Regulierung und Zertifizierung von KI in der Industrie: Ziele, Kriterien und Herausforderungen

Axel Mangelsdorf(✉), Nicole Wittenbrink, und Peter Gabriel
Institut für Innovation und Technik (iit), Berlin, Deutschland
mangelsdorf@iit-berlin.de, wittenbrink@iit-berlin.de
gabriel@iit-berlin.de

Zusammenfassung. Mit dem Vorschlag der Europäischen Kommission zur Regulierung von Künstlicher Intelligenz (dem Artificial Intelligence Act) ist die Zertifizierung von KI-Systemen in den Fokus gerückt. Gleichzeitig werden mit der Veröffentlichung der ersten Prüfkriterien die Anforderungen der Zertifizierung für Unternehmen konkreter. Der vorliegende Beitrag stellt den Regulierungsvorschlag der Kommission vor und fasst die dazugehörige Fachdebatte zusammen. Danach werden die Anforderungen eines Prüfkatalog zur Zertifizierung von KI-Systemen anhand eines ersten Vorschlags aus dem Projekt „Zertifizierte KI“ präsentiert und bewertet.

Schlüsselwörter: Künstliche Intelligenz · Zertifizierung · Europäische Union · Qualitätsinfrastruktur

1 Einleitung

Das Potenzial der Künstlichen Intelligenz (KI) in Industriebetrieben ist groß. KI-Systeme ermöglichen, dass sich Produktionsanlagen selbstständig im laufenden Betrieb optimieren, Maschinen bereits während der Fertigung Qualitätskontrollen vornehmen können oder im Rahmen der „Predictive Maintenance“ den optimalen Zeitpunkt für eine Wartung voraussagen. Werden KI-Systeme mit den Erfahrungswerten langjähriger Mitarbeiter trainiert, kann es sogar gelingen, implizites Wissen aus historischen Erfahrungswerten für zukünftige Generation nutzbar zu machen (Seifert et al. 2018, Frost & Sullivan 2018). Derzeit werden in Industriebetrieben traditionelle Automatisierungstechniken, das industrielle Internet der Dinge und neue innovative KI-Technologien mit dem Ziel eingesetzt, die Automatisierung weiter zu steigern, Prozesse zu optimieren und Kosten zu sparen. Derzeit sind Industrieunternehmen oftmals bemüht, KI-gestützte Softwaresysteme an die Stelle von traditionellen Automatisierungstechniken zu setzen, die noch immer eine Vielzahl von menschlichen Interaktionen erfordern. Diese KI-gestützten Systeme bieten zwar den Vorteil, dass sie sehr viel weniger menschliche Interaktion erfordern, aus Erfahrungen lernen und Entscheidungen selbstständig treffen können. Zugleich bringen sie aber auch einige Risiken mit sich (Müller-Quade et al. 2019). In der verarbeitenden Industrie können Produktionssteuerungssysteme sabotiert werden, in der Automobilindustrie entstehen

E. A. Hartmann (Hrsg.): *Digitalisierung souverän gestalten II,* S. 110–119, 2022.
https://doi.org/10.1007/978-3-662-64408-9_9

Haftungsfragen bei Unfällen und in der Pharmaindustrie können Daten oder Formeln gestohlen werden (Frost & Sullivan 2018).

Industriebetriebe benötigen deshalb für den Einsatz von KI-gestützten Softwaresystemen ein hohes Maß an Vertrauen. Bei der Implementierung von KI-Systemen müssen Datenschutz und Sicherheitsverstöße, Haftung und Verantwortung und ethische Bedenken adressiert werden. Ebenso erwarten Verbraucherinnen und Verbraucher, aber auch Arbeitnehmerinnen und Arbeitnehmer, dass KI-Systeme die Sicherheit persönlicher Daten sicherstellen und Benachteiligungen aufgrund von Alter, Geschlecht, Wohnort oder anderen Merkmalen verhindern (WEF 2021). Gelingt es, die Künstliche Intelligenz zu einer sicheren und verantwortungsbewussten Technologie weiterzuentwickeln, steigt auch die Wahrscheinlichkeit, dass Industrieunternehmen KI-Systeme in ihre Prozesse einbinden und von den Vorteilen profitieren.

Eine unabhängige Prüfung durch Dritte im Rahmen einer Zertifizierung stellt eine Möglichkeit dar, Sicherheit und Vertrauen in eben jene Systeme herzustellen (Matus und Veale 2021). Auch die Europäische Union hat in ihrem jüngst vorgelegten Gesetzesvorschlag zur Regulierung von KI-Systemen, dem „Artifical Intelligence Act“, die Konformitätsbewertung für Hochrisikosysteme vorgeschlagen. Zwar gibt es in Deutschland noch kein anerkanntes Verfahren für KI-Systeme. Derzeit werden die Bemühungen verschiedener Organisationen Zertifizierungsverfahren zu entwickeln aber immer konkreter. Dazu gehören unter anderem der Prüfkatalog des vom Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS getragenen Projekts „Zertifizierte KI“ (Poretschkin et al. 2021) und der Kriterienkatalog des Standards AIC4 des Bundesamts für Sicherheit in der Informationstechnik (BSI), der sich auf Cloud-basierte KI-Systeme fokussiert (Bundesamt für Sicherheit in der Informationstechnik 2021). Solche Zertifizierungen können wesentlich zur digitalen Souveränität des KI-Einsatzes in Unternehmen beitragen. Tun die aktuellen Vorschläge das aber auch wirklich?

Um darauf eine Antwort zu geben, wird im vorliegenden Beitrag zunächst der aktuelle Gesetzesvorschlag der Europäischen Union vorgestellt sowie die dazugehörige Debatte der Fachwelt zusammengefasst. Das geschieht exemplarisch anhand der Beiträge in der öffentlichen Konsultation zum Gesetzesvorschlag. Anschließend wird betrachtet, wie ein konkreter Prüfkatalog für eine KI-Zertifizierung aussehen könnte. Dafür dient beispielhaft der Prüfkatalog des Projekts „Zertifizierte KI“.

2 Der Artifical Intelligence Act der Europäischen Kommission

Ähnlich wie bei der Diskussion um den Datenschutz nimmt Europa auch bei der Debatte um die Zertifizierung von KI-Systemen im weltweiten Vergleich eine führende Rolle ein. Dänemark und Malta hatten im Jahr 2019 in ihren nationalen KI-Strategien Informationspflichten für KI-Systeme oder auch eine freiwillige Zertifizierung vorgesehen. In Deutschland wurde bereits ein Jahr zuvor, im Jahr 2018, eine Datenethikkommission eingesetzt, die in ihrem Gutachten unter anderem auch ein Zulassungsverfahren für KI-Software empfohlen hat. Um die Einheitlichkeit

des Binnenmarkts zu wahren, bemüht sich die Europäische Kommission um eine EU-weite Lösung. Das spiegelt sich im Weißbuch „Künstliche Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“ aus dem Jahr 2020, insbesondere aber im Gesetzesvorschlag der Kommission vom 21. April 2021 für einen „Artificial Intelligence Act (AIA)“ wider (Mangelsdorf et al. 2021).

Die EU-Kommission hat damit den weltweit ersten Vorschlag zu einem Rechtsrahmen für „vertrauenswürdige Künstliche Intelligenz“ vorgelegt. Der AIA sieht eine Einteilung von KI-Anwendungen in vier Risikoklassen vor: inakzeptables Risiko, hohes Risiko, begrenztes Risiko und minimales Risiko. Für Anwendungen mit inakzeptablen Risiko (zum Beispiel KI-getriebene Social-Scoring-Systeme) ist ein Verbot vorgesehen. Anwendungen der Klasse „hohes Risiko“ sollen vor der Markteinführung einer Reihe rechtlicher Verpflichtungen unterliegen – einschließlich einer Konformitätsbewertung. Unter anderem soll dadurch sichergestellt werden, dass der Betrieb der Anwendungen für die Nutzer hinreichend transparent ist. Jede Hochrisiko-KI-Anwendung ist mit einer Gebrauchsanleitung zu versehen, die ihre Merkmale, Fähigkeiten und Leistungsgrenzen einschließt. Darunter fallen neben der Beschreibung der Zweckbestimmung auch Angaben zur Genauigkeit, Robustheit und Cybersicherheit des Systems, den (Fehl-)Anwendungsrisiken sowie Spezifikationen in Hinblick auf die verwendeten Trainings-, Validierungs- und Testdatensätze. Im Rahmen der Konformitätsbewertung soll überprüft werden, ob die Mindestanforderungen an Hochrisiko-KI-Systeme in Hinblick auf Transparenz, Risikomanagement, Datensicherheit und Daten-Governance, technische Dokumentation, Aufzeichnungspflicht, menschliche Aufsicht sowie Genauigkeit, Robustheit und Cybersicherheit erfüllt werden. Die im AIA vorgeschlagenen Mindestanforderungen sind aus den im Jahr 2018 publizierten Ethik-Leitlinien der hochrangigen Expertengruppe für Künstliche Intelligenz (High-Level Expert Group, HLEG) abgeleitet (Hochrangige Expertengruppe für künstliche Intelligenz 2019).

Bereits vor der offiziellen Veröffentlichung des AIA hat sich abgezeichnet, dass das Interesse an dem Vorschlag groß und breit ist. Das Durchsickern mehrerer Entwurfsversionen in der Woche vor der Freigabe hat eine rege Diskussion in den sozialen und digitalen Medien entfacht. An der Diskussion beteiligen sich viele Akteursgruppen, die aktiv versuchen, auf die Gestaltung des Rechtsrahmens Einfluss zu nehmen, darunter Wirtschafts-, Unternehmens- und Verbraucherverbände, Gewerkschaften, Behörden, Forschungsinstitutionen, Nichtregierungsorganisationen (NRO) und EU-Bürgerinnen und -Bürger. Über 300 Akteure aus aller Welt sind unter anderem dem Aufruf der Kommission gefolgt, sich bis zum 6. August 2021 in einer öffentlichen Konsultation zu dem Vorschlag zu äußern und haben entsprechende Stellungnahmen eingereicht. Etwa die Hälfte der Stellungnahmen ging dabei in den letzten 24 h der ausgeschriebenen Feedback-Periode ein. Dies unterstreicht die Relevanz und Aktualität des Themas. Insbesondere von Seiten der Wirtschaft ist eine hohe Aktivität zu verzeichnen. Über die Hälfte der Stellungnahmen entfällt auf Wirtschafts- und Unternehmensverbände sowie auf Unternehmen mit starkem Industriebezug. Die Beiträge im Konsultationsprozess sind damit ein gutes Abbild der öffentlichen Debatte zum AIA.

Die Bemühung der EU-Kommission, einen einheitlichen Rechtsrahmen zu schaffen, wird von einer großen Mehrheit aller Akteursgruppen begrüßt. Dass ein

Rechtsrahmen erforderlich und sinnvoll ist, scheint grundsätzlich nicht zur Diskussion zu stehen. Ob seine jetzt vorliegende Ausgestaltung allen Wünschen gerecht wird, ist hingegen umstritten. Aus Sicht der NRO weist das Netz, dass der Rechtsrahmen für Anwendungen mit hohem Risiko aufspannt, noch zu viele Schlupflöcher auf. Von Seiten der Wirtschaft und Industrie wird im Gegensatz dazu die Befürchtung geäußert wird, dass das Netz zu engmaschig angelegt ist. Obwohl dies zunächst fundamental gegensätzlich erscheint, gibt es in Hinblick auf die zentralen Kritikpunkte und die geforderten Nachbesserungen tatsächlich einige Gemeinsamkeiten. Von beiden Seiten wird vor allem stark kritisiert, dass die Bewertung der Risikoklasse laut AIA ausschließlich im Voraus erfolgen soll. Dass sich eine Anwendung gravierend auf Individuen oder die Gesellschaft auswirkt, stellt sich aus Sicht der NRO unter Umständen erst zu einem späteren Zeitpunkt heraus. Aus der Perspektive der Wirtschaft ist jedoch andererseits auch denkbar, dass die Auswirkungen einer Anwendung weniger kritisch sind, als im Voraus angenommen. Beide Seiten sehen hier dringenden Bedarf für eine Überarbeitung des AIA.

Von Seiten der Wirtschaft, die als Anwender von KI-Systemen stark vom AIA betroffen wären, wurden auch weitere Kritikpunkte benannt:

Definition von KI
Die Definition von KI im AIA sei nicht hinreichend, da sie insgesamt zu weit gefasst sei. Sie beziehe auch rein statistische Methoden, Bayesische Schätzungen sowie Logik-und Wissensbasierte Ansätze mit ein. Eine klare Abgrenzung zu konventionellen Datenanalyseanwendungen sei daher nicht gegeben. Rein statistische und hochentwickelte KI-Anwendungen sollten nicht den gleichen Anforderungen unterliegen. Daher wäre eine Einengung der Definition zwingend erforderlich, um Rechtsunsicherheiten und unangemessene sowie nicht-gerechtfertigte zusätzliche Kosten zu vermeiden.

Mehrfachregulierung/-zertifizierung
Industrielle KI-Anwendungen unterlägen in der Regel bereits der Produktsicherheitsgesetzgebung und harmonisierten Sicherheitsvorschriften. Das zukünftige Zusammenspiel des AIA mit den bestehenden Vorschriften sollte klar definiert werden, um Mehrfach-Regulierung sowie -Zertifizierung zu vermeiden.

Komplexität der Anforderungen und Operationalisierung
Die Anforderungen an Hochrisiko-KI-Anwendungen seien in der vorliegenden Version sehr komplex, unter anderem der Aufbau eines Qualitätsmanagementsystems, fortlaufende technische Dokumentation, Konformitätsbewertung und Überwachung nach dem Inverkehrbringen. Sie reichten teilweise zu weit oder wären noch nicht ausreichend spezifiziert; insbesondere aber bleibe unklar, wie die Bewertungskriterien und Anforderungen operationalisiert werden sollen. Dadurch entstehe eine erhebliche Unsicherheit auf Seiten der Unternehmen.

Wirtschaftliche Belastung
Die Erfüllung der komplexen Anforderungen an Hochrisiko-Systeme sei mit hohen Kosten verbunden, unter anderem für das Einstellen von Experten/Auditoren zur

Eigen- oder Fremdbewertung, die insbesondere für Start-ups sowie kleine und mittelständische Unternehmen (KMU) nicht tragbar seien. Die bisher vorgesehenen Maßnahmen wie „regulatorische Sandboxes“ wären noch nicht ausreichend, um dem entgegenzuwirken. Eine Ausweitung sei erforderlich, um Nachteile für Start-ups und KMU zu vermeiden.

3 Der KI-Prüfkatalog des Projekts „Zertifzierte KI“

Die Prüfung von KI-Systemen zum Minimieren von Risiken verlangt nach konkreten Vorschriften für Qualität und Sicherheit. Bisherige Bemühungen zum Beispiel der Datenethikkommission der Bundesregierung, der EU-Kommission durch die hochrangige Expertengruppe für Künstliche Intelligenz oder von multinationalen Unternehmen (zum Beispiel Amazon, Apple, Baidu, Facebook, Google, IBM und Intel im Rahmen der „Partnership on AI“) haben bisher eher zu allgemeinen Absichtserklärungen und weniger zu konkreten Handlungsanweisungen für Industrieunternehmen und Prüfer geführt (Mangelsdorf et al. 2021).

Normen und Standards für IT-Sicherheit sind dagegen weit konkreter und geben sowohl der zu zertifizierenden Organisation als auch den Auditoren der unabhängigen Prüfgesellschaft Kataloge mit zu erfüllenden Kriterien an die Hand. Die ISO-Norm 27001 zu Informationssicherheits-Managementsystemen gehört beispielsweise zu den verbreitetsten Normen überhaupt und wird laut der Erhebung des Deutschen Normungspanels von 38 % aller Unternehmen mit mehr als 250 Mitarbeitern umgesetzt (Blind und Heß 2020). Konkrete Normen und Standards, die Qualitäts- oder Sicherheitsvorschriften für KI-Anwendungen definieren, fehlen noch weitgehend. Wenig überraschend stellt deshalb die Normungsroadmap KI einen Bedarf nach zertifizierbaren Normen für KI-Systeme fest (DIN/DKE 2020). Die Normungsroadmap KI fordert deshalb ein Programm, um standardisierte Prüfverfahren zu entwickeln, mit deren Hilfe Aussagen über die Qualität und Sicherheit von KI-Anwendungen getroffen werden können.

Die Forderungen der Normungsroadmap KI werden mit Stand August 2021 in verschiedenen Projekten auf der Ebene der deutschen Bundesländer umgesetzt. Das Hessische Ministerium für Digitale Strategie und Entwicklung fördert im Projekt „AI Quality & Testing Hub“ die Entwicklung eines sektorübergreifenden Ansatzes für Prüfungen von KI-Systemen. Das Projekt hat mit der Deutschen Kommission Elektrotechnik, Elektronik, Informationstechnik (DKE) enge Verknüpfungen zur nationalen, europäischen und internationalen Normung und mit dem VDE Prüf- und Zertifizierungsinstitut in Offenbach eine Institution für die Prüfung und Zertifizierung zukunftsträchtiger Produkte (VDE 2021).

Am weitesten fortgeschritten ist jedoch das Gemeinschaftsprojekt „Zertifizierte KI“ des Fraunhofer IAIS, der Universität Bonn und des Bundesamts für Sicherheit in der Informationstechnik (BSI). Das Projekt wird von Land Nordrhein-Westfalen gefördert und zielt darauf ab, Prüfkriterien für KI-Systeme zu entwickeln und zu standardisieren. Anhand dieser Arbeiten soll exemplarisch beschrieben werden, wie eine Zertifizierung in der Praxis aussehen kann und welche Herausforderungen sie mit sich bringt.

Innerhalb des Projekts wurde ein Prüfkatalog entwickelt, der die Dimensionen Fairness, Autonomie und Kontrolle sowie Transparenz, Verlässlichkeit, Sicherheit und Datenschutz abdeckt. Der Prüfkatalog enthält messbare Zielvorgaben sowie Maßnahmen, um die Ziele zu erreichen (Poretschkin et al. 2021). Er beschreibt die Dokumentationspflichten, die laut des Gesetzesentwurfs der EU für KI-Systeme mit hohem Risiko gelten. Die Dokumentation liefert Behörden die erforderliche Information über das KI-System, um dessen Konformität zu beurteilen. Der Prüfkatalog zeigt, wie Unternehmen eine solche technische Dokumentation anzufertigen haben. Im Folgenden werden die einzelnen Dimensionen, deren Messung sowie gegenseitige Abhängigkeiten zwischen den Dimensionen vorgestellt.

Der EU-Gesetzesentwurf fordert für KI-Systeme mit hohem Risiko strenge Vorgaben, die für den Marktzugang erfüllt sein müssen. Zu den Vorgaben gehören eine Reihe von technischen Dokumentationspflichten, die in Anhang IV des Gesetzentwurfes beschrieben sind. Der Prüfkatalog liefert für eine Vielzahl der geforderten Dokumentationspflichten einen Leitfaden für die technischen Dokumentationen. Konkret deckt der KI-Prüfkatalog die Punkte 1 (Beschreibung des KI-Systems), 2 (Beschreibung der Bestandteile des KI-Systems und seines Entwicklungsprozesses), 3 (Informationen über die Überwachung, Funktionsweise und Kontrolle des KI-Systems), 5 (Beschreibung aller an dem System während seines Lebenszyklus vorgenommenen Änderungen) und zum Teil 8 (Beschreibung des Systems zur Bewertung der Leistung des KI-Systems in der Phase nach dem Inverkehrbringen) der geforderten Inhalte ab. Im Folgenden werden die Prüfkatalog-Dimensionen Fairness, Autonomie und Kontrolle, Transparenz, Verlässlichkeit, Sicherheit und Datenschutz konkret beschrieben (Poretschkin et al. 2021).

Fairness soll verhindern, dass die KI-Anwendungen zu diskriminierenden Ergebnissen führt, die beispielsweise durch fehlerhafte Trainingsdaten entstehen. Um diese Dimension zu beurteilen, müssen die Unternehmen die KI-Anwendung genau dokumentieren. Dazu gehören Fragebögen zur Selbsteinschätzung oder Dokumentationen zur Entstehung und Verwendung von Trainingsdaten, die vom Prüfer eingesehen werden können. Zur notwendigen Dokumentation gehört zudem eine Beschreibung, welche Arten von Diskriminierung im Kontext der KI-Anwendung akzeptabel und welche ungerechtfertigt sind. Ebenso dokumentiert werden soll, wie beim jeweiligen KI-System Fairness gemessen wird, wie die Trainingsdaten auf Fairness überprüft worden sind und welche Maßnahmen zur Herstellung von Fairness in den Trainingsdaten vorgenommen worden sind. Ebenso müssen die Unternehmen dokumentieren, wie sich die KI-Anwendung beim Lernprozess auf die Fairness auswirkt und wie das Erreichen der Fairness im laufenden Betrieb überwacht wird. Der Prüfer stellt schließlich fest, ob die quantitativen Fairness-Kriterien erfüllt werden und ob das Unternehmen einen Prozess eingeführt hat, um fortlaufend die Fairness der KI-Anwendung zu überwachen. Arbeiten Industriebetriebe mit KI-Anwendungen, die keinen Einfluss auf Personen haben oder die ohne personenbezogene Daten arbeiten, ist die Dimension Fairness vernachlässigbar.

Die Dimension **Autonomie und Kontrolle** stellt sicher, dass die KI-Anwendung menschliche Eingriffs- und Aufsichtsmöglichkeiten hat und dass Nutzer über die Risiken der Anwendung aufgeklärt sind. Das Abschalten durch den Nutzer soll unabhängig von den Entscheidungen der KI-Anwendung selbst erfolgen,

insbesondere wenn die Sicherheit von Personen gefährdet ist. Der Prüfkatalog sieht auch für diese Dimension verschiedene Dokumentationspflichten vor. Es muss dokumentiert werden, welche Personengruppen bei der Entwicklung der KI-Anwendung beteiligt waren und welche Argumente für und gegen alternative Gestaltungsmöglichkeiten hinsichtlich der Autonomie vorlagen. Für den laufenden Betrieb muss unter anderem dokumentiert werden, welche Eingriffsmöglichkeiten Nutzer haben und welche Qualifikationen zur Aufsicht der KI erforderlich sind und wann ein Eingreifen in den laufenden Betrieb notwendig ist. Ein Nutzerhandbuch mit regelmäßige Aktualisierungen soll ebenfalls vorhanden sein.

Bei der Dimension **Transparenz** wird verlangt, dass die KI-Entscheidung von Menschen nachvollzogen werden können. Bei vielen KI-Anwendungen, die auf Black-Box-Modelle zurückgreifen, müssen nachgeschaltete Verfahren das Zustandekommen erklären. Der Prüfkatalog fordert deshalb für diese Dimension verschiedene Dokumentationen und Erklärungen von den KI-Herstellerunternehmen. Beispielsweise muss erklärt werden, warum sich die Entwickler für ein Modell entschieden haben. Bei Black-Box-Modellen muss zudem erklärt werden, welche Maßnahmen ergriffen werden, um das Zustandekommen der Ergebnisse zu erklären. Lernt das Modell im laufenden Betrieb anhand erhobener Daten weiter, sollen auch diese Daten zur späteren Einsicht gespeichert werden.

Die Dimension **Verlässlichkeit** soll unter anderem sicherstellen, dass die Ausgaben der KI-Anwendung korrekt sind, und dass die Ausgaben gegenüber manipulierten Eingaben robust sind. Wollen Unternehmen KI-Anwendungen bereitstellen, müssen sie dokumentieren und begründen, welche Performanz-Metrik eingesetzt wird, die dafür sorgt, dass die Anwendung verlässliche Ergebnisse liefert. Es soll tabellarisch dargestellt werden, wie Fehler gefunden und abgefangen werden. Zudem muss ein Prozess zur regelmäßigen Überprüfung der Verlässlichkeit installiert sein. Sollte es gegenseitige Abhängigkeiten zwischen der Dimension Verlässlichkeit und anderen Dimensionen geben, müssen diese erläutert werden.

Die Dimension **Sicherheit** bezieht sich auf den Schutz vor äußeren Gefährdungen bis hin zum funktionalen Versagen der KI-Anwendung. Dabei werden potenzielle Personen- und Sach- sowie finanzielle Schäden betrachtet. Im Gegensatz zu den anderen beschriebenen Dimensionen gibt es für die Dimension Sicherheit bereits bestehende Normen, die nun mit KI-spezifischen Anforderungen ergänzt werden. Zum Beispiel kann der Prüfkatalog auf die Elemente der Norm DIN EN ISO 10218 „Industrieroboter – Sicherheitsanforderungen“ zurückgreifen. Insgesamt sollen dokumentierte Tests zeigen, dass die KI-Anwendung ein vertretbares Unfallrisiko gewährleistet.

Die letzte Dimension des Prüfkatalogs bezieht sich auf den **Datenschutz.** Dabei geht es vor allem um den Schutz von sensiblen personenbezogene Daten bei der Entwicklung und Betrieb der KI-Anwendung sowie um Geschäftsgeheimnisse der Unternehmen. Letzteres bezieht sich auf die Möglichkeit durch gezielte Abfragen die Struktur und den Algorithmus des KI-Modells zu extrahieren und nachzubauen. Der Prüfkatalog schlägt zudem eine Reihe von Maßnahmen vor. Zum Schutz personenbezogener Daten sollen Unternehmen für die jeweilige KI-Anwendung dokumentieren, welche Daten das System verwendet und wie diese Daten mit anderen Datenquellen verknüpft werden können. Die Unternehmen müssen Beispieldaten für die

Dokumentation zur Verfügung stellen. Zudem muss angegeben werden, mit welchen Verfahren Daten anonymisiert wurden und welche Maßnahmen eingesetzt werden, um das ungewollte Abfließen von Informationen zur KI-Anwendung zu verhindern.

Für jede der beschriebenen Dimensionen müssen die Unternehmen abschließend argumentieren, dass die ergriffenen Maßnahmen ausreichend sind, um definierte Kriterien zu erfüllen. Dabei müssen jeweils die Restrisiken abgewogen und Zielkonflikte mit anderen Dimensionen benannt werden. Zum Umgang mit Zielkonflikten empfiehlt der Prüfkatalog, dass Unternehmen den Abwägungsprozess zum Umgang mit Zielkonflikten dokumentieren und ein unternehmensinternes Gremium installieren (ein sogenanntes „AI Ethics Review Board“), um die ethische Praxis zu diskutieren und das System kontinuierlich zu bewerten.

Mit dem Prüfkatalog des Projekts „Zertifizierte KI“ liegen nun zwar konkrete Prüfkriterien vor allem in Form von Dokumentationspflichten vor, offen bleiben jedoch Fragen zu Kosten und Dauer der Prüfung. Ebenfalls noch wenig Beachtung finden Fragen für die Prüfung und Zertifizierung der notwendigen Qualitätsinfrastruktur, also des Systems für Normung, Prüfdienstleistungen, Akkreditierung und Zertifizierung. Unter anderem ist bei bestehendem Fachkräftemangel im Feld der Künstlichen Intelligenz unklar, wie es Prüfunternehmen gelingen wird, ausreichend Auditoren mit KI-Kompetenz zur Konformitätsbewertung zu gewinnen, ohne dabei gleichzeitig den Fachkräftemangel bei den Unternehmen zu vergrößern. Im Bereich der staatlichen Förderung von Standards und Zertifizierungssystemen ist zudem kein auf einen Standard und ein Zertifikat erkennbares Vorgehen zu erkennen. Es ist zu erwarten, dass die parallellaufenden Projekte in Hessen („AI Quality & Testing Hub“) und Nordrhein-Westfalen („Zertifizierte KI“) auch zu unterschiedlichen Prüfkriterien und Standards führen. Die fördernden Institutionen stehen vor der Herausforderung, sinnvolle Prinzipien zum Umgang mit Künstlicher Intelligenz zu harmonisieren. Weltweit haben zwischen den Jahren 2015 und 2020 verschiedene Organisationen insgesamt 117 Ethikregeln entworfen und die Anzahl wächst weiterhin (ScienceBusiness 2021). Schon jetzt ähnelt die Zertifizierungslandschaft im Bereich der Künstlichen Intelligenz der Zertifizierungslandschaft im Nachhaltigkeitssektor, wo kostspielige multiple Zertifizierungen üblich sind (Matus und Veale 2021). Während im Nachhaltigkeitssektor Unternehmen mit wenig anspruchsvollen Zertifizierungen versuchen umweltfreundlicher zu erscheinen als es tatsächlich der Fall ist („greenwashing“), ist zu befürchten, dass Unternehmen mit KI-Anwendungen durch wenig anspruchsvolle Zertifizierungen („ethics-washing“) versuchen ihr Interesse an gerechten KI-Anwendungen zu übertreiben.

4 Zusammenfassung und Fazit

Systeme mit Künstlicher Intelligenz liefern auch für Industriebetriebe viele Vorteile. Zusammen mit dem industriellen Internet der Dinge können KI-Systeme die Automatisierung weiter steigern, Prozesse optimieren und Kosten sparen. Den Vorteilen stehen jedoch auch Risiken gegenüber. Produktionsanlagen mit KI-Steuerung können von außen sabotiert, Daten gestohlen oder bestimmte Verbraucherinnen- oder Verbrauchergruppen diskriminiert werden. Den Industriebetrieben können nicht nur

finanzielle Schäden entstehen. Auch die Reputation des Unternehmens insgesamt kann gefährdet werden. Industriebetriebe haben deshalb ein Interesse an qualitativ hochwertigen und sicheren KI-Systemen.

Mit dem Vorschlag des „Artificial Intelligence Act (AIA)“ will die Europäische Kommission Qualität und Sicherheit der KI-Systeme verbessern. Für bestimmte Risikoklassen sieht der Vorschlag auch die Konformitätsbewertung für KI-Systeme vor. Die öffentlichen Stellungnahmen der Wirtschafts- und Industrieverbände zeigen, dass zwar einerseits ein Rechtsrahmen aus Industriesicht erforderlich und sinnvoll ist. Andererseits machen die Stellungnahmen eine Reihe von Kritikpunkten deutlich. Bei der beabsichtigten Pflicht zur Konformitätsbewertung befürchten die Industrieverbände etwa, dass das Zusammenspiel aus bestehenden Regeln (zum Beispiel dem Produktsicherheitsgesetz) und dem neuen AIA zu Mehrfach-Regulierung oder –Zertifizierung führt. Ebenso erwarten Unternehmen hohe Kosten für die Einstellung von Experten, die notwendig sind, um die Kriterien der Regulierung einzuhalten und Kosten für Auditoren, um die Zertifizierung zu erhalten. Diese Kosten sind insbesondere für Start-ups und KMU laut den Stellungnahmen der Industrieverbände kaum tragbar.

Ein weiterer Kritikpunkt zum AIA aus Sicht der Industrie sind die bisher wenig ausformulierten Kriterien für KI-Systeme, die sich einer Konformitätsbewertung unterziehen müssen.

Mit dem Prüfkatalog des vom Land Nordrhein-Westfalen gefördert Projekts „Zertifizierte KI“ liegt nun erstmalig ein Leitfaden vor, der Unternehmen bei der Konformitätsbewertung unterstützt. Der Katalog zeigt Prüfkriterien für die Dimensionen Fairness, Autonomie und Kontrolle, Transparenz, Verlässlichkeit, Sicherheit und Datenschutz. Er zeigt für jede der Dimensionen, welche umfangreichen Dokumentation, Erklärungen und Begründungen für die Konformität mit dem AIA notwendig sind.

Wie schon beim AIA stellt sich aber auch hier die Frage nach den Kosten und der Dauer der Prüfung. Zudem bleibt unklar, ob die Prüfunternehmen überhaupt in der Lage sein werden, Auditoren mit ausreichender KI-Kompetenz einzustellen, insbesondere weil sie dabei im Wettbewerb zu den zu zertifizierenden Unternehmen stehen. Ein weiteres Risiko stellt die drohende Zersplitterung der Landschaft für KI-bezogene Ethiknormen und Prüfstandards dar. Ohne eine Harmonisierung kann es im schlimmsten Fall sogar zu einem Ausweichen auf weniger anspruchsvolle KI-Zertifizierungen („ethics washing“) kommen. Der digitalen Souveränität des KI-Einsatzes in Unternehmen wäre damit wenig gedient.

Literatur

Blind, K. Heß, P.: Deutsches Normungspanel. Indikatorenbericht. Berlin, Deutsches Institut für Normung e.V. (2020)

Bundesamt für Sicherheit in der Informationstechnik (Hrsg.): AI Cloud Service Compliance Criteria Catalogue (AIC4). Bonn. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/CloudComputing/AIC4/AI-Cloud-Service-Compliance-Criteria-Catalogue_AIC4.html (2021). Zugegriffen: 18. Aug. 2021

DIN/DKE: Normungsroadmap Künstliche Intelligenz. Berlin, Wolfgang Wahlster Christoph Winterhalter. (2020)

Frost & Sullivan: Artificial Intelligence in the factory floor „Künstliche Intelligenz in der Fertigung“. Frost & Sullivan. (2018)
Hochrangige Expertengruppe für künstliche Intelligenz: Ethik-Leitlinien für eine vertrauenswürdige KI. (Hrsg.) v. Europäische Kommission. Brüssel. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (2019). Zugegriffen: 18. Aug. 2021
Mangelsdorf, A., Gabriel, P., Weimer, M.: Die Zertifizierung von KI: Mehr Sicherheit für alle – oder unnötiger Ballast? Institut für Innovation und Technik. iit perspektive, Bd. 58. Berlin. https://www.iit-berlin.de/publikation/die-zertifizierung-von-ki-mehr-sicherheit-fuer-alle-oder-unnoetiger-ballast/ (2021). Zugegriffen: 16. Aug. 2021
Matus, K.J.M., Veale, M.: Certification Systems for Machine Learning: Lessons from Sustainability. *Regulation & Governance* (2021)
Müller-Quade, J., Meister, G., Holz, T., Houdeau, D.: Künstliche Intelligenz und IT-Sicherheit – Bestandsaufnahme und Lösungsansätze. Whitepaper aus der Plattform Lernende Systeme, München (2019)
Poretschkin, M., Schmitz, A., Akila, M.: Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz. www.iais.fraunhofer.de/ki-pruefkatalog (2021)
ScienceBusiness: Time to harmonise artificial intelligence principles, experts say. https://sciencebusiness.net/news/time-harmonise-artificial-intelligence-principles-experts-say (2021)
Seifert, I., Bürger, M., Wangler, L., Christmann-Budian, S., Rohde, M., Gabriel, P., Zinke, G.: Potenziale der Künstlichen Intelligenz im produzierenden Gewerbe in Deutschland. Studie im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi) im Rahmen der Begleitforschung zum Technologieprogramm PAiCE – Platforms | Additive Manufacturing | Imaging | Communication | Engineering. Institut für Innovation und Technik. Berlin. https://www.iit-berlin.de/publikation/potenziale-der-kuenstlichen-intelligenz-im-produzierenden-gewerbe-in-deutschland/ (2018). Zugegriffen: 18. Aug. 2021.
VDE: Hessische Ministerin für Digitale Strategie und Entwicklung und VDE planen Aufbau eines "AI Quality & Testing Hubs." https://www.vde.com/de/presse/pressemitteilungen/ai-quality-testing-hub (2021)
WEF: The global risks report 2021 16. Aufl. (2021)
WEF: The Global Risks Report 2021. Geneva, World Economic Forum. (2021)