

# PRESS RELEASE

PRESS RELEASE

31 May 2021 || Page 1 | 3

**White Paper: Dependable AI**

## Making AI ready for safety-critical applications

**Production planning, logistics, maintenance, quality control – there is a wide range of application areas for the use of Artificial Intelligence (AI) in industrial manufacturing. However, in practice, AI models have only been infrequently used up to now. The reason: It is incredibly challenging to empirically validate the reliability of such systems. New certification criteria could now ensure AI is ready to be deployed for safety-critical applications.**

The expectations could hardly be higher: Artificial Intelligence (AI) should make production more flexible, plan maintenance with foresight and optimize the flow of goods, while also automating logistics and quality control processes. “In fact, numerous promising AI algorithms and architectures have been developed over recent years – including at Fraunhofer IPA – such as computer vision, human-machine interfaces and networked robotics, for example”, explains Xinyang Wu from the Center for Cyber Cognitive Intelligence at Fraunhofer IPA. Practical application is all that is now missing. “There is a chasm between research and application. Industry has proven to be quite sluggish in implementing new AI applications. They are regarded as not reliable enough for safety-critical applications.”

Wu is very aware of user reservations: “When we speak with our industrial partners, it quickly becomes clear that the companies only really want to use autonomous and self-learning robots, for example, if they function absolutely reliably and when you can say without any doubt that the machines pose zero risk to humans.”

But it is precisely this which has been impossible to validate so far. There are neither norms nor standardized tests. Wu underlines that these are urgently needed: “The target has to be to make decisions taken by algorithms certifiable and transparent. For example, traceability must be guaranteed: When a machine independently makes decisions, then I have to be able – in retrospect at least – to work out why it has made an error in a certain situation. Only in this way, we can make sure that such a mistake is not repeated. Black-box models, which do not allow humans to trace algorithm-based decision paths, are from our perspective not directly suited to being used in safety-critical applications – unless the model has been certified by the correct method.”

But how could humans ensure the safety of Artificial Intelligence? The Fraunhofer IPA team for Cyber Cognitive Intelligence has now proposed a strategy aimed at resolving this issue and reported on the current status of the relevant technology in its white paper “Dependable AI – Using AI in safety-critical industrial applications”. The strategy is based on certification and transparency.

---

**Press kommunikation****Hannes Weik** | Phone +49 711 970-1664 | [presse@ipa.fraunhofer.de](mailto:presse@ipa.fraunhofer.de)Fraunhofer Institute for Manufacturing Engineering and Automation IPA | Nobelstrasse 12 | 70569 Stuttgart | [www.ipa.fraunhofer.de](http://www.ipa.fraunhofer.de)

**Criteria catalogue to improve safety**

“Generally speaking, the focus is first and foremost on finding rules that help us to evaluate the reliability of machine learning and the AI processes associated with this”, according to Wu. This research resulted in the establishment of five criteria that AI systems should meet in order to be regarded as safe:

- All algorithm-based decisions must be explainable for humans.
- The functionality of the algorithms must be reviewed using formal verification methods prior to being used.
- Moreover, statistical validation is required, particularly in cases where formal verification is not suited to certain application scenarios due to scalability issues. This can be checked by test runs with larger amounts of data or unit volumes.
- The uncertainties on which the decisions of neural networks are based must also be determined and quantified.
- During operation, the systems must be monitored, for example by using online monitoring processes. The important thing here is recording input and output – i.e. sensor data and the decisions made on the back of evaluating this.

Wu points out that the five criteria could form the basis for standardized checks in the future: “At IPA, we have already compiled various algorithms and methods for each of these points, which will allow us to empirically review the reliability of AI systems. We have even carried out checks of this kind for some of our customers already.”

**Transparency creates trust**

The second basic prerequisite for safe use of AI systems is that they are transparent. In line with the ethical guidelines of the High-Level Expert Group on Artificial Intelligence of the European Commission, or HLEG AI for short, this is one of the key elements for the realization of trustworthy AI. In contrast to the criteria that can be used to check reliability at algorithmic level, this transparency relates exclusively to human interaction at systematic level. Based on the HLEG AI guidelines, there are three points that transparent AI must fulfil: First, the decisions made by the algorithms must be traceable. Second, it must be possible for a person to explain these decisions at a full level of human understanding. And third, AI systems must communicate with a human to let them know what the algorithm is capable of, including tasks that are beyond its capabilities.

“Users will only trust AI – no matter if it’s being used in road traffic settings or manufacturing factories – when it is possible to test the reliability of self-learning, autonomous AI systems with standardized processes, also taking into account ethical considerations”, Wu predicts. “When this trust is in place, the chasm between research and application will be narrowed.”

FRAUNHOFER INSTITUTE FOR MANUFACTURING ENGINEERING AND AUTOMATION IPA



.....  
**PRESS RELEASE**

31 May 2021 || Page 3 | 3  
.....

**Dependable AI – Using AI in safety-critical industrial applications**

**Authors: Wu, Xinyang; El-Shamouty, Mohamed; Wagner, Philipp**

**Available to download in German and English at:**

<https://www.ki-fortschrittszentrum.de/de/studien/zuverlaessige-ki.html>

---

**Expert contact**

**Xinyang Wu** | Phone +49 711 970-3673 | [xinyang.wu@ipa.fraunhofer.de](mailto:xinyang.wu@ipa.fraunhofer.de) | Fraunhofer Institute for Manufacturing Engineering and Automation IPA | [www.ipa.fraunhofer.de](http://www.ipa.fraunhofer.de)

**Press communication**

**Jörg-Dieter Walz** | Telefon +49 711 970-1667 | [joerg-dieter.walz@ipa.fraunhofer.de](mailto:joerg-dieter.walz@ipa.fraunhofer.de)

With nearly 1000 employees, the **Fraunhofer Institute for Manufacturing Engineering and Automation IPA**, Fraunhofer IPA, is one of the largest institutes in the Fraunhofer-Gesellschaft. The total budget amounts to more than € 74 million. The institute's research focus is on organizational and technological aspects of production. We develop, test and implement not only components, devices and methods, but also entire machines and manufacturing plants. Our 15 departments are coordinated via six business units, which together conduct interdisciplinary work with the following industries: automotive, machinery and equipment industry, electronics and microsystems, energy, medical engineering and biotechnology as well as process industry. The research activities of Fraunhofer IPA aim at the economic production of sustainable and personalized products.