# PRESSEINFORMATION



**White Paper: Zuverlässige KI**

## KI fit machen für sicherheitskritische Anwendungen

**Produktionsplanung, Logistik, Wartung, Qualitätskontrolle – in der industriellen Fertigung gibt es viele Einsatzgebiete für Künstliche Intelligenz. In der Praxis werden KI-Modelle bisher allerdings noch wenig genutzt. Der Grund: Die Zuverlässigkeit ist schwer prüfbar. Neue Zertifizierungs-Kriterien können die KI fit machen für sicherheitskritische Anwendungen.**

Die Erwartungen sind kaum zu toppen: Künstliche Intelligenz soll die Produktion flexibilisieren, die Wartung vorausschauend planen, den Warenfluss optimieren, die Logistik automatisieren, die Qualitätskontrollen automatisieren. »Tatsächlich wurden in den letzten Jahren – auch am Fraunhofer IPA – zahlreiche vielversprechende KI-Algorithmen und -Architekturen entwickelt, beispielsweise für Computervision, Mensch-Maschine-Schnittstellen oder vernetzte Robotik«, berichtet Xinyang Wu vom Zentrum für Cyber Cognitive Intelligence am IPA. Was jetzt fehle, sei die praktische Umsetzung. »Zwischen Forschung und Anwendung klafft eine Lücke. In der Industrie werden die neuen KI-Anwendungen nur zögerlich eingesetzt. Sie gelten als nicht zuverlässig genug für sicherheitskritische Anwendungen.«

Die Vorbehalte der Anwender kennt Wu aus erster Hand: »Wenn wir mit unseren Partnern aus der Industrie sprechen, dann wird schnell klar, dass die Unternehmen beispielsweise autonome und selbstlernende Roboter nur dann nutzen wollen, wenn diese absolut zuverlässig arbeiten, und wenn man mit hundertprozentiger Sicherheit sagen kann, dass die Maschinen keine Gefahr für den Menschen darstellen.«

Genau das lässt sich bisher nicht beweisen. Es gibt weder Normen noch standardisierte Tests. Diese wären jedoch dringend nötig, betont Wu: »Das Ziel muss sein, die Entscheidungen, die von Algorithmen gefällt werden, zertifiziert und transparent zu machen. So muss zum Beispiel die Nachvollzierbarkeit gewährleistet sein: Wenn eine Maschine selbstständig Entscheidungen fällt, dann muss ich – zumindest im Nachhinein – herausfinden können, warum sie in einer bestimmten Situation einen Fehler gemacht hat. Nur so lässt sich verhindern, dass dieser wieder auftritt. Black-Box-Modelle, bei denen man die Entscheidung der Algorithmen nicht nachvollziehen kann, sind nach unserer Einschätzung für sicherheitskritische Anwendungen nicht direkt für den Einsatz geeignet – es sei denn das Modell wird durch die richtige Methode zertifiziert.«

Doch wie überprüft man Künstliche Intelligenz? Das IPA-Team am Zentrum für Cyber Cognitive Intelligence hat dafür jetzt eine Strategie vorgeschlagen und über den Stand der entsprechenden Technik in dem White Paper »Zuverlässige KI – KI in sicherheits-

**Pressekommunikation**
**Jörg-Dieter Walz** | Telefon +49 711 970-1667 | presse@ipa.fraunhofer.de
Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA | Nobelstraße 12 | 70569 Stuttgart | www.ipa.fraunhofer.de

kritischen industriellen Anwendungen einsetzen« berichtet: Die Strategie basiert auf Zertifizierbarkeit und Transparenz.

**Kriterien-Katalog für mehr Sicherheit**
»In erster Linie ging es uns erst einmal darum, Regeln zu finden, mit deren Hilfe sich die Zuverlässigkeit von Maschinellem Lernen und der dazugehörigen KI bewerten lässt«, berichtet Wu. Das Ergebnis dieser Recherche sind fünf Kriterien, die KI-Systeme erfüllen sollen, um als sicher zu gelten:

– Alle Entscheidungen der Algorithmen müssen für Menschen verständlich sein.

– Die Funktion der Algorithmen muss vor ihrem Einsatz mit Methoden der Formalen Verifikation geprüft werden.

– Darüber hinaus ist eine statistische Validierung notwendig, besonders wenn die Formale Verifikation wegen Skalierbarkeit für den bestimmten Anwendungsfall nicht nutzbar ist. Dies kann durch Testläufe mit größeren Datenmengen beziehungsweise Stückzahlen überprüft werden.

– Auch die Unsicherheiten, die den Entscheidungen Neuronaler Netze zu Grunde liegen, müssen ermittelt und quantifiziert werden.

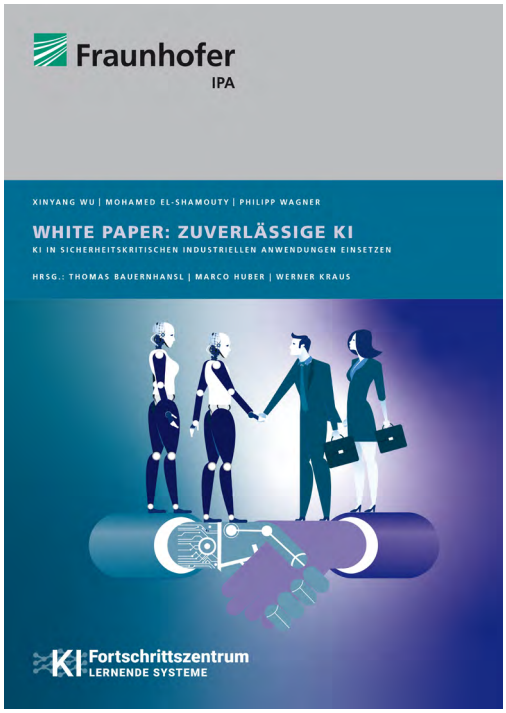
– Während des Betriebs müssen die Systeme permanent überprüft werden, beispielsweise durch Online-Monitoring. Wichtig ist dabei die Erfassung von Input und Output – also von Sensordaten und den aus deren Auswertung resultierenden Entscheidungen.

Die fünf Kriterien könnten die Grundlage bilden für eine – künftige – standardisierte Prüfung, betont Wu: »Am IPA haben wir bereits für jeden dieser Punkte unterschiedliche Algorithmen und Methoden zusammengestellt, mit denen sich die Zuverlässigkeit von KI-Systemen auch tatsächlich überprüfen lässt. Bei einigen unserer Kunden haben wir solche Prüfungen auch schon durchgeführt.«

**Transparenz schafft Vertrauen**
Die zweite Grundvoraussetzung für einen sicheren Einsatz der KI-Systeme ist deren Transparenz. Diese ist gemäß den ethischen Richtlinien der »High-Level Expert Group on Artificial Intelligence« der Europäischen Kommission, kurz HLEG AI, eines der Schlüsselelemente für die Realisierung einer vertrauenswürdigen KI. Diese Transparenz bezieht sich, anders als die Kriterien, mit denen die Zuverlässigkeit in der algorithmischen Ebene geprüft werden kann, ausschließlich auf die Interaktion mit dem Menschen in der systematischen Ebene. Drei Punkte sind dafür aufgrund der Richtlinien der HLEG AI zusammengefasst, die transparente KI erfüllen muss: Erstens müssen die von den Algorithmen gefällten Entscheidungen nachvollziehbar sein. Zweitens muss es für Menschen auf einer umfassenden Ebene des menschlichen Verständnisses möglich sein, die Entscheidungen zu erklären.

Und drittens müssen KI-Systeme mit dem Menschen kommunizieren und ihn informieren, welche Fähigkeiten die Algorithmen haben und wo sie an Grenzen stoßen.

»Nur wenn es gelingt, die Zuverlässigkeit selbstlernender, autonomer KI-Systeme mit standardisierten Verfahren zu testen und dabei auch ethische Aspekte zu berücksichtigen, werden die Anwender der KI vertrauen – egal ob im Straßenverkehr oder in der Fabrikhalle«, prognostiziert Wu. »Wenn dieses Vertrauen da ist, wird sich die Lücke zwischen Forschung und Anwendung schließen.«




**Zuverlässige KI – KI in sicherheitskritischen industriellen Anwendungen einsetzen**

**Autoren:**
**Wu, Xinyang**
**El-Shamouty, Mohamed**
**Wagner, Philipp**

**Download unter:**
https://www.ki-fortschrittszentrum.de/de/studien/zuverlaessige-ki.html

Weitere Informationen:
http://www.ki-fortschrittszentrum.de/studien

**Fachliche Ansprechpartner**
**Xinyang Wu** | Telefon +49 711 970-3673 | xinyang.wu@ipa.fraunhofer.de | Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA | www.ipa.fraunhofer.de
**Mohamed El-Shamouty** | Telefon +49 711 970-1660 | mohamed.el-shamouty@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