

# PRESSEINFORMATION

PRESSEINFORMATION

15. Oktober 2024 || Seite 1 | 2

## Whitepaper analysiert Status quo und Potenziale erklärbarer KI

**Erklärbare KI (XAI) kann tiefere Einblicke in KI-Systeme bieten und somit zu deren Absicherung und Zertifizierung beitragen. Auch wenn eine volle Absicherung nur durch erklärbare KI eher unrealistisch ist, wird XAI als wertvolles Instrument angesehen. Die Fraunhofer-Institute IAO und IPA haben in einem Whitepaper den Status quo und die Potenziale von XAI analysiert.**

Die Absicherung und Zertifizierung von Künstlicher Intelligenz (KI) sind essenziell, um das Vertrauen in die Leistungsfähigkeit von KI-basierten Produkten und Services zu gewährleisten. Eine vielversprechende Methode dafür bietet die erklärbare Künstliche Intelligenz oder XAI. Durch XAI können tiefere Einblicke in die Funktionsweise von KI-Systemen gewonnen werden. Denn Methoden der XAI können die Logiken, auf deren Basis Entscheidungen und Datenverknüpfungen in KI-Modellen erfolgen, nachvollziehbar machen.

XAI ist jedoch ein junges Forschungsfeld, dessen Potenziale und Herausforderungen noch nicht vollständig erschlossen sind. Forschende an den Fraunhofer-Instituten für Arbeitswirtschaft und Organisation IAO und für Produktionstechnik und Automatisierung IPA haben deshalb den aktuellen Stand sowie die Potenziale von XAI im Kontext der Absicherung und Zertifizierung von KI analysiert. Ihre Ergebnisse haben sie in einem neuen Whitepaper veröffentlicht.

### Vorteile von XAI auf mehreren Ebenen der KI-Entwicklung

XAI könnte erheblichen Mehrwert für die Entwicklung sicherer KI-Systeme bieten. Bereits im frühen Entwicklungsstadium können XAI-Methoden genutzt werden, um Fehler und Verzerrungen (Biases) in der Modellentwicklung zu identifizieren und das Verständnis der zugrunde liegenden Datenbasis zu verbessern. In späteren Entwicklungsphasen kann mittels XAI überprüft werden, ob das trainierte KI-Modell und die zugehörige Datenbasis rechtlichen Vorgaben und relevanten Normen entsprechen. In diesem Schritt könnte XAI ebenfalls genutzt werden, um offensichtliche Zusammenhänge im KI-Modell erkennbar zu machen, die dann wiederum als Grundlage der Zertifizierung des Modells dienen können. Im Schadensfall kann XAI helfen, die rechtliche Verantwortung zu klären. In der Studie wurden die Potenziale von XAI für die gerade genannten Anwendungsfälle näher erforscht.

### Volle Absicherung durch XAI eher unrealistisch

Eine umfassende Absicherung von KI durch XAI, insbesondere durch vollständige globale Erklärungen von Modellen, schätzen Expertinnen und Experten als unrealistisch ein. Die Vielzahl der Methoden und die Schwierigkeit, passende Tools für spezifische Anwendungsfälle auszuwählen, sind herausfordernd. Zudem sind viele Methoden für die Zielgruppen ungeeignet und schwer verständlich. Eine vollständige Erfassung aller

relevanten Informationen, die zu einer KI-Entscheidung führen, wird als nicht erreichbar angesehen. Dennoch kann XAI einen positiven Beitrag zur Absicherung und Zertifizierung leisten – allerdings nur in Kombination mit menschlicher Prüfung.

---

**PRESSEINFORMATION**

15. Oktober 2024 || Seite 2 | 2

---

**Weiterentwicklung von XAI gelingt nur in Zusammenarbeit**

Die Weiterentwicklung von XAI und deren Integration in Zertifizierungsprozesse erfordert eine interdisziplinäre Anstrengung. Eine stärkere Industrie- und Praxisorientierung könnte helfen, XAI-Methoden für bisher wenig erforschte Bereiche wie Zeitreihendaten oder Large Language Models (LLMs) zu entwickeln.

Nur durch die Zusammenarbeit von Forschung, Industrie und öffentlichen Akteuren können effektive und vertrauenswürdige Zertifizierungsstandards entwickelt werden, die den Herausforderungen moderner KI-Systeme gerecht werden und deren gesellschaftliche Akzeptanz fördern.



Link zum Paper:

<https://publica.fraunhofer.de/entities/publication/cb5bcc6e-3b56-4e3a-9286-e73d8c88aa1e>

---

**Fachlicher Ansprechpartner:**

**Benjamin Fresz** | Telefon +49 711 970-1404 | [benjamin.fresz@ipa.fraunhofer.de](mailto:benjamin.fresz@ipa.fraunhofer.de) | Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA | [www.ipa.fraunhofer.de](http://www.ipa.fraunhofer.de)

**Pressekommunikation**

**Dr. Karin Röhrich** | Telefon +49 711 970-3874 | [karin.roehricht@ipa.fraunhofer.de](mailto:karin.roehricht@ipa.fraunhofer.de)

Die **Fraunhofer-Gesellschaft** mit Sitz in Deutschland ist eine der führenden Organisationen für anwendungsorientierte Forschung. Im Innovationsprozess spielt sie eine zentrale Rolle – mit Forschungsschwerpunkten in zukunftsrelevanten Schlüsseltechnologien und dem Transfer von Forschungsergebnissen in die Industrie zur Stärkung unseres Wirtschaftsstandorts und zum Wohle unserer Gesellschaft.

Die 1949 gegründete Organisation betreibt in Deutschland derzeit 76 Institute und Forschungseinrichtungen. Die gegenwärtig knapp 32 000 Mitarbeitenden, überwiegend mit natur- oder ingenieurwissenschaftlicher Ausbildung, erarbeiten das jährliche Finanzvolumen von 3,4 Mrd. €. Davon fallen 3,0 Mrd. € auf den Bereich Vertragsforschung, der sich in drei Finanzierungssäulen gliedert: Einen Anteil davon erwirtschaftet Fraunhofer mit Aufträgen aus der Industrie und aus Lizenzertträgen, die sich auf insgesamt 836 Mio. € belaufen. Der hohe Anteil an Wirtschaftserträgen ist das Fraunhofer-Alleinstellungsmerkmal in der deutschen Forschungslandschaft. Ein weiterer Teil aus dem Bereich Vertragsforschung stammt aus öffentlich finanzierten Forschungsprojekten. Bund und Länder komplettieren die Vertragsforschung durch die Grundfinanzierung. Damit ermöglichen die Zuwendungsgeber, dass die Institute schon heute Problemlösungen entwickeln können, die in einigen Jahren für Wirtschaft und Gesellschaft relevant werden.