

PRESS RELEASE

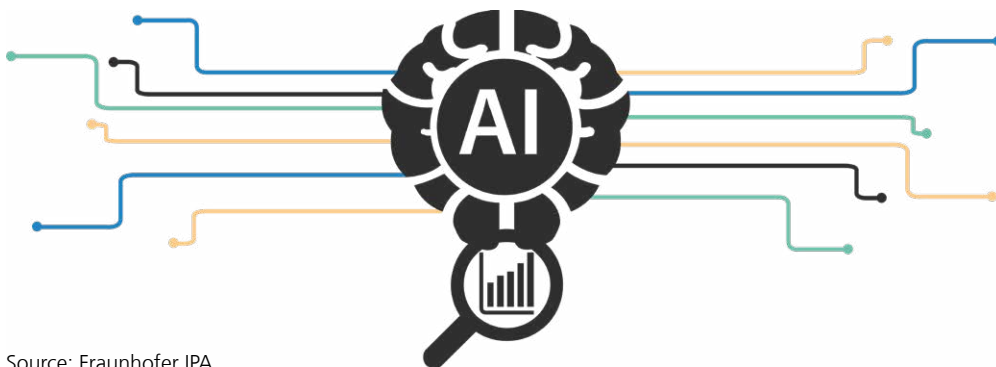
PRESS RELEASE

2 June 2021 || Page 1 | 3

Study “Explainable AI in practice – application-based evaluation of xAI methods”

Mastering Artificial Intelligence

In most cases, artificial intelligence has a black-box character. However, it is only through transparency that trust can be built. Specialized software is available in order to explain different AI models. Now, a study by Fraunhofer IPA has compared and evaluated various methods, which explain machine learning approaches.



Source: Fraunhofer IPA

Just a few decades ago, artificial intelligence (AI) was the stuff of science fiction, but has since become part of our daily lives. In manufacturing, it identifies anomalies in the production process. For banks, it makes decisions regarding loans. And on Netflix, it finds just the right film for every user. All of this is made possible by highly complex algorithms that work covertly in the background. The more challenging the problem, the more complex the AI model – and the more inscrutable.

But users want to be able to understand how a decision has been made, particularly with critical applications: Why was the work piece rejected? What caused the damage to my machine? It is only by understanding the reason for decisions that improvements can be made – and this increasingly applies to safety, too. In addition, the EU General Data Protection Regulation stipulates that decisions must be transparent.

Press communication

Hannes Weik | Phone +49 711 970-1664 | presse@ipa.fraunhofer.de

Fraunhofer Institute for Manufacturing Engineering and Automation IPA | Nobelstrasse 12 | 70569 Stuttgart | www.ipa.fraunhofer.de

Software comparison for xAI

PRESS RELEASE2 June 2021 || Page 2 | 3

In order to solve this problem, an entirely new area of research has arisen: “Explainable Artificial Intelligence”, or xAI for short. There are now numerous digital aids on the market which explain complex AI approaches. For example, in an image, they mark up the pixel which led to parts being rejected. The Stuttgart-based experts at the Fraunhofer Institute for Manufacturing Engineering and Automation IPA have now compared and evaluated nine common explanation techniques – including LIME, SHAP and Layer-Wise Relevance Propagation – by means of exemplary applications. There were three main criteria for this:

- **Stability:** The program should always generate the same explanation for the same problem. For example, an explanation should never flag up sensor A and then sensor B if the same anomaly arises in a production machine. This would harm trust in the algorithm and make it difficult to know what course of action to take.
- **Consistency:** At the same time, only slightly different input data should also receive similar explanations.
- **Fidelity:** Additionally, it is especially important that explanations are representative of how the AI model actually functions. For example, the explanation for the rejection of a bank loan should not be that the customer is too old when the reason was that their income was too low.

The use case is crucial

Conclusion of the study: All explanation methods researched were viable. However, as Nina Schaaf, who is responsible for the study at Fraunhofer IPA, explains: “There is no single perfect method.” For example, significant differences emerge in the time needed for generating explanations. The respective objective also largely determines which software is best to use. For example, Layer-Wise Relevance Propagation and Integrated Gradients are particularly good for image data. In summary, Schaaf says: “The target group is also important when it comes to explanations: An AI-developer will want and should receive an explanation phrased differently to the production manager, as both will draw different conclusions from the explanation.”

FRAUNHOFER INSTITUTE FOR MANUFACTURING ENGINEERING AND AUTOMATION IPA



.....
PRESS RELEASE

2 June 2021 || Page 3 | 3
.....

The study can be accessed in German and English
at [https://www.ki-fortschrittszentrum.de/de/
studien/erklarbare-ki-in-der-praxis.html](https://www.ki-fortschrittszentrum.de/de/studien/erklarbare-ki-in-der-praxis.html)

Further information

<https://www.ipa.fraunhofer.de/ai>

Expert contact

Nina Schaaf | Phone +49 711 970-1971 | nina.schaaf@ipa.fraunhofer.de | Fraunhofer Institute for Manufacturing Engineering and Automation IPA | www.ipa.fraunhofer.de

Press communication

Jörg-Dieter Walz | Phone +49 711 970-1667 | 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.