

Press release

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS

Silke Loh

04/03/2024

<http://idw-online.de/en/news831315>

Scientific Publications, Transfer of Science or Research
Economics / business administration, Information technology, Mechanical engineering
transregional, national



Lösungen für leistungsfähige und vertrauenswürdige Künstliche Intelligenz: Fraunhofer IAIS auf der Hannover Messe 2024

Die Nachfrage nach KI-Lösungen in der Industrie ist groß – dabei müssen sie leistungsfähig, vertrauenswürdig und sicher sein, um beispielsweise in der Serienproduktion oder Qualitätskontrolle eingesetzt zu werden. Die neuen Möglichkeiten der generativen KI werfen aber auch Fragen auf: Können Anwenderinnen und Anwender sich auf die Auskünfte eines Chatbots verlassen? Wie erkennt man bereits in der Entwicklung Schwachstellen eines KI-Modells? Auf der Hannover Messe 2024 präsentieren das Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS und die Institute der Fraunhofer-Allianz Big Data AI zwei Exponate und mehrere Use Cases rund um vertrauenswürdige KI-Lösungen.

KI besitzt großes Potenzial für die industrielle Produktion, beispielsweise für die Automatisierung, die Qualitätsprüfung oder die Optimierung von Prozessen: »Wir erhalten aktuell viele Anfragen von Unternehmen, die KI-Prototypen entwickelt haben und diese in die Serienproduktion bringen wollen. Damit die Skalierung gelingt, müssen diese KI-Lösungen systematisch getestet werden, um auch Schwachstellen erkennen zu können, die im Prototypen nicht zu Tage getreten sind«, erklärt Dr. Maximilian Poretschkin, Teamleiter KI-Absicherung und Zertifizierung am Fraunhofer IAIS.

Auf dem Fraunhofer Stand auf der Hannover Messe 2024 (22. bis 26. April, Halle 2, Stand B24) können Fachbesucherinnen und Fachbesucher sich anhand mehrerer Exponate und konkreter Use Cases über Anwendungsbeispiele und Lösungen für vertrauenswürdige KI informieren und erfahren, wie sie diese sicher in ihre Betriebsabläufe integrieren können.

Exponat 1: »KI-Zuverlässigkeit in der Produktion« – Prüftools zum systematischen Testen von KI-Modellen

Ein solches Anwendungsbeispiel ist ein Prüfwerkzeug für KI-Modelle, die in der Produktion sowie im Maschinen- und Anlagenbau eingesetzt werden. Das Tool ermöglicht die systematische Schwachstellensuche von KI-Systemen, um so deren Zuverlässigkeit und Robustheit sicherzustellen. »Unsere Methodik beruht darauf, den Anwendungsbereich des KI-Systems genau zu spezifizieren. Konkret geht es darum, den Raum der möglichen Eingaben, die das KI-System verarbeitet, zu parametrisieren und mit einer semantischen Struktur zu versehen. Mit den KI-Prüfwerkzeugen, die wir unter anderem im KI.NRW-Flagships-Projekt »Zertifizierte KI« entwickelt haben, lassen sich dann Fehler der KI-Systeme erkennen«, erklärt Poretschkin.

In dem Forschungsprojekt werden gemeinsam mit Partnern Prüfverfahren für die Zertifizierung von Systemen der Künstlichen Intelligenz entwickelt: Unternehmen erhalten mit dem KI-Prüfkatalog einen praxisorientierten Leitfaden, der sie befähigt, ihre KI-Systeme leistungsfähig und vertrauenswürdig zu gestalten. Ein aktuelles Whitepaper beschäftigt sich zudem mit der Frage, wie mit generativer KI und Foundation-Modellen entwickelte KI-Anwendungen

bewertet und gesichert werden können.

Exponat 2: »Who's deciding here?!« – Was kann Künstliche Intelligenz entscheiden und was noch nicht?

Ganz im Zeichen der vertrauenswürdigen Künstlichen Intelligenz (KI) steht auch das Gemeinschaftsexponat der Fraunhofer-Allianz Big Data AI auf der Hannover Messe 2024. Mit dem Titel »Who's deciding here?!« greift es das Thema »Freiheit« des Wissenschaftsjahrs 2024 des Bundesministeriums für Bildung und Forschung auf: Wie beeinflussen Technologien wie Künstliche Intelligenz unsere Entscheidungsfreiheit? Wie vertrauenswürdig sind die KI-Systeme, die in immer mehr sensiblen Anwendungen wie beispielsweise Kreditwürdigkeitsprüfungen zum Einsatz kommen sollen?

Das Exponat ist als interaktives Spiel gestaltet, bei dem Interessierte den Einsatz von KI im Alltag reflektieren können. Welche Entscheidungen können und wollen wir KI-Algorithmen überlassen und welche sollten wir besser selbst treffen? Themen sind hier unter anderem die Objekterkennung beim autonomen Fahren, die Gefahr von Diskriminierung durch Künstliche Intelligenz oder das Erkennen von Fake News. Sebastian Schmidt, Data Scientist für vertrauenswürdige KI am Fraunhofer IAIS, sagt: »Das Spiel macht für jeden erlebbar, wo KI ein hilfreicher Assistent sein kann und wo der Mensch besser mitentscheidet. Ziel ist immer, zur richtigen Entscheidung zu kommen, ohne dass der Mensch seine Entscheidungskompetenz und Souveränität verliert.« Das interaktive Spiel ist in diesem Jahr auch auf dem Ausstellungsschiff MS Wissenschaft erlebbar.

Use Cases für professionelle Anwendungen

Außerdem präsentieren die Institute der Fraunhofer-Allianz Big Data AI mehrere konkrete Anwendungsbeispiele aus verschiedenen Anwendungsbereichen. Die verbindende Fragestellung aller Use Cases: Wie lässt sich KI-Technologie kontrolliert und sicher einsetzen?

Sicheres Autofahren: Uncertainty Wrapper

Der verlässliche Umgang mit Unsicherheiten stellt bei der Nutzung von KI-Vorhersagen in vielen Anwendungsbereichen einen entscheidenden Faktor dar. Der Uncertainty Wrapper des Fraunhofer-Instituts für Experimentelles Software Engineering IESE prüft, wie sicher oder unsicher die Aussage der KI ist – etwa am Beispiel der Objekterkennung von Straßenschildern.

Fake News erkennen: DisCo

Fake News verbreiten sich über Social Media in rasender Geschwindigkeit. Mithilfe textforensischer Methoden des Fraunhofer-Instituts für Sichere Informationstechnologie SIT werden Texte für Journalistinnen und Journalisten so aufbereitet, dass prüfwürdige Textpassagen farbig hervorgehoben sind.

Große KI-Sprachmodelle für Unternehmen: OpenGPT-X

In dem vom Bundesministerium für Wirtschaft und Klimaschutz (BMWK) geförderten Konsortialprojekt OpenGPT-X trainieren Expertinnen und Experten aus Forschungsorganisationen und Wirtschaftsunternehmen unter Leitung des Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS und des Fraunhofer-Instituts für Integrierte Schaltungen IIS große KI-Sprachmodelle mit bis zu 24 europäischen Sprachen. Sie sollen Unternehmen künftig Open Source zur Verfügung zu stehen, um Anwendungen auf Basis generativer KI entwickeln und Prozesse optimieren zu können.

Vertrauenswürdigkeit sensibler Produktionsdaten: DigiWeld

Das Fraunhofer IPA hat eine Lösung entwickelt, die es Maschinen- und Anlagenbauern ermöglicht, die Vorteile Künstlicher Intelligenz zu nutzen, ohne sensible Produktionsdaten ihrer Kunden erfassen zu müssen.

Ethische Grundsätze und menschenzentrierte KI: ADA Lovelace Center

Im ADA Lovelace Center for Analytics, Data and Applications des Fraunhofer IIS arbeiten die Forscherinnen und Forscher an menschenzentrierten und ethischen Fragestellungen. Durch einen sozial- und verhaltenswissenschaftlichen Zugang soll insbesondere erklärt werden, warum Menschen KI akzeptieren (oder auch nicht) und welche Folgen die Nutzung für den Einzelnen mit sich bringt.

KI-Forschung für die technologische Souveränität Deutschlands

Mit all diesen Lösungen leisten die Fraunhofer-Forscherinnen und -Forscher einen wichtigen Beitrag, das Potenzial der KI auch konkret nutzbar zu machen. »Gerade die aktuelle Entwicklung von generativer KI zeigt das riesige Potenzial, aber auch die Herausforderungen dieser Zukunftstechnologie in Hinblick auf Sicherheit, Transparenz und Datenschutz auf. Viele Technologien und Lösungen kommen aktuell nicht aus Europa. Die Forschung der Fraunhofer-Gesellschaft trägt dazu bei, die technologische Souveränität und Unabhängigkeit deutscher Unternehmen zu erhalten und auszubauen«, sagt Dr. Sonja Holl-Supra, Geschäftsführerin der Fraunhofer-Allianz Big Data AI – einem Zusammenschluss aus über 30 Fraunhofer-Instituten, der die KI-Kompetenzen von Fraunhofer bündelt. Die vom EU-Parlament kürzlich verabschiedete europäische KI-Verordnung sieht ebenfalls die Durchführung von KI-Prüfungen vor. Für die Umsetzung dieser Anforderung sind entsprechende Prüftools hilfreich.

Für Fachbesucherinnen und -Besucher stehen sowohl das Exponat zur KI-Verlässlichkeit in der Produktion, das interaktive Spiel »Who's deciding here?!« und die Fraunhofer-Lösungen rund um vertrauenswürdige KI auf dem Gemeinschaftsstand der Fraunhofer-Gesellschaft auf der Hannover Messe 2024 (22. bis 26. April 2024, Halle 2, Stand B24) bereit.

URL for press release: <http://www.iais.fraunhofer.de/hannover-messe-2024> Veranstaltungshinweis Hannover Messe 2024

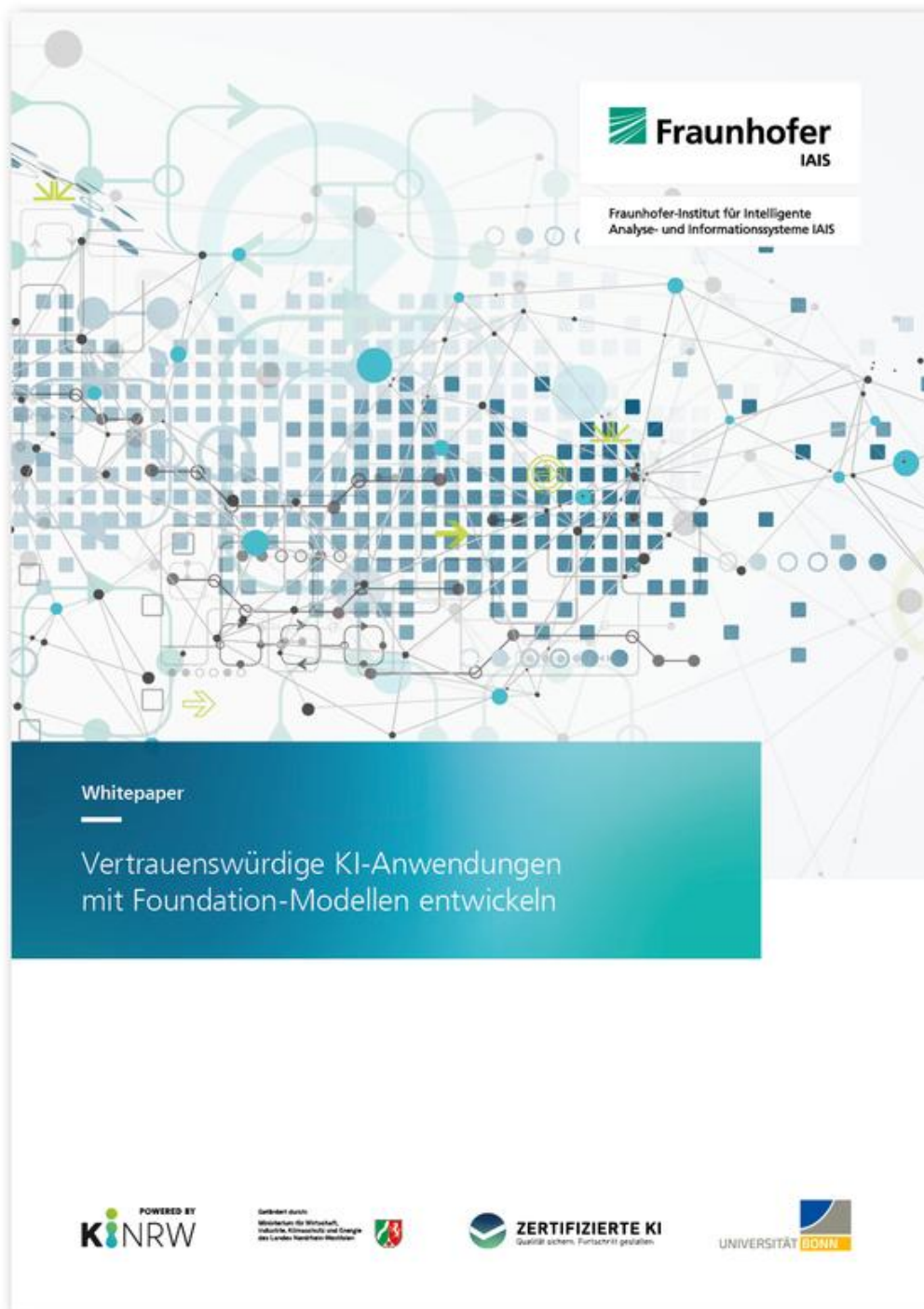
URL for press release: <http://www.bigdata-ai.fraunhofer.de/de/veranstaltungen/wissenschaftsjahr-2024.html>
Wissenschaftsjahr Freiheit

URL for press release: <http://www.iais.fraunhofer.de/de/presse/presseinformationen/presseinformationen-2024/press-information-240403.html> Zur Presseinformation



Ein systematisches Vorgehen in Kombination mit KI-Prüftools hilft dabei, Schwächen von KI-Modellen ausfindig zu machen.

Fraunhofer IAIS / Zertifizierte KI



Das Whitepaper »Vertrauenswürdige KI-Anwendungen mit Foundation-Modellen entwickeln« erläutert, wie sich spezielle Risiken der generativen Künstlichen Intelligenz auf die KI-Anwendung auswirken können.
Fraunhofer IAIS / Zertifizierte KI