PTB mitteilungen 1 2022

Metrologie und KI: Die KI-Strategie der PTB

DE

Metrology and AI: PTB's AI Strategy

EN

Physikalisch-Technische Bundesanstalt ■ Braunschweig und Berlin

Nationales Metrologieinstitut

PTB mitteilungen

Fachorgan für Wirtschaft und Wissenschaft, Amts- und Mitteilungsblatt der Physikalisch-Technischen Bundesanstalt Braunschweig und Berlin

132. Jahrgang, Heft 1, März 2022

Im Jahr 2021 wurde keine Ausgabe der
PTB-Mitteilungen veröffentlicht.

Metrologie und KI: Die KI-Strategie der PTB

Zusammenfassung

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 Diagnosen 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 Potenziale 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 Potenziale 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.

Autorinnen und Autoren und ihre Zugehörigkeit zu PTB-Fachbereichen

Sascha Meyne (1.3)
David Auerbach (3.1)
Tobias Klein (5.2)
Matthias Neuwirth (5.2)
Ulrike Ankerhold (6.2)
Mathias Anton (6.2)
Stefan Pojtinger (6.2)
Steffen Ketelhut (6.3)
Tobias Schäffter (8)
Hans Rabus (8.01)
Lukas Winter (8.1)
Andreas Kofler (8.1)
Christoph Kolbitsch (8.1)
Patrick Schünke (8.1)
Markus Bär (8.4)
Clemens Elster (8.4)
Lara Hoffmann (8.4)
Sebastian Heidenreich (8.4)
Stefan Haufe (8.4)
Martin Nischwitz (8.5)
Marko Esche (8.5)
Andreas Barthel (9.11)
Dirk Ratschko (9.2)
Harry Stolz (9.2)
Sascha Eichstädt (9.4)
Daniel Hutzschenreuter (9.4)
Julia Tesch (9.4)
Giacomo Lanza (Q.11)
Holger Israel (Q.11)
Daniel Lübbert (Q.4)

Inhalt

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 Potenziale, 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 BMWK 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

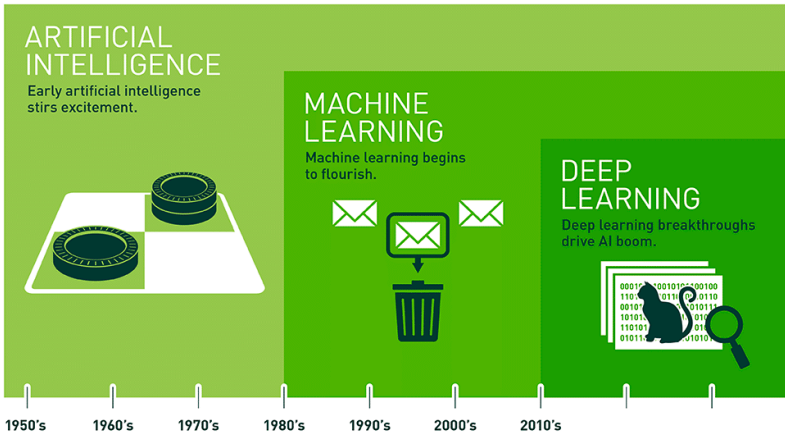
Die disruptive Schlüsseltechnologie KI hat längst das Nischenstadium in der Forschung verlassen (siehe Abb. 1 zur historischen Entwicklung) und drängt in Form verschiedenster Produkte und Dienstleistungen auf den Markt und damit in sämtliche Lebens- und Wirtschaftsbereiche. Um entsprechende Leitplanken 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.

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 mithilfe 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



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

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

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 (Deutsches Forschungszentrum für Künstliche Intelligenz) und das DLR (Deutsches Zentrum für Luft- und Raumfahrt) 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 zu 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, der konventionellen 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 zu 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 (*International Organization of Legal Metrology*) 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 potenziell von reinen KI-Entwicklungsaufgaben ab und verschafft der PTB einen Vorteil in der Personalgewinnung. Die Kompetenzen von Beschäftigten, wie metrologisches Domänenwissen, Datenkompetenz und Konzeptverständnis für maschinelles Lernen und Data Science werden durch grundlegende digitale Kompetenzen, wie den Umgang mit KI-Systemen (z. B. Prozess-, Problemlösungs- und Reflexionskompetenz), Erfahrung in der Gestaltung von Arbeitsprozessen (z. B. soziale, organisatorische und Selbstkompetenzen), ergänzt [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 mithilfe 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 essenzieller 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 (*Guide to the Expression of Uncertainty in Measurement*) 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 gleichen 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* (explainable AI, kurz: xAI) 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 sogenannten *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. Tests 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 Datenhandlings aufbauen, um bspw. die Anforderungen aus [29] und [30] abbilden zu können.

[1] EMPIR – European Metrology Programme for Innovation and Research

[2] BMBF – Bundesministerium für Bildung und Forschung

[3] BMWi – Bundesministerium für Wirtschaft und Energie (jetzt Bundesministerium für Wirtschaft und Klimaschutz)

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[1]-Projekt „*Metrology for the Factory of the Future*“ (Met-4FoF), dem BMBF[2]-Projekt “AAS-basierte Modellierung zur Analyse veränderlicher CPS“ (FAMOUS) und dem BMWi[3]-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 G*round-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 Potenziale, die es gezielt auszuschöpfen gilt. Nach einer Umfrage des EMN[4] 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.

[4] EMN – EURAMET's European Metrology Networks

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 mithilfe 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 Potenziale 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 datenanstelle von modellbasierten Vorhersagen, auch für die Wartungszeiträume experimenteller Aufbauten, ein großes Potenzial 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 Hardwarenah 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 kann 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 aufwendige 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 mithilfe 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 aufwendigen 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,

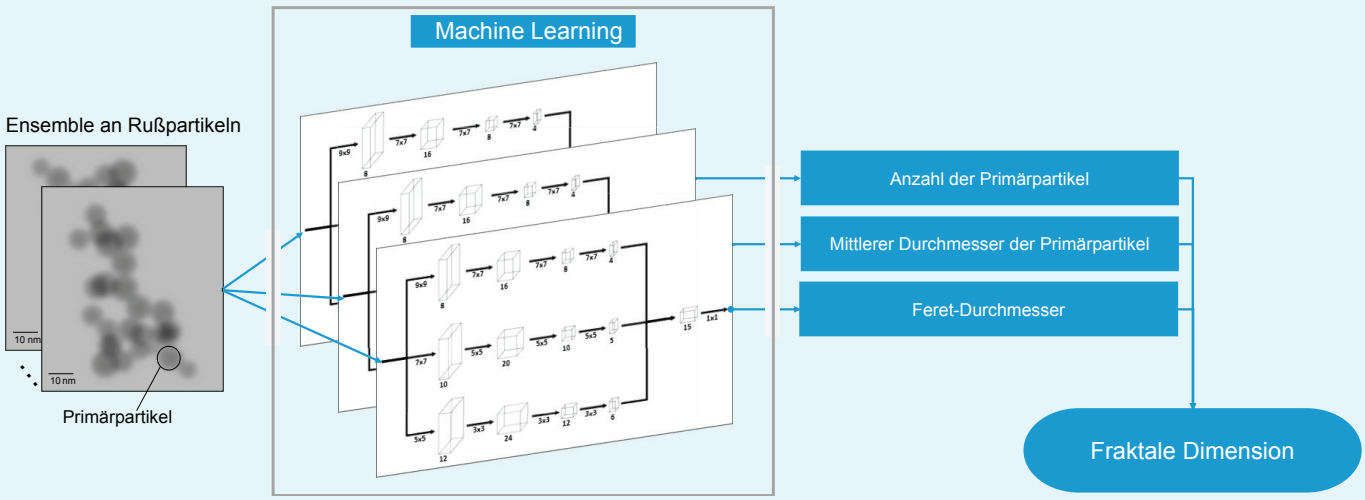


Abbildung 2: Die elektronenmikroskopische Aufnahme des Rußpartikels wird von drei Machine-Learning-Netzwerken ausgewertet, mit deren Hilfe verschiedene Parameter bestimmt werden, aus denen sich die fraktale Dimension ergibt

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.

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. C*onvergence of AI and High Performance Computing*" (HPC)) gesprochen[5]. 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 Aufbewah-

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

rung endgültiger Resultate (Isilon-Speicher, inkl. Backup und WORM[6]-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.

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 1700 CPU-cores (Hauptprozessoren) sind bisher nur 2 GPU-Knoten (Grafikprozessoren) mit Grafikbeschleunigern 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 Größenordnung 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.

[6] WORM – Write Once Read Many (schreibe einmal, lese vielfach)

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 potenziell auch auf bauliche und klimatechnische Anforderungen besser und kosteneffizienter reagieren, da entsprechende Infrastrukturen anders als bisher nicht einzeln bei wachsendem KI-Einsatz aufwendig 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 Datenzentrierte 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[7]-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, Umgebungsbedin-

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

gungen 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 Potenziale, 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 mithilfe 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]);
- 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, für routinemäßige aber situationsabhängige Analyseverfahren. Konkrete Beispiele dafür sind Anomaliedetektion oder Unsicherheitsschätzungen sowie komplexe, adaptive 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 (IZSM), welches im Schwerpunkt diese systemischen Herausforderungen adressiert.

KI kann Menschen außerdem bei Aufgaben der Prozesssteuerung 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 essenziell 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[8]:

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

[8] Vgl. 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 potenziell 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"[9]. 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[10]-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) 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.

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

[10] TraCIM Traceability for computational-intensive metrology (Rückführbarkeit für rechenintensive Metrologie)

Die wachsenden Einsatzmöglichkeiten von KI bieten einerseits ein großes wirtschaftliches Potenzial 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 essenziellen 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 für 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 (jetzt BMWK) 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 liegen dem BMWK bereits vor [57].

Grundsätzlich ist im Verordnungsentwurf auch die Definition des Begriffs 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 Rechtsrahmen nicht vorgesehen wäre. Kritik an dieser weiten Auslegung des KI-Begriffs wird an verschiedenen Stellen geäußert, unter anderem in der Stellungnahme des Zentralverbands 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 Bundesrats. 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[11], der Charité, dem BSI[12] 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 sozio-technische 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).

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 die DIN SPEC 92001-2 [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.

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.

[11] Fraunhofer HHI – Fraunhofer Heinrich-Hertz-Institut

[12] BSI – Bundesamt für Sicherheit in der Informationstechnik

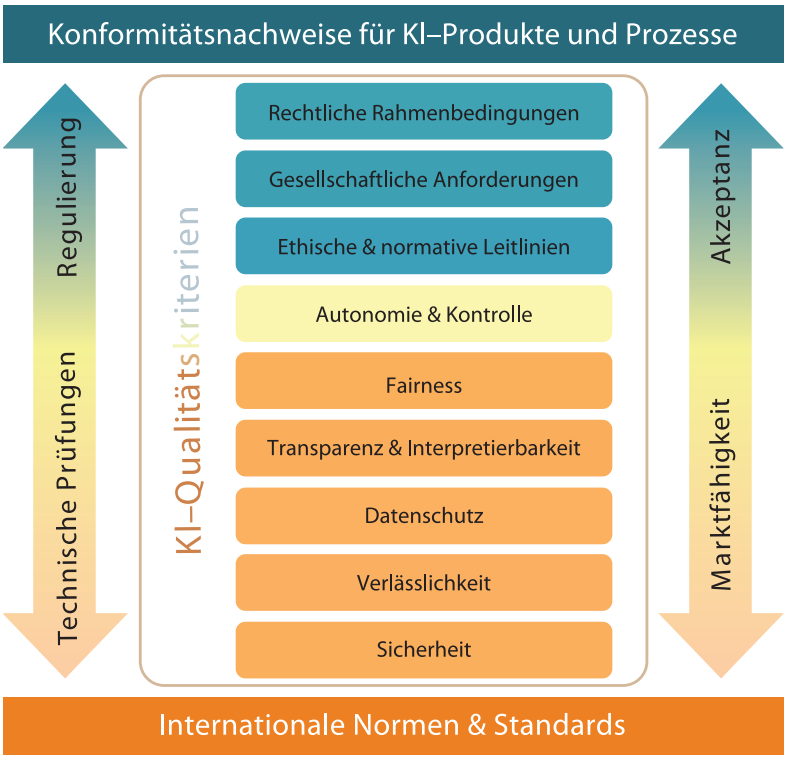


Abbildung 3: Kategorisierte Qualitätsdimensionen für die Bewertung von KI in der Konformitätsprüfung [10]

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 Protocol* (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 der 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[13] 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 (siehe Abb. 4). 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.

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 müsste stattdessen die Kompetenz des Personals der zu akkreditierenden Stelle festgestellt werden, 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

[13] Fraunhofer IPA – Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

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[14] 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

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

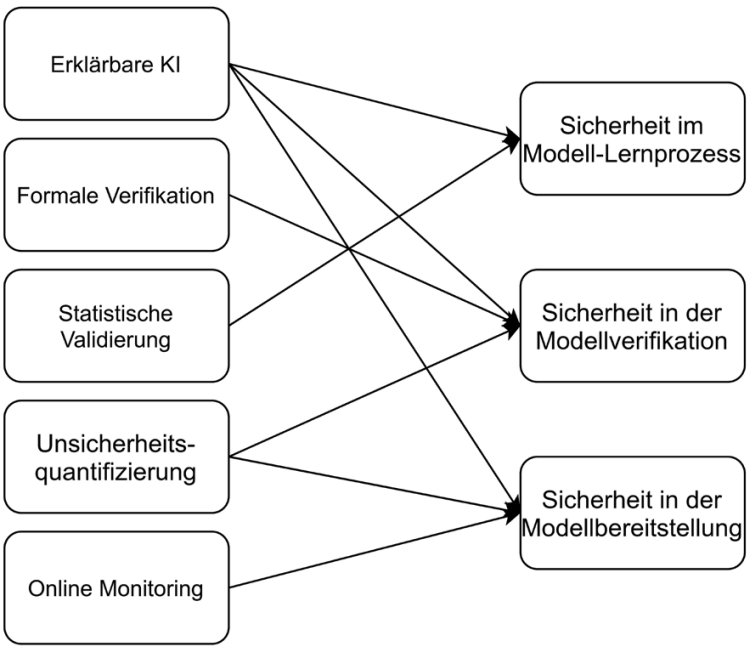


Abbildung 4: Methodenbaukasten für die Entwicklung zuverlässiger KI innerhalb der AMLAS-Sicherheitsargumentation [64]

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 Auftrags 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 mithilfe 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 der Qualität der Trainingsdaten als auch die Unsicherheit der Arbeitsdaten stellen mögliche Forschungsrichtungen dar. Dabei ergeben sich derzeit folgende Anwendungsszenarien für Soft Sensorik und damit eine entsprechend notwendige Hinterlegung der Thematik durch die PTB:

- 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. 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),
- 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.:

5. *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 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[15] (derzeit repräsentiert durch PTB-Beschäftigte) und OIML gebildet wird. Zukünftig könnte es die Aufgabe dieses Gremiums werden, ein Dokument zur Unsicherheitsermittlung im Zusammenhang mit KI-Methoden zu entwickeln.

6. 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.

7. *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.

[15] BIPM (Bureau International des Poids et Mesures), IEC (International Electrotechnical Commission), IFCC (International Federation of Clinical Chemistry and Laboratory Medicine), ILAC (International Laboratory Accreditation Cooperation), ISO (International Organization for Standardization), IUPAC (International Union of Pure and Applied Chemistry), IUPAP (International Union of Pure and Applied Physics)

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].

PTB mitteilungen

Impressum

Die PTB-Mitteilungen sind metrologisches Fachjournal der Physikalisch-Technischen Bundesanstalt, Braunschweig und Berlin. Als Fachjournal veröffentlichen die PTB-Mitteilungen wissenschaftliche Fachaufsätze zu metrologischen Themen aus den Arbeitsgebieten der PTB. Die PTB-Mitteilungen stehen in einer langen Tradition, die bis zu den Anfängen der Physikalisch-Technischen Reichsanstalt (gegründet 1887) zurückreicht.

Herausgeber
Physikalisch-Technische Bundesanstalt (PTB)
ISNI: 0000 0001 2186 1887
Postanschrift:
Postfach 33 45,
38023 Braunschweig
Lieferanschrift:
Bundesallee 100,
38116 Braunschweig

Redaktion / Layout
Presse- und Öffentlichkeitsarbeit, PTB
Dr. Julia Tesch (wissenschaftliche Redakteurin)
Dr. Dr. Jens Simon (verantwortlich)
Sabine Siems (Redaktion / Lektorat)
Sebastian Baumeister / stilsicher.design (Layout / Satz)
Telefon: (05 31) 592-82 02
Telefax: (05 31) 592-30 08
E-Mail: sabine.siems@ptb.de

Printed in Germany ISSN 0030-834X

Die fachlichen Aufsätze aus dieser Ausgabe der PTB-Mitteilungen sind auch online verfügbar unter:
doi: 10.7795/310.20220199

Bundesministerium für Wirtschaft und Klimaschutz

Die Physikalisch-Technische Bundesanstalt, das nationale Metrologieinstitut, ist eine wissenschaftlich-technische Bundesoberbehörde im Geschäftsbereich des Bundesministerium für Wirtschaft und Klimaschutz.

Metrology and AI: PTB's AI Strategy

PTB mitteilungen

Special Journal for the Economy and Science Official Information Bulletin of the Physikalisch-Technische Bundesanstalt Braunschweig and Berlin

Volume 132, No. 1, March 2022

No issue of PTB-Mitteilungen was published in 2021.

Metrology and AI: PTB's AI Strategy

Executive Summary

The increasing use of artificial intelligence (AI) is revolutionising the creation of value from (measurement) data. It opens up entirely new business areas and changes practically all areas of life and the economy. In smart homes and smart cities, smart meters and controllers enable demand-centred control and efficient billing of energy and water supply as well as optimisation of network utilisation. "Predictive maintenance" using AI significantly reduces production downtimes and maintenance costs in Industry 4.0. In the healthcare sector, AI-supported diagnosis and therapy planning improve the treatment of patients and thus significantly reduce absences from work and avoidable burdens on the healthcare system. The combination of widely used measurement technology and artificial intelligence methods therefore creates enormous benefits to both the economy and society.

This advance of the key technology AI is made possible by the progressive digitalisation of almost all our processes in the industrial and public sectors as well as the increasing availability of the associated data. Both digitalisation and the increasing use of AI create new potential for the market and fundamentally reshape the way we deal with products and services. In order to bring the benefits of AI applications to the broader society and to exploit the great economic potential of this technology, it is essential to build justified trust of users in the performance and results of AI technology and to ensure their safety when dealing with AI.

As the national metrology institute and highest-level authority for measurements as well as the associated measurement data, the Physikalisch-Technische Bundesanstalt (PTB) sees it as its mission to actively dedicate itself to this important task in interaction with the other players of the quality infrastructure (QI). Areas of action for metrology include, in particular, the evaluation of AI systems and the underlying data, the "measurement", so to speak, of AI and data quality. On the basis of suitable metrics, guidelines for the standardisation and certification of AI systems can be derived, which enable a trustworthy use of AI. Here, too, PTB aims to proactively contribute its metrological expertise to the design of the regulatory framework for AI. In addition, with regard to AI technology, it is important for PTB to re-evaluate existing metrological testing and assessment procedures with regard to their suitability for products and services with an AI component and to revise them where necessary.

In anticipation of future advances in the use of AI procedures, PTB recognises an increasing need for the provision of quality-assured, machine-usable data. Comparable to the internationally agreed standards for physical quantities, such as the prototype metre and prototype kilogram, PTB, as an essential anchor of trust for future technologies in metrology, also seeks to develop standards (e. g., "gold standards" or reference data sets) and benchmarks for the digital world and make them available to science, industry and society via a suitable infrastructure. On the one hand, these digital standards open up completely new business fields for PTB and, on the other hand, they form the backbone of customers' competitive technological innovations. It is PTB's endeavour to lead and coordinate these national digital standards internationally with the 102 member and associated states of the Metre Convention as well as with the organisations of the international quality infrastructure.

PTB is committed to strengthening the transfer of scientific results regarding AI into application and thus to always acting at the cutting edge of technology in metrological research and services. In order to advance this process, PTB itself already pilots some AI applications, among others in the optimisation of data analysis procedures, process automation, image reconstruction and for indirect measurements. These AI methods are used in a wide range of specialist departments and are successively being further implemented at PTB in suitable processes. Testing the safe use of AI at PTB ensures that the range of metrological, scientific and technical fields of application is expanded. Furthermore, AI competence is built up for research, services and administration and, at the same time, existing processes are optimised.

The combination of metrological domain knowledge, data and AI competence creates a

well-founded expertise for the challenges of the products and services of the future. This expertise is PTB's unique selling point within the AI research landscape. With its profound understanding of handling measurement data and data-driven processes, PTB contributes decisive know-how to the cooperation with AI research institutions and other QI players. Accordingly, it makes an essential contribution to the development of a reliable and confidence-inspiring evaluation, standardisation and certification of AI systems and data. The necessary skills development at PTB requires a planned and sustainable coordination of measures for networking, personnel recruitment and development. These efforts are flanked by a demand-centred expansion of the required infrastructure for computing and data organisation as well as technical support.

In order to achieve PTB's strategic objective for trust in AI, an appropriate policy framework is needed, which should be established through the following measures:

- Launching designated **flagship & pilot programmes for basic and applied research for AI** (including the development of suitable metrics for assessing the quality of AI and the data used) **involving metrological expertise**,
- Creation of an **innovation platform** for close and efficient cooperation of AI research institutions, companies, QI stakeholders and policymakers (e. g., within the framework of an innovation centre for systems metrology),
- Increase staff **resources for committee work and research tasks**, and promote education and training opportunities for the sustainable development of AI skills,
- Expansion and operation of corresponding **infrastructures** for AI research and application at PTB and
- Explicit anchoring of responsibility for metrological products and services with AI in **PTB's legal mandate.**

In the future, PTB is committed to continue the AI activities it has started and to significantly expand them both in the research area and in practical application. In addition to the development of a concrete implementation plan along the present strategic guidelines, PTB also aims to further increase its commitment and visibility within AI research and regulation. Among all players of the quality infrastructure, the research landscape and the economy, PTB would like to strengthen its reputation as a competent and proactive partner in questions concerning trustworthiness and reliability in the field of AI. In doing so, PTB also demands a more explicit responsibility for AI technologies within its legal mandate.

Authors and their affiliation with PTB's departments

Sascha Meyne (1.3)
David Auerbach (3.1)
Tobias Klein (5.2)
Matthias Neuwirth (5.2)
Ulrike Ankerhold (6.2)
Mathias Anton (6.2)
Stefan Pojtinger (6.2)
Steffen Ketelhut (6.3)
Tobias Schäffter (8)
Hans Rabus (8.01)
Lukas Winter (8.1)
Andreas Kofler (8.1)
Christoph Kolbitsch (8.1)
Patrick Schünke (8.1)
Markus Bär (8.4)
Clemens Elster (8.4)
Lara Hoffmann (8.4)
Sebastian Heidenreich (8.4)
Stefan Haufe (8.4)
Martin Nischwitz (8.5)
Marko Esche (8.5)
Andreas Barthel (9.11)
Dirk Ratschko (9.2)
Harry Stolz (9.2)
Sascha Eichstädt (9.4)
Daniel Hutzschenreuter (9.4)
Julia Tesch (9.4)
Giacomo Lanza (Q.11)
Holger Israel (Q.11)
Daniel Lübbert (Q.4)

Contents

Introduction

With the increasing availability of large volumes of data in all areas of life and the enormous technological advances in measurement technology during digitalisation, the use of artificial intelligence (AI) methods is also steadily increasing. The key technology AI is fundamentally revolutionising the understanding of products and services [1, 2] and thus acts as a catalyst for digital innovations. It is not only in Industry 4.0 where significant resources can be saved through the predictive maintenance of machines and systems using AI. New fields of application for AI are also constantly opening up in the intelligent control of supply systems in smart homes and smart cities, in self-learning diagnostic tools for personalised medicine, and in autonomous vehicles. Due to their versatility and inherent adaptability to problems of all kinds, AI systems offer outstanding economic potential as a component of products or as stand-alone items, which - if recognised and harnessed at an early stage - can decisively strengthen Germany's position on the global market and mean significant competitive advantages across the board, from start-ups to SMEs to large corporations.

In parallel to the growing scope of AI, however, there is also an increasing need for clear regulation that eliminates the associated risks of using AI, especially in critical areas such as the health or utilities sector or at least mitigates their consequences to an acceptable level. In order to sustainably strengthen the trust of customers and users in AI technology, a stringent quality infrastructure (QI) is indispensable for AI applications. Metrology is a recognised anchor of trust and an essential component of QI. This includes the characterisation of measurement technology and measurement methods, the evaluation of the quality of measurement data and the development of new measurement procedures. Principally, the legal mandate of the PTB within the framework of the EinhZG (§6 (3)), the MessEG (§ 45) and the Medizinproduktegesetz (§32 (2)) is formulated in a technology-agnostic way. At the same time, with new technological developments, it can always be assumed that there exists an expectation of PTB to continue to fulfil its legal mandate in a competent manner in the future.

PTB thus sees it as its duty to conduct fundamental research on the data quality and reliability of AI procedures. building on this research PTB seeks to advance the development of the legal framework for AI approval and regulation in cooperation with other QI stakeholders. In addition, PTB would also like to further the opportunities arising from the use of AI methods in the research and development environment and make them safely usable. With this approach, PTB follows the declared goal of the Federal Government in the "Fortschreibung der KI-Strategie" [3]:

"The Federal Government therefore advocates a suitable regulatory framework, adapted to AI-specific concerns, if necessary, in which the existing quality infrastructure is expanded and, if necessary, further developed. By setting clear rules as well as standards and norms, the fundamental rights of citizens can be protected, trust in AI can be strengthened, sustainable use as well as innovation and competition can be promoted."

The structure of the PTB's strategy paper presented here is also based on the continuation of the Federal Government's AI strategy and envisages its strategic interpretation in the field of metrology. In addition, PTB's current AI strategy complements the existing strategy for "AI in medicine", which was published in December 2020 and is being implemented as an integral part and Use Case of important initiatives such as "QI-Digital" of the BMWi (Federal Ministry for Economic Affairs and Energy). With this strategy paper, PTB addresses the opportunities and risks of using AI in the breadth of metrological research and application, while sharpening its understanding of the AI challenges and clarification needed for action in metrology.

Status quo

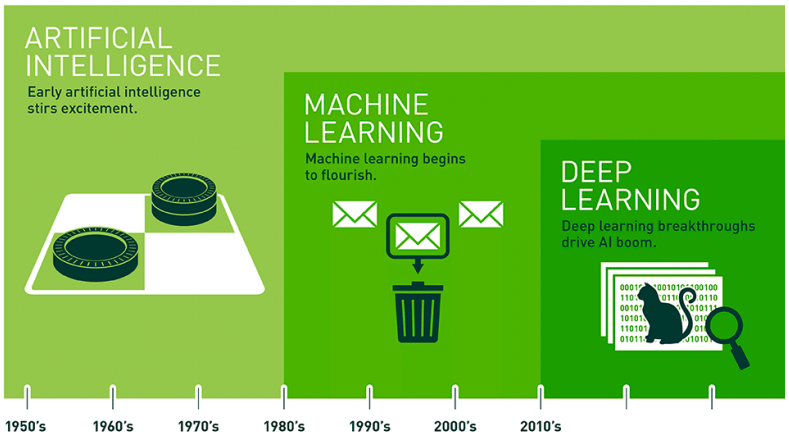
The disruptive key technology AI has long since left the niche stage in research (see Fig. 1 for historical development) and is pushing onto the market in the form of a wide variety of products and services and thus into all areas of life and the economy. In order to set corresponding guidelines for this dynamic, progressing development, the EU Commission published a White Paper on artificial intelligence [4] in February 2020. This paper builds on the European AI strategy of 2018 and places innovative, but - in contrast to the developments in the USA and China - strongly human-centred AI in the focus of further actions. This mission statement means that AI should benefit people and society while strengthening self-determined action and is often referred to as "European-style AI". The German government has also given the issue of AI high priority and embedded it in its strategic action with the AI Strategy of 2018 [5] and its update 2020 [3] as well as the statement on the EU's AI White Paper [6].

The Enquete Commission of the German Federal Government sees AI as the next stage of digitalisation driven by technological progress [7]. An essential element is the way in which these algorithms are developed. A conventional algorithm usually implements a previously described procedure in software. The procedure is based on mathematical, statistical, or other assumptions, theories and rules. In contrast, the algorithm of an AI method is trained with the help of data. These algorithms are characterised by a high complexity and a high-dimensional parameter space. Another feature is the very high adaptability of AI methods. However, this can lead to undesired features of the training data unintentionally being built into the algorithm. Therefore, in contrast to other software, a check of the algorithm based on the source code alone often no longer feasible.

Despite some very specific AI characteristics, so far, there is no uniform definition of AI and numerous other terms in the AI environment. While acknowledging the lack in universal definition, this policy paper follows the interpretation of the Federal Government of Germany and the National Institute of Standards and Technology (NIST), the US metrology institute, for a definition of terms:

AI refers to software and/or hardware that can learn to solve complex problems, make predictions and perform tasks that require "human" qualities and abilities such as (sensory) perception (e.g., vision, touch) through data acquisition, cognition, planning, learning, communication or even physical actions [8]. *AI is categorised as "strong" or "weak" AI. "Strong" AI is based on systems that equal or surpass the intellectual abilities of humans. "Weak" AI, on the other hand, refers to algorithm systems for solving concrete application problems based on methods from mathematics and computer science, whereby the developed systems are capable of self-optimisation* [5].

In addition to the question of the definition of the term, there are numerous other activities on AI in research and development - supported, among others, by the measures of the federal government and various federal initiatives, such as in the state of Lower Saxony [9]. Numerous publications on new methods, models and application examples are being published with increasing frequency. In addition, there has been a steadily growing trend for patents and licences for AI applications over the past decade. Numerous Fraunhofer Institutes, the DFKI and the DLR are massively expanding their activities in the field of AI



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Fig. 1: Historical development of AI research fields. Source: https://blogs.nvidia.com/wp-content/uploads/2016/07/Deep_Learning_Icons_R5_PNG.jpg.png

research. Currently, about 80 of the targeted 100 AI professorships have already been established. In standardisation, new working groups are being set up not only at ISO and IEC to deal with the new requirements of AI. DIN is also very active with the "Standardisation Roadmap AI" [10] as well as DIN SPEC 92001 [11, 12] and addresses the procedural changes for standardisation as well as possible fields of action. Similarly, topic-focused committees are emerging for AI related to digital health, autonomous mobility, etc., while AI continues to challenge existing certification bodies as a new component of products to be tested. An overview of the AI standardisation activities is shown in a publication of project "ExamAI – AI Testing & Auditing" [13].

PTB is already active in various areas of AI deployment and research as well as proactively finding its role and place in this dynamic environment. Large players with extensive AI expertise mainly deal with questions of feasibility and technical implementation (DFKI, etc.). PTB, on the other hand, relies on linking new AI competences with the rich technical expertise in the field of metrology and sensor technology, conventional methods in the field of simulation and data analysis as well as quality infrastructure. At the same time, PTB endeavours to contact political decision-makers and central bodies and associations at an early stage to be able to play a formative role in the further implementation of the Federal Government's AI strategy. In particular, the next steps announced by the Federal Government in [3]:

"Implementation of the roadmap defined in the AI standardisation roadmap: development of test criteria based on established and to-be-developed test technologies for testing the robustness, security, reliability, integrity, transparency, explainability, interpretability and non-discrimination of (hybrid) AI systems".

outline a predestined field of action for PTB and suitable partners.

In response to such calls for action, the first AI "nucleus" for "AI in Medicine" was established at PTB, supported by the economic stimulus programme "Fighting Corona Consequences, Securing Prosperity, Strengthening Future Capability". This acts as a competence centre with the character of a research training group and conducts both research on concrete applications and basic research on AI. Thematically, this centre will be organised around existing groups that have strong expertise, research experience and access to usable data in the field of medicine and supplemented by a new pool of ten scientists. These activities are designed in such a way that they propagate into other PTB research activities and are well interlinked with theoretical basic work and application research of other subject areas (e. g., autonomous driving, optics, etc.). Thus, a network is currently being created in which knowledge transfer and preservation are established. PTB's Digitalisation and Medicine Steering Committees as well as subordinate, topic-specific interest groups take on the coordination and prioritisation and provide guidance on the external representation of AI topics, especially to relevant decision-makers and bodies.

In the field of legal metrology, PTB is also dealing with the challenges posed by AI. Here, the national representatives, including PTB, actively coordinate activities regarding AI at an international level and are already finalising guidelines for the OIML on the use of AI in measuring instruments.

Focus areas

Human resources

PTB aims to complement its high level of metrological domain knowledge with essential AI skills in order to create and secure long-term trust in AI and act as a strong and competent authority in cooperation with partners.

In order to provide a metrological background for the steadily growing use of AI in almost all areas of life, PTB is dependent on a significant and sustainable expansion of AI competences. However, this increasing trend of AI applications and processes across all industrial sectors also means that PTB must compete with numerous companies and research institutions for the best minds to build up this competence.

In this competition, the interdisciplinarity of work at PTB is a special incentive for AI experts in the labour market. Synergistically, metrological domain knowledge from the fields of sensor technology, statistics, modelling, and QI is linked with AI skills; thus, creating practical added value. In particular, the metrological understanding of measurement uncertainties, error propagation and traceability to defined standards creates a new basis for a safe and confidence-inspiring use of AI. Due to its high economic and social relevance, this field of activity potentially stands out in the labour market from pure AI development tasks and gives PTB an advantage in recruiting personnel. In addition to metrological domain knowledge, data literacy and conceptual understanding of machine learning and data science, the skills development of employees is complemented by basic digital competences, competences in dealing with AI systems (e. g., process, problem-solving and reflection competences), and experience in designing work processes (e. g., social, organisational and self-competences) [14, 15].

Through the targeted promotion of young scientists, PTB can attract well-trained staff with comprehensive skills to contribute to AI research and development issues in the field of metrology at an early stage. Especially within the framework of joint appointments with universities or through supervised research projects. PTB inspires bachelor and master students as well as doctoral candidates in metrological AI research and development by imparting AI knowledge with metrological practical relevance. Correspondingly long-term perspectives for qualified researchers ensure the retention of competence and personnel.

The AI-specific qualification of existing PTB employees with the help of further education and training measures is also crucial. This way, they can specifically expand their AI competences, combine newly acquired knowledge with the physics and technology know-how of their research and development work and pass on their new knowledge with a metrological perspective. In addition to the interdepartmental and interdisciplinary project “Machine Learning for Medical Imaging (ML4MedIm)”, an important “nucleus” for the development of essential AI skills within PTB is the pool of ten (post-)doctoral employees of “Metrology for AI in Medicine”, which has been advertised and subsequently filled in 2021. Based on concrete Use Cases, fundamental questions on reliability, robustness and explainability of AI procedures as well as the necessary data quality are studied within this framework. The corresponding expertise is then transferred to other subject areas with AI application. In this context, the close scientific exchange within PTB on methodological procedures beyond purely technical disciplines is of particular importance. For this purpose, PTB must establish and ensure suitable structures and formats to preserve and transfer existing knowledge.

Through close national, European, and global networking with universities and other research institutions as well as funding networks, PTB offers an attractive working environment with challenging metrological issues for highly qualified scientists. In particular, by collaborating with large, established players in the AI landscape, PTB scientists can feed metrological expertise into the AI community and contribute considerably to solving the problems with significantly reduced human resources.

Research questions

PTB sets itself the goal of developing suitable metrics for the evaluation of AI and data in its metrological research mission, adapting existing measurement and testing processes to the use of AI and, at the same time, testing and expanding the safe application of AI for metrological research and services.

Despite its comparatively long history since the 1950s, which took place in several waves, the field of artificial intelligence still harbours a multitude of unanswered research questions regarding the fundamental understanding and practical application of this key technology. From a metrological perspective, two overarching complexes emerge for research on metrology and AI: On the one hand, AI itself as an object of scientific research up to the development of an evaluation of AI methods and the underlying data, and on the other hand, AI as a tool for improving metrological research and services. Accordingly, the focus area of research questions is divided into a presentation of the existing and planned activities of PTB for the development of a quality infrastructure for AI (including evaluation efforts of the data used for AI) and an overview of the possible applications of AI for metrology as a service and research activity. For each of these focus areas, two concrete Use Cases illustrate the individual research fields.

Quality Infrastructure for AI (QI 4 AI)

With its research activities in the field of AI, PTB wishes to ensure its ability to fulfil its legal mandate in the future. This includes the specifications for quality characteristics of AI set by standards as well as requirements of regulations and laws for the certification and conformity assessment of quality assurance methods for training and test data.

AI Assessment

PTB already conducts basic research and application studies for the development of evaluation procedures for AI methods. The focus here is on the development of quantitative measures for the evaluation of explainability, uncertainty, generalisability, and robustness.

For the evaluation of the functionality and performance of AI as a measure of quality, as called for in [11], the quantitative determination of the uncertainty of the AI's predictions is required. In data-based methods, the uncertainty is made up of three components [16]:

- Uncertainty due to inherent limitation in the model fit of the learning system
- Uncertainty due to data quality
- Uncertainty due to divergent training, testing and application contexts.

It is essential that the "measure" with which the uncertainty is measured is standardised, as only then uncertainties of different AI methods can be compared in their predictions at all, as required in [17]. Methods for the quantitative determination of measurement uncertainties play a central role in metrology, where there is now a globally recognised standard in the form of the GUM [18]. Such standardisation has been lacking in the field of AI, where there are a multitude of different approaches to quantifying uncertainty [10, 19, 20, 21]. The special challenge in contrast to classical measurement tasks is the strong dependence of the uncertainty estimation for AI methods on the individual problem. PTB is currently investigating the suitability of various approaches for quantifying the uncertainty of AI methods with the aim of developing a recommendation for possible standardisation. The studies include basic investigations as well as applied research [22]. From the point of view of metrology, it would be desirable if a standardisation of the uncertainty were in line with the principles of uncertainty evaluation in metrology, so that in applications where AI methods and classical methods operate in a similar way, the same uncertainties could also be assigned.

To ensure trust in AI methods, it is important to understand their behaviour and ensure that they do not merely respond to specific aspects of the training data [23] but use the relevant information in the data. Similar to uncertainty, there are now also many approaches to explainability, see e. g. [24, 25] and the references therein. One of PTB's goals in this area is ultimately to establish a standardised measure for the quantification of explainability. However, this topic requires further basic research: on the one hand, through the development of definitions for explainability and, on the other hand, through research on the inferences that explainability should allow. One option could be to define different classes of explainability, depending on the type of inferences allowed for a concrete problem. It is also conceivable that at the end of this research effort there may be no uniform measure of explainability, but instead a catalogue with concrete benchmarks for different applications. With regard to the

research question of explainability, a close cooperation of PTB with HHI (Fraunhofer Heinrich Hertz Institute) is planned. This cooperation is part of a project carried out at PTB to investigate AI methods in medical imaging from a metrological viewpoint.

The *robustness* and *generalisability* of AI methods to input data that deviates from the data used to train the method plays a major role, especially in medical technology or autonomous driving. Of particular importance are, for example, "out-of-distribution" errors, which arise because certain features are not represented in the training data. Attention must also be paid to so-called "adversarial attacks", in which "benign" input data are deliberately changed slightly in such a way that an AI method fails. To make the evaluation of robustness with regard to these influencing factors quantitatively comparable, several evaluation criteria have been proposed. PTB is investigating these criteria and has developed alternatives based on statistical approaches, which have shown very promising results in previous studies [26, 27].

Reference data and data quality assessment

All sources regarding evaluation, certification and conformity assessment of AI applications or products with AI components mention the need for reference data as well as generally accepted criteria for data quality and data handling. To perform its tasks, PTB must therefore build up expertise on these issues. In this context, the necessity of domain knowledge, which is also mentioned in [17], is important when deciding on suitable research projects. In particular, the representativeness of reference data "in itself" is not possible, but always depends on the background of a basic population. Instead, statistical criteria (e. g., test for equality of distributions) could come into play. Here, instructions for the construction of the (synthetic) reference data could also be considered a task for PTB. Metrology already deals with the assessment of data, but so far does so at the bottom-up level (GUM-like), based on the understanding of the underlying physics, rather than top-down via the properties of the data itself. In some areas, PTB already provides physical/chemical reference data. In the future, this could be further expanded with the aim of developing reference data specifically for the evaluation of AI methods. The development of methods for the generation of synthetic data sets, which are metrologically validated and traceable in a quality-assured manner, should also be considered. Especially this very typical metrological task of "synthetic reference data generation" combines the requirements of metrological domain knowledge with data skills and physical-technical understanding.

In the meantime, first examples exist for the automatic annotation of training data by combining different modalities. In [28], for example, an ML method was trained to combine tomographic images of the retina and co-registered fundus images to predict retinal thickness. As a result, the trained ML method was used to automatically annotate a dataset of 120 000 records. These in turn serve as a training dataset for ML methods to detect eye damage caused by diabetes potentially resulting in blindness. In an approval of such an ML procedure, not only the pure raw data have to be evaluated, but also the entire workflow for the use of these data. Accordingly, PTB would also have to build up competences in the field of data handling in order to be able to fulfil the requirements from, e. g., [29] and [30].

[1] EMPIR – European Metrology Programme for Innovation and Research

[2] BMBF – Federal Ministry of Education and Research

[3] BMWi – Federal Ministry for Economic Affairs and Energy (now Federal Ministry for Economic Affairs and Climate Action)

Use Case: Metrology for Autonomous Driving – Trust in AI

For the introduction of autonomously driving vehicles in road traffic, it is indispensable to test the associated functions in the course of approval procedures, which in turn requires suitable test catalogues on the one hand and technically suitable measuring equipment on the other. Due to the complexity of the problem, multi-stage certification procedures will have to be established according to current knowledge [31]. Approval must be granted both individually for the software and hardware as well as for the combined system. The metrological characterisation of individual sensors is already the subject of ongoing research – also at PTB. In vehicle use, however, the individual measurements of the most diverse sensors are aggregated and evaluated by (AI) algorithms. Only from this combination of measurement data a decision is derived autonomously for the behaviour of the vehicle.

The key task of metrology is to quantitatively evaluate the physical quantities measured by autonomous driving vehicles, the data generated, and the decisions made from them in terms of robustness, influence of measurement uncertainties and the associated effects on functionality. A long-term goal could be the creation of so-called gold standards, both on the physical measurement level and in relation to the input data of the AI decision logic. It is important that internal (e. g., ageing, technical defects) and external degradation effects (e. g., weather, dirt, precipitation, etc.) are considered. Only then can reliable statements be made in the future about the functional limits of a vehicle that shows signs of ageing, damage, or other disturbing phenomena.

Several AI research groups in Germany and around the world are working on the development of methods to implement autonomous driving (e. g., in the Competence Centre "Autonomous Driving" (AD) of the DFKI [32]). Fraunhofer IKS is also working on questions concerning the evaluation and validation of AI methods for autonomous driving. Mathematical-statistical issues are addressed as well as the appropriate implementation of the methods in software. Many other Fraunhofer activities relating to autonomous driving are bundled in the Fraunhofer Transport Alliance.

Following a review article on Deep Learning for autonomous driving [33], the following metrological questions can be formulated for the treatment of AI methods for autonomous driving:

- Understanding the impact of measurement uncertainties and sensor data degradation
- Understanding the context of data fusion and AI application within the overall system
- Definition of assumptions regarding the context in which the system will operate (operational design domain (ODD)) and how it will be tested
- Definition of basic requirements for data quality and AI procedures

From the point of view of metrology, this results in the following concrete tasks for PTB:

- Assessment of the suitability of measurement methods and procedures (e. g., camera vs. lidar use) for generating suitable sensor data
- Assessment of data from measurements and simulations for the training and testing of AI procedures; in particular, comparability of simulated and real data
- Consideration of measurement uncertainties in analysis and
- Handling the evaluation of AI method performance limits, e. g., through generalisation.

Basically, an autonomous driving vehicle is a mobile sensor network. Therefore, PTB's work in the field of AI for autonomous driving can initially build on the activities started in the subject area "Metrology for heterogeneous sensor networks". This includes work in the EMPIR[1] project "Metrology for the factory of the future" (Met4FoF), the BMBF[2] project "AAS-based modelling for the analysis of variable CPS" (FAMOUS) and the BMWi[3] project "Safe and robust calibrated measurement systems for the digital transformation" (GEMIMEG-II).

These projects already deal with methods for the use and propagation of measurement uncertainties in sensor networks as well as methods for feature extraction for machine learning considering uncertainties.

Use Case: Quality control for explainable AI in clinical diagnostics

AI is expected to play a particularly promising role in the medicine of the future, where the interaction of complex algorithms and increasingly extensive and better linked data sets can be used to solve clinically relevant questions in a targeted manner. These can be diagnoses or prognoses, for example. The most digitised areas of medicine at present are intensive care medicine and radiology. In neuroradiology, for example, it is desirable to detect early signs of neurological diseases (such as multiple sclerosis, MS, Parkinson's disease, PD, or Alzheimer's disease, AD) in the form of structural abnormalities of the brain (e. g., lesions, deposits, tissue atrophy). This is done, for example, by analysing structural magnetic resonance imaging (MRI) data, which are available from large, partly publicly available databases. Machine learning methods have recently achieved success in predicting Alzheimer's disease [34]. In addition to a high prediction quality, it is also increasingly required that the decisions of such models be "explainable" based on individual inputs (e. g., the MRI images of patients). Many "explainable AI" (xAI) methods have already been developed for this purpose [35]. However, a common problem with all these methods is their poor validation. There is no generally accepted formal definition of explainability, and the authors of most existing methods either do not provide instructions on how exactly the outputs of the method may be interpreted or provide insufficient evidence for the validity of the proposed interpretations. This situation is unsatisfactory in view of the fact that even commonly used interpretations of simple linear models are not formally tenable [36].

Due to these limitations, PTB will in future deal with both the theoretical foundations and the practical validation of explainability. In particular, formal definitions for explainability are to be developed. Synthetic data offer one possibility for this. In the neuroradiology application test case, a synthetic data set is to be produced based on real structural MRI images of healthy persons. These images are then to be controllably altered exhibiting lesions, deposits, ablations, and other structural anomalies in as realistic a manner as possible. Based on this set, prediction problems (e. g., the diagnosis or differential diagnosis of the different structural characteristics) are defined. The resulting data are suitable both as benchmarks for prediction models and their "explanations". The "ground-truth" for the latter methods results from the known positions of the structural anomalies. The quality of an explanation could then be quantified by comparing the image mask of the true anomalies and the output of the explanation method, the so-called "heat map". Metrics from image processing and signal detection theory such as Jaccard and Dice scores, as well as receiver operating characteristic (ROC) curves are suitable for this procedure. This project could provide a first step towards an objective and quantitative determination of explainability quality for this specific application.

AI for Metrology (AI4Metrology)

As in numerous other scientific fields, the use of AI methods also offers considerable potential for metrology, which should be exploited. According to a survey by EMN[4] Mathmet with responses from 13 national metrology institutes, the following areas of application of AI for metrology are of particular interest:

- Improvement of data evaluation
- New measurement possibilities
- Virtual measuring devices
- Dealing with large amounts of data (Big Data)
- Emergence of new technology areas during the digital transformation
- New services

Improved data analysis using AI includes the automation of evaluation processes, advanced methods of regression and classification as well as the optimisation of inline measurement technology. In many areas, AI-supported data evaluation can accelerate and thus reduce costs in the post-processing of measurement results. In addition, the use of AI opens a new way of dealing with the increasing amount of measurement data from both individual measuring devices and distributed sensor networks. Areas of application for AI would therefore also be multi-parameter modelling of large data sets, such as in metabolomics, or complex networks, such as in

[5] EMN – EURAMET's European Metrology Networks

the Internet of Things (IoT). Furthermore, ML methods can be used to generate and provide synthetic data sets for a wide range of uses, which is of great benefit for the enormous data demands of many evaluation methods.

In addition, AI opens up a wide field of new metrological applications. This is especially true in the field of imaging, analysis, and reconstruction (e. g., in medical imaging and microscopy), in complex sensor networks (e. g., environmental monitoring), in improved calibrations and the field of autonomous driving. Completely new metrological services, such as providing reference data sets, benchmark tests and infrastructures for trustworthy AI, as well as accelerated development cycles of existing products are also recognised as potentials of AI use.

In the field of virtual metrology, an optimisation of digital twins could be achieved from the interaction of in-silico models and AI, helping recreate real experiments in a virtual domain. In addition, data-based instead of model-based predictions, also for the maintenance periods of experimental setups (i.e., predictive maintenance), offer great potential for metrological research and application.

In some PTB working groups, AI is already being used for metrological research and services, e. g., in spectrometry, anomaly detection and embedded programming of sensor technology. The following Use Cases outline examples of applications of AI methods in metrology and highlight the applicability and transferability of these methods to related disciplines.

Use Case: AI for optical metrology – shape measurements and nanometrology

The importance of optical metrology ranges from the characterisation of surfaces and the measurement of dimensional quantities to the determination of optical properties. The objects to be examined are often irradiated with light (e. g., laser, synchrotron radiation) and scattered photons are detected. The measured quantities are not determined directly, but by a mathematical algorithm (solving an inverse problem).

One example is the measurement of optical aspheres and freeforms. An optical measurement method developed at PTB, the tilted-wave interferometer (TWI), is well-suited for this purpose and is based on an interferometric measuring principle that uses several sources to make the surface of the sample measurable at all points. The underlying design of a sample is generally known, given by the virtual topography parameters provided by the manufacturer. In order to determine the difference between a sample and its design, the interferograms belonging to the design are needed in addition to the observed interferograms of the sample. These are generated with a modelling of the TWI measurement setup and a simulation of the measurement process. The inverse problem to be solved involves tracing the differences from the simulated and measured data to the actual difference between the design topography and the sample and determining the real sample topography from this.

Another example is optical nanometrology. Shrinking structural dimensions and increased functionality requirements in the semiconductor industry pose demanding new challenges to established photonic measurement methods in the soft X-ray to IR wavelength range such as scatterometry, Mueller ellipsometry and reflectometry. Without traceability and rigorous uncertainty assessments, this will become a bottleneck for future technological developments. AI methods can help meet these challenges within a reasonable amount of time. In current practice, for example, the drift of device parameters cannot be sufficiently accounted for. In the future, neural networks should virtually map the processes and thus enable effective process control [37, 38].

Established AI methods are often aimed at applications in image recognition. For the application of optical metrology, or indirect measurements, these methods must be adapted and further developed. Ongoing work at PTB [39, 40, 41] focuses on the solution of the inverse problem by deep neural networks. In addition to the reconstructed topography, the model uncertainty is also estimated, i.e., the uncertainty of the prediction is quantified. There are various promising approaches for this.

In TWI, topography prediction and its uncertainty quantification is based on an ensemble method that scales well to high-dimensional problems. The method used is investigated using systematically introduced perturbations, e. g., in the form of a growing calibration error. In addition to uncertainty determination and reconstruction

accuracy, the generalisability of the proposed method to data outside the training range is analysed. The results are promising and show that the model uncertainty increases with increasing calibration error. This property could be used to determine when a recalibration of the virtual system is needed.

Another method was developed for optical nanometrology and is based on the use of invertible neural networks. These learn a so-called transport mapping to the target distribution and additionally, through a specially adapted optimisation, not only the desired measured quantities, such as line width, edge angle or the height of nanometre-sized lines, but also the associated measurement uncertainty [41].

The proposed deep-learning methods thus generate a significant time reduction for production facilities compared to existing conventional methods and can enable the use of these measurement methods in real time.

The application areas of AI methods developed in this Use Case are diverse and range from "computational imaging" [42], model calibration ("adaptive optics" [43]) to corrections of input parameters (e. g., position errors [44]). In general, the AI methods developed are expected to accelerate and improve complex evaluations of indirect measurements. Thus, the proposed deep-learning methods have a significant time advantage over existing conventional methods when used in production facilities and enable real-time applications. As a concrete example, AI methods are to be developed within the EMPIR project "20IND04 ATMOC" that can determine optical properties of thin-film systems or nanostructures with little computational effort.

Use Case: Soot particle characterisation with the help of AI

Soot particles are produced by combustion processes and not only pose a health hazard, but also have a climate-damaging effect due to their optical properties, whereby the fractal dimension of the particles in particular influences their behaviour. The morphology of soot particles has therefore long been studied by means of electron microscopy, with the images obtained representing two-dimensional projections of the particles. Only in a very elaborate experiment, which is not broadly suitable for applications, was the two-dimensional structure determined via coherent X-ray scattering at the Free Electron Laser at Stanford [45]. One is therefore faced with the problem of how to determine the fractal dimension as a property of the three-dimensional form (with values between 1, e. g., for a long chain, and 3 for spherical aggregates) from the usual projections. Several conventional approaches have been presented for this in the past, some of which are hardly suitable or only suitable for special cases [46]. The more suitable approaches are complex, time-consuming, and dependent on input from the users [47]. It has been shown that it is possible to replace large parts of the conventional algorithm with machine learning networks, which are shown schematically in Fig. 2.

For this purpose, the underlying machine learning network was trained with a data set obtained by Monte Carlo simulation of image generation in the scanning electron microscope. These simulations make it possible to generate extensive and high-quality training datasets in just a few days. This capability will not only enable new applications but will also foreseeably lead to further improvements of the described method, which is already on par with the conventional algorithm.

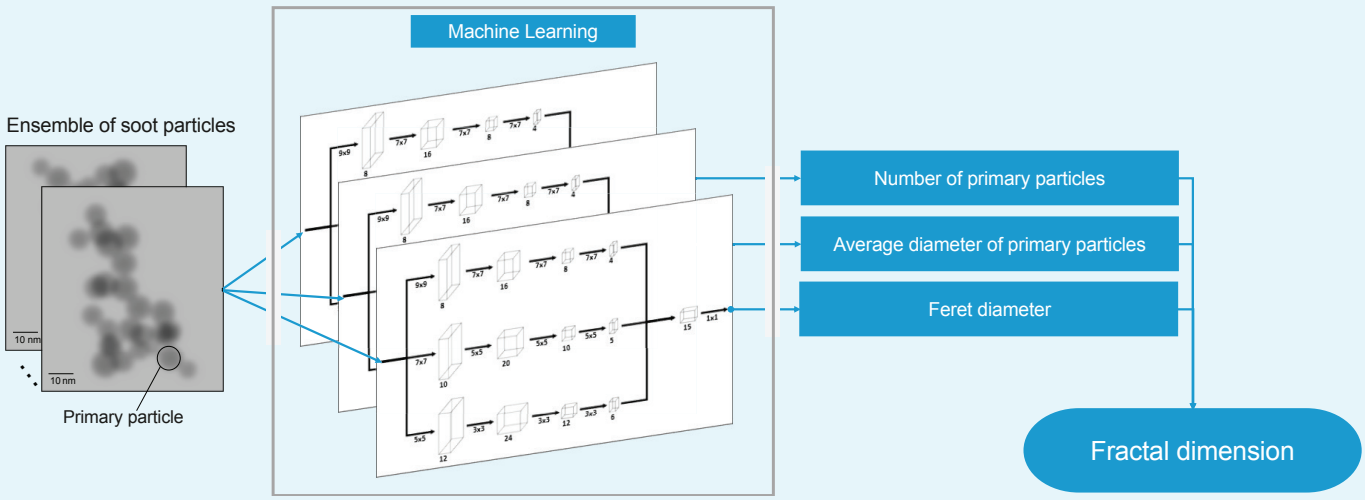


Fig. 2:
The electron microscope image of the soot particle is evaluated by three machine learning networks, which are used to determine various parameters from which the fractal dimension is derived.

The use of machine learning networks has already overcome some of the disadvantages of the conventional method [47]. Input by users is no longer necessary, which excludes any user bias, and the evaluation could be accelerated by a factor of about 10. This makes real-time use during the recording process on the electron microscope conceivable. The method is not only suitable for soot particles, but in principle for all fractal aggregates of which two-dimensional projections have been obtained, e. g., by means of microscopy. One example is dust particles in the so-called "protoplanetary disk", which played an important role in planet formation and are therefore the subject of current research [48].

Infrastructure & Data

PTB aims to establish well-organised and harmonised machine-usable data and AI methods as trust anchors for future technologies in metrology, to develop and provide digital standards (e. g., reference data sets) for metrology and to set up the necessary infrastructures.

Although AI has had limited use in various areas of application for several decades, it owes its breakthrough to the existence and constant growth of large data sets, so-called Big Data, and drastically increased computing capacities. Because this development is closely linked to advances in computing technology and capacities for processing very large amounts of data, experts even speak of the "convergence of AI and high-performance computing" (HPC)[5]. Accordingly, there is a need for sufficient computing infrastructure, which must be provided to use AI for metrological research and services at PTB, and for a comprehensive investigation of the properties of AI itself as a research object. Analogously, at a time when the data used are considered the "new oil", it becomes all the more relevant to take a critical look at the nature of data and to future-proof their provision and handling in the context of AI.

Computing Infrastructure

As mentioned at the beginning, the basic prerequisite for a successful engagement with AI topics is the availability of sufficiently dimensioned computing capacities and HPC resources for all employees in research and services. PTB can build on a good ten years of previous history: Since 2009, a cluster of Linux computers for high-performance computing has been set up at the Berlin-Charlottenburg site, to which all PTB employees have access upon request. The capacities have been successively expanded in three investment rounds so far. The latest generation of servers, which went into operation in 2017/2018, comprises of 60 computing nodes with a total of almost 1,700 CPU cores (processor "cores"). In addition, there is a fast connection network (Infiniband) for rapid data exchange between the servers, which is indispensable for distributed computing or parallel processing. The existing infrastructure is completed by a multi-level storage system, which enables both the fast storage of intermediate results via a parallel file system (scratch storage) and the permanently reliable storage of final results (Isilon storage, incl. backup and WORM[6] functionality).

Today, the existing HPC cluster is used by many users from the PTB's specialist departments: Up to now, the focus has been on classical simulations, such as finite element calculations for fluid dynamics or large Monte Carlo simulations for the determination of measurement uncertainties. In addition, the first AI-related applications have recently been developed, for example for the automated detection and characterisation of soot particles in microscopy images as presented in one of the previous Use Cases.

However, the currently existing HPC infrastructure is only suitable to a limited extent for a further intensification of AI activities. On the one hand, this is due to the limited capacities and the limited number of compute servers, which lead to longer waiting times when processing jobs on the HPC cluster (queue backlog). On the other hand, the quality of the previous HPC servers is not optimised for AI applications: In addition to the 1,700 CPU cores (main processors) mentioned above, only 2 GPU nodes (graphics processors) with graphics accelerators of the type TESLA-V100 (penultimate model generation of the well-known manufacturer Nvidia) are available. However, the use of graphics processors or GPU computing promises particularly high efficiency and performance increases in the context of data-intensive AI applications.

A comparison to the computing capacities of related institutions in the Berlin area (Robert Koch Institute, Fraunhofer Heinrich Hertz Institute),

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

[6] WORM – Write Once Read Many

which are currently intensifying their research efforts in the AI field, reinforces the conclusion that AI research must build on a higher proportion of GPU servers. Both institutes have recently invested large, five- to six-figure sums to procure TESLA-A100 GPUs (current model series from the manufacturer Nvidia). A ratio of CPUs to GPUs in the order of 2:1 to 3:1 is considered balanced.

In order to approach this target, PTB must above all significantly increase the number of GPUs. It seems sensible not to leave this to the individual research groups and departments, but to coordinate this centrally and to provide corresponding capacities for all PTB departments. As part of an expansion and replacement procurement, the existing CPUs should also be replaced by more powerful, up-to-date models. In the course of the upcoming investment cycle, it is also planned to modernise the storage infrastructure in order to significantly accelerate the reading and storage of training data sets for AI applications, and also to further increase the capacities for the longer-term storage of large amounts of data.

The basic alternative to the expansion of PTB's own computing capacities (on site / "on premise") would be the increased use of cloud computing offers. Cloud computing allows users to establish a connection to remote servers only during actual computing demand and have their computing jobs processed there. Such cloud computing offers have been on the market for several years and can by now be considered as well-established. Compared to upscaling PTB's own capacities, cloud computing has its advantages and disadvantages. The advantage of reducing investment costs and accessing theoretically unlimited capacities are offset by disadvantages in the form of higher running costs (billing according to consumption), restrictions on the transfer of larger data volumes and increased risks in terms of data protection and information security.

In order to find a suitable course of action for PTB, the following options should be considered:

1. The use of commercial cloud providers (such as Amazon Web Services, Microsoft Azure, Open Telekom Cloud, Oracle Cloud, among others).

2. The (co-)use of capacities to be publicly funded by the Federal and state governments and made available to the scientific community in Germany (NHR-Association).

3. The expansion of IT consolidation by the federal government, i.e., the planned consolidation of many IT capacities of the federal authorities with HPC or AI components as part of ITZ-Bund ("federal AI cloud").

4. Joint activities of selected federal agencies, in particular departmental research institutions (RFE), to build shared AI capacities ("RFE AI Cloud").

Option 1 is already available on the market today, but has been relatively expensive so far, and is therefore often only economically attractive for "peak loads". Option 2 is currently being developed, but according to the current decision, it will only be accessible to users from state institutions (universities), so that the PTB would only have access to it in the framework of university cooperation. Options 3 and 4 are not yet outlined in detail but could be tackled in the future.

It would be worth considering a bundling of the computing needs of federal departmental research institutions using AI (Option 4). Instead of an assessment of individual needs and lengthy individual upgrades of computing capacities as well as associated air-conditioning technology in the respective authorities, a joint "RFE AI Cloud" could potentially be set up. The computing capacities of the cloud would thus be flexibly available for all RFEs according to an allocation basis yet to be defined, and the shared use of the cloud could compensate for a temporal uneven distribution of the computing load in the RFEs. This way, free capacities of one authority would not remain unused but would be available to other participating authorities. With such an AI cloud, it would potentially also be possible to react better and more cost-efficiently to structural and air-conditioning requirements. Corresponding infrastructures would not have to be individually and expensively retrofitted as AI use grows and would not have to be made available for computing capacities that at times may not be fully utilised.

Data and AI

As data-driven methods, high-quality AI systems rely on high-quality as well as large data sets for training. While the AI models are trained on suitable training data, validation data independent of this data set is used to improve the AI model and equally independent (and unknown to the developers) test data is required for the final verification and evaluation of the AI functionality.

New trends to improve the functionality of AI methods also place a stronger focus on the

underlying data. Since the previous approach of optimising the model or the code to achieve better performance of the AI method often no longer achieves drastic improvements in many Use Cases, a data-centric approach has recently garnered a lot of attention. With this method, the training data are specifically selected according to strict guidelines (e. g., consistent labelling, discarding noisy data, coordinated handling and labelling of data sets that are difficult to evaluate) and can thus improve the performance of AI systems significantly, according to initial pilot studies [49]. In this context, too, the influence of the quality of the data and fundamental data requirements is becoming clearer than ever.

Data properties and quality

Today, measurement data is typically created in digital files and databases with user-specific proprietary formats. The transfer of this data into sustainable, interoperable, AI-usable formats usually requires manual conversion, which can only be done by the creators of the data (experts) with the necessary background knowledge of the type of data and its creation process ("data provenance"). This process often represents a great deal of additional work and therefore only takes place to a very limited extent or not at all. Digital tools for automatic data generation, which are increasingly emerging in all areas of metrology as part of the digital transformation, offer an elegant solution to circumvent the required additional effort for certain uses. New systems, however, should be developed in such a fashion from the beginning that they generate all metrology data in AI-capable formats at the "time of birth".

The development and establishment of AI-suitable digital formats for universal metrological core data based on the International System of Units (SI) is one of the long-term key objectives of the digitisation strategy of the International Committee on Weights and Measures (CIPM) [50]. Machine-usable representations for measured quantities, values, units, and measurement uncertainties are to be provided in a digital framework (SI Digital Framework), which allows the use of automatic analysis with AI methods directly and without human interaction. The technical implementation will be based on a combination of the application of FAIR[7]- principles with elementary metadata for metrological traceability of units and metrological comparability of units. The metrological principles of traceability and comparability are indispensable anchors of trust for the quality assessment and reproducibility of all measurement data worldwide.

These core metrological requirements are complemented by the following basic aspects for AI-suitable data quality:

- **Assured understanding of the background of data generation**
  Information on data generation provides an important basis for decision-making when nsidering issues of interoperability and the suitability of data for AI-based applications. This includes information from the data life cycle (time and place of origin, measuring device, validity period, etc.), on the quality of the measurement data (qualification of the laboratory, environmental conditions during the measurement, calibration, or conformity of the measuring device, etc.) as well as from the general context of the purpose of the data (research question, presence of measurement data or simulated data, etc.). The necessary metadata with the background of the data provenance are stored directly in the data (subject of the current EMPIR research project Met4FoF [51] on an annotated HDF5 dataset for ML applications).

- **Cross-domain semantics to expand the scope of AI interpretation**
  In digital data today, terms are often used that are defined in the form of controlled vocabulary lists or taxonomies in a very narrow application context and thus also only allow a rather narrow scope for interpretation. To achieve higher degrees of machine usability of data with AI as an interim solution, suitable additional semantics on the meaning of data and metadata as well as on their context from different domains is required (cf. DIN and DKE white paper on digital standards [52]).

Data organisation and handling

For the general handling of data, the use of AI offers diverse potentials that can be exploited through suitable structures and processes. This concerns the areas of data organisation, dynamic data evaluation, but also process control using AI. For the first time, AI makes it possible to automate certain data organisation processes to a significantly higher degree than conventional software. Tasks that are promisingly and efficiently completed with the help of AI include:

- data digitisation (capturing information from human-readable documents into a machine-readable database),

- data annotation (extraction and classification of metadata, e. g., demonstrated in the DiTraNo project for marking up DKE standards with machine learning methods [53]),

[7] FAIR – Findable, Accessible, Interoperable, Reusable

- data normalisation,
- assessing the quality of data,
- certain data completion and consolidation operations, such as value range detection and resolution,
- progressive, automatically learning processes of information organisation, e. g., measuring device management.

In addition, AI offers the possibility to develop dynamic and adaptive procedures for data evaluation. AI could be used, for example, in the (cross-repository) search and selection of suitable data for evaluation, routine but situation-dependent analysis procedures such as anomaly detection or uncertainty estimates, as well as in complex, adaptive analysis procedures such as the modelling of complex objects (e. g., biological, medical, sociological, ecological systems). The latter field offers an almost unlimited playground, on the one hand for use directly at PTB, but also in the environment of the planned Innovation Centre for Systems Metrology (IZSM), which will focus on addressing these systemic challenges.

AI can also support humans in process control tasks by recognising complex information contexts and initiating processes in a documentable way or by making the aggregated information available/knowledgeable to human operators (e. g., in the form of smart services). The process control areas of metrology cover a very broad spectrum of applications governed by different laws. Within this framework, questions of human supervision, legal and ethical responsibility and liability have great significance. Industrial metrology, first of all, represents a place of innovation for a short-term and medium-term emergence of AI methods to support automated processes and decisions. Initially, it is the end users of the measurement technology who will identify defective production parts from measurement data and recognise changes and anomalies. Furthermore, they make use of data of highly interconnected sensor networks of production and measurement devices (Industry 4.0) to better predict maintenance intervals. In legal metrology, with significantly stronger regulations and high-risk issues (e. g., medical technology and pharmaceuticals), the initial use of AI methods will require substantial development in the field of quality assurance of AI methods and their results. Erroneous results in AI-supported data analysis and evaluation, which would lead to incorrect dosing of medicines, for example, could have fatal consequences for doctors and patients. In particular, AI methods can only be used in critical decision-making processes if there is testable (accreditable) software and data for them (cf. EUROLAB position paper on the COM AI strategy [54]). To enable testing, even certification, of AI software results, it will be essential to develop digital standards in the form of high-quality reference data. These data standards make it possible to measure the accuracy and reliability of AI methods. Just as PTB today already assumes the sovereign task of providing physical standards for the national measurement quantities, the provision of national digital standards ("gold standard" data sets) is an important addition in the digital transformation of metrology.

PTB's field of activity will offer further areas of application for process-control AI, for example to support the generation of calibration certificates, to develop new procedures for co-calibration and to monitor data quality in the laboratory.

In addition to the use of AI in process-control, this key technology also places new demands on PTB's data handling. As a national metrology institute, PTB demands correct handling of research results in terms of reproducibility and traceability. It therefore adopts the current regulations on research data management (guidelines from the relevant research funding bodies; recommendations from the various FAIR Data initiatives; data strategy of the federal government) and acts proactively to bring a metrological understanding of data into the discourse. When setting up its own internal research data infrastructure, all aspects relevant to AI-compliant machine reusability are therefore taken into account and current fields of action are addressed. These include:

- Develop a procedure to provide data in a user-friendly and reliable way that can accept a wide range of input formats and capture information in a structured and coherent way.
- Develop a procedure for the user-friendly and reliable provision of metadata and the entire working documentation. Metadata should be extracted from the files both manually and automatically via crawler and be accessible via interfaces.
- Establishing appropriate working procedures for handling large amounts of data, possibly based on data compression or Git Large File Storage (https://git-lfs.github.com/), which references and retrieves data through persistent identifiers instead of "moving" it into a database.
- Protecting data from manipulation; ensuring its integrity and authenticity; with increased security measures such as encryption if required.

- Finding a balance between the need for open training and protected test data and data protection aspects, especially for medical data. This is achieved through close exchange with relevant expert initiatives (e. g., together with medical cooperation partners within the framework of AI4Health) and legal experts and is an ongoing process.

For its part, the recently updated Act governing the use of public sector data (Data Use Act – DUA) sets requirements for the availability, structuring and licensing of data of public relevance to enable their subsequent use. Among other things, it requires:

- Use of objective, proportionate, non-discriminatory licences justified by a public interest objective, which do not unnecessarily restrict the possibilities of use - including commercial use (Art. 4)

- Provision of data and metadata in open, machine-readable, interoperable formats, possibly language-independent, via suitable application programming interfaces and, where technically necessary, as a bulk download (§§ 7–9)

Data organisation for AI in data networks

At the intersection of the handling and quality assurance of research data in the sense of the FAIR principles, Open Data and AI applications, the National Research Data Infrastructure (NFDI) at the German level and the European Open Science Cloud (EOSC) at European level have central roles to play. The NFDI summarises its mission and added value as follows[8]:

"In the National Research Data Infrastructure (NFDI), valuable data from science and research are systematically accessed, networked, and made usable in a sustainable and qualitative manner for the entire German science system. Up to now, they have mostly been available on a decentralised, project-related or temporary basis. [...] The NFDI is intended to create a permanent digital repository of knowledge as an indispensable prerequisite for new research questions, findings and innovations. Relevant data should be made available according to the FAIR principles [...]."

The target group of the NFDI is primarily researchers at universities and other research institutions. There is thus a large overlap with the community potentially interested in using PTB's AI data and services. In its involvement in the NFDI, PTB focuses on data quality (in particular mechanisms of quality assurance and traceability of the quality of research data) as well as the documentation of research data through technically appropriate and semantically high-quality vocabularies and ontologies. At the same time, PTB is already perceived in the research world as a (possible) reference and role model regarding "Good Scientific Practice". The aspired role of PTB as a data trustee thus holds the chance to live up to this good reputation and to expand research cooperations as well as data services in Germany and Europe-wide.

Within this framework, PTB should take on a portfolio for research data and AI services that fits its mission as a long-term infrastructure task. As a supplement to the projects funded in specialist consortia, the establishment of a basic service consortium in the NFDI is planned for 2022, so that "the basic infrastructural supply for potentially all consortia is guaranteed and interoperability is permanently ensured."[9] This is necessary because the NFDI, although designed as a permanent infrastructure, is currently implemented in a project structure. PTB is particularly suited to contribute to such a necessary fundamental structure in the long term and to be perceived as a trustworthy actor for science and industry alike.

Data services for AI

In addition to the research-driven public provision of AI-suitable data free of charge, PTB will also create corresponding services within the scope of its sovereign tasks in industrial and legal metrology. Thiscourse of action will promote uniform quality standards for the development and establishment of AI in metrology. The following is a list of areas for which a great need can already be anticipated:

- Further development of existing services from legal metrology, where in-depth source code analyses of software are necessary, with additional procedures and evaluation criteria for software with AI (especially on traceability of AI).

- Use of the TraCIM[10] test system for automated online validation of AI software and procedures with high-quality reference data (quality label "QI-Digital for AI software").

- Use of the TraCIM test system for the automated online validation of data according to their suitability for further use with AI methods (quality label "QI-Digital for AI data").

[8] See https://www.nfdi.de/verein/#kurzinfo, Access: 07.09.2021

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

[10] TraCIM Traceability for computational-intensive metrology (Rückführbarkeit für rechenintensive Metrologie)

- Generating and providing high-quality reference data for AI applications for clients. Various methods are used to generate the data, such as development from existing data sets by reference software or simulated (artificially generated) data that have clear and unambiguous properties that are important for the development and testing of AI methods ("PTB Gold Standard" Data Sets).

- Custody and provision of high-quality data for AI for customers (PTB as data trustee). Especially in the focus of services, further trust features for digital data are of great importance. In addition, procedures must be installed to ensure authenticity (publisher), integrity (protection against manipulation), confidentiality (encryption, preservation of anonymity/pseudonymity) and the long-term storage and provision of data.

Regulatory framework

PTB aims to proactively shape its role as an important pillar of the quality infrastructure within a regulatory framework for AI, to revise processes based on the new requirements and capabilities, and to actively contribute its metrological expertise to the standardisation as well as the assessment and certification of AI.

On the one hand, the growing possibilities for using AI offer great economic potential for innovative technologies and increased competitiveness for Germany and Europe on the global market. On the other hand, they also pose new challenges for the existing national and international quality infrastructure. The further development of QI, considering the special characteristics of AI, is of crucial importance in order to secure people's trust in products and services and to create a clear safety and liability framework [1, 3, 55]. The basic idea of such an adapted regulatory framework is to provide comprehensive consumer protection and legal certainty for companies and thus to establish an early and sustainable acceptance of AI technology.

As a strong and essential partner in QI, the Federal Government also addresses the PTB with its call for the creation of a suitable regulatory framework adapted to AI-specific concerns [3]. This task is explicitly formulated in the update of the AI strategy [3]:

"In combination with metrology, accreditation, conformity assessment, market surveillance and environmental audits, rules, norms and standards form the quality infrastructure – the backbone of the "Made in Germany" brand. The quality infrastructure is thus an essential key to guaranteeing our economic success and trust and confidence in products and services. The Federal Government will promote the further development and bolstering of the national and European quality infrastructure in terms of the use and treatment of AI methods, in turn supporting market access, especially for SMEs in Europe and worldwide. Data quality assurance, for instance through benchmark tests, reference data, establishing and curating training data pools and setting up test data sets for validating algorithms, must also be ensured to enable the trustworthy application of AI methods. The involvement of users should also be considered."

In order to prevent nationally divergent regulation of AI applications in the EU internal market, to promote investments in innovations in the AI sector and at the same time to meet the high requirements for security and legal protection, the EU Commission published a draft for a harmonised European legal framework ("Artificial Intelligence Act") in April 2021 [55]. This draft treats AI systems according to defined risk classes:

- unacceptable risk (e. g., social scoring) by AI applications that are incompatible with citizens' fundamental rights and EU values,

- high risk for a list of AI applications (e. g., real-time biometric personal recognition, management, and operation of critical infrastructures, etc.) which are either used as security components of products subject to third-party conformity assessment in accordance with harmonised European legal acts or are listed separately due to their strong interference with fundamental rights,

- low risk (e. g., chatbots), which are subject to special transparency obligations,

- minimal risk by AI applications (e. g., spell check), whose security check is only suggested on a voluntary basis under the new legal framework.

The criteria for the classification into different risk classes are defined for this legal framework, but the risk assessment of individual Use Cases remains open to new technological developments and correspondingly changed risk assessments. For high-risk AI systems, there is a pre-market obligation for providers to undergo a conformity

assessment, which for certain products must be carried out by designated independent assessment bodies. This conformity assessment ensures the trustworthiness of the AI application in terms of data quality, technical documentation, transparency and information disclosure, human supervision, robustness, accuracy, and cybersecurity at the time of being placed on the market. In addition, providers of high-risk AI systems are required to establish advanced quality and risk management systems. These cover the entire AI product life cycle, i.e., they also ensure post-marketing feedback from users on the ongoing operation and possible misconduct of the AI systems. In the event of a significant change in the intended use of a high-risk AI system or of the system itself, a new conformity assessment is required. High-risk AI systems embedded in products already subject to conformity assessment under the EU's New Legislative Framework will be assessed for compliance within the existing conformity assessment procedure to avoid duplication and additional work for the relevant bodies. In particular, this concerns the interaction with the Machinery Regulation.

When implementing the legal framework at national level, it is up to the member states to designate corresponding competent authorities for the AI regulatory framework. With its strong role as a conformity assessment body for metrology, PTB therefore sees itself predestined to develop conformity tests as well as suitable test workflows for measuring instruments with AI components or AI systems. Based on existing structures and processes for non-AI products and services within the quality infrastructure; PTB is builds new competences and combines them with its domain knowledge to be able to adequately meet the requirements of the EU regulatory framework and future national requirements for AI.

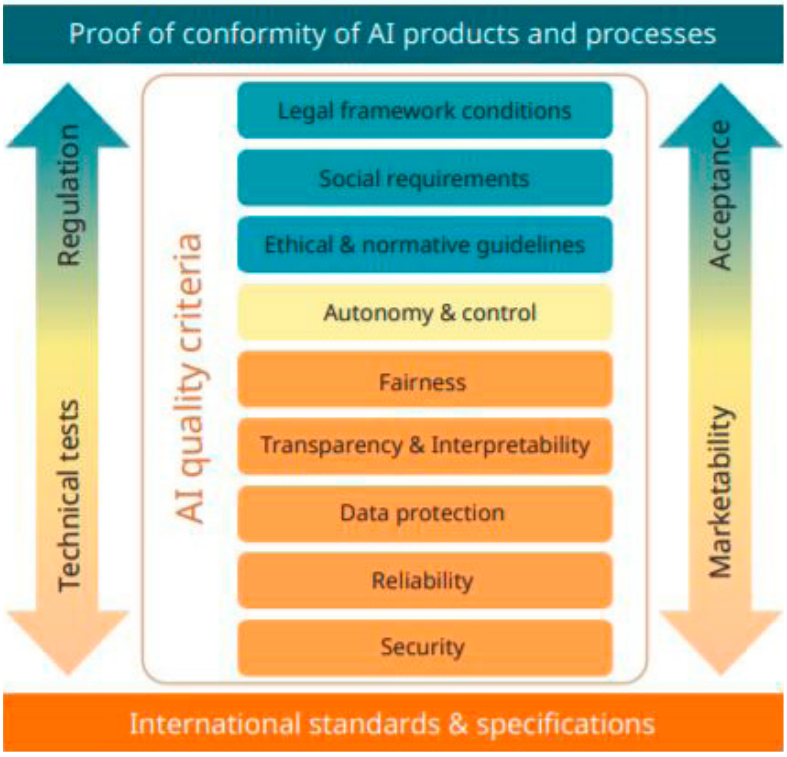


Fig. 3: Categorised quality dimensions for the assessment of AI in conformity assessment [10]

In support of these national activities, the regulation provides for the establishment of a European Committee for Artificial Intelligence, which consists of the designated national supervisory authorities for AI. Furthermore, the establishment of a platform for high-risk AI systems supervised by the European Commission is planned, in which providers must register their products. In the event of violations of the new legal framework, the regulation details appropriate sanction measures.

In addition to setting up committee structures, the regulation provides for regulatory "sandboxes" in which innovative AI systems are to be developed and tested. The responsible national supervisory authorities are explicitly assigned the task of creating innovation-friendly conditions for these sandboxes to enable a safe and forward-looking regulation of AI use. A corresponding concept for a Reallabore Act [56], which is intended to create such a universal and innovation-friendly framework in Germany, was recently presented by the Federal Ministry for Economic Affairs and Energy (BMWi). These experimental fields offer PTB an excellent opportunity to make a significant contribution to the basic understanding and quality assurance of AI applications based on measurement data; also in partnership with the "Innovation Centre for Systems Metrology" (IZSM) proposed by PTB. Together with the IZSM, other QI players and suitable companies, it will be possible, for example, to develop necessary evaluation metrics for the quality of data as well as "gold standard" data sets for training and testing. Another area of development will be benchmark tests for AI models in the subject areas of "autonomous driving", "digital medicine" and "city of the future". A corresponding concept for possible fields of action of the IZSM regarding AI as well as the complementary assignment of roles in the cooperation with the PTB has already been submitted to the BMWi [57].

In principle, the definition of AI in the draft regulation should also be questioned, as it is very broad and includes known statistical approaches, Bayesian estimation, search and optimisation methods. In the high-risk case, these application examples would be subject to possibly more stringent conformity assessments for AI according to the regulation, which would not be required for the conventional testing of these products in the existing legal framework. Criticism of this broad interpretation of the term AI is voiced by various institutions, including in the statement of the

German Electrical and Electronic Manufacturers' AssociationZVEI [58]. The implications of the AI regulation are also being discussed in the committees of the Federal Council of Germany and corresponding recommendations for adjustments are being formulated [59]. The very broad definition of AI systems is also met with criticism, if additional, economically inhibiting regulation should become necessary as a result. In principle, however, the EU-wide and risk-based approach as well as the focus on strengthening the economy and protecting citizens is supported by the Federal Council. Furthermore, it emphasises the opportunities arising from AI and the necessary protection of the economy from undue burdens due to excessive or non-transparent regulation. Inspection processes and documentation or transparency obligations are to be streamlined, thus avoiding double-burdening companies. It remains to be seen to what extent the regulation will be adapted considering these and other comments from the member states.

Standardisation and regulation of AI

The new EU legal framework for AI assigns a key role to standardisation [55]. At national level, the issues of standardisation, testability and auditability are addressed in a leading role by DIN (German Institute for Standardisation) and DKE (German Commission for Electrical, Electronic & Information Technologies). Standardisation is further represented at European level in CEN (European Committee for Standardization), CENELEC (European Committee for Electrotechnical Standardization) and ETSI (European Telecommunications Standards Institute) and internationally in ISO (International Organization for Standardization), IEC (International Electrotechnical Commission) and ITU (International Telecommunication Union). In the position paper [60] of DIN and DKE on the draft of the European AI Regulation, this important role is emphasised and a corresponding representation of the standardisation authorities in the planned European AI Board is demanded. In addition, the organisations urge that standardisation requests be formulated promptly, as preliminary work in standardisation will be indispensable for the implementation of the legal framework. This process is currently being led by the AI coordination group, which replaced the steering group of the AI standardisation roadmap, in suitable flagship projects with partners such as Fraunhofer HHI[11], Charité, BSI[12] and PTB.

An discussion paper by the Stiftung Neue Verantwortung [61] emphasises three features of AI that fundamentally reshape the approach of standardisation and certification to AI:

- technical standards are quickly outdated due to the high dynamic development of AI and require constant adaptation of the lengthy processes,
- the definition and verification of technical requirements is greatly complicated by the probabilistic nature of AI systems,
- the strong contextual dependency with a large social scope of AI as a socio-technical basic technology requires special consideration in standardisation and certification.

As a result of national activities, basic requirements and terminologies for the evaluation of AI methods have already been outlined. In a current white paper on auditable AI systems by TÜV-Association, BSI and Fraunhofer HHI [62], the quality dimensions are broken down in detail and, in addition to technical testing requirements, also consider regulatory criteria such as ethical guidelines as well as legal and social conditions (Fig. 3).

The more technically oriented DIN SPEC 92001-1 [11] defines three essential requirements for the quality of AI:

- Functionality and performance as an expression of AI's ability to complete the set task under specified conditions (reliability).
- Robustness as AI's ability to deal with erroneous, noisy, unknown or harmful input data.
- Explainability as an expression for the possibility to understand and comprehend the reasons for the result of an AI method.

DIN SPEC 92001-2 [12] distinguishes between adversarial robustness (AR) and corruption robustness (CR). The former refers to robustness against adversarial changes to the input data, the latter stands for robustness against noise or changes in the statistical properties of the input data. For the development of robust AI methods, [12] recommends a risk analysis-based approach. Methods for the targeted testing of AI methods are also mentioned (Fast Gradient Sign Method; Projected Gradient Descent). In general, scenario-based testing is recommended, which considers the later intended use and its characteristics. It is considered important [12] that the risk assessment of an AI method must be carried out continuously along the AI lifecycle.

[11] Fraunhofer HHI – Fraunhofer Heinrich-Hertz-Institut

[12] BSI – Bundesamt für Sicherheit in der Informationstechnik

In a recent discussion paper [30], the US Food and Drug Administration (FDA) also assumes the necessity of a "Total Product Lifecycle Regulatory Approach" (TPLC) for AI applications. During the certification of an AI-based medical device, the company's quality management in particular should be assessed with regard to the following aspects:

- Quality assurance in software development
- Testing and performance monitoring of the products

In doing so, the FDA establishes the following basic principles:

- Establishment of recognised "good machine learning (ML) practices"
- Consideration of the product life cycle in the approval of AI-based medical devices
- Expectation that manufacturers implement a risk-based approach to monitoring their AI-based medical devices for the entire product life cycle
- Transparent statements for customers and testers on actual performance and behaviour of AI-based medical devices by manufacturers

For the FDA, this also includes documenting the designated context of application (Software as a Medical Device (SaMD) Pre-Specification - SPS) as well as the (further) development in an "Algorithm Change Protocol" (ACP): Data management, re-training, performance evaluation, update procedures. SPS and ACP are then essential points in the approval of new products. The DIN AI standardisation roadmap [10] also calls for a graduated threshold for standardisation and approval depending on the planned area of use of the AI application via the so-called risk-adaptive criticality test.

According to the FDA [30], an important part of the quality assurance and management system is the choice of training and test data and the selection of user data for re-training. As part of data management, the FDA therefore requests

- Protocols for data collection
- Quality assurance systems for the data
- Determination of a reference standard
- Auditing and securing test and training data

by the manufacturers. The questionnaire of the German "Association of Notified Bodies for Medical Devices in Germany" (IG-NB) for the approval of AI for medical devices also devotes many questions to the selection and assessment of the data used [63]. Simultaneously, there is a lack of corresponding norms and standards as a basis for the assessment of the AI and the underlying data in [63], particularly regarding important questions.

AI certification

In the white paper by Fraunhofer IAIS [17], a certification for AI applications that can be operationally implemented by accredited auditors is discussed. According to this, a certificate for AI should:

- certify a certain standard of quality
- help to make AI applications verifiably legally compliant
- make AI applications comparable

In doing so, the IAIS states that domain knowledge and mathematical-statistical expertise are necessary for the assessment of AI reliability [17].

In the "White Paper: Dependable AI" [64], Fraunhofer IPA[13] names certification together with transparency at system level as key factors for dependable AI and outlines a corresponding security argumentation (AMLAS, "Assurance of Machine Learning for Use in Autonomous Systems" [65]) for the development of trustworthy AI procedures (see Fig. 4). In particular, [64] highlights various methods of current research that could contribute to certification in the context of an AMLAS:

- Explainable AI
- Formal verification
- Statistical validation
- Uncertainty quantification
- Online monitoring with boundary conditions

EUROLAB makes a clear distinction between indirect (indirect conformity assessment - ICA) and direct (direct conformity assessment - DCA) application of AI methods [54]. In ICA, the AI method serves as a support for decision-making. AI used for the evaluation of an X-ray measurement can be considered as an example of this distinction: Given an AI system that

determines a qualitative statement from the measurement data, e. g., about the patient's state of health. In this case, an assessment of the AI method itself would not be necessary for accreditation, but the competence of the staff of the body to be accredited would have to be determined. This would correspond to the procedure already practised today for any other non-linear numerical methods. However, if the AI method presents the result as part of the measurement (e. g., as an overlay) and thus gives the appearance of a "real result", the AI method itself must be understood as an "authority" and must be considered in the accreditation. EUROLAB currently recommends using DCA only in very uncritical areas (e. g., music quality assessment) until the methods are more mature. The EUROLAB position paper [54] also calls for a type of calibration for AI methods as with common measuring instruments. This should make the reliability of the AI method assessable by recording the ability of the method to reproduce the result of a "standard".

The white paper "Certification of AI Systems" [29] of Plattform Lernende Systeme states:

"Before a successful certification of AI systems can be established, open questions must […] be clarified. These concern the subject of certification, the test criteria, the timing and necessity of certification, the level of detail of certification, and how to deal with systems that continue to learn."

This assessment is also underlined by the white paper "Towards Auditable AI" [62] by TÜV-Association, BSI and Fraunhofer HHI, which proposes two strategic approaches to be introduced in parallel for the process of establishing successful audits of AI systems:

- Choice of appropriate (constrained) frameworks (e. g., complexity, scalability, generalisability) to achieve acceptable IT security, audit quality, robustness, and verifiability for individual applications
- Increased investment in AI research and development to gradually extend safe AI application to complex frameworks (increasing scalability, generalisability, etc.).

Furthermore, [29] also refers to the AI High Level Expert Group of the EU Commission (COM), which recommends the following criteria in the regulation of AI:

"Priority of human action and oversight, technical robustness and security, privacy and data quality management, transparency, diversity, non-discrimination and fairness as well as social and environmental well-being and accountability."

In particular, for high-risk AI applications (such as in the areas of health, biometric recognition and critical infrastructure), it is recommended check the following aspects during the conformity assessment:

- whether the training data is adequate for the intended use
- the results do not lead to discrimination in use
- data protection and privacy are respected
- relevant records on data sets, training methods and programming methods are available

These recommendations basically follow those of the FDA [30], which (as well as the federal government in its statement on the COM white paper) also considers a repeated examination of these criteria to be necessary for continuously learning AI systems.

The French metrology institute LNE has also chosen a process-oriented approach for the certification of AI comparable to the FDA's assessment. Instead of certifying the functionality of the AI system itself, the process steps along the design, development, evaluation, and operation of AI systems are set by the LNE for the first time in a certification standard [66].

In a recently launched flagship project of DIN, corresponding guidelines for AI certification along the AI life cycle are being developed with the focus

[13] Fraunhofer Institute for Manufacturing Engineering and Automation IPA

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

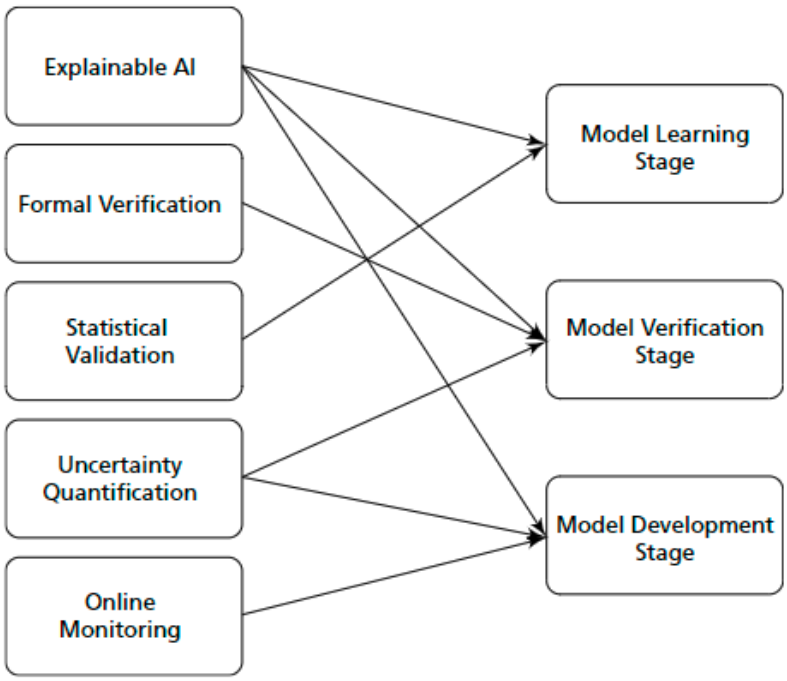


Fig. 4 Method kit for developing dependable AI within the AMLAS security argumentation [64]

on:

- Development of standardisation-ready test criteria and methods for AI systems
- Development of safeguarding methods and testing tools
- Transfer to commercial offers

For this purpose, a broad-based participation process is planned, in which PTB will actively participate with its expertise in testing and assessment according to its capabilities.

Outline of PTB's role

In accordance with its legal mandate, PTB already carries out conformity assessments and tests of measuring instruments and software on a large scale. Due to the diverse application possibilities of AI in measuring instruments and sensors of all kinds, it is foreseeable that AI will become a new and essential component of many products in the future. The responsibility for adapting the associated testing and assessment processes is formulated very clearly by Plattform Lernende Systeme:

"Existing national structures and procedures should be used for conformity assessment. Where no such authorities exist, there should be an obligation to establish one or to build up responsibilities in existing authorities." [29]

The Federal Government confirms this assessment in its statement [6] on the EU white paper. PTB, as an essential part of the quality infrastructure, particularly with its widely-recognised neutrality, is thus predestined to take on this role. In other words, PTB is already expected to fulfil its legal tasks, also for measuring instruments with AI components, by continuously building the corresponding expertise and actively contributing to standardisation with its domain knowledge. In its statement on the COM White Paper on AI, the Federal Government presents the following recommendations:

- Binding legal requirements should be considered for training data of AI systems. For this, requirements for test and evaluation data should also be considered.
- Legal requirements for quality parameters and requirements for training, test and evaluation data are also needed so that appropriate AI systems are developed with quantitatively sufficient and high-quality data sets.
- Robustness and accuracy requirements should cover predictable and realistic scenarios.
- If suitable procedures are available to check the AI results for representativeness and balance, access to the training/test data can be dispensed with.
- Central to this is ensuring the confidentiality, integrity, and availability of the AI system as such throughout its entire life cycle.
- For high-risk AI systems, going through an objective conformity assessment procedure should be mandatory. There is a need for repeated assessments of evolving adaptive AI systems.
- The accuracy and relevance of the data are crucial. Furthermore, it is necessary to ensure the provision of reference data, benchmark tests and the verification of algorithms using quality-assured, trustworthy reference data. In principle, the statement that data quality must be ensured throughout the entire lifecycle is supported. However, it should be noted that this can only be provided by the operator, who in turn is not an economic operator in the sense of product safety law.
- If a new product has been created as a result of a software change, this product must fully comply with the state of the art, as a "new" product is being placed on the market. This must be taken into account when considering a software change.

PTB's research activities in the field of AI must therefore ensure that these requirements and expectations can be met. Based on this, PTB must develop solutions to test and evaluate products with AI. This development is outlined exemplarily for various specialist areas of PTB.

In the field of dosimetry, the development of AI-based methods is still in its infancy, but the first fields of application are already emerging. The handling of such applications should be critically discussed in the relevant standardisation bodies for dosimetry (e. g., IEC/TC62/SC62C/WG 3) and can build on current efforts with more general harmonised standards (ISO) for AI application. A particular challenge in the area of diagnostic dosimeters already existed in the past, within the framework of conformity assessments; test specimens could potentially be too strongly adapted to PTB's test procedures by means of software. This problem would be further intensified with the use of AI software and would have to be considered in the corresponding

standard (DIN EN 61674), especially regarding generalisability. In the field of radiation therapy, a first manufacturer has already put AI-based measurement software for the determination of important dose measurands on the market. These measured quantities are relevant in the standards for reference dosimetry dealt with by PTB. Thus, strict criteria according to which such algorithms can be evaluated would be highly desirable. In addition, the future use of AI in the field of radiation planning has become apparent.

For imaging procedures that use ionising radiation, the Radiation Protection Act (StrlSchG §14 (1) 5 a) and the Radiation Protection Ordinance (StrlSchV §115 and §116) require testing procedures that allow optimisation of the ratio of image quality to patient dose. The latest generation of X-ray tomography (CT) equipment uses AI algorithms for image reconstruction. Since the training data as well as the image reconstruction methods used are in general not accessible for testing, the image quality can thus only be tested by considering the CT device as a "black box". Since AI algorithms are trained with the help of anatomical structures, the conventional approach of using technical test specimens (phantoms) to quantify the resolution capability and low-contrast detectability is no longer viable. Thus, alternative test methods and possibly new types of test specimens must be developed. Similarly in mammography, image processing methods without AI are used that are optimised for anatomical structures. They do not function in the same way as for technical phantoms, so that only a quality check of the raw images ("for processing") is possible, but not of the processed images presented to the radiologists ("for presentation"). With the introduction of tomosynthesis devices in mammography screening likely in the next few years, it must be expected that AI procedures will also be used there. Modern image reconstruction methods thus require the development of new ways of image quality assessment. A review of AI per se is completely out of the question in this context; the challenge is to develop "black-box" procedures that can cope with AI-systems and other undisclosed image reconstruction methods.

AI technologies are also likely to find their way into legal metrology in the long term. Isolated enquiries from measuring instrument manufacturers indicate that the use of AI for the purpose of calculating measured values (evaluation of sensor data) is of particular interest. In view of these developments, a subgroup of the OIML TC5/SC2/p4 under the leadership of PTB is currently dealing with globally harmonised software requirements for AI from the perspective of software and data security. It is the understanding of the international group that the currently existing requirements, especially the legal framework of the European Measuring Instruments Directive (MID), are already flexible enough for AI algorithms. The rationale lies in the view of AI as immutable software with highly variable parameters that define the behaviour of the software. This scenario is well known in legal metrology and leads to the situation that, from a software security point of view, questions regarding the transparent labelling of readings calculated by means of AI, logging, and traceability of any changes to the AI as well as software security testing of AI are of primary concern. The next step in standardisation is the international exchange of experiences and developments regarding the use of AI in legal metrology. Following this process, the requirements for AI currently in the draft stage must be examined with regard to their usability and adapted if necessary. It is expected that with this procedure, suitable requirements for AI can be sufficiently specified before such systems are brought onto the market in large numbers.

Building on activities for the evaluation of AI and the provision of reference data, an extension of the PTB service for the validation of algorithms (TraCIM) to AI procedures is generally conceivable. In the course of such a validation, the functionality of the AI would have to be tested by means of representative reference data sets. Furthermore, its robustness would have to be ensured, if necessary supplemented by a software test to safeguard against manipulation.

Another industry-related field of action for PTB in its role as national metrology institute is the quality assessment of "indirect measurements", in which AI methods are used to generate "measurement data" for quantities/locations/times for which no real measured values are available. In this context, one also speaks of "soft sensing", "sensor fusion" or "virtual sensing". It is of basic metrological interest to make quality statements about these data. The influence of the quality of the training data as well as the uncertainty of the working data represent possible research directions. At present, the following application scenarios for soft sensor technology present themselves and thus pose expectations for to be addressed by PTB:

- Monitoring of process parameters in hard-to-reach places (by using adjacent sensors + AI)
- Replacement of an expensive (special) sensor (by using cheaper sensors + AI)
- Continuous monitoring of a product quality parameter instead of widely spaced manual

measurements (by using existing sensors + manual measurements + AI)

For these and other fields of action in metrology, PTB sees itself as responsible for reacting proactively to the challenges of AI use to make innovative technologies available in a safe and trustworthy manner for the promotion of the economy and for the benefit of society. PTB's active participation in the relevant committees, networking with industry, users, and associations as well as at a political level plays a central role.

This understanding of the PTB's role leads to the realisation that the PTB must build know-how and skills in the field of AI and establish intensive cooperation with competent partners in order to fulfil its legal mandate in the long term. Due to the limited resources, it is desirable for PTB if in the future its mandate includes products and services with AI more explicitly and this responsibility is reflected within the legal frame. Especially in the new regulatory framework for AI, PTB sees itself as a supporting pillar within the (digital) quality infrastructure and strives for an early integration of its metrological expertise into QI processes. These include specific applications as well as fundamental research on AI and its embedding in QI. Furthermore, an AI-competent PTB is assumed to be an important building block for the establishment of an IZSM to actively participate in the research questions of systems metrology and subsequently ensure the fulfilment of PTB's permanent tasks.

Research cooperations

Even in the long term, PTB will not be able to compete with the scope of work and the capabilities of other large research associations - nor will it have to. For the most effective use of the available resources, PTB will therefore enter into targeted cooperations with national, European and international research partners.

A good overview of the current research landscape in Germany in the field of AI is provided, among other things, by the "Map AI" of the Plattform Lernende Systeme[15]. Particularly outstanding institutes for AI research and development are:

- German Research Center for Artificial Intelligence (DFKI)
- Fraunhofer Gesellschaft
- German Aerospace Center (DLR)
- Helmholtz Gemeinschaft (in particular Helmholtz AI Cooperation Unit)
- Max Planck Institute for Intelligent Systems
- Mevis Medical Solutions (Bremen)
- University hospitals (e. g., Charité) and medical faculties, e. g., University of Duisburg-Essen and University Medicine Essen (Institute for Artificial Intelligence in Medicine)
- Facilities in Cyber Valley in Tübingen as Europe's largest research consortium in the field of AI with partners from science and industry
- Robert Koch Institute (RKI) (establishment of a centre for the validation of AI algorithms in health research)
- TÜV/ DEKRA

In some cases, PTB aligns itself with organisational and structural decisions of large research institutions and is in close professional exchange with them through various cooperations. In numerous networks, PTB is also actively involved in the development of metrological contributions to the quality assurance of AI and the underlying data, among others:

5. *Working Group 1 on "Expression of Uncertainty in Measurement" of the Joint Committee for Guides in Metrology (JCGM)*: Uncertainties presented in certificates and calibration certificates, or CMCs (Calibration and Measurement Capabilities), are all based on the uncertainty evaluation methods of the Guide to the Expression of Uncertainty in Measurement (GUM). The maintenance and further development of the GUM is carried out by Working Group 1 of the Joint Committee for Guides in Metrology, which consists of representatives of BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP[15] (currently represented by PTB employees) and OIML. In the future, it could become the task of this body to develop a document on uncertainty evaluation in connection with AI methods.

6. The *"ITU-WHO-Focus Group "Artificial Intelligence for Health"* aims to set standards for the evaluation and validation of AI-based methods for healthcare. The PTB is in touch with the working groups "Clinical evaluation of AI for health (WG-CE) or "Ethical considerations on AI for health (WG-Ethics). In addition, there are thematic groups on various medical fields or specific diseases (e. g., TG-Radiology, TG-Malaria). The focus

group is currently preparing so-called trial audits in which concrete AI applications in medicine are evaluated based on a catalogue of criteria developed in advance with regard to quality aspects such as predictive quality, robustness, fairness and explainability.

7. *European Metrology Network on "Mathematics and Statistics" (EMN MATHMET)*: Many modern measurement methods use mathematical and statistical methods and the associated algorithms. Metrology is therefore increasingly dependent on advanced knowledge of modelling and simulation methods as well as statistical data analysis, especially for AI methods. PTB currently coordinates the newly founded network, which works at the interface of metrology and mathematics and promotes exchange within the European framework. AI is a focal point of MATHMET's emerging strategic research agenda, with a focus on the development of guidelines for the evaluation of algorithms, software, and reference data, with particular attention to uncertainties, robustness and explainability. Due to the overlap of personnel between MATHMET and numerous metrological standardisation committees, the outputs of MATHMET are efficiently fed into the standardisation work.

In its cooperation with partners, PTB generally focuses less on the development of new AI methods, but instead on the development of evaluation methods and the provision of reference data sets. An important unique selling point of PTB remains its profound domain knowledge and its neutrality.

[15] BIPM (Bureau International des Poids et Mesures), IEC (International Electrotechnical Commission), IFCC (International Federation of Clinical Chemistry and Laboratory Medicine), ILAC (International Laboratory Accreditation Cooperation), ISO (International Organization for Standardization), IUPAC (International Union of Pure and Applied Chemistry), IUPAP (International Union of Pure and Applied Physics)

Recommendations

The objectives outlined in the previous chapters will provide a basis for comprehensive strategic considerations regarding the practical implementation. As a starting point, the development of PTB activities in relation to AI will certainly require activity in the following focus areas:

I. Basic research on methodologies and tools for the assessment of large data sets and development of "good practice" examples regarding uncertainty, accuracy, representativeness and comparability

II. Development of reference datasets to assess the quality of AI

III. Upgrading PTB infrastructures for the provision of developed reference data sets (connection to the customer portal, etc.)

IV. Develop appropriate metrics for assessing AI performance, which includes robustness, explainability and dependability

V. Upgrading metrological services aiming at offering validation of AI algorithms (assessment of AI performance and software testing)

VI. Development of recommendations for annotation rules and use of metadata (especially units, uncertainties, measurement methods) in selected areas of application

VII. Further development of measurement methods and measurement data evaluation using AI

VIII. Transfer of scientific results on AI into application for metrological services, research and administration

IX. Use of AI methodologies for metrological scientific tasks and data organisation as well as process control.

References

[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

As detailed within this policy document, the terminology related to artificial intelligence is still the subject of ongoing research and therefore does not allow for a final definition at this time. However, to ensure a broad understanding of the content of the document, this glossary provides an explanation of the most important keywords.

Explainability of AI systems should provide persons with comprehensible justifications for the results of an AI model, thus ensuring comprehensibility [67]. Explainable AI can thus form the basis for human understanding of and trust in AI [68]. A distinction is made between local or data explainability of individual decisions and global or model explainability of the mechanisms of action [67]. Open questions of research in connection with explainability concern, among other things, the development of metrics for explainability as well as estimates for the validity of the statements of explainable AI (xAI).

Artificial intelligence (AI) comprises software and/or hardware that can learn to solve complex problems, make predictions and perform tasks that require "human" qualities and abilities such as (sensory) perception (e. g. vision, touch) through data acquisition, cognition, planning, learning, communication or even physical actions [8]. A distinction is made between "strong" and "weak" AI. "Strong" AI assumes systems that equal or surpass the intellectual abilities of humans. "Weak" AI, on the other hand, refers to algorithm systems for solving concrete application problems based on methods from mathematics and computer science, whereby the developed systems are capable of self-optimisation [5].

Robustness describes the ability of an AI system to deal with or compensate for erroneous, noisy, unknown or maliciously manipulated input data (stationarity). Robustness is therefore an important pillar in quality assurance for AI [12].

Transparency refers to the intelligibility of usually opaque AI model [67, 68, 69] as well as the statistics of the underlying training data and includes three partially hierarchically dependent aspects: Transparency of the overall model (simulability), at the level of the individual components (subdivisibility) and at the level of the training algorithm (algorithmic transparency) [69]. In addition to the model and data transparency described above, a system-level view of AI transparency is evolving that encompasses the totality of processes within the AI lifecycle [64, 30].

PTB mitteilungen

Imprint

The PTB-Mitteilungen are the metrological specialist journal and the official information bulletin of the Physikalisch-Technische Bundesanstalt. As a specialist journal the PTB-Mitteilungen publish original scientific contributions and general articles on metrological subjects from the areas of activities of the PTB. The Mitteilungen have a long tradition dating back to the beginnings of the Physikalisch-Technische Reichsanstalt (founded in 1887).

Publisher
Physikalisch-Technische Bundesanstalt (PTB)
ISNI: 0000 0001 2186 1887
Postal address:
P.O. Box 33 45,
38023 Braunschweig
Delivery address:
Bundesallee 100,
38116 Braunschweig

Editorial Staff / Layout
Press- and Information Office, PTB
Dr. Julia Tesch (Scientific Editor)
Dr. Dr. Jens Simon (Editor in Chief)
Sabine Siems (Editing / Proofreading)
Sebastian Baumeister / stilsicher.design (Layout and typesetting)
Phone: (05 31) 592-82 02
Fax: (05 31) 592-30 08
Email: sabine.siems@ptb.de

Printed in Germany ISSN 0030-834X

The technical articles from this issue of the PTB-Mitteilungen are also available online at:
doi: 10.7795/310.20220199

Federal Ministry for Economic Affairs and Climate Action

The Physikalisch-Technische Bundesanstalt, Germany's national metrology institute, is a scientific and technical higher federal authority falling within the competence of the Federal Ministry for Economic Affairs and Climate Action.

Bundesministerium
für Wirtschaft
und Klimaschutz

Die Physikalisch-Technische Bundesanstalt, das nationale Metrologieinstitut, ist eine wissenschaftlich-technische Bundesoberbehörde im Geschäftsbereich des Bundesministeriums für Wirtschaft und Energie.

PTB

Physikalisch-Technische Bundesanstalt
Braunschweig und Berlin
Nationales Metrologieinstitut

Bundesallee 100
38116 Braunschweig

Presse- und Öffentlichkeitsarbeit

Telefon: 0531 592-3006
Fax: 0531 592-3008
E-Mail: presse@ptb.de
www.ptb.de