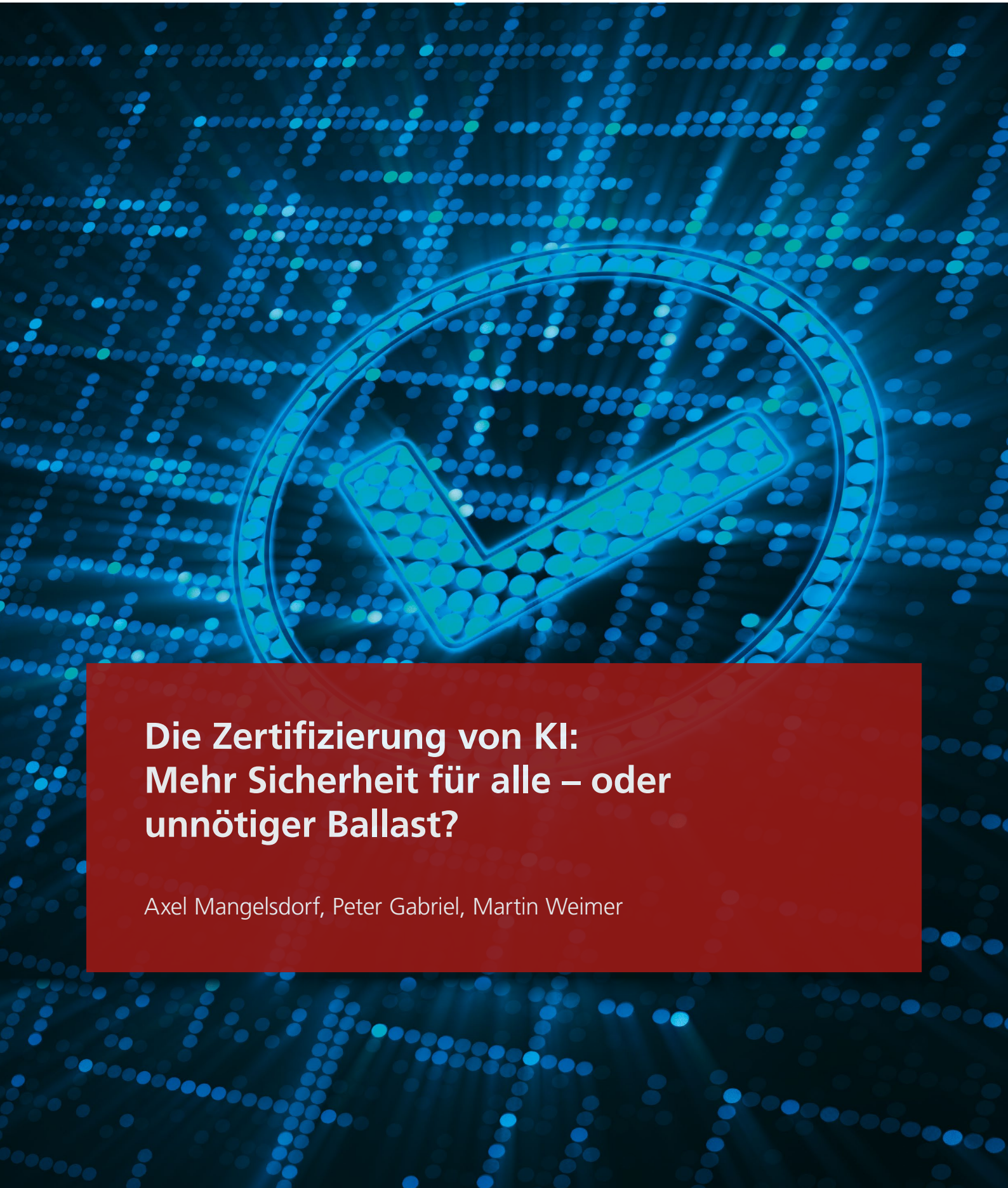
iit perspektive

Working Paper of the Institute for Innovation and Technology | Nr. 58



Die Zertifizierung von KI: Mehr Sicherheit für alle – oder unnötiger Ballast?

Axel Mangelsdorf, Peter Gabriel, Martin Weimer

iit perspektive

Working Paper des Instituts für Innovation und Technik in der VDI/VDE-IT | Nr. 58

Axel Mangelsdorf, Peter Gabriel, Martin Weimer

Die Zertifizierung von KI: Mehr Sicherheit für alle – oder unnötiger Ballast?

1 Einleitung

Systeme der Künstlichen Intelligenz (KI) beeinflussen immer mehr Lebensbereiche. Besonders im Fokus steht heutzutage das sogenannte maschinelle Lernen, bei dem mit vorhandenen Standardalgorithmen aus Trainingsdaten Modelle aufgebaut werden, die dann auf neue Eingangsdaten angewendet werden. Beispiele sind die Vorsortierung von potenziell besonders leistungsfähigen neuen Mitarbeitenden anhand einer Analyse ihrer Bewerbungsunterlagen durch die Personalabteilung oder die automatische Erkennung der Straße und Verkehrssituation durch ein teilautonomes Fahrassistenzsystem im Auto.

Zwei Besonderheiten zeichnen solche Verfahren des maschinellen Lernens vor allem aus:

- Die Verfahren sind häufig so vielschichtig und verflochten, dass ihre Entscheidungsfindung im konkreten Modell auch für Fachleute nicht mehr im Detail nachzuvollziehen ist. Das gilt insbesondere für die in vielen Anwendungsfällen sehr leistungsstarken neuronalen Netze.

- Das Ergebnis der Modelle hängt entscheidend von der Repräsentativität der Trainingsdaten ab. Die Modelle reproduzieren Muster und Strukturen aus einer limitierten Datengrundlage, auch solche, die den Entwickler:innen nicht bewusst sind. Ein bekanntes Beispiel ist das KI-Modell, das Pferde in Tierbildern nicht anhand ihrer Form, sondern an einem Wasserzeichen in den Pferdefotos erkannte.

KI-basierte Softwaresysteme, die Prognosen abgeben, Vorschläge für Entscheidungen machen oder sogar selbst Entscheidungen treffen, setzen ein hohes Maß an Vertrauen voraus. Anwender:innen und Verbraucher:innen erwarten, dass solche Systeme sicher und fair sind: Die persönlichen Daten sollen geschützt sein, und es soll keine Benachteiligung einer Person aufgrund von Alter, Geschlecht, Wohnort oder anderer Merkmale stattfinden. Für Unternehmen, die KI-basierte Produkte oder Dienstleistungen anbieten, ist diese Akzeptanz von hohem Wert.

Eine Möglichkeit, das Vertrauen in Sicherheit, Zuverlässigkeit und Fairness von KI-Anwendungen zu verbessern, ist die Prüfung durch unabhängige Dritte im Rahmen von Zertifizierungen. Zertifizierte KI-Systeme sollen nicht nur negative Konsequenzen verringern und eine faire Anwendung wahrscheinlicher machen, sondern überdies das Vertrauen in KI-Technologie allgemein steigern und damit zum Allgemeinwohl beitragen. Noch gibt es in Deutschland und der Europäischen Union keinen angemessenen Regulierungsrahmen und keine anerkannten Zertifizierungsverfahren für KI-Anwendungen. Eine Reihe von Organisationen und national geförderte Vorhaben entwickeln derzeit jedoch Prüfkriterien, Normen und Standards mit dem Ziel, KI-Systeme zu prüfen und zu zertifizieren.

Parallel hat sich eine sehr aktive Debatte um die Vor- und Nachteile dieses Ansatzes entwickelt: Eine KI-Zertifizierung, die durch akkreditierte Prüfer:innen nach einem standardisierten Verfahren vorgenommen wird, würde dem KI-System die Einhaltung rechtlicher Vorgaben, die technische Zuverlässigkeit und eine ethisch einwandfreie Anwendung bescheinigen und damit das Vertrauen erhöhen. Mit der unabhängigen Zertifizierung und der Etablierung vertrauenswürdiger Qualitätssiegel könnten einerseits, so die Befürworter:innen, internationale Wettbewerbsvorteile realisiert und die Interoperabilität verschiedener Systeme gefördert werden. Auf der anderen Seite könnten zu strenge regulatorische Vorgaben oder Prüfnormen, kostenintensive Prüfungen oder eine uneinheitliche europäische Zertifizierungslandschaft eine hemmende Wirkung für KI-Innovationen darstellen, insbesondere für Start-ups und kleine Unternehmen.

Mit dieser iit-Perspektive möchten die Autoren ein Schlaglicht auf die aktuelle Debatte werfen. Nach einer kurzen Einführung in die gängigen, für manche Leser und Leserinnen aber vermut-

lich nur ansatzweise bekannten Verfahren der Zertifizierung, werden kurz die aktuellen Initiativen für die KI-Zertifizierung mit Akteur:innen und jeweiligem Entwicklungsstand vorgestellt. Danach positionieren sich die Autoren pointiert mit einem Pro- bzw. einem Kontra-Beitrag. Sie verzichten auf eine abschließende Bewertung des Nutzens der Zertifizierung von KI. Dafür ist es angesichts der lebendigen Debatte noch viel zu früh. Und auch die Autoren behalten sich das Recht vor, ihre Position zu überdenken.

2 Die (oft) unbekannte Welt der Zertifizierung

Standards und Normen geben für viele Lebenslagen Regeln vor. Das gilt auch für die Informationstechnik, zu der die Künstliche Intelligenz gehört. Standards, wie etwa Datenformate für Office-Dokumente, können frei von Industriegremien oder einzelnen Unternehmen definiert werden. Normen, z. B. die ISO/IEC 27001 für das Informationssicherheitsmanagement, die von der International Organization for Standardization (ISO) und der International Electrotechnical Commission (IEC) herausgeben wird, werden dagegen in nationalen und internationalen Normungsgremien im Konsens und nach dem Stand von Technik und Wissenschaft erstellt. Wichtige Normungseinrichtungen in Deutschland sind das DIN Deutsche Institut für Normung und die Deutsche Kommission für Elektrotechnik (DKE). Normen sind zunächst unverbindlich. Sie haben jedoch oft eine sehr hohe Bedeutung, weil entweder in manchen Fällen der Gesetzgeber sie verpflichtend macht oder weil – was häufiger vorkommt – öffentliche, gewerbliche oder private Auftraggeber von ihren Dienstleistern und Zulieferern die Erfüllung der Normen erwarten. Dann wird eine Konformitätserklärung notwendig, die bescheinigt, dass ein Produkt, eine Dienstleistung, ein Prozess, ein technisches System oder eine Person einer Norm genügt.

Die Überprüfung der Anforderungen kann in Eigenregie durch das herstellende Unternehmen selbst („Selbsterklärung") oder durch einen Vertragspartner geschehen. Die höchste Stufe der Konformitätsbewertung ist die Zertifizierung, d. h. die Überprüfung von Normen (oder Standards) durch eine unabhängige dritte Partei. Die dritte Partei ist dabei meist eine akkreditierte Zertifizierungsstelle, z. B. eine der TÜV-Gesellschaften, die Dekra oder öffentliche Organisationen, wie die Bundesnetzagentur oder das Bundesamt für Sicherheit in der Informationstechnik. Damit sichergestellt wird, dass eine Zertifizierungsstelle über die notwendige Kompetenz verfügt, bestimmte Konformitätsbewertungen vorzunehmen, erfolgt die Überprüfung durch eine Akkreditierungsstelle wie die Deutsche Akkreditierungsstelle (DAkkS).

Die Zertifizierung nach international anerkannten Normen ist in Deutschland sehr verbreitet. Dabei sind nicht nur immer mehr Unternehmen zertifiziert, sondern auch die Anwendungsbereiche erweitern sich ständig auf neue Gebiete wie Fair-Trade-Handel oder ökologische Nachhaltigkeit. Bei der Anwendung von Normen zeigen sich deutliche Unterschiede zwischen großen und kleinen Unternehmen. So kommen etwa die sehr weit verbreiteten Normen ISO 9001 (Qualitätsmanagementsystem), ISO 14001 (Umweltmanagementsystem) und die bereits erwähnte ISO/IEC 27001 (Informationssicherheitsmanagementsystem) bei kleinen und mittleren Unternehmen seltener zum Einsatz als bei Großunternehmen (siehe Abb. 1). Die deutliche Zurückhaltung erklärt sich zum einen durch unterschiedliche Erwartung an den Nutzen für die Unternehmen, der durch interne Verbesserungen und verbesserte externe Beziehungen zu Auftragnehmern entsteht. Zum anderen bedeutet die Zertifizierung für kleinere Unternehmen auch im Verhältnis höhere Kosten durch Auditierung und Prüfung von Dritten. Kleine Unternehmen entscheiden sich nur dann für eine Zertifizierung, wenn der Nutzen die Kosten übersteigt.

3 Die aktuelle Debatte um die KI-Zertifizierung

Unternehmen, Forschungsorganisationen, Verbände und politische Akteure können bei der Erstellung von Standards und Normen mitwirken und eigene Zertifizierungssysteme erstellen (Heesen et al. 2020). In den vergangenen Jahren haben daher immer mehr Organisationen Leitlinien für Daten, ethische Gütesiegel oder ethische Leitlinien zur Anwendung von KI-Systemen entwickelt. Dabei ist bemerkenswert, dass eine Diskussion hinsichtlich einer KI-Zertifizierung durch Dritte außerhalb der Europäischen Union kaum stattfindet. Multinationale Unternehmen, z. B. Amazon, Apple, Baidu, Facebook, Google, IBM und Intel, haben im Rahmen der „Partnership on AI" zwar Richtlinien erstellt, die jedoch keine Zertifizierung durch Dritte vorsehen (Partnership on AI 2021).

Analog zur Debatte um den Datenschutz nimmt Europa auch in der Diskussion zur KI-Zertifizierung international eine führende Rolle ein. In Dänemark müssen seit Juni 2020 Unternehmen im Online-Bereich Informationen über ihre Richtlinien zur Datenethik veröffentlichen und auf Anfrage Informationen zu Algorithmen bereitstellen, die auf Plattformen verwendet werden (Danish Expert Group on Data Ethics 2019). Als weltweit erstes Land hat Malta bereits ein freiwilliges KI-Zertifizierungssystem eingeführt. Die Zertifizierung in Malta wird von unabhängigen Prüfer:innen vorgenommen und soll sicherstellen, dass das KI-System transparent, sozial verantwortlich und unter Berücksichtigung ethischer Leitlinien entwickelt wurde (Malta AI Taskforce 2019). Um eine Zersplitterung des Europäischen Binnenmarktes zu verhindern und mehr Vertrauen in die Siegel durch Drittzertifizierung zu erhalten, ist die Europäische Kommission sehr bemüht, einheitliche Regelungen zu finden.

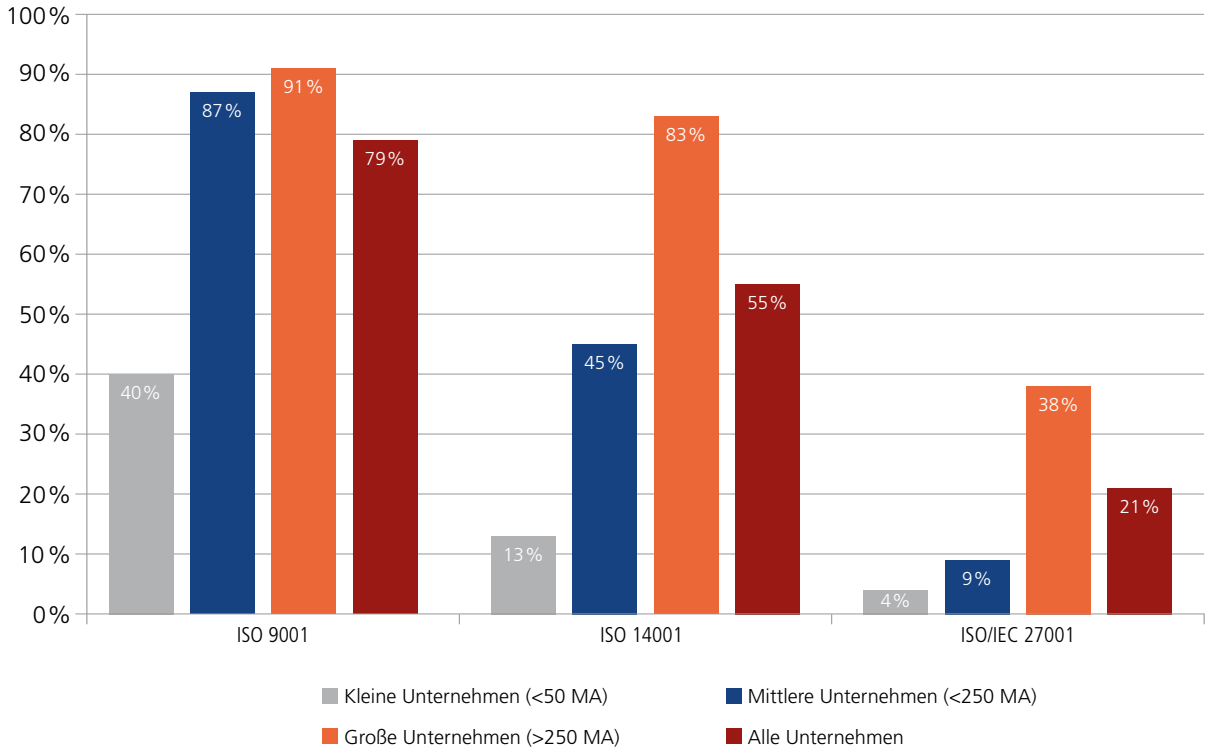


Abbildung 1: Anteil zertifizierter Unternehmen in Deutschland (eigene Darstellung nach Blind, Heß 2020, Deutsches Normungspanel. Indikatorenbericht 2020)

Datenethikkommission der Bundesregierung

Im Jahr 2018 hat die Bundesregierung eine Datenethikkommission eingesetzt, die einen ethisch fundierten „Entwicklungsrahmen für Datenpolitik, den Umgang mit Algorithmen, künstlicher Intelligenz und digitalen Innovationen" erarbeiten sollte. Im Mittelpunkt des Auftrags standen der Umgang mit „algorithmenbasierten Prognose- und Entscheidungsprozessen", KI-Verfahren und digitalen Daten (BMI, BMJV 2018). Die Kommission hat im Oktober 2019 ihr Gutachten vorgelegt (Datenethikkommission 2019). Die Kommissionsmitglieder unterscheiden bei ihren Betrachtungen und Empfehlungen nicht zwischen Software im allgemeinen und KI im Besonderen, sondern sprechen allgemein von „algorithmischen Systemen". Für diese Softwaresysteme wird eine risikobasierte Kritikalitätspyramide vorgeschlagen, die sich in fünf Stufen gliedert: von Anwendungen mit keinem oder geringem Schädigungspotenzial bis hin zu Anwendungen mit unvertretbarem Schädigungspotenzial. Bei Anwendungen ab der Stufe 3 („regelmäßiges oder deutliches Schadenspotenzial") empfiehlt die Kommission, auch Zulassungsverfahren vor Inbetriebnahme der Software in Betracht zu ziehen. Ob die Zulassung durch eine Behörde oder durch eine Zertifizierungsstelle nach einer (noch zu erstellenden) Norm geschehen soll, wird nicht näher thematisiert.

Europäische Kommission

Die EU-Kommissionn hat mit dem Weißbuch „Künstliche Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen (2020)" und mit dem Gesetzesvorschlag für die KI-Regulierung (April 2021) unter anderem auch einen Ansatz zur Konformitätsbewertung von KI-Anwendungen vorgeschlagen. Eindeutiger als die Datenethikkommission tritt die EU-Kommission damit für ein Zertifizierungsverfahren ein. Im Mittelpunkt steht die Unterscheidung zwischen verschiedenen Risikoklassen für KI-Anwendungen. Die Kommission folgt damit prinzipiell dem risikobasierten Ansatz der deutschen Datenethikkommission, unterscheidet jedoch nur zwischen zwei Risikostufen. Die Zertifizierung durch Dritte sollte nach Meinung der Kommission für KI-Systeme mit „hohem Risiko" verpflichtend sein (Europäische Kommission 2020, 2021). Für weniger risikoreiche Anwendungen kann ein freiwilliges Gütezeichen verwendet werden. Die Bundesregierung begrüßt den Vorschlag der EU-Kommission, betont aber in einer Stellungnahme zum Weißbuch die Notwendigkeit einer weiteren Risikodifferenzierung, die sich am jeweiligen Schadenspotenzial einer KI-Anwendung orientiert. Inhaltlich soll sich die Konformitätsbewertung u. a. an den Kriterien technische Sicherheit, Privatsphäre, Transparenz und Nichtdiskriminierung ausrichten (BMJV 2020).

EU-Kommission und Bundesregierung erkennen beide an, dass durch die Kosten der Konformitätsbewertung vor allem für kleinere Unternehmen erhebliche Belastungen entstehen. Vorgeschlagen wird deshalb, kleine und mittlere Unternehmen (KMU) sowie Non-Profit-Organisationen zu unterstützen und Ausnahmen von der Verpflichtung etwa für die Forschung zu erlauben. Wie diese Unterstützung von KMU konkret gestaltet wird, ist zum jetzigen Zeitpunkt noch unklar.

Normen und Standards

Bisher existieren kaum Standards und Normen, die eine Grundlage für die Zertifizierung von KI-Systemen bilden können. Erste Versuche zur Vereinheitlichung finden sich in der Normung von Fachbegriffen, z. B. in der Norm ISO/IEC 2382:2015 (Information technology – Vocabulary). Hier werden IT-bezogene Definitionen zusammengefasst, die zuvor in mehr als 30 Normen verteilt waren.

Die Normungsroadmap KI von DIN und DKE schlägt vor, wie der Strategieplan für die Entwicklung von zertifizierbaren KI-Normen aussehen kann. Demnach sollen zuerst Normen für die technische Prüfung von KI-Algorithmen (Verlässlichkeit, Robustheit, Leistungsfähigkeit und funktionale Sicherheit) entwickelt werden. Im Anschluss sollten auch Normen für rechtliche und ethische Anforderungen erstellt werden. Die Autor:innen der Normungsroadmap KI erwarten jedoch, dass Normen und Standards vor allem für die technische Prüfung entwickelt werden können. Ethische und gesellschaftliche Fragestellungen sollten dagegen eher Gegenstand verpflichtender Regulierungen oder gesellschaftlicher Diskurse sein.

Für die Erprobung eines Zertifizierungssystems für KI-Anwendungen schlagen die Autor:innen ein nationales Umsetzungsprogramm vor. Das Umsetzungsprogramm soll auf der deutschen Prüfinfrastruktur mit seinen Bestandteilen Normung, Prüfdienstleistungen, Akkreditierung und Zertifizierung aufbauen. In Pilotprüfungen sollen Standards entwickelt und in die internationale Normung eingebracht werden. Ein Ziel dabei ist die Vorstellung, dass deutsche Stakeholder führend bei der Entwicklung international anerkannter KI-Zertifizierungsverfahren werden. Pilotprojekte sollen sicherstellen, dass die entwickelten Normen von der Wirtschaft breit akzeptiert werden.

Die Normungsroadmap KI schlägt vor, dass die Kooperation CERTIFIED AI des Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS und des Bundesamtes für Sicherheit in der Informationstechnik (BSI) die Basis für eine KI-Zertifizierung bildet. In das Projekt sollen die Forschungsaktivitäten des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI) eingehen. Existierende internationale Normen und Standards sowie deren Umsetzung durch das BSI im Bereich der IT-Sicherheit sollen ebenfalls mit in das Projekt aufgenommen werden (DIN, DKE 2020).

CERTIFIED AI

Die Kooperation CERTIFIED AI zwischen Fraunhofer IAIS und BSI im Rahmen der Kompetenzplattform Künstliche Intelligenz Nordrhein-Westfalen (KI.NRW) zielt darauf ab, technische Prüfverfahren für die Zertifizierung von Systemen der Künstlichen Intelligenz zu entwickeln. Zudem sollen Grundlagen und Anforderungen für die Überprüfung der zugesicherten Eigenschaften im Rahmen des Zertifizierungsprozesses entstehen. Die Entwicklungen sollen anschließend in Use Cases validiert werden. Ein breiter Beteiligungsprozess soll sicherstellen, dass die konkreten Bedarfe der beteiligten Unternehmen aus verschiedenen Branchen aufgenommen und die entwickelten Prüfverfahren in der Praxis angenommen werden.

Bisher ist aus dem Projekt ein Whitepaper entstanden, in dem die interdisziplinären Grundlagen für die Zertifizierung beschrieben werden. Sechs Handlungsfelder wurden definiert: Fairness, Transparenz, Autonomie und Kontrolle, Datenschutz sowie Sicherheit und Verlässlichkeit (Cremers et. al. 2019). Das BSI hat parallel dazu im Februar einen Prüfkatalog für die Überprüfung der Sicherheit von cloudbasierten KI-Diensten vorgestellt. Der AIC4 (Artificial Intelligence Cloud Services Compliance Criteria Catalogue) (BSI 2021) genannte Standard baut auf früheren Versionen auf und wurde laut einer Pressemeldung von Huawei „zusammen mit der Industrie entwickelt" (Huawei 2021). Die Einhaltung der im Standard versprochenen Eigenschaften soll in einem Auditprozess überprüft werden, d. h., der Standard wäre durch Dritte zertifizierbar.

Die im Standard genannten Kriterien orientieren sich an Sicherheit und Robustheit, Funktionalität und Performanz, Zuverlässigkeit, Datenmanagement, Erklärbarkeit und Bias. Dabei wird unter Sicherheit und Robustheit verstanden, die Manipulation eines KI-Systems durch externe Angriffe zu verhindern. Funktionalität und Performanz bedeutet nach AIC4 zu testen, ob die vorgegebenen Leistungsindikatoren des Algorithmus erfüllt werden. Zuverlässigkeit umfasst die Untersuchung möglicher Fehler und Ausfälle sowie die Etablierung von Prozessen für den kontinuierlichen Betrieb des KI-Systems, z. B. Fehlerbehandlung und Back-ups. Im Rahmen der Datenqualität erfordert der Standard das Vorhandensein von Richtlinien für die Datenverarbeitung. Mit Datenmanagement sind Kriterien gemeint, die zur strukturierten Erfassung und Übernahme von Trainingsdaten beitragen. Unter Erklärbarkeit werden Maßnahmen verstanden, die Entscheidungen der KI nachvollziehbar machen sollen und transparent machen, welche Elemente der KI nicht gänzlich erklärbar sind. Schließlich soll im Rahmen der Bias-Überprüfung gezeigt werden, dass die Trainingsdaten keine statistische Ver-

zerrung beinhalten, mit der je nach Anwendung einzelne Personengruppen diskriminiert würden.

Die Anforderungen an die Auditor:innen orientieren sich an denen für die Zertifizierung gemäß ISO 27001 bzw. ISO 27001 auf Basis IT-Grundschutz (BSI-Grundschutz) oder an denen eines CSA-Personenzertifikats. Besondere Kenntnisse von KI-Verfahren werden von den Auditor:innen nicht verlangt.

4 Pro: KI-Zertifizierung führt zu mehr Sicherheit und bietet einen Wettbewerbsvorsprung

Eine Zertifizierung von KI-Systemen führt zu Vorteilen für die Unternehmen, die Anwender:innen und der Gesellschaft im Allgemeinen. Deshalb ist es wenig überraschend, dass die Entwicklung einer Zertifizierung von KI-Systemen in den Strategien zur KI, z. B. der Europäischen Kommission, eine wichtige Rolle spielt.

Zertifizierung schafft Rechtssicherheit für Entwickler:innen und Anwender:innen

Für Unternehmen, die KI-Anwendungen bereitstellen, bedeutet eine gesetzlich vorgeschriebene KI-Zertifizierung den Zugang zum Markt, vergleichbar mit der bereits gesetzlich vorgeschriebenen Zertifizierung von Medizinprodukten. Im Fall von Rechtsstreitigkeiten, etwa bei Fragen zu Haftung und Entschädigung im Schadensfall, bietet die KI-Zertifizierung Rechtssicherheit für Anwender:innen und Unternehmen. Auf der einen Seite wird das Haftungsrisiko für die Unternehmen begrenzt, wenn sie mit der Zertifizierung nachweisen können, dass sie sich an den Stand der Technik gehalten haben. Die Anwender:innen der KI-Systeme profitieren auf der anderen Seite von der Einhaltung der Regeln für IT-Sicherheit, Datenschutz und Fairness, was durch ein externes Audit versichert wird.

Zertifizierung stellt Vertrauen bei den Nutzer:innen her und stärkt den Einsatz von KI durch deutsche sowie europäische Unternehmen

KI-Zertifizierungen können dazu beitragen, die Wahrscheinlichkeit fehlerhafter KI-Modelle zu begrenzen. Auf der gesamtgesellschaftlichen Ebene hilft die KI-Zertifizierung damit, das Nutzenpotenzial der KI-Technologie zu realisieren, indem sie auch den – gerade in Deutschland bei neuen Technologien oft skeptischen – Anwender:innen Vertrauen in die Qualität und Sicherheit der KI-Anwendungen ermöglicht. Durch erhöhtes Vertrauen in KI steigt die Wahrscheinlichkeit, dass diese Schlüsseltechnologie in immer mehr hiesigen Unternehmen Anwendung findet. So steigt auch die Wettbewerbsfähigkeit deutscher und europäischer Unternehmen.

Zertifizierung macht den Einsatz von KI auch in sicherheitskritischen Anwendungen möglich

Die sich bereits abzeichnenden Prüfkriterien eines KI-Zertifizierungssystems sind besonders sinnvoll für Anwendungen mit hohem Risiko. Es besteht zwar durchaus noch die Herausforderung, im Detail festzulegen, welche Anwendungen ein hohes Risiko aufweisen und wie diesem mit einem geeigneten Risikomanagement zu begegnen ist. Letztendlich ist das jedoch eine bekannte und schon gelöste Aufgabe. Für den Einsatz von Software in sicherheitskritischen Anwendungen, etwa in der Medizintechnik oder im Flugzeug- und Fahrzeugbau, gibt es bereits etablierte Verfahren für die Definition von Risikoklassen und Risikomanagement. Mit einer Zertifizierung gibt es die Möglichkeit, KI-Verfahren auch in sicherheitskritischen Anwendungen einzusetzen. Das schafft nicht nur Vertrauen bei den Nutzer:innen, sondern gibt den Entwickler:innen auch die heute noch oft vermissten Leitlinien für den Einsatz von KI vor.

Zertifizierung schafft Vorteile im Wettbewerb

KI-Zertifizierungen machen zudem verschiedene Angebote für gewerbliche und private Anwender:innen hinsichtlich der Güte ihrer KI-Modelle vergleichbar und steigern damit die Transparenz im Wettbewerb. Eine europäische oder deutsche KI-Zertifizierung („Made in Europe“ bzw. „Made in Germany“) kann ein Ausweis für qualitativ hochwertige KI-Anwendungen aus Europa bzw. Deutschland sein und stellt dann einen Vorteil im internationalen Wettbewerb mit China und den USA dar.

Dieser Weg, bewusst hohe Anforderungen an die eigenen IT-Produkte und -Systeme zu stellen und damit einen Wettbewerbsvorteil zu erzeugen, hat sich beim Datenschutz schon bewährt. Durch die Entwicklung von Standards wie des Kriterienkatalogs AIC4 sind erstmals von einer öffentlichen Stelle zertifizierbare Anforderungen entwickelt worden, die zu vertrauenswürdigen, sicheren und nachvollziehbaren KI-Systemen führen können. Nach dem Leitsatz „Wer die Norm macht, macht auch den Markt“ können sich deutsche und europäische KI-Anbieter eine Vorreiterposition im internationalen Wettbewerb schaffen.

5 Contra: KI-Zertifizierung leistet zu wenig und ist dafür zu teuer

Die aktuell diskutierten Konzepte für eine KI-Zertifizierung folgen hehren Zielen, sind aber wenig praxistauglich, zu teuer und verzerren den Wettbewerb.

Die Zertifizierungskriterien sind kaum oder gar nicht prüfbar

Die in den aktuellen Zertifizierungskonzepten genannten Anforderungen an KI-Systeme sind wohlbegründet und sinnvoll.

Sie lassen sich aber nur sehr schwer als allgemein gültige, für Dritte ohne viel Aufwand nachvollziehbare und effizient prüfbare Kriterien definieren. Schon die allgemeine Differenzierung von mehreren mittleren Risikoklassen zwischen nicht vertretbarem Schädigungspotenzial und keinem oder nur geringem Schädigungspotenzial dürfte kaum gelingen –außer in klar abgrenzbaren Anwendungsbereichen wie der Medizintechnik oder dem Fahrzeugbau.

Auch die wichtige Eigenschaft der diskriminierungsfreien Trainingsdaten kann nicht in ein von einem Auditor bzw. einer Auditorin prüfbares Kriterium gezwängt werden. Das Verständnis, welche Merkmale ein Datensatz aufweisen muss, um eine Diskriminierung einzelner Gruppen zu vermeiden, wird in einer gesellschaftlichen Debatte erreicht und unterliegt einem ständigen Wandel. Es wird auch Fälle geben, in denen überhaupt kein allgemeines Verständnis zustande kommt, ob Trainingsdaten frei von Diskriminierung sind oder nicht.

Für sicherheitskritische Systeme gibt es bereits funktionierende Zertifizierungsverfahren

Für Software in sicherheitskritischen Anwendungen, etwa in der Medizintechnik oder der Luft- und Raumfahrt, gibt es bereits etablierte Standards, Normen und Zertifizierungsverfahren. Der Einsatz, oder eben auch der Nichteinsatz, von KI-Algorithmen lässt sich anhand der Kriterien der etablierten Zertifizierungsverfahren entscheiden. Wenn die bestehenden Prüfkriterien für den Einsatz von KI-Modellen angepasst oder erweitert werden müssen, etwa um Gütemasse für Trainingsdaten aufzustellen, ist das für den einzelnen Anwendungsbereich deutlich leichter umsetzbar als anwendungsübergreifend für viele verschiedene Einsatzzwecke.

Eine Zertifizierung sollte technologieneutral sein

Die in der Debatte genannten Kriterien Verlässlichkeit, Robustheit, Leistungsfähigkeit, funktionale Sicherheit und Diskriminierungsfreiheit sollten prinzipiell für alle Softwaresysteme gelten. Es ist nicht einsehbar, warum nur für eine bestimmte Klasse von Algorithmen besondere Anforderungen aufgestellt werden. Auch mit herkömmlichen Programmiermethoden ohne den Einsatz von KI wurden in der Vergangenheit schon diskriminierende Scoring-Verfahren für die Prüfung der persönlichen Bonität umgesetzt. Die in manchen Anwendungsbereichen, wie etwa dem Personen-Scoring, durchaus sinnvolle Zertifizierung von Software muss von der Notwendigkeit einer Regulierung der konkreten Anwendung her gedacht werden – und nicht von der Art der eingesetzten Algorithmen. Zudem sind die Begriffe Künstliche Intelligenz und maschinelles Lernen nicht eindeutig definiert, auch weil sich das Feld zurzeit so dynamisch entwickelt. Die Abgrenzung, welche Software tatsächlich zertifiziert werden muss, wäre damit auf absehbare Zeit unklar.

Kosten und Nutzen stehen in keinem ausgeglichenen Verhältnis

Für Unternehmen bringt eine KI-Zertifizierung Kosten mit sich, wie andere Zertifizierungen auch. Diese entstehen durch die Prüfung und Auditierung. Neben den direkten monetären Kosten für Prüfung und Auditierung müssen die Unternehmen ebenfalls personelle und zeitliche Ressourcen einsetzen. Mit weniger personellen und zeitlichen Kapazitäten fehlen den Unternehmen Ressourcen für andere gewinnbringende Aktivitäten wie Marketing- oder Forschungs- und Entwicklungsaktivitäten; d. h., die KI-Zertifizierung bringt relevante Opportunitätskosten mit sich. Da anzunehmen ist, dass KI als Basistechnologie, ähnlich wie Datenbanksysteme oder Elektrizität, in immer mehr Produkten und Dienstleistungen Einzug halten wird, führen die Kosten für KI-Zertifizierung gleichfalls zu hohen volkswirtschaftlichen Kosten. Das gilt vor allem dann, wenn der Kreis der zu zertifizierenden KI-Anwendungen weit ausgelegt wird und deren Zertifizierung gegebenenfalls auch gesetzlich vorgeschrieben wird. Angesichts des sehr begrenzten Nutzens einer gesonderten Zertifizierung von KI-Anwendungen ist ihr Aufwand weder betriebswirtschaftlich noch volkswirtschaftlich gerechtfertigt.

KMU, vor allem IT-Start-ups, werden durch die Aufwände der Zertifizierung benachteiligt

Die negative Korrelation zwischen Zertifizierungshäufigkeit und Unternehmensgröße legt nahe, dass bei den meisten kleinen Unternehmen die Kosten der Zertifizierung den Nutzen übersteigt. Führt die geplante KI-Zertifizierung zu ähnlichen Effekten, werden eher größere Unternehmen die Zertifizierung in Anspruch nehmen. Kleinere Unternehmen sind dann im Einsatz innovativer KI-Methoden eingeschränkt, was ihre Wettbewerbsfähigkeit mindert. Gerade für einheimische IT-Start-ups mit ihren oft sehr begrenzten personellen und finanziellen Ressourcen kann das ein erheblicher Nachteil im internationalen Wettbewerb werden, der mittlerweile stark durch den Einsatz von KI-Verfahren geprägt ist.

Ein deutscher Sonderweg fragmentiert den Binnenmarkt

Ein Vorpreschen durch einen deutschen Alleingang bei der KI-Zertifizierung, so wie er sich jetzt gelegentlich unter dem Schlagwort „KI-Zertifizierung made in Germany“ andeutet (KI.NRW 2020), fragmentiert den EU-Binnenmarkt, weil dann in Deutschland strengere Regeln gelten als in den anderen Mitgliedsstaaten. Eine zu strenge KI-Zertifikation könnte Auswirkung auf Märkte für Produkte und Dienstleistungen mit einem hohen Anteil an KI-Software haben, wie etwa sprachbasierte Assistenzsysteme.

6 Ein Zwischenfazit

Die Autoren können kein endgültiges Fazit ziehen, ob sich eine KI-Zertifizierung lohnt oder nicht. Dafür ist es bei weitem noch zu früh. Ohne ausgearbeitete Vorschläge für Risikoklassen, Prüfkriterien und Zertifizierungsverfahren und deren Test in Pilotprojekten kann es derzeit keine fundierte Entscheidung geben. Gerade deswegen haben Praxisprojekte wie CERTIFIED AI eine hohe Bedeutung: Sie bringen die Debatte vom notwendigen, aber abstrakten Niveau der White Paper und Ethikkommissionen auf die konkrete, alltägliche Ebene der Konformitätsprüfung von Standards und Normen.

Ob es eine KI-Zertifizierung geben wird und wie diese im Detail aussieht, weiß momentan niemand. Allein die Debatte um die Zertifizierung hilft jedoch schon, besser zu verstehen, wie KI-Algorithmen gleichzeitig nutzbringend und verantwortungsvoll eingesetzt werden können.

Dank

Die Autoren danken ihren Kolleg:innen Lene Ganschow, Tom Kraus, Marieke Rohde, Inessa Seifert und Nicole Wittenbrink aus dem iit für die gemeinsame Diskussion, die zur Entstehung dieser iit-Perspektive geführt hat, und für die wertvolle Kommentierung der ersten Fassung.

Literaturverzeichnis

Bundesamt für Sicherheit in der Informationstechnik (2021): AI Cloud Service Compliance Criteria Catalogue (AIC4), Bonn, www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/CloudComputing/AIC4/AI-Cloud-Service-Compliance-Criteria-Catalogue_AIC4.pdf;jsessionid=04767F9267F9743AD51A6D8BD9EBF88A.internet082?__blob=publicationFile&v=4 [09.04.2021]

Bundesministerium der Justiz und für Verbraucherschutz (2020): Bundesregierung übermittelt Stellungnahme zum Weißbuch für Künstliche Intelligenz an die Europäische Kommission. www.bmjv.de/SharedDocs/Artikel/DE/2020/062920_BReg_Weissbuch_KI.html [09.04.2021]

Bundesministerium des Innern, für Bau und Heimat, Bundesministerium der Justiz und für Verbraucherschutz (2018): Leitfragen der Bundesregierung an die Datenethikkommission. https://datenethikkommission.de/arbeitsauftrag-und-leitfragen [09.04.2021]

Cremers, Armin B.; Englander, Alex; Gabriel, Markus et. al. (2019): Vertrauenswürdiger Einsatz von künstlicher Intelligenz. Handlungsfelder aus philosophischer, ethischer, rechtlicher und technologischer Sicht als Grundlage für eine Zertifizierung von künstlicher Intelligenz, Sankt Augustin: Fraunhofer IAIS, www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_KI-Zertifizierung.pdf [09.04.2021]

Danish Expert Group on Data Ethics (2019): Denmark decides on ethical principles for businesses working with AI, https://investindk.com/insights/denmark-decides-on-ethical-principles-for-ai [09.04.2021]

Datenethikkommission der Bundesregierung (2019): Gutachten der Datenethikkommission, Berlin https://datenethikkommission.de/gutachten [09.04.2021]

Deutsches Institut für Normung, Deutsche Kommission für Elektrotechnik (2020): Deutsche Normungsroadmap Künstliche Intelligenz, Berlin, Frankfurt am Main, https://www.din.de/resource/blob/772438/6b5ac6680543eff9fe372603514be3e6/normungsroadmap-ki-data.pdf [09.04.2021]

Europäische Kommission (2020): Weißbuch zur Künstlichen Intelligenz – Ein europäisches Konzept für Exzellenz und Vertrauen, Brüssel. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_de.pdf [09.04.2021]

Europäische Kommission (2021): Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts, Brüssel. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence [29.04.2021]

Heesen, Jessica; Müller-Quade, Jörn; Wrobel, Stefan et al. (2020): Zertifizierung von KI-Systemen – Kompass für die Entwicklung und Anwendung vertrauenswürdiger KI-Systeme. Whitepaper aus der Plattform Lernende Systeme, München: Acatech, www.plattform-lernende-systeme.de/publikationen-details/zertifizierung-von-ki-systemen-kompass-fuer-die-entwicklung-und-anwendung-vertrauenswuerdiger-ki-systeme.html [09.04.2021]

Huawei (2021): AIC4-Katalog des BSI: Huawei unterstützt Vertrauenswürdigkeitsprüfung von KI-Diensten. www.huawei.com/de/news/de/2021/huawei-unterstuetzt-vertrauenswuerdigkeits-pruefung-von-ki-diensten [09.04.2021]

KI.NRW (2020): KI-Zertifizierung „made in Germany", www.ki.nrw/flagships/zertifizierung [09.04.2021]

Malta AI Taskforce (2019): Malta The Ultimate AI Launchpad – A Strategy and Vision for Artificial Intelligence in Malta 2030. www.ai4eu.eu/news/malta-ultimate-ai-launchpad [09.04.2021]

Partnership on AI (2021): Research, Publications & Initiatives, www.partnershiponai.org/research-lander [09.04.2021]

Herausgeber

Prof. Dr. Volker Wittpahl
Institut für Innovation und Technik (iit)
in der VDI/VDE Innovation + Technik GmbH
Steinplatz 1
10623 Berlin

Kontakt

Dr. Axel Mangelsdorf
Tel: +49 (0)30 310078-5559
E-Mail: Mangelsdorf@iit-berlin.de

Peter Gabriel
Tel: +49 (0)30 310078-206
E-Mail: Gabriel@iit-berlin.de

Dr. Martin Weimer
Tel: +49 (0)89 5108963-015
E-Mail: Weimer@iit-berlin.de

iit perspektive Nr. 58

April 2021
Layout: Poli Quintana

ISBN: 978-3-89-750232-1

Bildnachweise:
Adobe Stock/Sikov (Titelbild)

ISBN: 978-3-89-750232-1