systems could be discussed in the future.

Imprint
Editor: Lernende Systeme – Germany's Platform for Artificial Intelligence | Managing Office | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de |
Follow us on Twitter: @LernendeSysteme | Status: November 2020 | Image credit: Tierney/Adobe Stock/Title

This executive summary is based on the white paper *Certification of AI systems – Compass for the development and application of trusted AI systems*, Munich, 2020. The authors are members of the working group IT Security, Privacy, Legal and Ethical Framework and the working group Technological Enablers and Data Science of Plattform Lernende Systeme. The original version of this publication is available at: https://www.plattform-lernende-systeme.de/publikationen.html

SPONSORED BY THE

Federal Ministry of Education and Research

acatech
NATIONAL ACADEMY OF SCIENCE AND ENGINEERING