Fraunhofer
IAIS

FRAUNHOFER-INSTITUT FÜR INTELLIGENTE ANALYSE- UND INFORMATIONSSYSTEME IAIS

AI TRUSTWORTHINESS CHECK

ÜBERPRÜFUNG DER VERTRAUENSWÜRDIGKEIT VON KI-ANWENDUNGEN

Die Herausforderung: Potenziale und Risiken von KI-Anwendungen erkennen

Der Einsatz von KI-Anwendungen bietet enormes Innovationspotenzial und wird zunehmend als Schlüssel zur Kostensenkung und Produktoptimierung gesehen. Jedoch werden KI-Anwendungen oft von ihren Anbietern und Entwickler*innen in ihrer Leistungsfähigkeit überschätzt und selten kritisch hinterfragt. Nutzer*innen können die potenziellen Schwächen von KI-Anwendungen häufig erst nach ihrer Einführung und kostenintensiven Erprobung feststellen. So entstehen Unsicherheiten und die Potenziale von Künstlicher Intelligenz werden oft nicht voll ausgeschöpft.

- Sie überlegen, ob der Einsatz einer KI-Anwendung in Ihrem Unternehmen im Hinblick auf alle relevanten Fragestellungen sinnvoll und vertretbar ist?
- Sie ziehen den Erwerb oder Einsatz einer KI-Anwendung in Betracht und brauchen fundierte Beratung bei der Auswahl?
- Sie entwickeln selbst KI-Anwendungen und wollen diese – im Hinblick auf zukünftige Kundenanforderungen oder auch für den internen Einsatz – von neutralen, anerkannten Expert*innen begutachten lassen?

Unsere Lösung: Unabhängige Bewertung von KI-Anwendungen vor ihrer Einführung

Wir analysieren Ihre KI-Anwendung systematisch im Hinblick auf Vertrauenswürdigkeit und Einsatzfähigkeit. Mit dem »AI Trustworthiness Check« bewerten wir Ihre KI-Anwendung nachvollziehbar und kostengünstig schon vor der Einführung im konkreten Use Case.

Unsere erfahrenen Projektleiter*innen erstellen gemeinsam mit Ihren und unseren Fachleuten ein spezifisches Anforderungsprofil für Ihren Use Case. Dazu analysieren wir KI-Risiken, die sich speziell aus dem vorliegenden Anwendungskontext sowie den gewählten Lernverfahren ergeben. Anschließend untersuchen wir die KI-Anwendung systematisch auf Schwachstellen in allen relevanten Handlungsfeldern und analysieren dabei:

- die Datengrundlage
- Design und Architektur im Hinblick auf die Erfüllung zugesicherter Eigenschaften (z. B. »privacy-by-design«)
- Anforderungen an die Einbettung bzw. die Kompatibilität der KI-Anwendung mit dem Use Case
- Performanz und Vertrauenswürdigkeit der in der KI-Komponente implementierten Lernverfahren durch gezielte Tests

Die von uns entwickelten Prüfwerkzeuge erlauben einen schnellen und systematischen Einblick in die von der KI-Anwendung erzielten Ergebnisse. Auf Wunsch führen wir auch Tests durch, die die Generalisierungsfähigkeit beurteilen, d. h. wie gut die KI-Anwendung ihre antrainierte Funktionalität in der Betriebsumgebung mit neuen Daten umsetzen wird.

Basierend auf unseren Analyseergebnissen erhalten Sie eine detaillierte und fundierte Übersicht über die Stärken und Schwächen der KI-Anwendung zusammen mit konkreten, herstellerneutralen Handlungsempfehlungen.

SIE HABEN INTERESSE?

Wir beraten Sie kostenlos und unverbindlich dazu, wie Ihr »AI Trustworthiness Check« aussehen könnte.

Dr. Maximilian Poretschkin
Telefon +49 2241 14-1984
maximilian.poretschkin@iais.fraunhofer.de

www.iais.fraunhofer.de/zuverlässige-ki

ÜBER DAS FRAUNHOFER IAIS

Als Teil der größten Organisation für anwendungsorientierte Forschung in Europa ist das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS mit Sitz in Sankt Augustin bei Bonn eines der führenden Wissenschaftsinstitute auf den Gebieten Künstliche Intelligenz, Maschinelles Lernen und Big Data in Deutschland und Europa. Mit seinen mehr als 300 Mitarbeitenden unterstützt das Institut Unternehmen bei der Optimierung von Produkten, Dienstleistungen, Prozessen und Strukturen sowie bei der Entwicklung neuer digitaler Geschäftsmodelle. Damit gestaltet das Fraunhofer IAIS die digitale Transformation unserer Arbeits- und Lebenswelt. Einen wichtigen Schwerpunkt stellt die Absicherung von KI-Systemen und die Entwicklung geeigneter Prüfverfahren dar. Das Fraunhofer IAIS steht im Zentrum eines starken Forschungsnetzwerks und koordiniert unter anderem seit 2014 als geschäftsführendes Institut die Fraunhofer-Allianz Big Data und Künstliche Intelligenz. Daneben bestehen langjährige enge Kooperationen in Forschung und Lehre. Im Jahr 2018 hat das Fraunhofer IAIS sein strategisches Netzwerk weiter ausgebaut und ist auf Landes-, Bundes, und EU-Ebene in führender Rolle an wichtigen Initiativen beteiligt. So leitet das Fraunhofer IAIS die Geschäftsstelle der Kompetenzplattform KI.NRW, ist gemeinsam mit der TU Dortmund führender Partner im Kompetenzzentrum Maschinelles Lernen Rhein-Ruhr (ML2R) – einem von sechs bundesweiten Knotenpunkten für Spitzenforschung und Transfer im Maschinellen Lernen, und hat eine tragende Rolle innerhalb der EU-Initiative »A European AI On-Demand Platform and Ecosystem« (AI4EU) inne.

Die Methodik des »AI Trustworthiness Check«

Handlungsfelder

Zur Prüfung der Vertrauenswürdigkeit einer KI-Anwendung werden alle relevanten Handlungsfelder untersucht.

Autonomie & Kontrolle	Ist eine selbstbestimmte, effektive Nutzung der KI möglich?
Fairness	Behandelt die KI alle Betroffenen fair?
Datenschutz	Schützt die KI die Privatsphäre und sonstige sensible Informationen?
Sicherheit	Ist die KI sicher gegenüber Angriffen, Unfällen und Fehlern?
Transparenz	Sind Funktionsweise und Entscheidungen der KI nachvollziehbar?
Verlässlichkeit	Funktioniert die KI zuverlässig und ist sie robust?

Prüfumfang

Wir verfolgen einen risikobasierten Prüfansatz. Dabei wird das Risiko der KI-Anwendung für jedes dieser Handlungsfelder unter Berücksichtigung des gesamten Entwicklungs- und Betriebszyklus der KI-Anwendung bewertet.

Design		Durch Konzeption und Architektur der KI-Anwendung können bestimmte Eigenschaften bereits »by design« sichergestellt werden, z. B. »privacy-by-design«, »safety-by-design«.
Entwicklung	**Daten**	Wahl, Augmentierung und Vorverarbeitung von Trainings-, Test- und Inputdaten haben einen essentiellen Anteil an der Qualität des KI-Systems.
	KI-Komponente	Die Auswahl einer Methode / eines Algorithmus, das Training und Testen bzw. Validieren der Modelle, Aspekte zu Transparenz und Erklärbarkeit. Die Implementierung in Software.
	Einbettung	Die Aktionen und Entscheidungen der KI-Anwendung basieren auf Ihrer KI-Komponente, die mit der Einbettung interagiert. Die Einbettung kann unter anderem zur Erfüllung von Sicherheitseigenschaften beitragen.
Betrieb		Das Testen und das Evaluieren der Qualität der KI-Anwendung während des Betriebs sind wichtige Faktoren für die Vertrauenswürdigkeit.

KI-Absicherung und -Zertifizierung
Dr. Maximilian Poretschkin
Telefon +49 2241 14-1984
maximilian.poretschkin@iais.fraunhofer.de

Fraunhofer-Institut für
Intelligente Analyse- und
Informationssysteme IAIS

Schloss Birlinghoven
53757 Sankt Augustin

www.iais.fraunhofer.de/
zuverlässige-ki