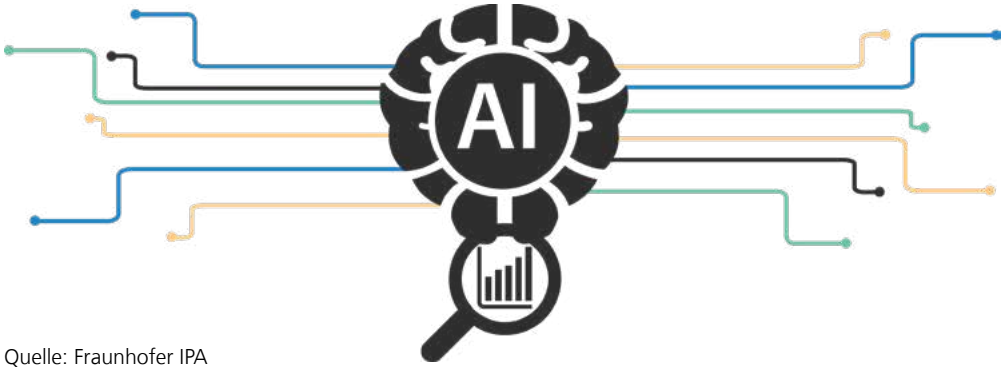
PRESSEINFORMATION

Studie »Erklärbare KI in der Praxis – Anwendungsorientierte Evaluation von xAI-Verfahren«

Künstliche Intelligenz beherrschen

Künstliche Intelligenz hat meistens Black-Box-Charakter. Doch nur Transparenz kann Vertrauen schaffen. Um den jeweiligen Lösungsweg zu erklären, gibt es spezielle Software. Eine Studie des Fraunhofer IPA hat jetzt unterschiedliche Methoden verglichen und bewertet, die maschinelle Lernverfahren erklärbar machen.



Quelle: Fraunhofer IPA

Künstliche Intelligenz, vor wenigen Jahrzehnten noch Sciencefiction, ist inzwischen im Alltag angekommen. In der Fertigung erkennt sie Anomalien im Produktionsprozess, in Banken entscheidet sie über Kredite und bei Netflix findet sie für jeden Kunden den passenden Film. Dahinter stecken hochkomplexe Algorithmen, die im Verborgenen agieren. Je anspruchsvoller das Problem, desto komplexer das KI-Modell – und damit auch undurchschaubarer.

Doch die Nutzer wollen insbesondere bei kritischen Anwendungen verstehen, wie eine Entscheidung zustande kommt: Warum wurde das Werkstück als fehlerhaft aussortiert? Wodurch wird der Verschleiß meiner Maschine verursacht? Nur so sind Verbesserungen möglich, die zunehmend auch die Sicherheit betreffen. Zudem zwingt die europäische Datenschutzgrundverordnung dazu, Entscheidungen nachvollziehbar zu machen.

Pressekommunikation
Hannes Weik | Telefon +49 711 970-1664 | presse@ipa.fraunhofer.de
Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA | Nobelstraße 12 | 70569 Stuttgart | www.ipa.fraunhofer.de

Softwarevergleich für xAI

Um dieses Problem zu lösen, ist ein ganzes Forschungsfeld entstanden: die »Explainable Artificial Intelligence«, die erklärbare Künstliche Intelligenz, kurz xAI. Auf dem Markt gibt es inzwischen zahlreiche digitale Hilfen, die komplexe KI-Lösungswege erklärbar machen. Sie markieren etwa in einem Bild diejenigen Pixel, die dazu geführt haben, dass fehlerhafte Teile aussortiert wurden. Experten des Fraunhofer-Instituts für Produktionstechnik und Automatisierung IPA aus Stuttgart haben nun neun gängige Erklärungsverfahren – wie LIME, SHAP oder Layer-Wise Relevance Propagation – miteinander verglichen und mithilfe von beispielhaften Anwendungen bewertet. Dabei zählten vor allem drei Kriterien:

- **Stabilität**: Bei gleicher Aufgabenstellung soll das Programm stets dieselbe Erklärung liefern. Es darf nicht sein, dass für eine Anomalie in der Produktionsmaschine einmal Sensor A und dann Sensor B verantwortlich gemacht wird. Das würde das Vertrauen in den Algorithmus zerstören und das Ableiten von Handlungsoptionen erschweren.

- **Konsistenz**: Gleichzeitig sollten nur geringfügig unterschiedliche Eingabedaten auch ähnliche Erklärungen erhalten.

- **Wiedergabetreue**: Besonders wichtig ist auch, dass Erklärungen tatsächlich das Verhalten des KI-Modells abbilden. Es darf nicht passieren, dass die Erklärung für die Verweigerung eines Bankkredits ein zu hohes Alter des Kunden benennt, obwohl eigentlich das zu geringe Einkommen ausschlaggebend war.

Ausschlaggebend ist der Anwendungsfall

Fazit der Studie: Alle untersuchten Erklärungsmethoden haben sich als brauchbar erwiesen. »Doch es gibt nicht die eine perfekte Methode«, sagt Nina Schaaf, die beim Fraunhofer IPA für die Studie verantwortlich ist. Große Unterschiede gibt es beispielsweise bei der Laufzeit, die ein Verfahren benötigt. Die Auswahl der besten Software ist zudem maßgeblich von der jeweiligen Aufgabenstellung abhängig. So sind etwa Layer-Wise Relevance Propagation und Integrated Gradients für Bilddaten besonders gut geeignet. »Und schließlich ist immer auch die Zielgruppe einer Erklärung wichtig: Ein KI-Entwickler möchte und sollte eine Erklärung anders dargestellt bekommen als der Produktionsleiter, denn beide ziehen jeweils andere Schlüsse aus den Erklärungen«, resümiert Schaaf.

Studie zum Download:
www.ki-fortschrittszentrum.de/de/studien/erklaerbare-ki-in-der-praxis.html

Weitere Informationen

www.ipa.fraunhofer.de/ki
www.ki-fortschrittszentrum.de/studien

Fachliche Ansprechpartnerin
Nina Schaaf | Telefon +49 711 970-1971 | nina.schaaf@ipa.fraunhofer.de | Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA | www.ipa.fraunhofer.de

Pressekommunikation
Jörg-Dieter Walz | Telefon +49 711 970-1667 | joerg-dieter.walz@ipa.fraunhofer.de

Das **Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA**, kurz Fraunhofer IPA, ist mit annähernd 1000 Mitarbeiterinnen und Mitarbeitern eines der größten Institute der Fraunhofer-Gesellschaft. Der gesamte Haushalt beträgt über 74 Mio €. Organisatorische und technologische Aufgaben aus der Produktion sind Forschungsschwerpunkte des Instituts. Methoden, Komponenten und Geräte bis hin zu kompletten Maschinen und Anlagen werden entwickelt, erprobt und umgesetzt. 15 Fachabteilungen arbeiten interdisziplinär, koordiniert durch 6 Geschäftsfelder, vor allem mit den Branchen Automotive, Maschinen- und Anlagenbau, Elektronik und Mikrosystemtechnik, Energie, Medizin- und Biotechnik sowie Prozessindustrie zusammen. An der wirtschaftlichen Produktion nachhaltiger und personalisierter Produkte orientiert das Fraunhofer IPA seine Forschung.