PTB

Physikalisch-Technische Bundesanstalt
Nationales Metrologieinstitut

KI-Strategie der PTB

Executive Summary

Der zunehmende Einsatz von Verfahren künstlicher Intelligenz (KI) revolutioniert die Wertschöpfung aus (Mess-)Daten, eröffnet dabei gänzlich neue Geschäftsfelder und verändert praktisch sämtliche Lebens- und Wirtschaftsbereiche. In Smart Homes und Smart Cities ermöglichen intelligente Zähler und Controller eine bedarfszentrierte Steuerung und effiziente Abrechnung der Energie- und Wasserversorgung sowie eine Optimierung der Netzauslastung. „Predictive Maintainance“, d. h. vorausschauende Instandhaltung mittels KI, reduziert in der Industrie 4.0 Produktionsausfälle und Wartungsaufwendungen um ein Vielfaches. Und auch im Gesundheitssektor verbessern KI-gestützte Diagnose und Therapieplanungen die Behandlung der Patientinnen und Patienten und vermindern somit maßgeblich Ausfallzeiten und vermeidbare Belastungen des Gesundheitssystems. Gerade aus der Kombination aus breit eingesetzter Messtechnik und Verfahren der künstlichen Intelligenz entsteht also ein enormer wirtschaftlicher und gesellschaftlicher Mehrwert.

Möglich wird dieser Vormarsch der Schlüsseltechnologie KI aufgrund der fortschreitenden Digitalisierung nahezu all unserer Prozesse im industriellen und im zivilen Bereich sowie der steigenden Verfügbarkeit der damit verbundenen Daten. Sowohl die Digitalisierung als auch der zunehmende Einsatz von KI schaffen neue Potentiale für den Markt und gestalten den Umgang mit Produkten und Dienstleistungen grundlegend neu. Um die Vorteile von KI-Anwendungen in die Breite der Gesellschaft zu tragen und die großen wirtschaftlichen Potentiale dieser Technologie auszuschöpfen, ist es unabdingbar, gerechtfertigtes Vertrauen der Nutzerinnen und Nutzer in die Funktionsweise und die Ergebnisse der Technologie aufzubauen und deren Sicherheit im Umgang mit KI zu gewährleisten.

Als nationales Metrologieinstitut und oberste Instanz für das Messen sowie die einhergehenden Messdaten versteht es die Physikalisch-Technische Bundesanstalt (PTB) als ihren Auftrag, sich dieser wichtigen Aufgabe im Zusammenspiel mit den anderen Akteuren der Qualitätsinfrastruktur (QI) aktiv zu widmen. Handlungsfelder für die Metrologie bestehen insbesondere bei der Bewertung von KI-Systemen und der zugrundeliegenden Daten, sozusagen dem „Messen“ der KI- und der Daten-Qualität. Auf Grundlage hierfür geeigneter Metriken lassen sich Leitlinien für Standardisierung und Zertifizierung von KI-Systemen ableiten, welche einen vertrauenswürdigen KI-Einsatz ermöglichen. Auch hierbei setzt es sich die PTB zum Ziel, ihre messtechnische Expertise proaktiv in die Gestaltung des regulatorischen Ordnungsrahmens für KI einzubringen. Zudem gilt es mit Blick auf die Schlüsseltechnologie KI für die PTB, bestehende metrologische Prüf- und Bewertungsverfahren auf ihre Tauglichkeit für Produkte und Dienstleistungen mit KI-Anteil hin neu zu bewerten und wo nötig zu überarbeiten.

Im Zuge des fortschreitenden Einsatzes von KI-Verfahren erkennt die PTB einen steigenden Bedarf für die Bereitstellung qualitätsgesicherter, maschinennutzbarer Daten. Vergleichbar zu den international abgestimmten Normalen für physische Größen, wie z. B. Ur-Meter und Ur-Kilogramm, will die PTB als wesentlicher Vertrauensanker für Zukunftstechnologien in der Messtechnik auch für die digitale Welt Normale (z. B. „Goldstandards“ oder Referenzdatensätze) und Benchmarks entwickeln und diese über eine geeignete Infrastruktur der Wissenschaft, Wirtschaft und Gesellschaft bereitstellen. Diese digitalen Normale eröffnen dabei einerseits gänzlich neue Geschäftsfelder für die PTB und bilden andererseits das Rückgrat für wettbewerbsfähige technologische Innovationen bei Kundinnen und Kunden. Es ist das Bestreben der PTB, diese nationalen digitalen Normale international mit den 102 Mitglieds- und assoziierten Staaten der Meterkonvention sowie mit den Organisationen der internationalen Qualitätsinfrastruktur in führender Position abzustimmen.

Um den Transfer wissenschaftlicher Ergebnisse zu KI in die Anwendung zu stärken und damit auch bei der metrologischen Forschung und Dienstleistung stets auf dem neusten Stand der Technik zu agieren, pilotiert die PTB selbst bereits einige KI-Anwendungen, u. a. bei der Optimierung von

Datenanalyseverfahren, der Prozessautomatisierung, bei der Bildrekonstruktion und für indirekte Messungen. Diese KI-Methoden finden Anwendung in verschiedensten Fachabteilungen und werden an der PTB in geeigneten Prozessen sukzessive weiter etabliert. Die Erprobung des sicheren Einsatzes von KI in der PTB sorgt dafür, die Bandbreite metrologischer, wissenschaftlich-technischer Anwendungsfelder zu erweitern, KI-Kompetenz für Forschung, Dienstleistung und Verwaltung aufzubauen und zugleich bestehende Prozessabläufe zu optimieren.

Durch die Kombination aus metrologischem Domänenwissen, Daten- und KI-Kompetenz entsteht eine fundierte Expertise für die Herausforderungen der Produkte und Dienstleistungen der Zukunft. Diese Expertise bildet dabei das Alleinstellungsmerkmal der PTB innerhalb der KI-Forschungslandschaft. Mit ihrem sie auszeichnenden fundierten Verständnis im Umgang mit Messdaten und darauf aufbauenden Daten-getriebenen Verfahren bringt die PTB entscheidendes Know-How in die Kooperation mit KI-Forschungseinrichtungen und anderen QI-Akteuren ein. Damit leistet sie einen wesentlichen Beitrag für die Entwicklung einer verlässlichen und vertrauensstiftenden Bewertung, Standardisierung und Zertifizierung von KI-Systemen und Daten. Die notwendige Kompetenzentwicklung und nachhaltige -absicherung an der PTB erfordert dabei eine planvolle Koordination von Maßnahmen zur Vernetzung, Personalgewinnung und -entwicklung. Flankiert werden diese Bestrebungen durch einen bedarfszentrierten Ausbau der benötigten Infrastruktur für Computing und Datenorganisation sowie Unterstützung in Fragen der technischen Kompetenz.

Zur Erreichung der strategischen Zielsetzung der PTB für Vertrauen in KI bedarf es eines entsprechenden politischen Rahmens, welcher durch folgende Maßnahmen bereitet werden sollte:

- Einrichtung designierter **Leuchtturm- & Pilotprogramme zur Grundlagen- und Anwendungsforschung für KI** (u. a. zur Erarbeitung geeigneter Metriken für eine Bewertung der Qualität von KI und der verwendeten Daten) unter **gezielter Einbindung messtechnischer Expertise**,
- Schaffung einer **Innovationsplattform** für die enge und effiziente Kooperation von KI-Forschungseinrichtungen, Unternehmen, QI-Akteuren und Regelsetzern (z. B. im Rahmen eines Innovationszentrums für systemische Metrologie),
- Aufstockung personeller **Ressourcen für Gremienarbeit und Forschungsaufgaben** sowie Förderung von **Aus- und Weiterbildungsangeboten** zum nachhaltigen Aufbau von KI-Kompetenzen,
- Ausbau und Betrieb entsprechender **Infrastrukturen** für KI-Forschung und -Anwendung an der PTB und
- explizite Verankerung der Zuständigkeit für messtechnische Produkte und Dienstleistungen mit KI im **gesetzlichen Auftrag der PTB**.

Für die Zukunft plant die PTB, die begonnenen KI-Aktivitäten engagiert weiterzuführen und sowohl im Forschungsbereich als auch im praktischen Einsatz deutlich auszubauen. Neben der Erarbeitung einer konkreten Umsetzungsplanung entlang der vorliegenden strategischen Leitplanken steckt sich die PTB zudem das Ziel, ihr Engagement und ihre Sichtbarkeit innerhalb der KI-Forschung und -Regulierung weiter zu steigern. Bei allen Akteuren der Qualitätsinfrastruktur, der Forschungslandschaft und der Wirtschaft möchte die PTB ihren Ruf als kompetente und proaktive Partnerin bei Fragen rund um die Vertrauenswürdigkeit und Verlässlichkeit auch im Bereich KI bestärken und fordert dabei auch eine explizitere Verantwortung für KI-Technologien innerhalb ihres gesetzlichen Auftrages.

Einleitung

Mit steigender Verfügbarkeit großer Datenmengen in allen Lebensbereichen und den enormen technologischen Fortschritten in der Messtechnik im Zuge der Digitalisierung nimmt auch der Einsatz von Methoden der künstlichen Intelligenz (KI) stetig zu. Die Schlüsseltechnologie KI revolutioniert das Produkt- und Dienstleistungsverständnis grundlegend [1, 2] und wirkt damit als Katalysator für digitale Innovationen. Nicht nur in der Industrie 4.0 lassen sich durch vorausschauende Instandhaltung (sogenannte „Predictive Maintenance“) von Maschinen und Anlagen mittels KI erhebliche Ressourcen einsparen. Auch bei der intelligenten Steuerung der Versorgungssysteme in Smart Homes und Smart Cities, bei selbstlernenden Diagnosetools für die personalisierte Medizin bis hin zum autonom fahrenden Fahrzeug eröffnen sich stetig neue Einsatzfelder für KI. Durch ihre Vielseitigkeit und inhärente Anpassungsfähigkeit an Problemstellungen aller Art, bieten KI-Systeme als Bestandteil von Produkten oder als eigenständige Artikel herausragende wirtschaftliche Potentiale, die – frühzeitig erkannt und nutzbar gemacht – die Stellung Deutschlands auf dem Weltmarkt entscheidend stärken und in der Breite, von Startups über KMU zu großen Konzernen, signifikante Wettbewerbsvorteile bedeuten können.

Parallel zum wachsenden Anwendungsbereich von KI steigt jedoch auch die Notwendigkeit für klare Regeln, die die einhergehenden Risiken des Einsatzes von KI insbesondere in kritischen Bereichen wie z. B. dem Gesundheits- oder Versorgungssektor ausräumen oder deren Folgen auf ein akzeptables Maß abmildern. Um das Vertrauen der Kund*innen und Nutzer*innen in die Schlüsseltechnologie KI nachhaltig zu festigen, ist eine stringente Qualitätsinfrastruktur (QI) auch für KI-Anwendungen unabdingbar. Metrologie ist ein anerkannter Vertrauensanker und wesentlicher Bestandteil der QI. Dazu gehört die Charakterisierung der Messtechnik und Messmethoden, die Bewertung der Qualität von Messdaten und die Entwicklung neuer Messverfahren. Grundsätzlich ist die gesetzliche Beauftragung der PTB im Rahmen des EinhZG (§ 6 Abs. 3), des MessEG (§ 45) und des Medizinproduktegesetzes (§ 32 Abs. 2) sehr technologieoffen formuliert. Insofern ist die PTB selbst auch kontinuierlich aufgefordert, ihre eigene Rolle im Sinne der gesetzlichen Beauftragung angesichts technologischer Entwicklungen, insbesondere derart disruptiver mit weitreichenden Konsequenzen der Anwendung wie KI, zu bewerten und zu hinterfragen. Gleichzeitig ist bei neuen technologischen Entwicklungen immer davon auszugehen, dass ein Erwartungsdruck gegenüber der PTB besteht, ihrem gesetzlichen Auftrag auch zukünftig in kompetenter Weise gerecht zu werden.

Die PTB versteht es damit als ihre Aufgabe, grundlegende Forschung zur Datenqualität und Verlässlichkeit von KI-Verfahren zu leisten und die Entwicklung der rechtlichen Rahmenbedingungen für Zulassung und Regelsetzung im Zusammenspiel mit anderen Akteuren der QI voranzubringen. Zudem möchte die PTB ebenso die durch den Einsatz von KI-Methoden entstehenden Chancen im Forschungs- und Entwicklungsumfeld ausbauen und sicher nutzbar machen. Mit diesem Vorgehen folgt die PTB dem erklärten Ziel der Bundesregierung in der „Fortschreibung der KI-Strategie“ [3]:

> *„Die Bundesregierung setzt sich deshalb für einen geeigneten, ggf. an KI-spezifischen Belangen angepassten Ordnungsrahmen ein, in dem die bestehende Qualitätsinfrastruktur ausgebaut und wenn nötig weiterentwickelt wird. Durch das Setzen klarer Regeln sowie Standards und Normen können die Grundrechte von Bürgerinnen und Bürgern geschützt, Vertrauen in die KI gestärkt, ein nachhaltiger Einsatz sowie Innovation und Wettbewerb gefördert werden.“*

Das hier vorgelegte Strategiepapier der PTB orientiert sich auch in der Struktur an der Fortschreibung der KI-Strategie der Bundesregierung und projiziert deren strategische Auslegung auf das Aufgabengebiet der Metrologie. Zudem ergänzt die aktuelle KI-Strategie der PTB die bestehende Strategie für „KI in der Medizin“, welche im Dezember 2020 veröffentlicht wurde und bereits als

integraler Bestandteil und Use Case wichtiger Initiativen wie z. B. „QI-Digital“ des BMWi etabliert und in der Umsetzung befindlich ist. Mit diesem Strategiepapier adressiert die PTB die Chancen und Risiken des Einsatzes von KI in der Breite der metrologischen Forschung und Anwendung, schärft ihr Verständnis zu den Herausforderungen von KI und klärt Handlungsbedarfe und Aktionsfelder für die Metrologie.

Status quo

Die disruptive Schlüsseltechnologie KI hat längst das Nischenstadium in der Forschung verlassen (siehe Abb. 1 zur historischen Entwicklung) und drängt in Form verschiedenster Produkte und Dienstleistungen auf den Markt und damit in sämtliche Lebens- und Wirtschaftsbereiche. Um entsprechende Leitplanken für diese dynamisch fortschreitende Entwicklung festzusetzen, veröffentlichte die EU-Kommission im Februar 2020 ein Weißbuch zur künstlichen Intelligenz [4], welches auf die europäische KI-Strategie von 2018 aufbaut und eine innovative, aber – in Abgrenzung von den Entwicklungen in den USA und China – unbedingt menschenzentrierte KI in den Fokus der weiteren Handlungen stellt. Dieses Leitbild bedeutet, dass die KI dem Menschen und der Gesellschaft nutzen und dabei ein selbstbestimmtes Handeln stärken sollte, und wird oft als „KI mit europäischer Prägung“ bezeichnet. Auch von Seiten der Bundesregierung wird die Thematik KI mit der KI-Strategie von 2018 [5] und deren Fortschreibung 2020 [3] sowie der Stellungnahme zum KI-Weißbuch der EU [6] hoch priorisiert und in das strategische Handeln eingebettet.

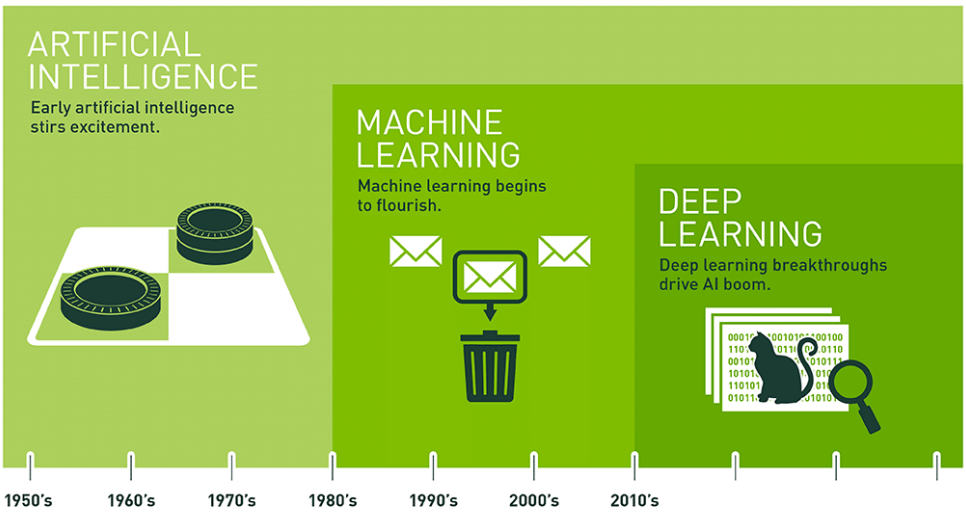

Abb. 1 Historische Entwicklung der KI-Forschungsbereiche. Quelle: https://blogs.nvidia.com/wp-content/uploads/2016/07/Deep_Learning_Icons_R5_PNG.jpg.png

Die Enquete-Kommission der Bundesregierung sieht in der KI die nächste Stufe einer durch technologischen Fortschritt getriebenen Digitalisierung [7]. Ein wesentliches Element ist dabei die Art, wie diese Algorithmen entwickelt werden. Ein klassischer Algorithmus setzt in der Regel ein vorher beschriebenes Verfahren in Software um. Das Verfahren basiert dabei auf mathematischen, statistischen oder anderen Annahmen, Theorien und Regeln. Im Gegensatz dazu wird der Algorithmus einer KI-Methode mit Hilfe von Daten trainiert. Diese Algorithmen zeichnen eine hohe Komplexität und einen sehr hochdimensionalen Parameterraum aus. Ein weiteres Merkmal ist die sehr hohe Anpassungsfähigkeit von KI-Methoden. Diese kann jedoch dazu führen, dass auch unerkannte Merkmale der Trainingsdaten ungewollt in den Algorithmus einfließen. Daher ist, im Gegensatz zu anderer Software, eine Prüfung des Algorithmus auf Basis des Quellcodes allein kaum durchführbar.

Trotz einiger sehr spezifischer KI-Merkmale besteht bisher keine einheitliche Definition von KI sowie zahlreicher weiterer Begrifflichkeiten im KI-Umfeld. Dieses Strategiepapier orientiert sich entsprechend an der Auslegung der Bundesregierung sowie dem National Institute of Standards and Technology (NIST), dem US-amerikanischen Metrologieinstitut, für eine Begriffserklärung:

> *KI bezeichnet Software und/oder Hardware, welche lernen kann komplexe Probleme zu lösen, Vorhersagen zu treffen und Aufgaben zu verrichten, die „menschliche" Qualitäten und Fähigkeiten wie (Sinnes-)Wahrnehmung (z. B. Sehen, Berührung) durch Datenerfassung, Kognition, Planen, Lernen, Kommunikation oder auch physische Handlungen erfordern [8]. Man unterteilt sie in „starke" und „schwache" KI. „Starke" KI geht dabei von Systemen aus, die den intellektuellen Fähigkeiten der Menschen gleichkommen oder diese übertreffen. „Schwache" KI bezeichnet hingegen Algorithmensysteme zur Lösung konkreter Anwendungsprobleme auf Basis von Methoden aus der Mathematik und Informatik, wobei die entwickelten Systeme zur Selbstoptimierung fähig sind [5].*

Neben der Frage der Begriffsdefinition sind zahlreiche weitere Aktivitäten zu KI im Bereich der Forschung und Entwicklung zu verzeichnen – u. a. gestützt durch die Maßnahmen der Bundesregierung und diverser föderaler Initiativen wie beispielweise im Land Niedersachsen [9]. Etliche Publikationen zu neuen Methoden, Modellen und Anwendungsbeispielen werden in zunehmender Häufigkeit veröffentlicht. Zudem ist innerhalb des vergangenen Jahrzehnts ein stetig anwachsender Trend für Patente und Lizenzen zu KI-Anwendungen zu verzeichnen. Zahlreiche Fraunhofer Institute, das DFKI und das DLR bauen ihre Aktivitäten im Bereich der KI-Forschung massiv aus. Derzeit sind bereits etwa 80 der angestrebten 100 KI-Professuren eingerichtet worden. In der Normung und Standardisierung werden nicht nur bei ISO und IEC neue Arbeitsgruppen eingerichtet, um die neuen Anforderungen durch KI zu behandeln. Auch das DIN ist mit der „Normungsroadmap KI" [10] und der DIN SPEC 92001 [11, 12] sehr aktiv und adressiert die prozessualen Veränderungen für die Standardisierung sowie mögliche Handlungsfelder. Ebenso entstehen themengebundene Gremien für KI mit Bezug zu digitaler Gesundheit, autonomer Mobilität, etc. oder KI wird als neuer Bestandteil zu prüfender Produkte zur Herausforderung bestehender Gremien. Eine Übersicht über die KI-Normungs- und -Standardisierungsaktivitäten zeigt eine Veröffentlichung aus dem Projekt „ExamAI – KI Testing & Auditing" [13].

Auch die PTB ist in diversen Bereichen des KI-Einsatzes und der -Forschung bereits aktiv und ist nun dabei in diesem dynamischen Umfeld proaktiv ihre Rolle und ihren Platz finden. Große Player mit weitreichender KI-Expertise bearbeiten hauptsächlich Fragen der Machbarkeit und der technischen Umsetzung (DFKI, etc.). Die PTB hingegen setzt auf die Verknüpfung neuer KI-Kompetenzen mit der reichhaltigen fachlichen Expertise im Bereich der Metrologie und Sensorik, konventioneller Methoden im Bereich Simulation und Datenanalyse sowie der Qualitätsinfrastruktur. Gleichzeitig ist die PTB bemüht, frühzeitig mit politischen Entscheidungsträgern und zentralen Gremien und Verbänden in Kontakt treten, um eine gestaltende Rolle der weiteren Umsetzung der KI-Strategie der Bundesregierung einnehmen zu können. Konkret wird in [3] als weiterer Schritt der Bundesregierung angekündigt:

> *„Umsetzung der in der Normungsroadmap KI definierten Roadmap: Entwicklung von Prüfkriterien auf der Basis etablierter und zu entwickelnder Prüftechnologien zur Prüfung der Robustheit, Sicherheit, Verlässlichkeit, Integrität, Transparenz, Erklärbarkeit, Interpretierbarkeit und Nichtdiskriminierung von (hybriden) KI-Systemen."*

und damit ein prädestiniertes Handlungsfeld für die PTB und geeignete Partner umrissen.

Als Reaktion auf derartige Handlungsaufforderungen wurde in der PTB unterstützt durch das Konjunkturprogramm „Corona-Folgen bekämpfen, Wohlstand sichern, Zukunftsfähigkeit stärken" eine erste KI-„Keimzelle" für „KI in der Medizin" eingerichtet. Diese agiert als Kompetenzzentrum mit Charakter eines Graduiertenkollegs und betreibt sowohl Forschung zu konkreten Anwendungsfällen

als auch Grundlagenforschung zu KI. Thematisch wird dieses Zentrum um bestehende Gruppen organisiert, die eine starke Kompetenz, Forschungserfahrung und Zugriff auf verwendbare Daten im Bereich Medizin aufweisen, und um einen neuen Stellenpool von zehn Wissenschaftler*innen ergänzt. Diese Aktivitäten sind so angelegt, dass sie auch in andere Wirkbereiche der PTB ausstrahlen und gut mit theoretischen Grundlagenarbeiten und Anwendungsforschung anderer Themenfelder (z. B. autonomes Fahren, Optik, etc.) verzahnt werden. So entsteht derzeit ein Netzwerk, in dem Wissenstransfer und -erhalt begründet werden. Die Lenkungskreise Digitalisierung und Medizin sowie untergeordnete, themenspezifische Interessengruppen übernehmen hierbei die Koordinierung und Priorisierung und geben Hinweise zur Außenvertretung der KI-Themen insbesondere zu relevanten Entscheidungsträgern und Gremien.

Auch im Bereich des gesetzlichen Messwesens beschäftigt sich die PTB mit den Herausforderungen durch KI. So stimmen sich die nationalen Vertretungen, darunter die PTB, auf internationaler Ebene aktiv zur Behandlung von KI-Anwendungen ab und finalisieren für die OIML bereits Leitlinien zur Verwendung von KI in Messgeräten.

Themenkomplexe

Köpfe

Die PTB setzt es sich zum Ziel, ihr hohes metrologisches Domänenwissen um entscheidende KI-Kompetenzen zu erweitern, um als starke und kompetente Instanz im Zusammenspiel mit anderen Partnern Vertrauen in KI zu schaffen und langfristig zu sichern.

Um den stetig wachsenden Einsatz von KI in nahezu allen Lebensbereichen metrologisch zu hinterlegen, ist die PTB auf einen signifikanten und nachhaltigen Ausbau von KI-Kompetenzen angewiesen. Dieser ansteigende Trend von KI-Anwendungen und -verfahren über alle Branchengrenzen hinweg hat jedoch gleichzeitig zur Folge, dass die PTB diesen Kompetenzaufbau im Wettbewerb um die besten Köpfe mit zahlreichen Unternehmen und Forschungseinrichtungen bestreiten muss.

Die Interdisziplinarität der Arbeit an der PTB ist in diesem Wettbewerb ein besonderer Anreiz für KI-Expertinnen und -Experten auf dem Arbeitsmarkt. Synergetisch wird metrologisches Domänenwissen aus den Bereichen der Sensorik, Statistik, Modellierung und der QI mit den KI-Kompetenzen verknüpft und schafft dadurch praktischen Mehrwert. Insbesondere das metrologische Verständnis für Messunsicherheiten, Fehlerfortpflanzung und Rückführung auf festgelegte Standards erschließt neue Grundlagen für einen sicheren und vertrauensbildenden Einsatz von KI. Durch seine hohe wirtschaftliche und gesellschaftliche Relevanz hebt sich dieses Betätigungsfeld der PTB auf dem Arbeitsmarkt potentiell von reinen KI-Entwicklungsaufgaben ab und verschafft der PTB einen Vorteil in der Personalgewinnung. Neben metrologischem Domänenwissen, Datenkompetenz und Konzeptverständnis für maschinelles Lernen und Data Science ergänzen die Kompetenzentwicklung von Beschäftigten weiterhin grundlegende digitale Kompetenzen, Kompetenzen im Umgang mit KI-Systemen (z. B. Prozess-, Problemlösungs- und Reflexionskompetenz), sowie Erfahrung in der Gestaltung von Arbeitsprozessen (z. B. soziale, organisatorische und Selbstkompetenzen) [14, 15].

Durch gezielte Nachwuchsförderung kann die PTB gut ausgebildete Kräfte mit umfassenden Basiskompetenzen frühzeitig für KI-Forschungs- und Entwicklungsfragen im Bereich der Metrologie gewinnen. Insbesondere im Rahmen von gemeinsamen Berufungen mit Universitäten oder betreuten Forschungsarbeiten begeistert die PTB Bachelor- und Masterstudierende sowie Promovierende für metrologische KI-Forschung und Entwicklung, indem sie KI-Wissen mit metrologischem Praxisbezug vermittelt. Entsprechend langfristige Perspektiven für qualifizierte Forschende sichern den Personal- und Kompetenzerhalt nachhaltig ab.

Entscheidend ist auch die KI-spezifische Befähigung von Beschäftigten der PTB mit Hilfe von Fort- und Weiterbildungsmaßnahmen. So können sie gezielt ihre KI-Kompetenzen ausbauen, neu erworbenes Wissen mit den physikalisch-technischen Anwendungsfeldern ihrer Forschungs- und Entwicklungsarbeit verknüpfen und die gewonnenen Erkenntnisse innerhalb der PTB mit Praxisbezug weitergeben. Eine wichtige „Keimzelle“ für den Aufbau essentieller KI-Kompetenzen innerhalb der PTB bildet neben dem abteilungs- und fachbereichsübergreifenden Projekt „Machine Learning for Medical Imaging (ML4MedIm)“ der 2021 ausgeschriebene und besetzte Stellenpool mit zehn (Post-)Doktorand*innen für „KI in der Medizin“. Ausgehend von konkreten Use Cases werden in diesem Rahmen grundlegende Fragen zu Verlässlichkeit, Robustheit und Erklärbarkeit von KI-Verfahren sowie der notwendigen Datenqualität erarbeitet und diese Kompetenzen anschließend auf weitere

Themenfelder mit KI-Einsatz übertragen. Besondere Bedeutung gewinnt dabei der enge wissenschaftliche Austausch innerhalb der PTB zu methodischen Verfahren über rein fachliche Disziplinen hinweg. Dafür muss die PTB geeignete Strukturen und Formate zum Kompetenzerhalt und Wissenstransfer etablieren und sicherstellen.

Durch eine enge nationale, europäische und globale Vernetzung an universitäre und außeruniversitäre Forschungseinrichtungen und -fördernetzwerke bereitet die PTB ein attraktives Arbeitsumfeld mit metrologisch herausfordernden Fragestellungen für hochqualifizierte Wissenschaftlerinnen und Wissenschaftler. Insbesondere durch Kooperationen mit großen, etablierten Akteuren der KI-Landschaft können die Wissenschaftlerinnen und Wissenschaftler der PTB metrologisches Fachwissen in die KI-Community einspeisen und mit deutlich geringerem personellen Aufwand signifikant zur Lösung der Fragestellungen beitragen.

Forschungsfragen

Die PTB setzt es sich zum Ziel, geeignete Metriken zur Bewertung von KI und Daten in ihrem metrologischen Forschungsauftrag zu erarbeiten, bestehende Mess- und Prüfprozesse auf den Einsatz von KI anzupassen und gleichzeitig die sichere Anwendung von KI für metrologische Forschung und Dienstleistung zu prüfen und auszubauen.

Trotz der vergleichsweise langen und in mehreren Wellen ablaufenden Geschichte seit den 1950er Jahren birgt das Gebiet der künstlichen Intelligenz noch eine Vielzahl unbeantworteter Forschungsfragen in Bezug auf das grundsätzliche Verständnis und die praktische Anwendung dieser Schlüsseltechnologie. Aus metrologischer Sichtweise ergeben sich für die Forschung zu Metrologie und KI dabei zwei übergeordnete Komplexe: Einerseits KI selbst als Gegenstand wissenschaftlicher Forschung bis hin zur Erarbeitung einer Bewertung von KI-Methoden und der zugrundeliegenden Daten und andererseits KI als Werkzeug zur Verbesserung metrologischer Forschung und Dienstleistung. Entsprechend ist der Themenkomplex Forschungsfragen unterteilt in eine Darstellung der bestehenden und geplanten Aktivitäten der PTB zur Erarbeitung einer Qualitätsinfrastruktur für KI (einschließlich einer Bewertung der verwendeten Daten) und eine Übersicht über die Einsatzmöglichkeiten von KI für die Metrologie als Dienstleistungs- und Forschungstätigkeit. Für jeden dieser Themenkomplexbereiche verdeutlichen zwei konkrete Use Cases die Forschungsfelder.

Qualitätsinfrastruktur für KI (QI4AI)

Die PTB versetzt sich mit ihren Forschungsaktivitäten im Bereich KI in die Lage, ihrem gesetzlichen Auftrag auch in Zukunft gerecht werden zu können. Dazu gehören die durch Normen und Standards gesetzten Vorgaben für Qualitätsmerkmale von KI ebenso wie Anforderungen aus Verordnungen und Gesetzen für die Zertifizierung und Konformitätsbewertung von Qualitätssicherungsmethoden für Trainings- und Testdaten.

Bewertung von KI

Die PTB führt bereits Grundlagenuntersuchungen und Anwendungsstudien für die Ermittlung von Bewertungsverfahren für KI-Methoden durch. Dabei steht die Entwicklung quantitativer Maße für die Bewertung von *Erklärbarkeit*, *Unsicherheit*, *Generalisierbarkeit* und *Robustheit* im Mittelpunkt.

Für die in [11] geforderte Bewertung der Funktionalität und Performance von KI als ein Maß für die Qualität wird die quantitative Bestimmung der *Unsicherheit* der Vorhersagen der KI benötigt. Die Unsicherheit setzt sich bei datenbasierten Verfahren aus drei Komponenten zusammen [16]:

- Unsicherheit aufgrund inhärenter Beschränkung im Modell-Fit des lernenden Systems
- Unsicherheit aufgrund der Datenqualität
- Unsicherheit aufgrund abweichender Trainings-, Test- und Anwendungskontexte

Wesentlich ist, dass das „Maß“, mit dem die Unsicherheit gemessen wird, standardisiert ist, da nur dann Unsicherheiten verschiedener KI-Methoden in ihren Vorhersagen überhaupt verglichen werden können wie in [17] gefordert. Methoden zur quantitativen Ermittlung von Messunsicherheiten spielen eine zentrale Rolle in der Metrologie, wo es mittlerweile mit dem GUM einen weltweit anerkannten Standard gibt [18]. Eine solche Standardisierung fehlt bisher im Bereich der KI, wo es eine Vielzahl unterschiedlicher Ansätze zur Quantifizierung der Unsicherheit gibt [10, 19, 20, 21]. Die besondere Herausforderung im Kontrast zu klassischen Messaufgaben ist die starke Abhängigkeit der

Unsicherheitsschätzung für KI-Verfahren von der individuellen Problemstellung. Die PTB untersucht derzeit die Eignung aktueller Ansätze zur Quantifizierung der Unsicherheit von KI-Methoden mit dem Ziel, eine Empfehlung für eine mögliche Standardisierung zu erarbeiten. Die Untersuchungen beinhalten grundlegende Untersuchungen sowie Anwendungsbeispiele [22]. Aus Sicht der Metrologie wäre es erstrebenswert, wenn eine Standardisierung der Unsicherheit im Einklang mit den Prinzipien der Unsicherheitsermittlung in der Metrologie stehen würde und so in Anwendungen, bei denen KI Methoden und klassische Verfahren in ähnlicher Weise operieren, auch gleiche Unsicherheiten zugeordnet werden könnten.

Um Vertrauen in die KI-Methoden zu gewährleisten ist es wichtig, deren Verhalten zu verstehen und sicherzustellen, dass diese nicht etwa nur auf spezielle Aspekte der Trainingsdaten reagieren [23], sondern die relevante Information in den Daten verwenden. Ähnlich wie bei der Unsicherheit gibt es auch bei der *Erklärbarkeit* mittlerweile eine Vielzahl an Ansätzen, siehe z. B. [24, 25] und die Referenzen darin. Ein Ziel der PTB in diesem Bereich ist es letztlich auch, für die Quantifizierung der Erklärbarkeit ein standardisiertes Maß festzulegen. Dafür bedarf es jedoch noch weiterer Grundlagenforschung: Einerseits durch die Erarbeitung von Definitionen für Erklärbarkeit und andererseits durch Forschung zu den Rückschlüssen, welche Erklärbarkeit erlauben soll. Denkbar wäre eine Definition verschiedener Klassen von Erklärbarkeit, je nach Art der erlaubten Rückschlüsse und der konkreten Problemstellung. Möglicherweise steht am Ende dieser Forschungsarbeit auch kein einheitliches Maß für Erklärbarkeit, sondern stattdessen ein Katalog mit konkreten Benchmarks für verschiedene Anwendungen. Zur Forschungsfrage der Erklärbarkeit ist eine enge Kooperation der PTB mit dem HHI (Fraunhofer Heinrich-Hertz-Institut) geplant. Diese Zusammenarbeit ist Teil eines an der PTB durchgeführten Projekts zur Untersuchung von KI-Methoden bei der medizinischen Bildgebung aus Sicht der Metrologie.

Die *Robustheit* und *Generalisierbarkeit* von KI-Methoden gegenüber Eingangsdaten, die von den zum Trainieren der Methode benutzten Daten abweichen, spielt insbesondere in der Medizintechnik oder beim autonomen Fahren eine große Rolle. Von Bedeutung sind hierbei zum Beispiel „out-of-distribution“-Fehler, die dadurch entstehen, dass gewisse Merkmale nicht in den Trainingsdaten abgebildet sind. Eine große Bedeutung kommt auch den sog. „adversarial attacks“ zu, bei denen „gutartige“ Eingangsdaten gezielt geringfügig so geändert werden, dass eine KI-Methode versagt. Um die Bewertung der Robustheit bezüglich dieser Einflussfaktoren quantitativ vergleichbar zu machen, sind mehrere Bewertungskriterien vorgeschlagen worden. Die PTB untersucht diese Kriterien, und hat auf Basis statistischer Ansätze Alternativen entwickelt, die in bisherigen Untersuchungen sehr gute Eigenschaften aufweisen [26, 27].

Referenzdaten und Bewertung von Datenqualität

In allen Quellen zur Bewertung, Zertifizierung und Konformitätsbewertung von KI-Anwendungen oder Produkten mit KI-Anteilen wird die Notwendigkeit von Referenzdaten sowie allgemein anerkannten Kriterien für Datenqualität und Datenhandling genannt. Für die Wahrnehmung ihrer Aufgaben muss die PTB demnach Kompetenzen zu diesen Fragen aufbauen. Dabei ist die bspw. auch in [17] genannte Notwendigkeit von Domänenwissen wichtig bei der Entscheidung für geeignete Forschungsvorhaben. So ist insbesondere die Repräsentativität von Referenzdaten „aus sich selbst“ nicht möglich, sondern immer nur kontextbezogen vor dem Hintergrund einer Grundpopulation. Stattdessen könnten statistische Kriterien (z. B. Test auf Gleichheit der Verteilungen) zum Zuge kommen. Hier könnten auch Anleitungen zur Konstruktion der (synthetischen) Referenzdaten als Aufgabe für die PTB hinzukommen. Die Metrologie beschäftigt sich bereits mit der Beurteilung von Daten, aber tut das bisher eher auf dem bottom-up level (GUM-like), basierend auf dem Verständnis der zugrundeliegenden Physik, als top-down über die Eigenschaften der Daten selbst. In einigen Bereichen stellt die PTB bereits physikalische/chemische Referenzdaten zur Verfügung. In Zukunft könnte dies

weiter ausgebaut werden mit dem Ziel, Referenzdaten gezielt für die Bewertung von KI-Methoden zu entwickeln. Dabei sollte auch die Entwicklung von Methoden für die Erzeugung synthetischer Datensätze, die metrologisch validiert und qualitätsgesichert rückgeführt sind, berücksichtigt werden. Gerade diese sehr typische Metrologie-Aufgabe „synthetische Referenzdaten-Erzeugung“ kombiniert die Erfordernisse der metrologischen Domänenkompetenz mit Datenkompetenz und physikalisch-technischem Verständnis.

Inzwischen existieren erste Beispiele für die automatische Annotation von Trainingsdaten durch die Kombination verschiedener Modalitäten. So wurde in [28] in einem ersten Schritt ein ML-Verfahren darauf trainiert, Tomografie-Aufnahmen der Retina und co-registrierte Fundus-Aufnahmen zu einer Prädiktion der Retina-Dicke zu kombinieren. Als Ergebnis wurde das trainierte ML-Verfahren dazu verwendet, einen Datenbestand von 120 000 Datensätzen automatisch zu annotieren. Diese dienen dann wiederum als Trainingsdatensatz für ML-Verfahren zur Detektion von durch Diabetes hervorgerufenen Augenschädigungen mit drohender Blindheit. In einer Zulassung solch eines ML-Verfahrens sind dann nicht mehr nur die reinen Rohdaten zu bewerten, sondern auch der gesamte Workflow zur Verwendung dieser Daten. Entsprechend müsste die PTB auch Kompetenzen im Bereich des Datenhandling aufbauen, um bspw. die Anforderungen aus [29] und [30] abbilden zu können.

Use Case: Metrologie für das autonome Fahren – Vertrauen in KI

Für die Einführung autonom fahrender Fahrzeuge im Straßenverkehr ist es unabdingbar, die damit zusammenhängenden Funktionen im Zuge von Zulassungsverfahren zu testen, was wiederum geeignete Prüfkataloge einerseits und technisch geeignete Messeinrichtungen andererseits erfordert. Aufgrund der Vielschichtigkeit und Komplexität der Problemstellung werden hier nach derzeitigem Kenntnisstand mehrstufige Verfahren zur Zertifizierung etabliert werden müssen [31]. Die Zulassung muss dabei jeweils sowohl einzeln für die Soft- und Hardware erfolgen als auch im Verbund als Gesamtsystem. Die Einzelsensoren metrologisch zu charakterisieren ist bereits Inhalt laufender Forschungsarbeiten – auch in der PTB. Im Fahrzeugeinsatz jedoch werden die Einzelmessungen verschiedenster Sensoren aggregiert und durch (KI)-Algorithmen ausgewertet. Erst aus dieser Kombination von Messdaten wird autonom eine Entscheidung für das Verhalten des Fahrzeugs abgeleitet.

Die Aufgabe der Metrologie ist es, die von autonom fahrenden Fahrzeugen gemessenen physikalischen Größen, die damit erzeugten Daten und die daraus getroffenen Entscheidungen in Bezug auf Robustheit, Einfluss von Messunsicherheiten und die damit einhergehenden Auswirkungen auf die Funktionalität quantitativ zu evaluieren. Ein Fernziel kann hierbei die Erstellung sogenannter Goldstandards sein, und zwar sowohl auf der physikalischen Messebene als auch in Bezug auf die Eingangsdaten der KI-Entscheidungslogik. Wichtig ist dabei, dass interne (z. B. Alterung, technischer Defekt) und von außen einwirkende Degradationseffekte (z. B. Witterung, Verschmutzung, Niederschlag etc.) berücksichtigt werden. Nur so können in Zukunft verlässliche Aussagen gemacht werden zu Funktionsgrenzen eines Fahrzeugs, welches Alterungs-, Beschädigungs- oder sonstige störende Erscheinungen aufweist.

Mehrere KI-Forschungsgruppen in Deutschland und der Welt beschäftigen sich mit der Entwicklung von Methoden, um das autonome Fahren zu realisieren (z. B. im Kompetenzzentrum „Autonomes Fahren“ (AD) des DFKI [32]). Auch das Fraunhofer IKS beschäftigt sich bereits mit Fragen zur Bewertung und Absicherung von KI-Methoden für das autonome Fahren. Dabei werden sowohl mathematisch-statistische Fragestellungen behandelt als auch die geeignete Implementierung der Methoden in Software. Viele weitere Fraunhofer-Aktivitäten in Bezug auf das autonome Fahren werden in der Fraunhofer-Allianz „Verkehr“ gebündelt.

Angelehnt an einen Übersichtsartikel zu Deep Learning für autonomes Fahren [33] können folgende metrologische Fragestellungen für die Behandlung von KI-Methoden für das autonome Fahren formuliert werden:

- Verständnis der Auswirkung von Messunsicherheiten und der Beeinträchtigung von Sensordaten
- Verständnis des Kontexts von Datenfusion und KI-Anwendung innerhalb des Gesamtsystems
- Definition von Annahmen bzgl. des Kontexts, in welchem das System operieren (operational design domain (ODD)) und wofür es geprüft/getestet werden soll
- Definition grundlegender Anforderungen an Datenqualität und KI-Verfahren

Aus Sicht der Metrologie ergeben sich dadurch für die PTB konkret folgende Aufgabenfelder:

- Abschätzung zur Eignung von Messmethoden und -verfahren (z. B. Kamera- vs. Lidar-Verwendung) zur Erzeugung geeigneter Sensordaten
- Beurteilung von Daten aus Messungen und Simulationen für das Training und Testen von KI-Verfahren; insbesondere auch Vergleichbarkeit von simulierten und realen Daten
- Berücksichtigung von Messunsicherheiten bei Analyse und
- Umgang mit Grenzfällen in der Bewertung von KI-Methoden, bspw. durch Generalisierung.

Im Grunde ist ein autonom fahrendes Fahrzeug ein mobiles Sensornetzwerk. Daher können die Arbeiten der PTB im Bereich KI für autonomes Fahren initial auf den begonnenen Aktivitäten im Themenfeld „Metrologie für heterogene Sensornetzwerke“ aufbauen. Dazu gehören Arbeiten im EMPIR-Projekt “Metrology for the factory of the future” (Met4FoF), dem BMBF-Projekt “AAS-basierte Modellierung zur Analyse veränderlicher CPS“ (FAMOUS) und dem BMWi-Projekt „Sichere und robuste kalibrierte Messsysteme für die digitale Transformation“ (GEMIMEG-II).

In diesen Projekten werden bereits Methoden zur Verwendung und Fortpflanzung von Messunsicherheiten in Sensornetzwerken sowie Methoden zur Feature Extraction für das maschinelle Lernen unter Berücksichtigung von Unsicherheiten behandelt.

Use Case: Qualitätskontrolle für erklärbare KI in der klinischen Diagnostik

Eine besonders vielversprechende Rolle wird der KI in der Medizin der Zukunft zugedacht, wo sich im Zusammenspiel von komplexen Algorithmen und immer umfangreicheren und besser verknüpften Datenmengen gezielt klinisch relevante Fragen lösen lassen. Dies können z. B. Diagnosen oder Prognosen sein. Die gegenwärtig am stärksten digitalisierten Bereiche der Medizin sind die Intensivmedizin und die Radiologie. In der Neuroradiologie ist es beispielsweise erstrebenswert, frühe Anzeichen neurologischer Krankheiten (wie der Multiplen Sklerose, MS, des Morbus Parkinson, PD, oder der Alzheimer’schen Krankheit, AD) in Form struktureller Auffälligkeiten des Gehirns (z. B. Läsionen, Ablagerungen, Gewebeschwund) zu erkennen. Dies geschieht z. B. durch Analyse struktureller Magnetresonanztomografiedaten (MRT), welche in großen, teils öffentlich verfügbaren Datenbanken vorliegen. Methoden des Maschinellen Lernens haben in der jüngsten Vergangenheit Erfolge z. B. bei der Vorhersage der Alzheimer Krankheit erzielt [34]. Neben einer hohen Vorhersagegüte wird aber auch immer öfter gefordert, dass die Entscheidungen solcher Modelle auf individuelle Eingaben (z. B. den MRT-Aufnahmen von Patient*innen) „erklärbar“ sind. Hierzu wurden bereits eine Vielzahl von „explainable AI“-Methoden (xAI) entwickelt [35]. Ein gemeinsames Problem all dieser Methoden ist jedoch, dass sie nur unzureichend validiert sind. Es existiert keine allgemein akzeptierte formale Definition von Erklärbarkeit, und die Autoren der meisten existierenden Methoden liefern entweder keine Anweisungen, *wie* genau die Ausgaben der Methode interpretiert

werden dürfen oder liefern nur unzureichende Evidenz für die Validität der vorgeschlagenen Interpretationen. Dieser Zustand ist unbefriedigend vor dem Hintergrund, dass selbst allgemein übliche Interpretationen einfacher linearer Modelle formal nicht haltbar sind [36].

Aufgrund dieser Limitationen wird sich die PTB in Zukunft sowohl mit den theoretischen Grundlagen als auch der praktischen Validierung von Erklärbarkeit befassen. Insbesondere sollen formale Definitionen für Erklärbarkeit erarbeitet werden. Eine Möglichkeit dazu bieten synthetische Daten. Im Anwendungsbeispiel der Neuroradiologie soll dazu ein synthetischer Datensatz auf Basis realer struktureller MRT-Bilder gesunder Personen hergestellt werden. Diese Bilder sollen dann kontrolliert und in möglichst realistischer Art und Weise mit Läsionen, Ablagerungen, Ablationen und anderen strukturellen Anomalien versehen werden. Darauf basierend werden Vorhersageprobleme (z. B. die Diagnose oder Differentialdiagnose der unterschiedlichen strukturellen Charakteristika) definiert. Die so entstandenen Daten eignen sich sowohl als Benchmarks für Vorhersagemodelle als auch deren „Erklärungen". Die „ground-truth" für letztere Methoden ergibt sich durch die bekannten Positionen der strukturellen Anomalien. Die Güte einer Erklärung könnte dann durch den Vergleich der Bildmaske der wahren Anomalien und der Ausgabe der Erklärmethode, der sogenannten „heat map", quantifiziert werden. Hierzu eignen sich Metriken aus der Bildverarbeitung und Signalentdeckungstheorie wie Jaccard und Dice scores, sowie receiver operating characteristic (ROC)-Kurven. Somit wäre ein erster Schritt einer objektiven und quantitativen Bestimmung der Erklärgüte für diesen konkreten Anwendungsfall vollbracht.

KI für die Metrologie (AI4Metrology)

Wie in zahlreichen anderen Wissenschaftsbereichen bietet der Einsatz von KI-Verfahren auch für die Metrologie erhebliche Potentiale, die es gezielt auszuschöpfen gilt. Nach einer Umfrage des EMN Mathmet mit Antworten aus 13 nationalen Metrologieinstituten liegen folgende Einsatzbereiche von KI für die Metrologie besonders im Fokus:

- Verbesserung der Datenauswertung
- Neue Messmöglichkeiten
- Virtuelle Messgeräte
- Umgang mit großen Datenmengen (Big Data)
- Entstehung neuer Technologiebereiche im Zuge der digitalen Transformation
- Neue Dienstleistungen

Die verbesserte Datenanalyse durch Einsatz von KI beinhaltet dabei sowohl die Automatisierung der Auswertungsprozesse, erweiterte Methoden der Regression und Klassifizierung sowie die Optimierung von Inline-Messtechnik. In vielen Bereichen kann KI-gestützte Datenauswertung eine Beschleunigung und damit eine Kostenreduktion in der Nachverarbeitung von Messergebnissen bewirken. Zudem erschließt der Einsatz von KI einen neuen Umgang mit der zunehmenden Menge an Messdaten sowohl von Einzelmessgeräten als auch von verteilten Sensornetzwerken. Einsatzgebiete für KI wären somit also auch Multi-Parameter-Modellierungen großer Datensätze, wie beispielsweise in der Metabolomik, oder komplexer Netzwerke, wie im Internet of Things (IoT). Des Weiteren können mit Hilfe von ML-Verfahren synthetische Datensätze für verschiedenste Nutzungsbereiche erzeugt und bereitgestellt werden, was für den enormen Datenbedarf vieler Auswertungsverfahren von großem Nutzen ist.

Zudem eröffnet die KI ein weites Feld neuer metrologischer Anwendungen. Dies gilt insbesondere im Bereich der Bildgebung, -analyse und -rekonstruktion (z. B. in der medizinischen Bildgebung und der Mikroskopie), in komplexen Sensornetzwerken (z. B. dem Umweltmonitoring), bei verbesserten

Kalibrierungen und dem Bereich des autonomen Fahrens. Auch vollständig neue metrologische Dienstleistungen, wie die Bereitstellung von Referenzdatensätzen, Benchmark-Tests und Infrastrukturen für vertrauenswürdige KI, sowie beschleunigte Entwicklungszyklen bestehender Produkte werden als Potentiale des KI-Einsatzes erkannt.

Im Gebiet der virtuellen Metrologie könnte aus dem Zusammenspiel von in-silico-Modellen und KI eine Optimierung digitaler Zwillinge erreicht werden, mit deren Hilfe reale Experimente virtuell nachgebildet werden können. Zudem bieten daten- anstelle von modellbasierten Vorhersagen, auch für die Wartungszeiträume experimenteller Aufbauten, ein großes Potential für die metrologische Forschung und Anwendung.

In einigen Arbeitsgruppen der PTB findet KI auch bereits Anwendung für metrologische Forschung und Dienstleistung, z. B. in der Spektrometrie, der Anomaliedetektion und auch Hardware-nah im Bereich der Sensorik. Die folgenden Use Cases skizzieren exemplarisch den konkreten Einsatz von KI-Verfahren in metrologischen Fragestellungen und beleuchten für diese Methoden die Anwendbarkeit in und Übertragbarkeit auf angrenzende Fachbereiche.

Use Case: KI für optische Metrologie – Formmessungen und Nanometrologie

Die Bedeutung der optischen Metrologie reicht von der Charakterisierung von Oberflächen, der Vermessung dimensioneller Größen bis zu der Bestimmung von optischen Eigenschaften. Dabei werden die zu untersuchenden Objekte oft mit Licht bestrahlt (z. B. Laser, Synchrotronstrahlung) und die gestreuten Photonen detektiert. Die Messgrößen werden hierbei nicht direkt bestimmt, sondern durch einen mathematischen Algorithmus (das Lösen eines inversen Problems).

Ein Beispiel ist die Vermessung von optischen Asphären und Freiformen. Ein dafür gut geeignetes an der PTB entwickeltes, optisches Messverfahren, das Tilted-Wave-Interferometer (TWI), basiert auf einem interferometrischen Messprinzip, welches mehrere Quellen benutzt, um die Prüflingsoberfläche an allen Stellen messbar zu machen. Das einem Prüfling zugrundeliegende Design ist allgemein bekannt, es sind die virtuell gegebenen Topografieparameter des Herstellers. Um die Differenz zwischen einem Prüfling und seinem Design zu ermitteln, werden neben den beobachteten Interferogrammen des Prüflings auch die zum Design gehörenden Interferogramme benötigt. Diese werden mit einer Modellierung des TWI-Messaufbaus und einer Simulation des Messprozesses erzeugt. Das zu lösende inverse Problem besteht nun darin, die Unterschiede von den simulierten und gemessenen Daten auf die tatsächliche Differenz zwischen der Designtopografie und dem Prüfling zurückzuführen und daraus die reale Prüflingstopografie zu bestimmen.

Ein anderes Beispiel ist die optische Nanometrologie. Schrumpfende Strukturdimensionen und gesteigerte Funktionalitätsanforderungen in der Halbleiterindustrie stellen etablierte photonische Messmethoden im weichen Röntgen- bis IR-Wellenlängenbereich wie Scatterometrie, Mueller-Ellipsometrie und Reflektometrie zunehmend vor neue Herausforderungen. Ohne Rückführbarkeit und strenge Unsicherheitsabschätzungen wird dies zu einem Engpass für zukünftige technologische Entwicklungen werden. KI-Methoden können helfen, diesen Herausforderungen mit einem vertretbaren Zeitaufwand zu begegnen. So können in der derzeitigen Praxis beispielsweise nicht ausreichend die Drift von Geräteparametern berücksichtigt werden. Neuronale Netze sollen hier künftig die Prozesse virtuell abbilden und somit eine effektive Prozesskontrolle ermöglichen [37] [38].

Etablierte KI Verfahren zielen oft auf Anwendungen in der Bilderkennung ab. Für die Anwendung auf die optische Metrologie, bzw. auf indirekte Messungen müssen diese Verfahren angepasst und weiterentwickelt werden. Aktuelle Arbeiten an der PTB [39, 40, 41] fokussieren sich auf die Lösung des inversen Problems durch tiefe neuronale Netze. Zusätzlich zur rekonstruierten Topografie wird auch

die Modellunsicherheit mitgeschätzt, d. h. die Unsicherheit der Vorhersage wird quantifiziert. Hierbei gibt es verschiedene vielversprechende Ansätze.

Beim TWI basiert die Topografievorhersage und ihre Unsicherheitsquantifizierung auf einer Ensemble-Methode, die gut auf hochdimensionale Probleme skaliert. Die benutzte Methode wird anhand von systematisch eingeführten Störungen, z. B. in Form eines wachsenden Kalibrierfehlers, untersucht. Neben der Unsicherheitsbestimmung und der Rekonstruktionsgenauigkeit wird hierbei auch die Generalisierbarkeit der vorgeschlagenen Methode auf Daten, die außerhalb des Trainingsbereiches liegen, analysiert. Die Ergebnisse sind vielversprechend und zeigen, dass die Modellunsicherheit mit steigendem Kalibrierfehler wächst. Diese Eigenschaft könnte benutzt werden, um zu bestimmen, wann eine Rekalibrierung des virtuellen Systems benötigt wird.

Eine weitere Methode wurde für die optische Nanometrologie entwickelt und basiert auf dem Einsatz von invertierbaren neuronalen Netzen. Diese lernen eine sogenannte Transportabbildung auf die Zielverteilung und zusätzlich, durch eine speziell angepasste Optimierung, nicht nur die gewünschten Messgrößen, wie Linienbreite, Kantenwinkel oder die Höhe von nanometergroßen Linien, sondern auch die dazugehörige Messunsicherheit [41].

Die vorgeschlagenen Deep-Learning-Methoden erzeugen damit für Produktionsstätten einen erheblichen Zeitvorteil gegenüber den existierenden konventionellen Methoden und können den Einsatz dieser Messmethoden in Echtzeit ermöglichen.

Die Einsatzbereiche von KI-Methoden, die in diesem Use Case entwickelt werden, sind vielfältig und reichen von „computational imaging“ [42], Modellkalibrierung („adaptive optics“ [43]) bis hin zu Korrekturen von Inputparametern (z. B. Positionsfehler [44]). Ganz allgemein wird erwartet, dass die hier entwickelten KI-Methoden aufwändige komplexe Auswertungen von indirekten Messungen beschleunigen und verbessern werden. Damit haben die vorgeschlagenen Deep-Learning-Methoden bei ihrem Einsatz in Produktionsstätten einen erheblichen Zeitvorteil gegenüber den existierenden konventionellen Methoden und ermöglichen Echtzeitanwendungen. Als ein konkretes Beispiel sollen innerhalb des EMPIR-Projekts „20IND04 ATMOC“ KI-Verfahren entwickelt werden, die optische Eigenschaften von Dünnschichtsystemen oder Nanostrukturen mit geringem Rechenaufwand bestimmen können.

Use Case: Rußpartikelcharakterisierung mit Hilfe von KI

Rußpartikel entstehen durch Verbrennungsprozesse und stellen nicht nur eine Gefahr für die Gesundheit dar, sondern wirken aufgrund ihrer optischen Eigenschaften auch klimaschädlich, wobei insbesondere die fraktale Dimension der Partikel ihr Verhalten beeinflusst. Die Morphologie von Rußpartikeln wird daher schon lange mittels Elektronenmikroskopie untersucht, wobei die auf diese Weise gewonnenen Aufnahmen zweidimensionale Projektionen der Partikel darstellen. Lediglich in einem sehr aufwändigen Experiment, das sich nicht in der Breite für Anwendungen eignet, wurde die zweidimensionale Struktur über kohärente Röntgenstreuung am Freie Elektronen Laser in Stanford ermittelt [45]. Man steht daher vor dem Problem, wie sich aus den üblichen Projektionen die fraktale Dimension als Eigenschaft der dreidimensionalen Form (mit Werten zwischen 1 z. B. für eine lange Kette und 3 für kugelförmige Aggregate) bestimmt werden kann. Dazu wurde in der Vergangenheit eine Reihe von konventionellen Ansätzen vorgestellt, von denen ein Teil kaum oder nur für Spezialfälle geeignet ist [46]. Die geeigneteren Ansätze sind komplex, zeitaufwendig und abhängig von Eingaben der Nutzenden [47]. Es konnte gezeigt werden, dass es möglich ist, große Teile des konventionellen Algorithmus durch Machine-Learning-Netzwerke zu ersetzen, die schematisch in Abb. 2 dargestellt sind.

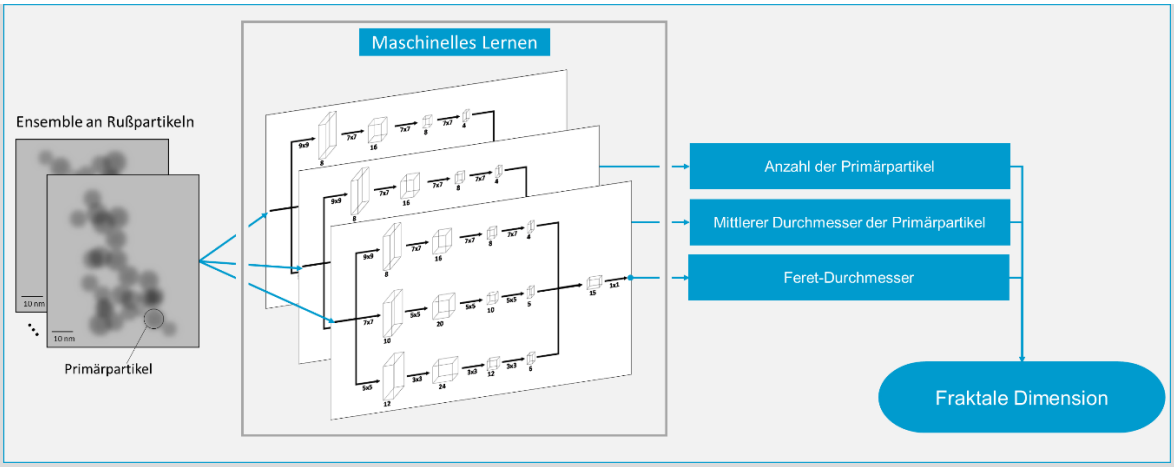

Abb. 2 Die elektronenmikroskopische Aufnahme des Rußpartikels wird von drei Machine Learning-Netzwerken ausgewertet, mit deren Hilfe verschiedene Parameter bestimmt werden, aus denen sich die fraktale Dimension ergibt.

Das zugrunde liegende Machine-Learning-Netzwerk wurde dazu mit einem Datensatz trainiert, der durch Monte-Carlo-Simulation der Bildentstehung im Rasterelektronenmikroskop gewonnen wurde. Durch diese Simulationen ist es möglich, in nur wenigen Tagen umfangreiche und qualitativ hochwertige Trainingsdatensätze zu generieren. Diese Fähigkeit wird nicht nur neue Anwendungen ermöglichen, sondern absehbar auch zu weiteren Verbesserungen des beschriebenen Verfahrens führen, das bereits heute dem konventionellen Algorithmus ebenbürtig ist.

Dabei werden durch den Einsatz von Machine-Learning-Netzwerken schon jetzt einige Nachteile der herkömmlichen Methode [47] überwunden: Eingaben durch Nutzende sind nicht mehr notwendig, was einen etwaigen User Bias ausschließt, und die Auswertung konnte ca. um den Faktor 10 beschleunigt werden. Dadurch ist ein Einsatz in Echtzeit während des Aufnahmeprozesses am Elektronenmikroskop denkbar. Die Methode eignet sich nicht nur für Rußpartikel, sondern prinzipiell für alle fraktalen Aggregate, von denen zweidimensionalen Projektionen z. B. mittels Mikroskopie gewonnen wurden. Ein Beispiel sind Staubpartikel in der sogenannten „protoplanetaren Scheibe“, die eine wichtige Rolle bei der Planetenentstehung gespielt haben und daher Gegenstand aktueller Forschung sind [48].

Infrastruktur & Daten

Die PTB setzt es sich als Ziel, sorgfältig aufeinander abgestimmte maschinennutzbare Daten und KI-Methoden als Vertrauensanker für Zukunftstechnologien in der Messtechnik zu etablieren, digitale Normale (z. B. Referenzdatensätze) für das Messwesen zu entwickeln und bereitzustellen sowie die dafür benötigten Infrastrukturen einzurichten.

Wenngleich KI schon seit mehreren Jahrzehnten in verschiedensten Anwendungsbereichen punktuell ihren Einsatz gefunden hat, so hat doch erst das Vorhandensein und das stetige Anwachsen großer Datensätze, sogenannter Big Data, und drastisch gesteigerter Rechenkapazitäten KI zum Durchbruch verholfen. Weil dieser eng mit Fortschritten in der Rechentechnik und Kapazitäten zur Verarbeitung sehr großer Datenmengen verbunden ist, wird in der Fachwelt sogar von „Konvergenz von Künstlicher Intelligenz und Hochleistungsrechnen“ (engl. „convergence of AI and high performance computing“ (HPC)) gesprochen[1]. Entsprechend bedarf es einer hinreichenden Recheninfrastruktur, die bereitgestellt werden muss, um KI für die metrologische Forschung und Dienstleistung an der PTB gewinnbringend einzusetzen und die Eigenschaften von KI selbst als Forschungsgegenstand umfassend zu untersuchen. Analog wird es in einer Zeit, in der die verwendeten Daten als das „neue Öl“ betrachtet werden, umso relevanter, einen kritischen Blick auf die Beschaffenheit von Daten zu werfen und deren Bereitstellung und Handling im Kontext von KI zukunftssicher aufzustellen.

Recheninfrastruktur

Grundvoraussetzung für eine erfolgreiche Befassung mit KI-Themen ist wie eingangs erwähnt die Verfügbarkeit ausreichend dimensionierter Rechenkapazitäten und HPC-Ressourcen für alle Beschäftigten in Forschung und Dienstleistung. Die PTB kann dabei auf eine gut zehnjährige Vorgeschichte aufbauen: Seit 2009 wurde am Standort Berlin-Charlottenburg ein Cluster aus Linux-Rechnern für das Hochleistungsrechnen aufgebaut, zu dem alle PTB-Beschäftigten auf Antrag Zugang erhalten. Die Kapazitäten wurden in bisher drei Investitionsrunden sukzessive erweitert. Die letzte Generation von Servern, die 2017/2018 in Betrieb genommen wurde, umfasst 60 Rechen-Knoten mit insgesamt knapp 1.700 CPU-Kernen (Prozessor-„cores“). Hinzu kommt ein schnelles Verbindungsnetzwerk (Infiniband) für den raschen Datenaustausch zwischen den Servern, welches insbesondere beim verteilten Rechnen bzw. der parallelen Bearbeitung unverzichtbar ist. Komplettiert wird die bestehende Infrastruktur durch ein mehrstufiges Speichersystem, das sowohl das schnelle Abspeichern von Zwischenergebnissen über ein paralleles Dateisystem (Scratch-Speicher) als auch die dauerhaft-verlässliche Aufbewahrung endgültiger Resultate (Isilon-Speicher, inkl. Backup und WORM-Funktionalität) ermöglicht.

Genutzt wird der bestehende HPC-Cluster heute von einer Vielzahl an Nutzenden aus den Fachabteilungen der PTB: Im Vordergrund stehen bisher klassische Simulationen, wie etwa Finite-Elemente-Berechnungen zur Strömungsdynamik oder aber auch große Monte-Carlo-Simulationen zur Bestimmung von Messunsicherheiten. Hinzu kommen in jüngster Zeit erste Anwendungen mit KI-Bezug, etwa zur automatisierten Erkennung und Charakterisierung von Rußpartikeln in Mikroskopie-Bildern wie im Use Case vorgestellt.

[1] Vgl. https://www.intel.de/content/www/de/de/high-performance-computing/hpc-artificial-intelligence.html

Für eine weitere Intensivierung der KI-Aktivitäten ist die aktuell bestehende HPC-Infrastruktur allerdings nur bedingt geeignet. Das liegt einerseits an den limitierten Kapazitäten bzw. der begrenzten Anzahl an Compute-Servern, die zu längeren Wartezeiten bei der Bearbeitung von Aufträgen auf dem HPC-Cluster führen (Queue-Rückstau). Andererseits ist aber auch die Qualität der bisherigen HPC-Server teils nicht für KI-Anwendungen optimiert: Neben den erwähnten 1.700 CPU-cores (Hauptprozessoren) sind bisher nur 2 GPU-Knoten (Grafikprozessoren) mit Graphikbeschleunigern des Typs TESLA-V100 (vorletzte Modell-Generation des bekannten Herstellers Nvidia) verfügbar. Gerade die Verwendung von Grafikprozessoren bzw. das GPU-Computing versprechen aber besonders hohe Effizienzgewinne und Performance-Steigerungen im Rahmen datenintensiver KI-Anwendungen.

Ein Vergleich der Rechenkapazitäten verwandter Institutionen im Berliner Raum (Robert Koch-Institut, Fraunhofer Heinrich-Hertz-Institut), die ihre Forschungsarbeiten im KI-Bereich derzeit stark intensivieren, bestärkt die Schlussfolgerung, dass KI-Forschung auf einen höheren Anteil an GPU-Servern aufbauen muss: Beide Institute haben jüngst größere, fünf- bis sechsstellige Beträge investiert, um Grafikprozessoren des Typs TESLA-A100 (aktuelle Modellreihe des Herstellers Nvidia) zu beschaffen. Als ausgewogen wird dabei ein Zahlenverhältnis von CPUs zu GPUs in der Größenordnungen von etwa 2:1 bis 3:1 betrachtet.

Um sich dieser Zielgröße anzunähern, muss die PTB vor allem die Anzahl an GPUs signifikant erhöhen. Es erscheint sinnvoll, dies nicht den einzelnen Forschungsgruppen und Fachbereichen zu überlassen, sondern zentral zu koordinieren und entsprechende Kapazitäten für alle PTB-Abteilungen bereitzustellen. Im Rahmen einer Erweiterung und Ersatzbeschaffung sollen auch die vorhandenen CPUs durch leistungsstärkere, aktuelle Modelle abgelöst werden. Im Zuge der anstehenden Reinvestition ist zudem geplant, auch die Speicher-Infrastruktur zu modernisieren, um das Einlesen und Abspeichern von Trainingsdatensätzen für KI-Anwendungen signifikant zu beschleunigen, und auch die Kapazitäten für die längerfristige Aufbewahrung großer Datenmengen weiter zu erhöhen.

Die grundsätzliche Alternative zum Ausbau eigener PTB-Rechenkapazitäten (vor Ort / „on premise“) läge in der verstärkten Inanspruchnahme von Cloud-Computing-Angeboten. Solche Angebote, bei denen Nutzende erst im Moment des konkreten Rechenbedarfs eine Verbindung zu entfernten Servern aufbauen und ihre Rechenaufträge dort abarbeiten lassen können, werden seit einigen Jahren auf dem Markt angeboten und können inzwischen als etabliert gelten. Im Vergleich zum Aufbau eigener Kapazitäten haben sie Vor- und Nachteile: Dem Vorteil vermiedener Investitionskosten und theoretisch unbegrenzter Kapazitäten stehen Nachteile in Form von höheren laufenden Kosten (Abrechnung nach Verbrauch), Einschränkungen beim Transfer größerer Datenmengen sowie erhöhte Risiken in Bezug auf Datenschutz und Informationssicherheit gegenüber.

Konkret sind insbesondere folgende Varianten in Erwägung zu ziehen:

1. Die Nutzung kommerzieller Angebote (Anbieter wie u. a. Amazon Web Services, Microsoft Azure, Open Telekom Cloud, Oracle Cloud).
2. Die (Mit-)Nutzung von Kapazitäten, die von Bund und Ländern öffentlich finanziert und der Wissenschaftsgemeinschaft in Deutschland zur Verfügung gestellt werden sollen (NHR-Verbund).
3. Die Erweiterung der IT-Konsolidierung im Bund, also der geplanten Zusammenführung vieler IT-Kapazitäten der Bundesbehörden beim ITZ-Bund um HPC- bzw. KI-Komponenten („Bundes-KI-Cloud“)
4. Eigene, gemeinsame Aktivitäten ausgewählter Bundesbehörden, insbesondere weiterer Ressortforschungseinrichtungen (RFE), zum Aufbau geteilter KI-Kapazitäten („RFE-KI-Cloud“)

Option 1 ist bereits heute am Markt verfügbar, bisher jedoch relativ kostspielig, und deshalb oft nur für „Spitzenlast“ wirtschaftlich attraktiv. Option 2 befindet sich im Aufbau, wird nach derzeitiger Beschlusslage jedoch nur für Nutzende aus Landeseinrichtungen (Hochschulen) zugänglich sein, sodass die PTB allenfalls im Rahmen von Kooperationen darauf Zugriff hätte. Optionen 3 und 4 sind bisher nicht konkret geplant, könnten jedoch künftig in Angriff genommen werden.

Überlegenswert wäre insbesondere eine Bündelung der Computing-Bedarfe von KI-nutzenden Ressortforschungseinrichtungen des Bundes (Option 4). Anstelle der individuellen Bedarfsermittlung und langwieriger Einzelaufrüstung von Rechenkapazitäten und zugehöriger Klimatechnik in den jeweiligen Behörden könnte beispielsweise eine gemeinsame *„RFE-KI-Cloud“* eingerichtet werden. Die Rechenkapazitäten der Cloud wären damit flexibel für alle RFE entsprechend eines festzulegenden Nutzungsschlüssels verfügbar und die gemeinsame Nutzung der Cloud könnte eine zeitliche Ungleichverteilung der Rechenauslastung in den RFE kompensieren. Damit würden freie Kapazitäten einer Einzelbehörde nicht ungenutzt verfallen, sondern stünden anderen beteiligten Behörden zur Verfügung. Mit einer solchen *KI-Cloud* ließe sich potentiell auch auf bauliche und klimatechnische Anforderungen besser und kosteneffizienter reagieren, da entsprechende Infrastrukturen anders als bisher nicht einzeln bei wachsendem KI-Einsatz aufwändig nachgerüstet und für ggf. zeitweise nicht komplett ausgeschöpfte Rechenkapazitäten bereitgestellt werden müssten.

Daten und KI

Als Daten-getriebene Verfahren sind KI-Systeme mit hoher Qualität auf qualitativ hochwertige wie auch umfangreiche Datensätze für das Training angewiesen. Während die KI-Verfahren auf geeigneten Trainingsdaten trainiert werden, benötigt es von diesem Datensatz unabhängige Validierungsdaten zur Verbesserung des KI-Modells und ebenso unabhängige (und für die Entwickelnden unbekannte) Testdaten zur schlussendlichen Überprüfung und Bewertung der KI-Funktionalität.

Neue Trends zur Verbesserung der Funktionalität von KI-Verfahren setzen auch einen stärkeren Fokus auf die zugrundeliegenden Daten. Nachdem der bisherige Ansatz, am reinen Modell bzw. dem Code zu optimieren, um eine bessere Performance der KI-Methode zu erreichen, in vielen Anwendungsfällen keine drastischen Verbesserungen mehr erzielt, verstärkt sich aktuell eine Daten-zentrierte Herangehensweise („Data-centric AI“). Bei dieser Methode werden die Trainingsdaten entsprechend konkreter Prämissen (z. B. konsistentes Labeln, Aussortieren von Rauschdaten, koordiniertes Handling schwierig bewertbarer Datensätze) gezielt ausgewählt und können so die Leistungsfähigkeit der KI-Systeme laut erster Pilotstudien sprunghaft verbessern [49]. Auch in diesem Zusammenhang wird der Einfluss der Datenbeschaffenheit und grundlegender Anforderungen an die Datenqualität deutlicher denn je.

Datenbeschaffenheit und Datenqualität

Heute entstehen Messdaten typisch in digitalen Dateien und Datenbanken mit anwenderspezifischen proprietären Formaten. Die Übertragung dieser Daten in zukunftsfähige, interoperable, KI-nutzbare Formate erfordert in der Regel händische Konversionsarbeit, die meist nur die Erstellenden der Daten (Expert*innen) mit dem nötigen Hintergrundwissen zur Art der Daten und deren Entstehungsprozess (engl. „Data Provenance“) leisten können. Dieser Vorgang stellt oft einen großen Mehraufwand dar und findet deshalb bisweilen nur sehr begrenzt oder gar nicht statt. Digitale Werkzeuge zur automatischen Datenerzeugung, die zunehmend im Rahmen der digitalen Transformation in allen Bereichen der Messtechnik aufkommen, bieten eine elegante Lösung, den bisherigen Mehraufwand mittelfristig zu umgehen. Neue Systeme können von Anfang an so entwickelt werden, dass diese zum „Geburtszeitpunkt“ alle Metrologiedaten in KI-fähigen Formaten erzeugen.

Die Entwicklung und Etablierung KI-geeigneter digitaler Formate für universelle metrologische Kerndaten auf der Basis des Internationalen Einheitensystems (SI) gehört zu den langfristigen Kernzielen der Digitalisierungsstrategie des Internationalen Komitee für Maße und Gewichte (CIPM) [50]. Maschinennutzbare Darstellungen für Messgrößen, Werte, Einheiten und Messunsicherheiten sollen in einem digitalen Rahmenwerk (SI Digital Framework) bereitgestellt werden, welche die Nutzung durch und automatische Analyse mit KI-Methoden direkt und ohne menschliche Interaktion erlaubt. Bei der technischen Realisierung wird dabei auf eine Kombination der Anwendung von FAIR[2]-Prinzipien mit elementaren Metadaten zur metrologischen Rückführbarkeit der Einheiten und metrologische Vergleichbarkeit von Einheiten gesetzt. Die metrologischen Prinzipien zur Rückführbarkeit und Vergleichbarkeit sind dabei unumgängliche Vertrauensanker für die Qualitätsbewertung und Reproduzierbarkeit aller Messdaten weltweit.

Diese metrologischen Kernanforderungen werden durch die folgenden Grundaspekte für die KI-geeignete Datenbeschaffenheit ergänzt:

- **Sichergestelltes Verständnis zu den Hintergründen der Datenerzeugung**
  Informationen zur Datengenerierung liefern eine wichtige Entscheidungsgrundlage bei der Betrachtung von Fragestellungen zur Interoperabilität und der Eignung von Daten für KI-basierte Anwendungen. Hierzu zählen Informationen aus dem Datenlebenszyklus (Zeit und Ort der Entstehung, Messgerät, Gültigkeitsdauer, usw.), zur Messdatenqualität (Qualifikation des Labors, Umgebungsbedingungen bei der Messung, Kalibrierung bzw. Konformität des Messgeräts, usw.) sowie aus dem allgemeinen Kontext zum Zweck der Daten (Fragestellung der Untersuchung, Vorliegen von Messdaten oder simulierter Daten, usw.). Die notwendigen Metadaten mit den Hintergründen der Datenerzeugung werden direkt in den Daten hinterlegt (Gegenstand des aktuellen EMPIR-Forschungsprojektes Met4FoF [51] zu einem annotierten HDF5-Datensatz für ML-Anwendungen).

- **Domänenübergreifende Semantik zur Erweiterung des KI-Interpretationsspielraums**
  In digitalen Daten werden heute oft Begriffe genutzt, die in der Form von kontrollierten Vokabularlisten oder Taxonomien in einem sehr engen Anwendungskontext definiert sind und damit auch nur einen eher engen Interpretationsspielraum erlauben. Um mittelfristig höhere Grade der maschinellen Nutzbarkeit von Daten mit KI zu erreichen, ist eine geeignete zusätzliche Semantik zur Bedeutung von Daten und Metadaten sowie zu deren Kontext aus verschiedenen Domänen erforderlich (vgl. DIN und DKE Whitepaper zu digitalen Normen [52]).

Datenorganisation und Umgang mit Daten

Für den allgemeinen Umgang mit Daten bietet der Einsatz von KI vielfältige Potentiale, die durch geeignete Strukturen und Prozesse gewinnbringend gehoben werden können. Dies betrifft die Bereiche der Datenorganisation, der dynamischen Datenauswertung, aber auch der Prozesssteuerung mittels KI. So ermöglicht KI erstmals die Automatisierung bestimmter Datenorganisationsprozesse, in einem deutlich höheren Maß als herkömmliche Software. Zu den Aufgaben, die vielversprechend mit Hilfe von KI ressourcenschonender und schneller erledigt werden, zählen

- die Datendigitalisierung (Erfassung von Informationen aus menschenlesbaren Dokumenten in eine maschinenlesbare Datenbank);
- die Datenannotation (Extraktion und Klassifikation von Metadaten, z. B. demonstriert im DiTraNo-Projekt zur Auszeichnung von DKE-Normen mit Machine-Learning-Methoden [53]);

[2] FAIR – Findable, Accessible, Interoperable, Reusable (dt. auffindbar, zugänglich, interoperabel, wiederverwendbar)

- die Datennormalisierung;
- die Bewertung der Qualität von Daten;
- gewisse Operationen zur Datenergänzung und -konsolidierung, wie Erkennung von Wertebereichen und Auflösung;
- progressive automatisch lernende Prozesse der Informationsorganisation, bspw. das Messgerätemanagement.

Darüber hinaus bietet die KI die Möglichkeit, zeitlich veränderliche Prozeduren der Datenauswertung dynamisch und adaptiv zu entwickeln. Einsatz finden könnte KI z. B. bei der (repositoriumsübergreifenden) Suche und Auswahl geeigneter Daten für eine Auswertung, routinemäßigen aber situationsabhängigen Analyseverfahren wie Anomaliedetektion oder Unsicherheitsschätzungen sowie bei komplexen, adaptiven Analyseverfahren wie die Modellbildung für komplexe Objekte (z. B. biologische, medizinische, soziologische, ökologische Systeme). Letzteres Feld bietet eine nahezu unbeschränkte Spielwiese einerseits für den Einsatz direkt an der PTB, aber auch im Umfeld des geplanten Innovationszentrums für Systemische Metrologie, welches im Schwerpunkt diese systemischen Herausforderungen adressiert.

KI kann Menschen außerdem bei Aufgaben der Prozessteuerung unterstützen, indem sie komplexe Informationszusammenhänge erkennt und in dokumentierbarer Weise Prozesse einleitet bzw. die aggregierte Information menschlichen Operator*innen zur Verfügung / Kenntnis stellt (u. a. in Form von Smart Services). Die Prozesssteuerungsbereiche der Metrologie umfassen ein sehr breites Spektrum von gesetzlich unterschiedlich regulierten Anwendungen. In diesem Rahmen haben Fragen nach der menschlichen Aufsicht, rechtlicher und ethischer Verantwortung sowie Haftung große Bedeutung. Die industrielle Messtechnik ist hier zunächst ein Innovationsort für eine kurzfristige und mittelfristige Entstehung von KI-Methoden zur Unterstützung automatisierter Prozesse und Entscheidungen. Zunächst sind es die Endnutzenden von Messtechnik, die mit Messdaten von Produktionsteilen Ausschussteile identifizieren und mit Daten aus stark vernetzter Sensorik an Fertigungs- und Messgeräten (Industrie 4.0) Änderungen und Anomalien erkennen, um Wartungsintervalle besser vorherzusagen. Im gesetzlichen Messwesen mit deutlich stärkeren Regularien und hohen Risikothematiken (z. B. Medizintechnik und Pharmazie) wird der erste Einsatz von KI-Methoden mittelfristig substanzielle Entwicklung im Feld der Qualitätssicherung von KI-Methoden und deren Ergebnissen erfordern. Fehlerhafte Ergebnisse bei KI-gestützter Datenanalyse und Auswertung, die beispielsweise zu Fehldosierung von Medikamenten führen würden, hätten für Mediziner*innen und Patient*innen fatale Folgen. Insbesondere können in kritischen Entscheidungsprozessen erst dann KI-Methoden zum Einsatz kommen, wenn es prüfbare (akkreditierbare) Software und Daten dafür gibt (vgl. EUROLAB Positionspapier zur KI-Strategie der COM [54]). Um eine Prüfung, sogar Zertifizierung, von KI-Software-Ergebnissen zu ermöglichen, wird es essentiell sein, digitale Normale in der Form von hochwertigen Referenzdaten zu entwickeln. Diese Datenstandards ermöglichen es, die Genauigkeit und Zuverlässigkeit von KI-Methoden zu messen. Wie die PTB heute bereits die hoheitliche Aufgabe übernimmt, physikalische Normale für die nationalen Messgrößen bereitzustellen, so ergibt sich mit der Bereitstellung nationaler digitaler Normale („goldene Datensätze“) eine wichtige Ergänzung im Zuge der digitalen Transformation in der Messtechnik.

Im Aufgabenfeld der PTB bieten sich mittelfristig weitere Einsatzgebiete für prozessbegleitende KI, um beispielsweise die Erzeugung von Kalibrierzertifikaten zu unterstützen, neue Verfahren zur Kokalibrierung zu entwickeln und die Datenqualität im Labor zu verfolgen.

Neben dem reinen Prozesseinsatz von KI stellt die Schlüsseltechnologie auch neue Anforderungen an das Datenhandling der PTB. Als nationales Metrologieinstitut fordert die PTB einen korrekten Umgang

mit Forschungsergebnissen im Sinne der Reproduzierbarkeit und Nachvollziehbarkeit. Daher übernimmt sie die geltenden Regelungen zum Forschungsdatenmanagement (Richtlinien von den einschlägigen Forschungsförderern; Empfehlungen der verschiedenen *FAIR Data*-Initiativen; Datenstrategie der Bundesregierung) und agiert proaktiv, um ein metrologisches Verständnis für Daten in den Diskurs zu bringen. Beim Aufbau der eigenen internen Forschungsdateninfrastruktur werden deswegen alle für die maschinelle, KI-konforme Nachnutzbarkeit relevanten Aspekte beachtet und wesentliche aktuelle Handlungsfelder angesprochen. Dazu zählen:

- Erarbeitung einer Prozedur zur nutzungsfreundlichen und zuverlässigen Bereitstellung von Daten, die eine breite Palette an Input-Formaten akzeptieren kann und die Informationen strukturiert und kohärent erfasst.
- Erarbeitung einer Prozedur zur nutzungsfreundlichen und zuverlässigen Bereitstellung von Metadaten und der gesamten Arbeitsdokumentation. Metadaten sollen sowohl manuell als auch automatisiert per Crawler aus den Dateien extrahiert werden und über Schnittstellen zugänglich sein.
- Die Etablierung geeigneter Arbeitsverfahren für die Handhabung großer Datenmengen, ggf. auf Basis einer Datenkompression oder des *Git Large File Storage* (https://git-lfs.github.com/), das die Daten durch persistente Identifikatoren referenziert und aufruft statt sie in eine Datenbank zu „schieben".
- Schutz der Daten vor Manipulation; Gewährleistung von deren Integrität und Authentizität; bei Bedarf mit erhöhten Sicherheitsmaßnahmen wie z. B. Verschlüsselung
- Finden einer Balance zwischen dem Bedarf nach offenen Trainings- und Testdaten und datenschutzrechtlichen Aspekten, insbesondere bei medizinischen Daten. Das wird durch engen Austausch mit einschlägigen Fachinitiativen (z. B. gemeinsam mit medizinischen Kooperationspartnern im Rahmen von AI4Health) und Rechtsexperten (Justiziariat) bewerkstelligt und ist ein laufender Prozess.

Das kürzlich aktualisierte Gesetz für die Nutzung von Daten des öffentlichen Sektors (Datennutzungsgesetz – DNG) stellt seinerseits Anforderungen an die Verfügbarkeit, Strukturierung, Lizenzierung von Daten öffentlicher Relevanz, um deren Nachnutzung zu ermöglichen. Unter anderem fordert es:

- Verwendung objektiver, verhältnismäßiger, nichtdiskriminierender und durch ein im Allgemeininteresse liegendes Ziel gerechtfertigter Lizenzen, die Nutzungsmöglichkeiten nicht unnötig einschränken – auch nicht die kommerziellen (§ 4)
- Bereitstellung von Daten und Metadaten in offenen, maschinenverständlichen, interoperablen Formaten, womöglich sprachenunabhängig, über geeignete Anwendungsprogrammierschnittstellen und, falls technisch erforderlich, als Massen-Download (§§ 7-9).

Datenorganisation für KI im Verbund

Im Schnittfeld der Handhabung und Qualitätssicherung von Forschungsdaten im Sinne der FAIR-Prinzipien, Open Data und KI-Anwendungen kommen der Nationalen Forschungsdateninfrastruktur (NFDI) auf deutscher Ebene und der European Open Science Cloud (EOSC) auf europäischer Ebene zentrale Rollen zu. Ihren Auftrag und Mehrwert fasst die NFDI wie folgt zusammen[3]:

> *„In der Nationalen Forschungsdateninfrastruktur (NFDI) werden wertvolle Datenbestände von Wissenschaft und Forschung für das gesamte deutsche Wissenschaftssystem systematisch erschlossen, vernetzt und nachhaltig sowie*

[3] https://www.nfdi.de/verein/#kurzinfo, Zugriff: 07.09.2021

qualitativ nutzbar gemacht. Bislang sind sie zumeist dezentral, projektbezogen oder auf Zeit verfügbar. […] Mit der NFDI soll ein dauerhafter digitaler Wissensspeicher als unverzichtbare Voraussetzung für neue Forschungsfragen, Erkenntnisse und Innovationen geschaffen werden. Relevante Daten sollen nach den FAIR-Prinzipien […] zur Verfügung gestellt werden."

Zielgruppe der NFDI sind also in erster Linie Forschende an universitären wie außeruniversitären Forschungseinrichtungen. Somit ergeben sich große Überschneidungen mit der potentiell an der Nutzung von KI-Daten und -Diensten der PTB interessierten Community. Die PTB setzt in ihrem Engagement in der NFDI einen Schwerpunkt auf Datenqualität (insbesondere Mechanismen der Qualitätssicherung und der Rückführbarkeit der Qualität von Forschungsdaten) sowie der Dokumentation von Forschungsdaten durch fachlich passgenaue und semantisch hochwertige Vokabulare und Ontologien. Gleichzeitig wird die PTB in der Forschungswelt bereits als (mögliche) Referenz und Vorbild in Fragen der „Guten Wissenschaftlichen Praxis" wahrgenommen. Eine angestrebte Rolle der PTB als Datentreuhänderin birgt also die Chance, diesem guten Ruf gerecht zu werden und Forschungskooperationen sowie Datendienstleistungen in Deutschland und europaweit auszubauen.

In diesem Rahmen sollte die PTB ein zu ihrem Auftrag passendes Portfolio an Forschungsdaten- und KI-Diensten als langfristige Infrastruktur-Aufgabe übernehmen: Als Ergänzung zu den in fachlichen Konsortien geförderten Vorhaben ist für 2022 der Aufbau eines Basisdienst-Konsortiums in der NFDI geplant, damit „die infrastrukturelle Grundversorgung für potenziell alle Konsortien gewährleistet und Interoperabilität dauerhaft gesichert wird"[4]. Dies ist notwendig, da die NFDI zwar als dauerhafte Infrastruktur konzipiert, jedoch gegenwärtig in einer Projektstruktur umgesetzt ist. Die PTB ist besonders geeignet, zu einer solchen notwendigen Grundversorgung dauerhaft beizutragen und als für Wissenschaft und Wirtschaft gleichermaßen vertrauenswürdiger Akteur wahrgenommen zu werden.

Datendienstleistungen für KI

Neben der forschungsgetriebenen öffentlichen und kostenlosen Bereitstellung KI-geeigneter Daten wird die PTB auch im Rahmen ihrer hoheitlichen Aufgaben im industriellen und gesetzlichen Messwesen entsprechende Dienstleistungsangebote schaffen, um einheitliche Qualitätsstandards für die Entwicklung und Etablierung von KI im Messwesen zu fördern. Nachstehend sind Bereiche gelistet, für die bereits heute ein großer Bedarf abzusehen ist.

- Weiterentwicklung bestehender Dienstleistungen aus dem gesetzlichen Messwesen, bei denen tiefgreifende Quelltextanalysen von Software nötig sind, mit zusätzlichen Verfahren und Bewertungskriterien bei Software mit KI (insbesondere zu Nachvollziehbarkeit von KI).
- Nutzung des TraCIM-Testsystems für die automatisierte Online-Validierung von KI-Software und Prozeduren mit qualitativ hochwertigen Referenzdaten (Siegel „QI-Digital für KI-Software").
- Nutzung des TraCIM-Testsystems für die automatisierte Online-Validierung von Daten nach ihrer Eignung für eine Weiternutzung mit KI-Methoden (Siegel „QI-Digital für KI-Daten").
- Erzeugung und Bereitstellung hochwertiger Referenzdaten für KI-Anwendungen für Kunden. Zur Erzeugung der Daten kommen verschiedene Methoden zum Einsatz wie die Entwicklung aus bestehenden Datensätzen durch Referenzsoftware oder simulierte (künstlich generierte)

[4] https://www.dfg.de/foerderung/info_wissenschaft/2021/info_wissenschaft_21_37/index.html, Zugriff: 27.09.2021

Daten, die klare und eindeutige Eigenschaften haben, auf die es bei der Entwicklung und Prüfung von KI-Methoden ankommt („PTB – Goldene Datensätze“).

- Verwahrung und Bereitstellung hochwertiger Daten für KI für Kund*innen (PTB als Datentreuhänder).

Besonders im Fokus der Dienstleistungen sind weitere Vertrauensmerkmale für digitale Daten von großer Bedeutung. Zudem müssen Verfahren zur Sicherstellung der Authentizität (Herausgeber), Integrität (Manipulationsschutz), Vertraulichkeit (Verschlüsslung, Wahrung der Anonymität/ Pseudonymität) sowie zur langfristigen Aufbewahrung und Bereitstellung von Daten installiert werden.

Ordnungsrahmen

Die PTB setzt es sich zum Ziel, ihre Rolle als wichtige Säule der Qualitätsinfrastruktur innerhalb eines Ordnungsrahmens für KI proaktiv zu gestalten, Prozessabläufe auf Grundlage der neuen Anforderungen und Möglichkeiten zu überarbeiten und in der Standardisierung sowie der Bewertung und Zertifizierung von KI ihre metrologische Expertise engagiert einzubringen.

Die wachsenden Einsatzmöglichkeiten von KI bieten einerseits ein großes wirtschaftliches Potential für innovative Technologien und eine gesteigerte Wettbewerbsfähigkeit Deutschlands und Europas auf dem globalen Markt, aber sie stellen andererseits auch neue Herausforderungen an die bestehende nationale und internationale Qualitätsinfrastruktur. Die Weiterentwicklung der QI unter Berücksichtigung der besonderen Eigenschaften von KI ist von elementarer Bedeutung, um das Vertrauen der Menschen in Produkte und Dienstleistungen zu sichern und einen klaren Sicherheits- und Haftungsrahmen zu schaffen [1, 3, 55]. Grundgedanke eines derart angepassten Ordnungsrahmens ist es, umfassenden Verbraucherschutz und Rechtssicherheit für Unternehmen zu bieten und damit eine frühzeitige und nachhaltige Akzeptanz von KI-Technik zu begründen.

Als starken und essentiellen Partner in der QI adressiert die Bundesregierung also auch die PTB mit ihrer Aufforderung zur Schaffung eines geeigneten, an KI-spezifische Belange angepassten Ordnungsrahmens [3]. Explizit formuliert wird dieser Auftrag in der Fortschreibung der KI-Strategie [3]:

> *„Zusammen mit Metrologie, Akkreditierung, Konformitätsbewertung, Marktüberwachung und Umweltprüfungen bilden Regeln, Normen und Standards die Qualitätsinfrastruktur – das Rückgrat der Marke „Made in Germany“. Die Qualitätsinfrastruktur ist somit ein wesentlicher Garant unseres wirtschaftlichen Erfolges und des Vertrauens in Produkte und Dienstleistungen. Die Bundesregierung wird die Weiterentwicklung und Stärkung der nationalen und europäischen Qualitätsinfrastruktur hinsichtlich der Nutzung und Behandlung von KI-Methoden fördern, um damit den Marktzugang insbesondere von KMU in Europa und weltweit zu unterstützen. Auch die Qualitätssicherung der Daten, zum Beispiel durch Benchmark-Tests, Referenzdaten, Aufbau und Kuratierung von Trainingsdatenpools und Einrichtung von Testdatensätzen zur Validierung von Algorithmen ist sicherzustellen, damit eine vertrauenswürdige Anwendung von KI-Methoden ermöglicht wird. Die Einbindung der Anwendenden sollte ebenfalls berücksichtigt werden.“*

Um national abweichende Regulierung von KI-Anwendungen im EU-Binnenmarkt zu verhindern, Investments in Innovationen im KI-Bereich zu befördern und gleichzeitig den hohen Anforderungen an Sicherheit und Rechtsschutz gerecht zu werden, veröffentlichte die EU-Kommission im April 2021 einen Entwurf für einen harmonisierten europäischen Rechtsrahmen („Artificial Intelligence Act“) [55]. Dieser Entwurf behandelt KI-Systeme entsprechend definierter Risikoklassen:

- unannehmbares Risiko (z. B. Social Scoring) bei KI-Anwendungen, die unvereinbar mit den Grundrechten der Bürger*innen und den Werten der EU sind,
- hohes Risiko für eine Liste von KI-Anwendungen (z. B. biometrische Personenerkennung in Echtzeit, Management und Betrieb kritischer Infrastrukturen etc.), welche entweder als Sicherheitskomponente von Produkten verwendet werden, die entsprechend der

harmonisierten europäischen Rechtsakte einer Konformitätsbewertung durch Dritte unterliegen, oder wegen ihres starken Eingriffs in die Grundrechte gesondert gelistet sind,
- geringes Risiko (z. B. Chatbots), welche besonderen Verpflichtungen zur Transparenz unterliegen, sowie
- minimales Risiko bei KI-Anwendungen (z. B. Rechtschreibprüfung), deren Sicherheitsprüfung gemäß des Rechtsrahmens lediglich auf freiwilliger Basis angeraten wird.

Die Kriterien für die Einteilung in verschiedene Risikoklassen werden für diesen Rechtsrahmen verbindlich festgeschrieben, die Risikobewertung einzelner Anwendungsfälle bleibt jedoch offen neue technologische Entwicklungen und entsprechend veränderte Risikoabschätzungen. Für Hochrisiko-KI-Systeme besteht vor dem Inverkehrbringen die Verpflichtung der Anbieter zu einer Konformitätsbewertung, die für bestimmte Produkte durch benannte, unabhängige Bewertungsstellen erfolgen muss. Mit dieser Konformitätsbewertung wird die Vertrauenswürdigkeit der KI-Anwendung in Bezug auf Datenqualität, technische Dokumentation, Transparenz und Informationsauskunft, menschliche Aufsicht, Robustheit, Genauigkeit und Cybersicherheit zum Zeitpunkt des Inverkehrbringens sichergestellt. Zusätzlich sind Anbieter von KI-Systemen mit hohem Risiko verpflichtet, erweiterte Qualitäts- und Risikomanagementsysteme einzurichten. Diese umspannen den gesamten KI-Produktlebenszyklus, d. h. sie gewährleisten auch nach dem Inverkehrbringen eine Rückkopplung der Nutzenden zum laufenden Betrieb und möglichem Fehlverhalten der KI-Systeme. Bei wesentlicher Veränderung des Einsatzzwecks eines Hochrisiko-KI-Systems oder auch des Systems an sich, wird eine erneute Konformitätsbewertung erforderlich. Hochrisiko-KI-Systeme eingebettet in Produkte, welche nach dem New Legislative Framework der EU bereits einer Konformitätsbewertung unterliegen, werden auf die Einhaltung des neuen Rechtsrahmens für KI innerhalb des bestehenden Konformitätsbewertungsverfahrens geprüft, um Doppelung und Mehraufwand für die relevanten Stellen zu vermeiden. Insbesondere betrifft dies das Zusammenspiel mit der Maschinenverordnung.

Bei der Umsetzung des Rechtsrahmens auf nationaler Ebene, obliegt es den Mitgliedstaaten, entsprechende zuständige Behörden für den KI-Ordnungsrahmen zu benennen. Mit ihrer starken Rolle als Konformitätsbewertungsstelle für das Messwesen sieht sich die PTB daher prädestiniert, auch für Messgeräte mit KI-Komponenten oder KI-Gesamtsysteme Konformitätsprüfungen sowie geeignete Prüfprozessabläufe zu entwickeln. Ausgehend von bestehenden Strukturen und Prozessen für Produkte und Dienstleistungen ohne KI innerhalb der Qualitätsinfrastruktur, baut die PTB neue Kompetenzen auf und verknüpft diese mit ihrem Domänenwissen, um die Anforderungen des EU-Rechtsrahmens und zukünftiger nationaler Vorgaben für KI angemessen erfüllen zu können.

Unterstützend zu diesen nationalen Aktivitäten, sieht die Verordnung vor, einen europäischen Ausschuss für künstliche Intelligenz einzurichten, welcher sich aus den benannten, nationalen Aufsichtsbehörden für KI zusammensetzt. Des Weiteren ist die Einrichtung einer von der europäischen Kommission beaufsichtigten Plattform für Hochrisiko-KI-Systeme geplant, in der Anbieter ihre Produkte registrieren müssen. Im Falle von Verstößen gegen den neuen Rechtsrahmen detailliert die Verordnung entsprechende Sanktionsmaßnahmen.

Zudem sieht die Verordnung regulatorische „Sandboxes“ (sogenannte Reallabore) vor, in denen innovative KI-Systeme entwickelt und geprüft werden sollen. Den zuständigen nationalen Aufsichtsbehörden wird dabei explizit die Aufgabe zugedacht, innovationsfreundliche Rahmenbedingungen für diese Experimentierfelder zu schaffen, um eine sichere und vorausschauend regulierte Nutzung von KI zu ermöglichen. Ein entsprechendes Konzept für ein Reallabore-Gesetz [56], das derartige, einheitliche und innovationsfreundliche Rahmenbedingungen in Deutschland schaffen soll, hat das BMWi kürzlich vorgelegt. Diese Experimentierfelder bieten der PTB eine hervorragende

Möglichkeit, auch in Partnerschaft mit dem von der PTB geplanten „Innovationszentrum für Systemische Metrologie“ (IZSM) einen signifikanten Beitrag zum Grundverständnis und zur Qualitätssicherung von auf Messdaten aufbauenden KI-Anwendungen zu leisten. Gemeinsam mit dem IZSM, anderen Akteuren der QI sowie geeigneten Unternehmen können z. B. für die Themenfelder „autonomes Fahren“, „digitale Medizin“ und „Stadt der Zukunft“ unbedingt erforderliche Bewertungsgrundlagen für die Qualität von Daten, „goldene Datensätze“ für Training und Testen sowie Benchmarktests für KI-Verfahren entwickelt werden. Ein entsprechendes Konzept für mögliche Handlungsfelder des IZSM im Bereich KI sowie die komplementäre Rollenverteilung in der Zusammenarbeit mit der PTB liegt dem BMWi bereits vor [57].

Grundsätzlich ist im Verordnungsentwurf auch die Definition des Begriffes KI zu hinterfragen, da diese sehr breit gefasst ist und bekannte statistische Ansätze, Bayes‘sche Schätz-, Such- und Optimierungsmethoden einschließt. Im Hochrisiko-Fall würden diese Anwendungsbeispiele entsprechend der Verordnung somit ggf. verschärften Konformitätsbewertungen für KI unterliegen, die für die konventionelle Prüfung dieser Produkte im bestehenden Rechtrahmen nicht vorgesehen wäre. Kritik an dieser weiten Auslegung des KI-Begriffs wird an verschiedenen Stellen geäußert, unter anderem in der Stellungnahme des Zentralverbandes der Elektrotechnik- und Elektronikindustrie [58]. Auch in den Ausschüssen des Bundesrats werden die Implikationen der KI-Verordnung diskutiert und entsprechende Empfehlungen zu Anpassungen formuliert [59]. Die sehr weite Definition von KI-Systemen stößt auch hier auf Kritik, falls dadurch zusätzliche, wirtschaftshemmende Regulierung nötig werden sollte. Grundsätzlich findet die EU-einheitliche und risikobasierte Herangehensweise sowie die Ausrichtung auf die Stärkung der Wirtschaft und den Schutz der Bürger*innen jedoch den Zuspruch des Bundesrates. Des Weiteren betont er die Ausnutzung der Chancen von KI und den unbedingt nötigen Schutz der Wirtschaft vor unangemessenen Belastungen durch übermäßige oder intransparente Regulierung. Prüfprozesse und Dokumentations- bzw. Transparenzpflichten sollen verschlankt und eine Doppelbelastung vermieden werden. Es bleibt abzuwarten, inwiefern die Verordnung unter Berücksichtigung dieser und weiterer Anmerkungen aus den Mitgliedsstaaten noch angepasst wird.

Standardisierung und Regulierung von KI

Der neue Rechtsrahmen der EU für KI spricht der Standardisierung eine Schlüsselrolle zu [55]. Auf nationaler Ebene werden die Fragen der Normung, Prüfbarkeit und Auditierbarkeit federführend von DIN und DKE adressiert und auf europäischer Ebene in CEN, CENELEC und ETSI sowie international in ISO, IEC und ITU vertreten. Im Positionspapier [60] von DIN und DKE zum Entwurf der europäischen KI-Verordnung wird diese gewichtige Rolle betont und eine entsprechende Repräsentation der Standardisierungsbehörden im vorgesehenen europäischen KI-Board gefordert. Zudem drängen die Organisationen darauf, zeitnah Standardisierungsanfragen zu formulieren, da Vorarbeiten in der Standardisierung für die Umsetzung des Rechtsrahmens unabdingbar sein werden. Diese erfolgen derzeit unter anderem gesteuert durch eine, die Steuerungsgruppe der Normungsroadmap KI ablösende, KI-Koordinierungsgruppe in geeigneten Leuchtturmprojekten mit Partnern wie dem Fraunhofer HHI, der Charité, dem BSI und der PTB.

Ein Impulspapier der Stiftung Neue Verantwortung [61] betont insbesondere drei Merkmale von KI, welche die Herangehensweise von Standardisierung und Zertifizierung an KI grundlegend neu gestalten:

- technische Standards sind durch die hohe Entwicklungsdynamik von KI schnell überholt und erfordern ständige Anpassung der teils langwierigen Prozesse,
- die Definition und Überprüfung technischer Anforderungen wird durch die probabilistische Natur der KI-Systeme stark erschwert und

- die starke Kontextabhängigkeit mit großer gesellschaftlicher Trageweite von KI als soziotechnische Basistechnologie erfordert eine besondere Berücksichtigung in Standardisierung und Zertifizierung.

Als Ergebnis nationaler Aktivitäten sind grundsätzliche Anforderungen und Terminologien für die Bewertung von KI-Methoden bereits in Grundzügen erarbeitet. In einem aktuellen Whitepaper zu auditierbaren KI-Systemen von TÜV-Verband, BSI und Fraunhofer HHI [62] werden die Qualitätsdimensionen detailliert aufgeschlüsselt und betrachten neben technischen Prüfanforderungen auch regulatorische Kriterien wie ethische Leitlinien sowie rechtliche und gesellschaftliche Rahmenbedingungen (Abb. 3).

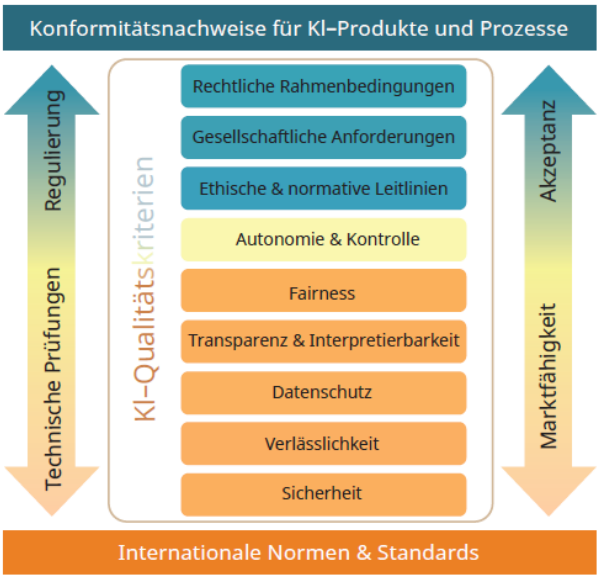

Abb. 3 Kategorisierte Qualitätsdimensionen für die Bewertung von KI in der Konformitätsprüfung [10].

Die stärker technisch ausgerichtete DIN SPEC 92001-1 [11] definiert drei wesentliche Anforderungen an die Qualität von KI:

- *Funktionalität und Performance* als Ausdruck der Fähigkeit der KI, die gestellte Aufgabe unter festgelegten Bedingungen zu erledigen (Verlässlichkeit);
- *Robustheit* als Fähigkeit der KI mit fehlerhaften, verrauschten, unbekannten oder schädlichen Eingangsdaten umgehen zu können;
- *Erklärbarkeit* als Ausdruck für die Möglichkeit, die Gründe für das Ergebnis einer KI-Methode verstehen und nachvollziehen zu können.

Die DIN SPEC 92001-2 [12] konkretisiert den Begriff der Robustheit und unterscheidet zwischen *adversarial robustness* (AR) und *corruption robustness* (CR). Erstere bezeichnet die Robustheit gegenüber schadhaften (adversarial) Änderungen an den Eingangsdaten, letzteres steht für Robustheit gegenüber Rauschen oder Veränderung der statistischen Eigenschaften der Eingangsdaten. Für die Entwicklung von robusten KI-Methoden empfiehlt [12] einen Risikoanalyse-basierten Ansatz. Dazu werden auch Methoden zum gezielten Testen von KI-Methoden genannt (Fast Gradient Sign Method; Projected Gradient Descent). Allgemein wird ein Szenario-basiertes Testen empfohlen, welches den späteren Einsatzzweck und dessen Eigenschaften mit einbezieht. Als wichtig wird erachtet [12], dass die Risiko-Bewertung einer KI-Methode eigentlich kontinuierlich durchgeführt werden muss, wie in Abb. 4 dargestellt.

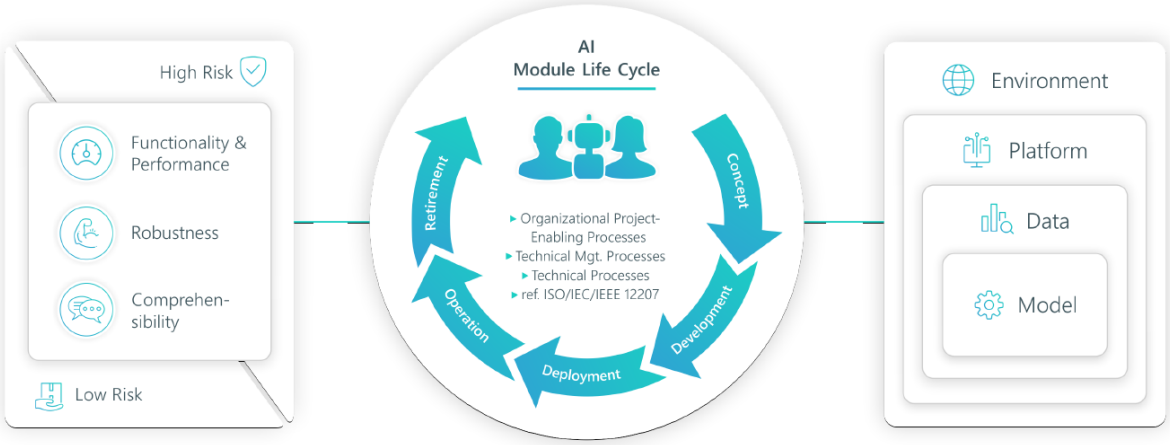

Abb. 4 KI-Lebenszyklus im Qualitätsmetamodell der DIN SPEC 92001-2 [12].

Auch die US-amerikanische *Food and Drug Administration* (FDA) geht in einem aktuellen Diskussionspapier [30] von der Notwendigkeit eines „Total Product Lifecycle Regulatory Approach“ (TPLC) für KI-Anwendungen aus. Im Zuge der Zertifizierung eines KI-basierten Medizinprodukts soll dabei insbesondere das Qualitätsmanagement des Unternehmens begutachtet werden hinsichtlich

- Qualitätssicherung in der Softwareentwicklung und
- Tests und Performance-Monitoring der Produkte.

Dabei stellt die FDA folgende grundsätzliche Prinzipien auf:

- Etablierung anerkannter „good machine learning (ML) practices“;
- Berücksichtigung des Produktlebenszyklus bei der Zulassung KI-basierter Medizinprodukte;
- Erwartung, dass die Hersteller einen Risiko-basierten Ansatz für das Monitoring ihrer KI-basierten Medizinprodukte für den gesamten Produktlebenszyklus realisieren;
- Transparente Aussagen für Kunden und Prüfer zu tatsächlicher Leistung und Verhalten von KI-basierten Medizinprodukten durch die Hersteller.

Dazu gehört für die FDA auch das Dokumentieren des geplanten Einsatzbereiches (Software as a Medical Device (SaMD) Pre-Specification – SPS) sowie der (Weiter-)Entwicklung in einem „Algorithm Change Protokoll“ (ACP): Data management, Re-training, Performance evaluation, Update procedures. SPS und ACP sind dann wesentliche Punkte bei der Zulassung neuer Produkte. Auch die Normungsroadmap KI des DIN [10] fordert eine abgestufte Schwelle für Normung und Zulassung abhängig vom geplanten Einsatzbereich der KI-Anwendung über die sogenannte risikoadaptive Kritikalitätsprüfung.

Ein wichtiger Teil des Qualitätssicherungs- und -Managementsysteme ist laut Aussage der FDA [30] die Wahl der Trainings- und Testdaten sowie die Auswahl von Anwenderdaten für das Re-Training. Beim Datenmanagement sieht die FDA daher

- Protokolle zur Datenerhebung,
- Qualitätssicherungssysteme für die Daten,
- Bestimmung eines Referenzstandards und
- Auditierung und Sicherung von Test- und Trainingsdaten

durch die Hersteller vor. Auch der Fragenkatalog der deutschen „Interessengemeinschaft der Benannten Stellen für Medizinprodukte in Deutschland“ (IG-NB) für die Zulassung von KI bei

Medizinprodukten widmet viele Fragen der Auswahl und Beurteilung der verwendeten Daten [63]. Gleichzeitig fehlen gerade zu den wichtigen Fragen für die Bewertung der KI und der zugrundeliegenden Daten in [63] entsprechende Normen und Standards als Grundlage für die Prüfung.

Zertifizierung von KI

Im Whitepaper des Fraunhofer IAIS [17] wird eine durch akkreditierte Prüfer operativ durchführbare Zertifizierung für KI-Anwendungen diskutiert. Demzufolge soll ein Zertifikat für KI

- einen gewissen Qualitätsstandard bescheinigen,
- dabei helfen, KI-Anwendungen überprüfbar rechtskonform zu gestalten und
- KI-Anwendungen vergleichbar machen.

Dabei stellt das IAIS fest, dass für die Bewertung der Verlässlichkeit Domänenwissen und mathematisch-statistische Expertise notwendig sind [17].

Das Fraunhofer IPA benennt im „White Paper: Zuverlässige KI“ [64] die Zertifizierung gemeinsam mit Transparenz auf Systemebene als Schlüsselfaktoren für zuverlässige KI und skizziert eine entsprechende Sicherheitsargumentation (AMLAS, „Assurance of Machine Learning for Use in Autonomous Systems“ [65]) für die Entwicklung vertrauenswürdiger KI-Verfahren. Insbesondere beleuchtet [64] verschiedene Methoden der aktuellen Forschung, die zur Zertifizierung im Rahmen eines AMLAS beitragen könnten:

- Erklärbare KI
- formale Verifikation
- Statistische Validierung
- Unsicherheitsquantifizierung
- Online-Monitoring mit Randbedingungen.

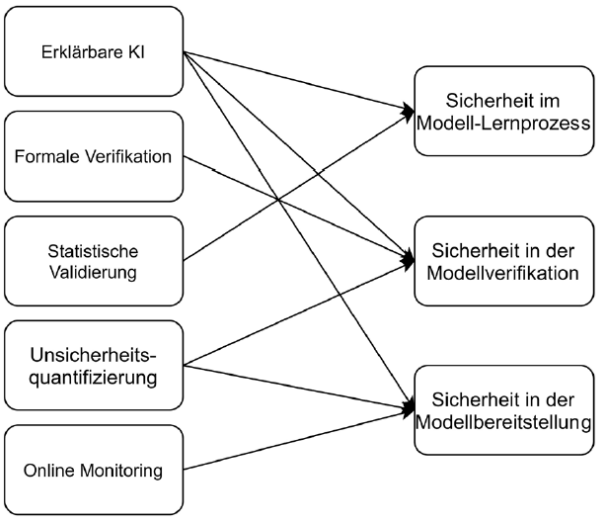

Abb. 5 Methodenbaukasten für die Entwicklung zuverlässiger KI innerhalb der AMLAS-Sicherheitsargumentation [64].

EUROLAB macht eine deutliche Unterscheidung zwischen indirekter (indirect conformity assessment – ICA) und direkter (direct conformity assessment – DCA) Anwendung von KI-Methoden [54]. Bei ICA dient die KI-Methode als Unterstützung für die Entscheidungsfindung. Hier wird das Beispiel einer KI-Methode für die Auswertung einer Röntgen-Messung verwendet: Die KI-Methode ermittelt aus den Messdaten eine qualitative Aussage, z. B. über den Gesundheitszustand des Patienten. In diesem Fall wäre bei einer Akkreditierung keine Bewertung der KI-Methode an sich notwendig, sondern es würde die Kompetenz des Personals der zu akkreditierenden Stelle festgestellt werden müssen mit der

Methode umzugehen. Das entspräche dem bereits heute praktiziertem Vorgehen für beliebige andere nichtlineare numerische Verfahren. Wenn die KI-Methode jedoch das Ergebnis als Teil der Messung präsentiert (z. B. als Overlay) und damit den Anschein eines „echten Ergebnisses“ erweckt, ist die KI-Methode selbst als „Autorität“ zu verstehen und in der Akkreditierung mit zu beachten. EUROLAB empfiehlt derzeit, DCA nur in sehr unkritischen Bereichen einzusetzen (z. B. Bewertung von Musikqualität), bis die Methoden ausgereifter sind. Im Positionspapier von EUROLAB [54] wird auch für KI-Methoden eine Art Kalibrierung wie für gängige Messmittel gefordert. Diese sollte die Verlässlichkeit der KI-Methode bewertbar machen, indem die Fähigkeit der Methode erfasst wird, das Ergebnis eines „Standards“ zu reproduzieren.

Im Whitepaper „Zertifizierung von KI-Systemen“ [29] der Plattform Lernende Systeme heißt es:

> *„Bevor eine gelungene Zertifizierung von KI-Systemen etabliert werden kann, sind daher noch offene Fragen zu klären. Diese betreffen den Gegenstand der Zertifizierung, die Prüfkriterien, den Zeitpunkt und die Notwendigkeit der Zertifizierung, den Detailgrad der Zertifizierung sowie den Umgang mit weiterlernenden Systemen.“*

Diese Einschätzung unterstreicht auch das Whitepaper „Towards Auditable AI“ [62] von TÜV-Verbund, BSI und Fraunhofer HHI und schlägt für den Prozess zur Etablierung erfolgreicher Prüfungen von KI-Systemen zwei parallel einzuleitende strategische Herangehensweisen vor:

- Wahl geeigneterer (eingeschränkter) Rahmenbedingungen (z. B. Komplexität, Skalierbarkeit, Generalisierbarkeit), um akzeptable IT-Sicherheit, Audit-Qualität, Robustheit und Verifizierbarkeit für einzelne Anwendungen zu erreichen,
- verstärkte Investition in KI-Forschung und -Entwicklung, um sichere KI-Anwendung graduell auf komplexe Rahmenbedingungen auszuweiten (Erhöhung von Skalierbarkeit, Generalisierbarkeit u. a.).

Außerdem verweist [29] auch auf die AI *High Level Expert Group* der EU-Kommission (COM), welche folgende Kriterien bei der Regulierung von KI empfiehlt

> *„Vorrang menschlichen Handelns und menschlicher Aufsicht, technische Robustheit und Sicherheit, Privatsphäre und Datenqualitätsmanagement, Transparenz, Vielfalt, Nichtdiskriminierung und Fairness sowie gesellschaftliches und ökologisches Wohlergehen und Rechenschaftspflicht.“*

Insbesondere für KI-Anwendungen mit hohem Risiko (wie in den Bereichen Gesundheit, biometrische Erkennung und kritische Infrastrukturen) wird empfohlen, bei der Konformitätsbewertung zu prüfen,

- ob die Trainingsdaten adäquat sind für den geplanten Einsatzzweck;
- die Ergebnisse bei der Nutzung nicht zu Diskriminierung führen;
- Datenschutz und Privatsphäre beachtet werden;
- Relevante Aufzeichnungen zu Datensätzen, Trainingsmethoden und Programmiermethoden vorliegen.

Damit folgen diese Empfehlungen im Grunde denen der FDA [30], die (ebenso wie die Bundesregierung in ihrer Stellungnahme zum COM Whitepaper) außerdem eine wiederholte Prüfung dieser Kriterien bei lernenden – also sich verändernden – KI-Systemen als notwendig erachtet.

Das französische Metrologieinstitut LNE wählt für die Zertifizierung von KI vergleichbar zur Einschätzung der FDA ebenfalls eine prozessorientierte Herangehensweise. Anstatt die Funktionalität des KI-Systems an sich zu zertifizieren, werden die Prozessschritte entlang des Designs, der

Entwicklung, der Evaluierung und des Betriebs von KI-Systemen durch das LNE in einem Zertifizierungsstandard [66] erstmals geregelt.

Auch in einem kürzlich aufgesetzten Flagship-Projekt des DIN werden entsprechende Leitlinien für KI-Zertifizierung entlang des KI-Lebenszyklus erarbeitet mit den Schwerpunkten:

- Erarbeitung standardisierungsreifer Prüfkriterien und -methoden für KI-Systeme,
- Entwicklung von Absicherungsmethoden und Prüfwerkzeugen sowie
- Transfer in kommerzielle Angebote.

Geplant ist dafür ein breit angelegter Beteiligungsprozess, in den sich die PTB mit ihrer Expertise zu Prüfung und Bewertung entsprechend ihrer Möglichkeiten aktiv einbringen wird.

Schlussfolgerungen für die Zuständigkeit der PTB

Entsprechend ihres gesetzlichen Auftrages führt die PTB bereits in großem Umfang Konformitätsbewertungen und Prüfungen von Messgeräten und Software durch. Durch die vielfältigen Einsatzmöglichkeiten von KI in Messgeräten und Sensoren aller Art ist abzusehen, dass KI zukünftig ein neuer und wesentlicher Bestandteil vieler Produkte wird. Die Verantwortlichkeit für die Anpassung der zugehörigen Prüfungs- und Bewertungsprozesse formuliert die Plattform Lernende Systeme sehr klar:

> *„Für die Konformitätsbewertung sollte auf bestehende nationale Strukturen und Verfahren zurückgegriffen werden. Sofern es keine solche Behörden gibt, sollte es eine Pflicht zum Aufbau einer solchen Behörde oder zum Aufbau von Zuständigkeiten in bestehenden Behörden geben.“ [29]*

Die Bunderegierung bekräftigt in ihrer Stellungnahme [6] zum EU-Whitepaper diese Einschätzung. Die PTB als wesentlicher Bestandteil der Qualitätsinfrastruktur und mit anerkannter Neutralität ist damit also prädestiniert, diese Rolle zu übernehmen. Konkret ist die PTB bereits heute aufgefordert, ihren gesetzlichen Aufgaben auch für Messgeräte mit KI-Anteilen gerecht zu werden, indem sie entsprechende Kompetenzen kontinuierlich aufbaut und sich auch in der Normung mit ihrem Domänenwissen aktiv einbringt. In ihrer Stellungnahme zum Weißbuch KI der COM legt die Bundesregierung dabei folgende Empfehlungen vor:

- Für Trainingsdaten von KI-Systemen sollten verbindliche rechtliche Anforderungen in Betracht gezogen werden. Dafür sollten konsistent auch Anforderungen für Test- und Evaluierungsdaten in Betracht gezogen werden.
- Auch rechtliche Anforderungen für Qualitätsparameter und -anforderungen für Trainings-, Test- und Evaluierungsdaten sind erforderlich, damit entsprechende KI-Systeme mit quantitativ ausreichenden und qualitativ hochwertigen Datensätzen entwickelt werden.
- Anforderungen bzgl. Robustheit und Genauigkeit sollten erkennbare und realistische Szenarien abdecken.
- Wenn geeignete Verfahren zur Prüfung der KI-Ergebnisse auf Repräsentativität und Ausgewogenheit zur Verfügung stehen, kann auch auf den Zugriff auf die Trainings-/Testdaten verzichtet werden.
- Zentral ist die Sicherstellung der Vertraulichkeit, Integrität und Verfügbarkeit des KI-Systems als solches über dessen gesamten Lebenszyklus.
- Für KI-Systeme mit hohem Risiko sollte das Durchlaufen eines objektiven Konformitätsbewertungsverfahrens verbindlich vorgeschrieben werden. Dabei besteht eine Notwendigkeit wiederholter Bewertungen von sich weiterentwickelnden lernfähigen KI-Systemen.

- Entscheidend sind die Genauigkeit und Relevanz der Daten. Darüber hinaus ist es notwendig, die Bereitstellung von Referenzdaten, Benchmarktests und die Überprüfung von Algorithmen anhand von qualitätsgesicherten, vertrauenswürdigen Referenzdaten zur Verfügung zu stellen. Grundsätzlich wird die Aussage unterstützt, dass die Datenqualität während der gesamten Nutzungsdauer gewährleistet sein muss. Es ist aber zu beachten, dass dies nur vom Betreiber geleistet werden kann, der wiederum kein Wirtschaftsakteur im Sinne des Produktsicherheitsrechts ist.
- Für den Fall, dass aufgrund einer Software-Änderung ein neues Produkt entstanden ist, muss dieses Produkt vollumfänglich dem Stand der Technik entsprechen, da ein neues Inverkehrbringen vorliegt. Dies muss bei der Betrachtung einer Software-Änderung mitberücksichtigt werden.

Die Forschungsaktivitäten der PTB im Bereich KI müssen demnach sicherstellen, dass diese Anforderungen und Erwartungen erfüllt werden können. Darauf aufbauend muss die PTB Lösungen entwickeln, um Produkte mit KI zu prüfen und zu bewerten. Beispielhaft wird im Folgenden diese Entwicklung für verschiedene fachliche Bereiche der PTB skizziert.

Im Bereich der Dosimetrie befindet sich die Entwicklung KI-gestützter Methoden noch in den Anfängen, doch zeichnen sich bereits erste Anwendungsfelder ab. Der Umgang mit solchen Anwendungen sollte in den relevanten Normungsgremien zur Dosimetrie (z. B. IEC TC62 SC62C WG 3) kritisch diskutiert werden und kann auf aktuelle Arbeiten zu allgemeineren harmonisierten Normen (ISO) für die KI-Anwendung aufsetzen. Eine besondere Herausforderung im Bereich der Diagnostikdosimeter bestand bereits in der Vergangenheit darin, dass im Rahmen der Konformitätsbewertungen Prüflinge potenziell mittels Software zu stark auf die Prüfverfahren der PTB abgestimmt sein könnten. Diese Problematik würde sich mit dem Einsatz von KI-Software weiter verstärken und müsste in der entsprechenden Norm (DIN EN 61674) insbesondere im Hinblick auf die Generalisierbarkeit Berücksichtigung finden. Im Bereich der Strahlentherapie hat bereits ein erster Hersteller eine KI-basierte Messsoftware zur Bestimmung wichtiger Dosismessgrößen auf den Markt gebracht. Diese Messgrößen gehen in die von der PTB bearbeiteten Normen für die Referenzdosimetrie ein, bei denen strikte Kriterien, nach welchen solche Algorithmen bewertet werden können, wünschenswert wären. Zudem zeichnet sich der zukünftige Einsatz von KI im Bereich der Bestrahlungsplanung ab.

Für bildgebende Verfahren, die ionisierende Strahlung einsetzen, verlangen das Strahlenschutzgesetz (StrlSchG §14 (1) 5 a) und die Strahlenschutzverordnung (StrlSchV §115 and §116) Prüfverfahren, die eine Optimierung des Verhältnisses von Bildqualität zu Patientendosis ermöglichen. Die neueste Generation von Röntgentomographie-Geräten (CT) verwendet KI-Algorithmen für die Bildrekonstruktion. Da die Trainingsdaten wie im Allgemeinen auch die verwendeten Bildrekonstruktionsverfahren nicht für die Prüfung zugänglich sind, kann die Bildqualität somit nur geprüft werden, indem das CT-Gerät als "black box" betrachtet wird. Da KI-Algorithmen mit Hilfe anatomischer Strukturen trainiert werden, ist der bisherige Weg über technische Prüfkörper (Phantome) zur Quantifizierung des Auflösungsvermögens und der Niedrigkontrast-Detektierbarkeit nicht mehr gangbar, sodass alternative Prüfverfahren und ggf. neuartige Prüfkörper entwickelt werden müssen. Auch in der Mammographie werden bereits ohne KI für anatomische Strukturen optimierte Bildbearbeitungsverfahren eingesetzt, die beim Einsatz technischer Phantome nicht auf gleiche Weise funktionieren, sodass lediglich eine Qualitätsprüfung der Rohbilder ("for processing"), nicht aber der den Radiologen vorgelegten bearbeiteten Bilder ("for presentation") möglich ist. Mit der für die nächsten Jahre erwarteten Einführung von Tomosynthese-Geräten im Mammographie-Screening muss damit gerechnet werden, dass auch dort KI-Verfahren zum Einsatz kommen werden. Moderne Bildrekonstruktionsverfahren erfordern damit die Entwicklung neuer Wege der

Bildqualitätseinschätzung. Eine Überprüfung der KI an sich steht in diesem Zusammenhang völlig außer Diskussion; die Herausforderung besteht darin, "black-box"-Verfahren zu entwickeln, die mit KI-Verfahren und anderen nicht offen gelegten Bildrekonstruktionsmethoden zurechtkommen.

Auch im gesetzlichen Messwesen werden voraussichtlich langfristig KI-Technologien Einzug halten. Vereinzelte Anfragen von Messgeräteherstellern weisen darauf hin, dass vor allem der Einsatz von KI zum Zweck der Messwertberechnung (Auswertung von Sensordaten) von großem Interesse ist. Vor diesem Hintergrund beschäftigt sich derzeit eine Untergruppe des OIML TC5/SC2/p4 unter Leitung der PTB mit weltweit harmonisierten Software-Anforderungen an KI aus Perspektive der Software- und Datensicherheit. Es ist das Verständnis der international besetzten Gruppe, dass die momentan bestehenden Anforderungen, insbesondere auch der Rechtsrahmen der europäischen Messgeräterichtlinie (MID), bereits flexibel genug für KI-Algorithmen sind. Die Begründung liegt in der Betrachtung der KI als unveränderliche Software mit hochgradig veränderlichen Parametern, die das Verhalten der Software definieren. Dieses Szenario ist im gesetzlichen Messwesen hinlänglich bekannt und führt dazu, dass aus Sicht der Softwaresicherheit Fragen rund um die Kennzeichnung von mittels KI berechneten Messwerten, Protokollierung und Nachverfolgbarkeit jeglicher Änderungen der KI sowie Fragen der Softwaresicherheitsprüfung von KI im Vordergrund stehen. Der nächste Schritt in der Normung ist der internationale Austausch von Erfahrungen und Entwicklungen bezüglich der Verwendung von KI im gesetzlichen Messwesen. Daran anschließend müssen die derzeit im Entwurfsstadium befindlichen Anforderungen an KI hinsichtlich ihrer Verwendbarkeit geprüft und ggf. angepasst werden. Es wird erwartet, dass mit diesem Vorgehen passende Anforderungen an KI hinreichend konkretisiert werden können, bevor solche Systeme in großer Stückzahl auf den Markt gebracht werden.

Aufbauend auf Aktivitäten zur Bewertung von KI und der Bereitstellung von Referenzdaten wäre zukünftig eine Ausweitung der PTB-Dienstleistung zur Validierung von Algorithmen (TraCIM) auf KI-Verfahren grundsätzlich denkbar. Im Zuge einer solchen Validierung müsste die Funktionalität der KI mittels repräsentativer Referenzdatensätze getestet und ihre Robustheit ggf. ergänzt durch eine Softwareprüfung zur Absicherung gegenüber Manipulation sichergestellt werden.

Ein weiteres industrienahes Handlungsfeld für die PTB in ihrer Rolle als nationales Metrologieinstitut ist die Qualitätsbewertung „indirekter Messungen“, bei denen KI-Methoden dafür eingesetzt werden, „Messdaten“ für Größen/Orte/Zeiten zu generieren, zu denen keine real gemessenen Werte vorliegen. Man spricht in diesem Zusammenhang auch von „Soft Sensorik“, „Sensorfusion“ oder „virtual sensing“. Es ist hier von metrologischem Grundinteresse, Qualitätsaussagen über diese Angaben zu treffen. Der Einfluss von Qualität der Trainingsdaten als auch die Unsicherheit der Arbeitsdaten stellen mögliche Forschungsrichtungen dar. Dabei stellen sich derzeit folgende Anwendungsszenarien für Soft Sensorik und damit eine entsprechend notwendige Hinterlegung der Thematik durch die PTB dar:

- Überwachung von Prozessparametern an schwer zugänglichen Stellen (durch Nutzung von angrenzenden Sensoren + KI);
- Ablösung/Ersatz eines teuren (Spezial-)Sensors (durch Nutzung günstiger Sensoren + KI);
- Kontinuierliche Überwachung eines Produktqualitätsparameters statt zeitlich weit auseinanderliegender manueller Messungen (durch Nutzung vorhandener Sensoren + manuelle Messungen + KI).

Für diese und weitere Handlungsfelder des Messwesens sieht sich die PTB in der Verantwortung, entsprechend proaktiv auf die Herausforderungen des KI-Einsatzes zu reagieren, um innovative Technologien sicher und vertrauenswürdig zur Förderung der Wirtschaft und zum Wohle der Gesellschaft zugänglich zu machen. Eine zentrale Rolle spielt hierbei die aktive Mitarbeit der PTB in

den relevanten Gremien, die Vernetzung mit Industrie, Anwendern und Verbänden sowie mit der politischen Ebene.

Dieses Rollenverständnis führt zu der Erkenntnis, dass die PTB gezielt und in beträchtlichem Umfang Kompetenzen im Bereich KI aufbauen und vielfältige Kooperation mit kompetenten Partnern etablieren muss, um ihrem gesetzlichen Auftrag nachhaltig gerecht zu werden und diesen vorausschauend zu gestalten. Aufgrund der begrenzten Ressourcenlage ist es für die PTB langfristig erstrebenswert, ihre Beauftragung für Produkte und Dienstleistungen mit KI expliziter auszuformulieren und auch rechtlich klarer zu verankern. Insbesondere beim neuen Ordnungsrahmen für KI sieht sich die PTB als eine tragende Säule innerhalb der (digitalen) Qualitätsinfrastruktur und strebt eine frühzeitige Integration ihrer metrologischen Expertise in QI-Prozesse an. Diese umfassen sowohl spezifische Anwendungen als auch grundlegende Forschung zu KI und deren Einbettung in die QI. Des Weiteren wird eine KI-kompetente PTB als ein wichtiger Baustein für die Gründung eines IZSM vorausgesetzt, um aktiv an den Forschungsfragen der systemischen Metrologie mitzuwirken und anschließend die Erfüllung von Daueraufgaben sicherzustellen.

Forschungskooperationen

Die PTB wird auch auf lange Sicht nicht mit dem Umfang der Arbeiten und den Möglichkeiten anderer großer Forschungsverbünde konkurrieren können – und das auch nicht müssen. Für einen möglichst effektiven Einsatz der verfügbaren Ressourcen wird die PTB daher gezielte Kooperationen mit nationalen, europäischen und internationalen Forschungspartnern eingehen.

Eine gute Übersicht über die aktuelle Forschungslandschaft Deutschlands im Bereich KI bietet u. a. die „Landkarte KI“ der Plattform Lernende Systeme[5]. Besonders herausragende Institute für KI-Forschung und -Entwicklung sind

- Deutsches Forschungszentrum KI (DFKI)
- Fraunhofer Gesellschaft
- Deutsches Zentrum für Luft- und Raumfahrt (DLR)
- Helmholtz Gemeinschaft (insbesondere Helmholtz KI-Kooperationseinheit)
- Max-Planck-Institut für Intelligente Systeme
- Mevis Medical Solutions (Bremen)
- Uni-Kliniken (z. B. Charité) und medizinische Fakultäten, bspw. der Universität Duisburg-Essen und Universitätsmedizin Essen (Institut für Künstliche Intelligenz in der Medizin)
- Einrichtungen im Cyber Valley in Tübingen als Europas größtes Forschungskonsortium im Bereich KI mit Partnern aus Wissenschaft und Industrie
- Robert Koch Institut (RKI) (Aufbau eines Zentrums zu Validierung von KI-Algorithmen in der Gesundheitsforschung)
- KI-Absicherung
- TÜV/ DEKRA

Teilweise orientiert sich die PTB auch für organisatorische und strukturelle Entscheidungen an großen Forschungseinrichtungen und steht über verschiedenste Kooperationen mit ihnen in engem fachlichen Austausch. In diversen Netzwerken ist die PTB zudem aktiv an der Erarbeitung metrologischer Beiträge zur Qualitätssicherung von KI und der zugrundeliegenden Daten beteiligt, u. a.:

1. Working Group 1 on “Expression of Uncertainty in Measurement” des Joint Committee for Guides in Metrology (JCGM): Unsicherheitsangaben in Zertifikaten und Kalibrierscheinen, oder auch CMCs (Calibration and Measurement Capabilities), basieren alle auf den Methoden zur

[5] https://www.plattform-lernende-systeme.de/ki-landkarte.html

Unsicherheitsermittlung des Guide to the Expression of Uncertainty in Measurement (GUM). Die Pflege und Weiterentwicklung des GUMs geschieht durch die Working Group 1 des Joint Committee for Guides in Metrology, die aus Vertretern von BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP (derzeit repräsentiert durch PTB-Beschäftigten) und OIML gebildet wird. Zukünftig könnte es die Aufgabe dieses Gremiums werden, ein Dokument zur Unsicherheitsermittlung im Zusammenhang mit KI-Methoden zu entwickeln.

2. Die ITU-WHO-Focus Group „Artificial Intelligence for Health“ hat das Ziel, Standards zu setzen für die Bewertung und Validierung von KI-basierten Methoden für das Gesundheitswesen. Die PTB hält Verbindung zu den Arbeiten der Gruppen „Clinical evaluation of AI for health (WG-CE) oder auch „Ethical considerations on AI for health (WG-Ethics). Zudem gibt es thematische Gruppen zu verschiedenen medizinischen Fachgebieten bzw. Krankheitsbildern (z.B. TG-Radiology, TG-Malaria). Aktuell bereitet die Fokusgruppe sogenannte trial-audits vor, in denen konkrete KI-Anwendungen in der Medizin anhand eines vorab erarbeiteten Kriterienkataloges hinsichtlich Qualitätsaspekten wie Vorhersagegüte, Robustheit, Fairness und Erklärbarkeit bewertet werden.
3. European Metrology Network on “Mathematics and Statistics” (EMN MATHMET): Viele moderne Messverfahren nutzen mathematische und statistische Methoden und die zugehörigen Algorithmen. Die Metrologie ist daher zunehmend auf fortgeschrittene Kenntnisse von Modellierung und Simulationsverfahren sowie in statistischer Datenanalyse angewiesen insbesondere für KI-Verfahren. Derzeit koordiniert die PTB das neugegründete Netzwerk, das an der Grenzfläche von Messwesen und Mathematik arbeitet und den Austausch im europäischen Rahmen fördert. KI ist ein Schwerpunkt der entstehenden strategischen Forschungsagenda von MATHMET, im Fokus stehen die Erarbeitung von Guidelines zur Bewertung von Algorithmen, Software und Referenzdaten unter besonderer Berücksichtigung von Unsicherheiten, Robustheit und Erklärbarkeit. Aufgrund der personellen Überschneidungen zwischen MATHMET und zahlreichen metrologischen Normungsgremien gelangen die Outputs von MATHMET effizient in die Normungsarbeit.

In der Kooperation mit Partnern steht für die PTB grundsätzlich weniger die Entwicklung neuer KI-Methoden im Vordergrund, sondern sie legt stattdessen den Fokus auf die Entwicklung von Bewertungsmethoden sowie die Bereitstellung von Referenzdatensätzen. Ein wichtiges Alleinstellungsmerkmal der PTB ist dabei ihr Domänenwissen sowie ihre Neutralität.

Empfehlungen

Aus den Zielsetzungen der vorangegangenen Kapitel ergeben sich umfassendere strategische Überlegungen für die praktische Ausgestaltung. Grundsätzliche Leitplanken für die Entwicklung der PTB-Aktivitäten in Bezug zu KI finden sich in verschiedenen Themenfeldern:

I. Grundlagenforschung zu Methodiken und Werkzeugen für die Bewertung großer Datensätze und Entwicklung von „good practice“ Beispielen hinsichtlich Unsicherheit, Genauigkeit, Repräsentativität und Vergleichbarkeit
II. Entwicklung von Referenzdatensätzen zur Bewertung der Qualität von KI
III. Ertüchtigung der PTB-Infrastrukturen zur Bereitstellung erarbeiteter Referenzdatensätze (Anbindung an das Kundenportal etc.)
IV. Entwicklung geeigneter Metriken zur Beurteilung der KI-Leistungsfähigkeit, die auch Robustheit, Erklärbarkeit und Vorhersagesicherheit einschließt
V. Ertüchtigung der metrologischen Dienstleistungen hin zur Validierung von KI-Algorithmen (Beurteilung von KI-Leistungsfähigkeit und Softwareprüfung)
VI. Erarbeitung von Empfehlungen für Annotationsvorschriften und Verwendung von Metadaten (insbesondere Einheiten, Unsicherheiten, Messverfahren) in ausgewählten Anwendungsbereichen
VII. Weiterentwicklung von Messverfahren und Messdatenauswertung durch den Einsatz von KI
VIII. Transfer wissenschaftlicher Ergebnisse zu KI in die Anwendung für metrologische Dienstleistungen, Forschung und Verwaltung
IX. Einsatz von KI-Methodiken für die Bearbeitung von metrologischen, wissenschaftlichen Fragestellungen und die Datenorganisation sowie Prozesssteuerung.

Literaturverzeichnis

[1] Europäische Kommission, „Bericht der Kommission über die Auswirkungen künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung," 2020.

[2] T. Jürgensohn, C. Platho, D. Stegmaier, M. Hartwig, M. Krampitz, L. Funk, T. Plass und H. Ehrlich, „Rechtliche Rahmenbedingungen für die Bereitstellung autonomer und KI-Systeme," Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund, 2021.

[3] Bundesregierung, „Strategie Künstliche Intelligenz der Bundesregierung - Fortschreibung 2020," 2020.

[4] Europäische Kommission, „Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen," Europäische Kommission, Brüssel, 2020.

[5] Bundesregierung, „Strategie Künstliche Intelligenz der Bundesregierung," Berlin, 2018.

[6] Bundesregierung, „Stellungnahme der Bundesregierung der Bundesrepublik Deutschland zum Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen," 2020.

[7] Enquete-Kommission Künstliche Intelligenz, „Bericht der Enquete-Kommission Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale," Bundestagsdrucksache 19/2978, 2020.

[8] NIST, „U.S. LEADERSHIP IN AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools," Prepared in response to Executive Order 13859, 2019.

[9] MW MWK Niedersachsen, „KI-Working Paper Niedersachsen," MW MWK Niedersachsen, 2020.

[10] DIN, „Normungsroadmap Künstliche Intelligenz," 2020.

[11] DIN SPEC 92001-1, *Artificial Intelligence - Life Cycle Processes and Quality Requirements - Part 1: Quality Metamodel.*

[12] DIN SPEC 92001-2, *Künstliche Intelligenz - Life Cycle Prozesse und Qualitätsanforderungen - Teil 2: Robustheit.*

[13] N. Becker, P. Junginger, L. Martinez und D. Krupka, „KI in der Arbeitswelt: Übersicht einschlägiger Normen und Standards," Gesellschaft für Informatik e.V. (GI), Berlin, 2021.

[14] Plattform Lernende Systeme, „Kompetenzentwicklung für künstliche Intelligenz: Veränderungen, Bedarfe und Handlungsoptionen," PLS, 2021.

[15] BMWi, „KI-Bedarfe der Wirtschaft am Standort Deutschland," 2020.

[16] M. Kläs, „Towards identifying and managing sources of uncertainty in AI and machine learning models-an overview," *arXiV:1811.11669,* 2018.

[17] Fraunhofer IAIS, *Whitepaper: Vertrauenswürdiger Einsatz von Künstlicher Intelligenz.*

[18] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, OIML, Evaluation of measurement data – Guide to the expression of uncertainty in measurement, Joint Committee for Guides in Metrology, JCGM 100:2008.

[19] Y. Gal, „Uncertainty in deep learning," *University of Cambridge,* 2016.

[20] A. Kendall und Y. Gal, „What uncertainties do we need in Bayesian deep learning for computer vision?," *Advances in neural information processing systems,* pp. 5574-5584, 2017.

[21] D. Kingma, T. Salimans und M. Welling, „Variational dropout and the local reparameterization trick," *Advances in neural information processing systems,* pp. 2575-2583, 2015.

[22] T. Kretz, K. Müller, T. Schaeffter und C. Elster, „Mammography Image Quality Assurance Using Deep Learning," *IEEE Transactions on Biomedical Engineering,* 2020.

[23] S. Lapuschkin, S. Wäldchen, A. Binder, G. Montavon, W. Samek und K. R. Müller, „Unmasking clever hans predictors and assessing what machines really learn," *Nature communications 10(1),* pp. 1-8, 2019.

[24] R. Muller, S. Kornblith und G. Hinton, „When does label smoothing help?," *Advances in neural Information processing systems,* pp. 4694-4703, 2019.

[25] I. Goodfellow, J. Shlens und C. Szegedy, „Explaining and Harnessing Adversarial Examples," in *International Conference on Learning Representations*, 2015.

[26] J. Martin und C. Elster, „Inspecting adversarial examples using the Fisher information," *Neurocomputing 382,* pp. 80-86, 2020.

[27] J. Martin und C. Elster, „Detecting unusual input to neural networks," *Applied Intelligence,* 2020.

[28] Holmberg et al., „Self-supervised retinal thickness prediction enables deep learning from unlabelled data to boost classification of diabetic retinopathy," *Nature Machine Intelligence,* pp. 719-726, 2020(2).

[29] Plattform Lernende Systeme, Whitepaper: Zertifizierung von KI-Systemen, 2020.

[30] FDA, „Proposed Regulatory Framework for Modifications to Artificial lntelligence/Machine Learning [Al/ML] Based Software as a Medical Device [SaMD]".

[31] PEGASUS, „Projekt zur Etablierung von generell akzeptierten Gütekriterien, Werkzeugen und Methoden sowie Szenarien und Situationen zur Freigabe hochautomatisierter Fahrfunktionen," 2016.

[32] DFKI Kompetenzzentrum, „Autonomes Fahren," [Online]. Available: https://www.dfki.de/web/forschung/kompetenzzentren/autonomes-fahren/.

[33] Grigorescu et al., „A survey of deep learning techniques for autonomous driving".

[34] Klöppel, Stefan, et al., „Automatic classification of MR scans in Alzheimer's disease," *Brain,* Bd. 131.3, pp. 681-689, 2008.

[35] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen und K. R. (. Müller, „Explainable AI: interpreting, explaining and visualizing deep learning," in *Vol. 11700*, Springer Nature, 2019.

[36] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, S. Haynes, J. D. Blankertz und F. Bießmann, „On the interpretation of weight vectors of linear models in multivariate neuroimaging," *Neuroimage 87,* pp. 96-110, 2014.

[37] M.-H. Hung, T.-H. Lin, F.-T. Cheng und R.-C. Lin, „A novel virtual metrology scheme for predicting CVD thickness in semiconductor manufacturing," *IEEE/ASME Transactions on mechatronics 12(3),* pp. 308--316, 2007.

[38] Yung-Cheng, J. Chang und F.-T. Cheng, „Application development of virtual metrology in semiconductor industry, IECON 2005," in *31st Annual Conference of IEEE Industrial Electronics Society*, IEEE, 2005.

[39] L. Hoffmann und C. Elster, „Deep neural networks for computational optical form measurements," *Journal of Sensors and Sensor Systems 9(2),* pp. 301-307, 2020.

[40] L. Hoffmann, I. Fortmeier und C. Elster, „Uncertainty Quantification by Ensemble Learning for Computational Optical Form Measurements," *arXiv preprint arXiv:2103.01259,* 2021.

[41] A. Andrle, N. Farchmin, P. Hagemann, S. Heidenreich, V. Soltwisch und G. Steidl, „Invertible Neural Networks Versus MCMC for Posterior Reconstruction in Grazing Incidence X-Ray Fluorescence in Scale Space and Variational Methods in Computer Vision," in *Lecture Notes in Computer Vision*, Springer International Publishing, 2021.

[42] G. Barbastathis, A. Ozcan und G. Situ, „On the use of deep learning for computational imaging," *Optica 6(8),* pp. 921-943, 2019.

[43] G. V. Vdovin, „Model of an adaptive optical system controlled by a neural network," *Optical Engineering 34(11),* pp. 3249-3253, 1995.

[44] L. Zhang, S. Zhou, J. Li und B. Yu, „Deep neural network based calibration for freeform surface misalignments in general interferometer," *Optics express 27(23),* pp. 33709-33723, 2019.

[45] Loh et al., „Fractal morphology, imaging and mass spectrometry of single aerosol particles in flight," *Nature,* p. 513–517, 2012.

[46] D. A. Lack, H. Moosmüller, G. R. McMeeking, R. K. Chakrabarty und D. Baumgardner, „Characterizing elemental, equivalent black, and refractory black carbon aerosol particles: a review of techniques, their limitations and uncertainties," *Anal Bioanal Chem 406,* p. 99–122, 2014.

[47] Cortés, D. et al., „Effect of Fuels and Oxygen Indices on the Morphology of Soot Generated in Laminar Coflow Diffusion Flames," *Energy Fuels 32,* p. 11802–11813, 2018.

[48] J. Blum, „ Dust Evolution in Protoplanetary Discs and the Formation of Planetesimals," *Space Sci Rev 214 (52),* 2018.

[49] A. Ng, „Data-centric AI: Real World Approaches," DeepLearning.AI, 2021. [Online]. Available: https://https-deeplearning-ai.github.io/data-centric-comp/.

[50] CIPM Task Group on the "Digital-SI", „Draft of the Grand Vision - Transforming the International System of Units for a Digital World (Version 3.4),“ 2020. [Online]. Available: https://www.bipm.org/documents/20126/46590079/WIP+Grand_Vision_v3.4.pdf/aaeccfe3-0abf-1aaf-ea05-25bf1fb2819f.

[51] T. Dorst, M. Gruber und A. P. Vedurmudi, „Sensor data set of one electromechanical cylinder at ZeMA testbed (ZeMA DAQ and Smart-Up Unit) [Data set],“ 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5185953.

[52] DIN und DKE, „Whitepaper: Szenarien zur Digitalisierung der Normung und Normen,“ DIN und DKE, 2021.

[53] DKE, Frauenhofer IAIS, ICMS GmbH, „DiTraNo - Die digitale Transformation der Normung – Schaffung informationstechnischer Voraussetzungen, um die zukünftigen Herausforderungen der Normung erfüllen zu können,“ [Online]. Available: https://www.dke.de/de/normen-standards/digitalisierung-normung-digitalstrategie-dke-transformation/digitale-transformation-normung.

[54] EUROLAB, „Position paper in response to EC report COM(2020) 65 final“.

[55] Europäische Kommission, „Verordnung des europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union,“ COM, Brüssel, 2021.

[56] BMWi, „Neue Räume, um Innovationen zu erproben,“ BMWi, 2021.

[57] PTB (Kurzfassung online), „Innovationszentrum für Systemische Metrologie,“ [Online]. Available: www.izsm.eu.

[58] ZVEI, „ZVEI Stellungnahme zum Vorschlag der EU-Kommission für eine Verordnung zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz ("AI Act"),“ Zentralverband Elektrotechnik- und Elektronikindustrie e. V., 2021.

[59] Bundesrat, „Empfehlungen der Ausschüsse,“ Drucksache 488/1/21, 2021.

[60] DIN & DKE, „Position Paper on the EU “Artificial Intelligence Act”,“ DIN & DKE, 2021.

[61] L. Beining, „Vertrauenswürdige KI durch Standards?,“ Stiftung Neue Verantwortung, 2020.

[62] TÜV Verband, BSI, Fraunhofer HHI, „Towards Auditable AI Systems,“ 2021.

[63] IG-NB, „Fragenkatalog „Künstliche Intelligenz bei Medizinprodukten“,“ 2020.

[64] Fraunhofer IPA, „White Paper: Zuverlässige KI,“ 2020.

[65] R. Hawkins, C. Paterson, C. Picardi, Y. Jia, R. Calinescu und I. Habli, „Guidance on the Assurance of Machine Learning in Autonomous Systems (AMLAS),“ *arXiv preprint arXiv:2102.01564,* 2021.

[66] LNE, „Certification standard of processes for AI: Design, development, evaluation and maintenance in operational conditions,“ LNE, Paris, 2021.

[67] BMWi, „Erklärbare KI: Anforderungen, Anwendungsfälle und Lösungen," Technologieprogramm KI-Innovationswettbewerb des Bundesministeriums für Wirtschaft und Energie, Berlin, 2021.

[68] A. B. Arrieta et al., „Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion 58,* pp. 82-115, 2020.

[69] Z. C. Lipton, „The mythos of model interpretability," *Queue 16(3),* p. 30:31–30:57, 2018.

[70] J. Caldeira und B. Nord, „Deeply uncertain: comparing methods of uncertainty quantification in deep learning algorithms," *Machine Learning: Science and Technology 2(1),* p. 015002, 2020.

Appendix: Glossar

Wie innerhalb des vorliegenden Strategiepapiers im Detail ausgeführt, ist die Terminologie im Zusammenhang mit künstlicher Intelligenz noch Gegenstand laufender Forschung und erlaubt daher zu diesem Zeitpunkt noch keine abschließende Definition. Um dennoch ein weitreichendes Verständnis für die Inhalte des Dokumentes sicherzustellen, bietet dieses Glossar eine Begriffserklärung der wichtigsten Schlagworte.

Erklärbarkeit von KI-Systemen soll dazu dienen, Zielpersonen verständliche Begründungen für die Ergebnisse eines KI-Modells zu liefern, um damit eine Nachvollziehbarkeit zu gewährleisten [67]. Erklärbare KI kann somit die Grundlage für menschliches Verständnis für und Vertrauen in KI bilden [68]. Man unterscheidet zwischen lokaler bzw. Datenerklärbarkeit von Einzelentscheidungen und globaler bzw. Modellerklärbarkeit der Wirkmechanismen [67]. Offene Fragen der Forschung im Zusammenhang mit Erklärbarkeit betreffen u. a. die Entwicklung von Metriken für Erklärbarkeit sowie Abschätzungen für die Belastbarkeit der Aussagen erklärbarer KI (xAI).

Künstliche Intelligenz (KI) umfasst Software und/oder Hardware, welche lernen kann komplexe Probleme zu lösen, Vorhersagen zu treffen und Aufgaben zu verrichten, die „menschliche“ Qualitäten und Fähigkeiten wie (Sinnes-)Wahrnehmung (z. B. Sehen, Berührung) durch Datenerfassung, Kognition, Planen, Lernen, Kommunikation oder auch physische Handlungen erfordern [8]. Man unterscheidet sie in „starke“ und „schwache“ KI. „Starke“ KI geht dabei von Systemen aus, die den intellektuellen Fähigkeiten der Menschen gleichkommen oder diese übertreffen. „Schwache“ KI bezeichnet hingegen Algorithmensysteme zur Lösung konkreter Anwendungsprobleme auf Basis von Methoden aus der Mathematik und Informatik, wobei die entwickelten Systeme zur Selbstoptimierung fähig sind [5].

Robustheit beschreibt die Fähigkeit eines KI-Systems, mit fehlerhaften, verrauschten, unbekannten oder schädlich manipulierten Eingangsdaten umzugehen bzw. diese zu kompensieren (Stationarität). Robustheit bildet daher eine wichtige Säule in der Qualitätssicherung für KI [12].

Transparenz bezeichnet die Durchsichtigkeit eines KI-Modells [67, 68, 69] sowie der Statistik zugrundeliegender Trainingsdaten und beinhaltet drei teilweise hierarchisch abhängige Aspekte: Transparenz des Gesamtmodells (Simulierbarkeit), auf Ebene der Einzelkomponenten (Unterteilbarkeit) und auf Ebene des Trainingsalgorithmus (Algorithmische Transparenz) [69]. Neben der beschriebenen Modell- und Datentransparenz entwickelt sich eine Betrachtung von KI-Transparenz auf Systemebene, welche die Gesamtheit der Prozesse innerhalb des KI-Lebenszyklus umspannt [64, 30].