Fraunhofer IPA

Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

Ein AI Innovation Seed des KI-Fortschrittszentrums

Erklärbare KI: Evaluation von Erklärbarkeitsverfahren in der Anwendung

Ausgangssituation

In vielen Anwendungsfällen bietet der Einsatz von KI-Verfahren ein enormes Potenzial. Allerdings ist oftmals eine hohe Vorhersagegenauigkeit von KI-Modelle alleine nicht ausreichend, etwa beim Einsatz von KI in stark regulierten Branchen wie der Pharmaindustrie. Hier sollten die eingesetzten KI-Modelle bestenfalls nachvollziehbar sein. Eine Lösung hierfür bieten Verfahren der erklärbaren KI (explainable AI – XAI), mit deren Hilfe KI-Modelle und deren Entscheidungen interpretiert werden können.

Bisher sind diese Ansätze jedoch noch sehr forschungsnah. Das bedeutet, dass XAI-Techniken oftmals lediglich an öffentlich verfügbaren Benchmark Datensätzen erprobt werden. Zudem sind bisher wenige Techniken entwickelt worden, mit deren Hilfe die XAI-Verfahren selbst sowie die generierten Erklärungen evaluiert werden können. Alles in allem fehlen Praxisbeispiele und Ansätze zur Evaluation der Eignung von XAI-Verfahren in realen Use Cases.

Erklärbarkeitsverfahren in der Anwendung

Genau hier setzt das Vorhaben an. Im Projekt steht die methodische Konzeption für den Einsatz von XAI-Verfahren für reale Anwendungsfälle im Fokus. Dabei soll einerseits die Erklärung von KI-Modellen und deren Vorhersagen betrachtet werden. Darüber hinaus wird ein weiteres Augenmerk auf der Bewertung der Sicherheit von Modellentscheidungen liegen.

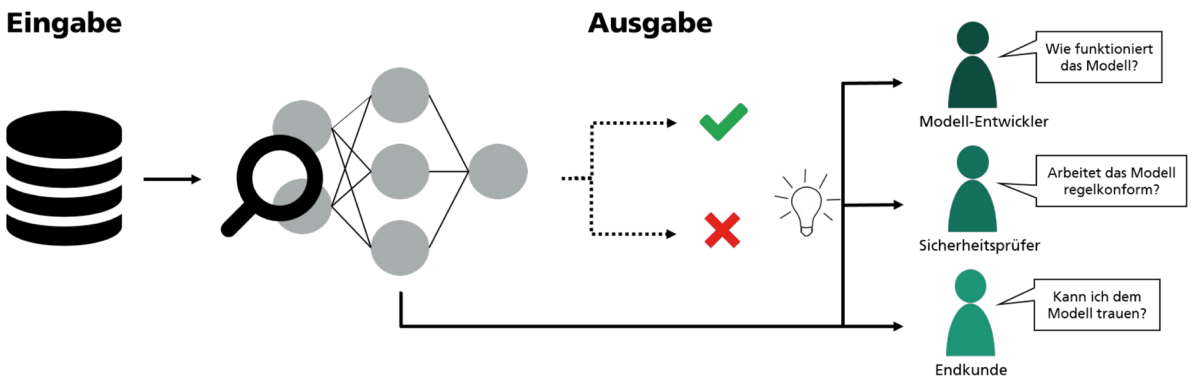

Abbildung 1: Beispiel »Erklärung von KI-Modellentscheidungen« für unterschiedliche Stakeholder, Quelle: Fraunhofer IPA

KI Fortschrittszentrum
LERNENDE SYSTEME UND KOGNITIVE ROBOTIK

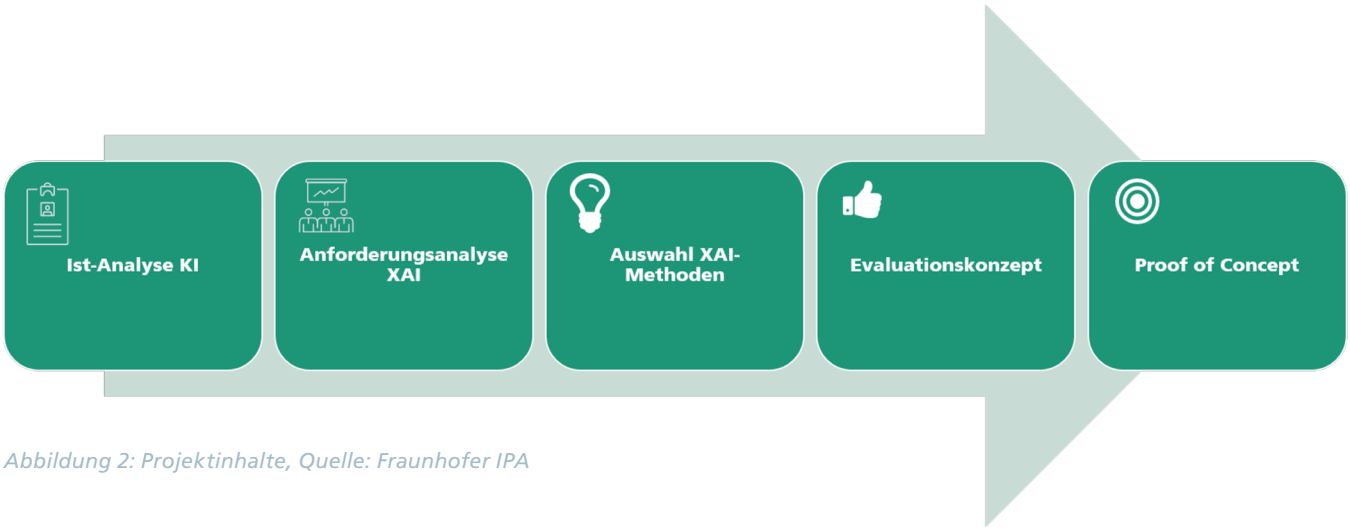

Abbildung 2: Projektinhalte, Quelle: Fraunhofer IPA

Folgende Fragestellungen werden im Projekt adressiert:

- Welche Limitierungen gibt es derzeit beim Einsatz von KI in stark regulierten Branchen?
- Welche Anforderungen an die Nachvollziehbarkeit von KI-Modellen existieren?
- Welche XAI-Methoden sind geeignet für welche Anwendungsszenarien?
- Wie kann der Einsatz von XAI-Methoden in realen Anwendungsfällen begleitet und evaluiert werden?

Ergebnis der gemeinsamen Arbeiten sind Best-Practice-Ansätze in Form eines Vorgehensmodells für den Einsatz von Erklärbarkeitsansätzen für reale Anwendungsfälle. Zudem wird das Vorgehensmodell im Rahmen des Projekts prototypisch erprobt.

Zielgruppe

Das Projekt »Evaluation von Erklärbarkeitsverfahren in der Anwendung« steht allen Wirtschaftspartnern aus Baden-Württemberg offen, die Interesse an der Nachvollziehbarkeit von KI-Modellen haben und sich aktiv einen Wettbewerbsvorteil erarbeiten möchten. Im Idealfall existieren in ihrem Unternehmen bereits Anwendungsfälle bei denen KI produktiv eingesetzt wird.

Für Sie als Partner hat eine Teilnahme am Projekt zahlreiche Vorteile:

Wissensvorsprung
- durch exklusive Forschungsergebnisse
- durch Aufbau eines Netzwerks von Experten und Anwendern
- durch Best-Practice Ansätze
- für die strategische Frühausrichtung

Kommunikation
- Starke öffentliche Wahrnehmung
- Innovationsführerschaft

Projektablauf

Wie in Abbildung 2 dargestellt, soll zunächst im Rahmen eines Workshops erarbeitet werden, ob in ihrem Unternehmen bereits KI-Applikationen eingesetzt werden, die aber aufgrund regulatorischer Limitierungen nicht vollumfänglich genutzt werden können. Des Weiteren sollen KI-Anwendungsfälle erarbeitet werden, die sich für das Projektformat eignen, falls in ihrem Unternehmen KI noch nicht produktiv zum Einsatz kommt. Aus dem ausgewählten Use-Case leitet sich im nächsten Schritt auch die Anforderungsanalyse an die XAI-Verfahren und deren generierte Erklärungen ab. Die ausgewählten Methoden zur Extraktion von Erklärungen aus dem KI-Modell werden daraufhin an den Use-Case angepasst und hinsichtlich der zuvor definierten Anforderungen evaluiert. Finalisiert wird das Projekt mit einem Proof-of-Concept, der vollumfänglich sinnvolle und nachvollziehbare Erklärungen für den gewählten Anwendungsfall generiert und die KI-Applikation somit transparenter und vertrauenswürdiger werden lässt.

Kontakt

Sind Sie an einer Teilnahme interessiert?
Sprechen Sie uns gerne an!

Danilo Brajovic
Telefon: +49 711 970-3647
danilo.brajovic@ipa.fraunhofer.de

Philipp Wagner
Telefon: +49 711 970-1988
philipp.wagner@ipa.fraunhofer.de

Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA
Nobelstraße 12
70569 Stuttgart

www.ipa.fraunhofer.de

Kontakt:
info@ki-fortschrittszentrum.de

Weitere Informationen unter:
www.ki-fortschrittszentrum.de/erklaerbarkeitsverfahren

Fraunhofer

KI-Fortschrittszentrum »Lernende Systeme und Kognitive Robotik«

Eine Kooperation der Fraunhofer-Institute für Arbeitswirtschaft und Organisation IAO und für Produktionstechnik und Automatisierung IPA

Das KI-Fortschrittszentrum »Lernende Systeme und Kognitive Robotik« unterstützt Firmen dabei, die wirtschaftlichen Chancen der Künstlichen Intelligenz und insbesondere des Maschinellen Lernens für sich zu nutzen. In anwendungsnahen Forschungsprojekten und in direkter Kooperation mit Industrieunternehmen arbeiten die Stuttgarter Fraunhofer-Institute Produktionstechnik und Automatisierung IPA sowie für Arbeitswirtschaft und Organisation IAO daran, Technologien aus der KI-Spitzenforschung in die breite Anwendung der produzierenden Industrie und der Dienstleistungswirtschaft zu bringen. Finanzielle Förderung erhält das Zentrum vom Ministerium für Wirtschaft, Arbeit und Tourismus Baden-Württemberg.

Europas größte Forschungskooperation auf dem Gebiet der KI

Das KI-Forschungszentrum ist Forschungspartner des Cyber Valley, einem Konsortium aus den renommierten Universitäten Tübingen und Stuttgart, dem Max-Planck-Institut für intelligente Systeme und einigen führenden Industrieunternehmen. In gemeinsamen Forschungslabors werden Grundlagenforschung und anwendungsorientierte Entwicklung zu aktuellen wie auch zukünftigen Bedarfen behandelt und vorangetrieben.

Menschzentrierte KI

Alle Aktivitäten des Zentrums verfolgen das Ziel, eine menschzentrierte KI zu entwickeln, der die Menschen vertrauen und die sie akzeptieren. Nur wenn Menschen mit neuen Technologien intuitiv interagieren und vertrauensvoll zusammenarbeiten, kann ihr Potenzial optimal ausgeschöpft werden. Daher konzentrieren sich die Forschungsaktivitäten unter anderem auf die Themen Erklärbarkeit, Datenschutz, Sicherheit und Robustheit von KI-Technologien.

www.ki-fortschrittszentrum.de

Kontakt

Prof. Dr. Marco Huber
Telefon +49 711 970-1960
marco.huber@ipa.fraunhofer.de

Dr. Matthias Peissner
Telefon +49 711 970-2311
matthias.peissner@
iao.fraunhofer.de

Dr. Werner Kraus
Telefon +49 711 970-1049
werner.kraus@ipa.fraunhofer.de

Kooperationspartner

S·TEC

Cyber Valley

Gefördert durch

Baden-Württemberg
MINISTERIUM FÜR WIRTSCHAFT, ARBEIT UND TOURISMUS

KI Fortschrittszentrum
LERNENDE SYSTEME UND KOGNITIVE ROBOTIK