DIN DKE

DEUTSCHE NORMUNGSROADMAP
KÜNSTLICHE INTELLIGENZ

Gefördert durch:

Bundesministerium
für Wirtschaft
und Energie

aufgrund eines Beschlusses
des Deutschen Bundestages

HERAUSGEBER
Wolfgang Wahlster
Christoph Winterhalter

DIN e. V.

Burggrafenstr. 6
10787 Berlin
Tel.: +49 30 2601-0
E-Mail: presse@din.de
Internet: www.din.de

DKE Deutsche Kommission Elektrotechnik
Elektronik Informationstechnik in DIN und VDE
Stresemannallee 15
60596 Frankfurt am Main
Tel.: +49 69 6308-0
Fax: +49 69 08-9863
E-Mail: standardisierung@vde.com
Internet: www.dke.de

Bildnachweise:
Titelbild: LightFieldStudios – istockphoto.com
Kapiteleingangsgrafiken: kras99 (S. 9, 33), assistant (S. 23),
Thitichaya (S. 27), Maxim (S. 35), Shutter2U (S. 63),
gunayaliyeva (S. 79), peshkov (S. 97), LuckyStep (S. 117),
kaptn (S. 127), ryzhi (S. 135), Alex (S. 143),
pickup (S. 151) – stock.adobe.com

Stand: November 2020

VORWORT

links: **Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster**
Leiter der Steuerungsgruppe, CEA DFKI

rechts:
Christoph Winterhalter
Vorsitzender des Vorstandes, DIN

Sehr geehrte Leserinnen und Leser,

Mit der Normungsroadmap Künstliche Intelligenz legt Deutschland als erstes Land weltweit eine umfassende Analyse des Bestands und des Bedarfs an internationalen Normen und Standards für diese Schlüsseltechnologie vor. Dabei werden in dieser ersten Ausgabe der Deutschen Normungsroadmap in einem breit angelegten interdisziplinären Ansatz nicht nur die technischen sondern gleichwertig auch die ethischen und gesellschaftlichen Aspekte von Normen in der KI ausführlich berücksichtigt.

Mit der Erstellung dieser Roadmap wurde eines der zwölf Handlungsziele der KI-Strategie der Bundesregierung von 2018 umgesetzt, die unter Aufgabe 10 „Standards setzen“ ein gemeinsames Projekt mit DIN zu diesem Zweck vorsieht. DIN und DKE haben dann im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi) am 16. Oktober 2019 in einer Auftaktveranstaltung mit über 300 Teilnehmerinnen und Teilnehmern aus Wirtschaft, Wissenschaft, Zivilgesellschaft und Politik die Arbeit an der Roadmap offiziell gestartet.

Hierbei handelt es sich um ein laufend fortzuschreibendes Dokument, das regelmäßig aktualisiert werden muss, um der enormen Entwicklungsdynamik der KI-Technologien und ihren sich rasant ausweitenden Anwendungsgebieten gerecht zu werden. Obwohl alle bisher veröffentlichten Normen und Standards im Bereich der KI dokumentiert und die zahlreichen laufenden Normungsaktivitäten in der Roadmap aufgezeigt sind, wurden viele „weiße Flecken“ auf der KI-Normungslandkarte identifiziert, die für die nächste Version erschlossen werden müssen.

Die Roadmap wurde in sieben Arbeitsgruppen erstellt, die neben den Grundlagen und den drei für Deutschland besonders wichtigen KI-Anwendungsfeldern Industrielle Automation, Mobilität/Logistik und Medizin wichtige Fragen und Handlungsempfehlungen zur Ethik, Qualität/Konformitätsbewertung/Zertifizierung und IT-Sicherheit als horizontale Themen ausgearbeitet haben.

KI bildet derzeit die Speerspitze der Digitalisierung, weil durch KI zahlreiche kognitive Leistungen, die bislang nur mit menschlicher Intelligenz erbracht werden konnten, erstmals automatisierbar werden. KI-Systeme werden als reine Software oder als cyber-physische Systeme realisiert, die aber stets mit anderen aktuellen IT-Komponenten verknüpft werden müssen, um in der Praxis einsetzbar zu sein. Wir betrachten KI daher im Kontext anderer Digitalisierungstrends wie Cloud-, Edge-, GPU- und Quanten-Computing, dem Internet der Dinge und 5G, Industrie 4.0 und der Plattform-Ökonomie.

Seit der ersten Welle der Digitalisierung sind die meisten Daten maschinenlesbar, denn sie resultierte in der umfassenden Ablösung der analogen Informationsverarbeitung und zum nahezu vollständigen digitalen Erfassen, Speichern, Übertragen und Speichern von Daten. Dabei haben zahlreiche Normen und Standards geholfen. Aber die zweite Welle der Digitalisierung, die als Treiber von einem großen Spektrum von KI-Technologien ausgelöst wurde, führt in die neue Ära der maschinenverstehbaren Daten. Dabei werden die digitalen Daten durch KI-Systeme inhaltlich interpretiert,

klassifiziert, angereichert mit Meta-Daten und veredelt, um dann neue Schlussfolgerungen daraus ziehen zu können, neuartige Entscheidungsvorschläge zu erarbeiten oder durch autonomes Verhalten ein vom Menschen vorgegebenes Ziel zu erreichen. Bei den notwendigen Standards und Normen für diese neue Ära der Digitalisierung stehen wir aber heute noch am Anfang, den unsere Roadmap erstmals dokumentiert und mit Handlungsempfehlungen für die nächsten Schritte verbindet.

Als Leitbild bei der Erstellung der Roadmap stand die menschzentrierte KI im Vordergrund, die für alle KI-Systeme neben deren Erklärungsfähigkeit, Robustheit und Resilienz, auch eine strikte Berücksichtigung europäischer Werte wie die Diskriminierungsfreiheit und den Schutz der Privatsphäre einfordert. Insgesamt leistet die Standardisierung einen entscheidenden Beitrag zur technischen Souveränität und Interoperabilität für KI-Anwendungen, die in Zukunft für alle Branchen große Relevanz haben.

Es bietet sich an, KI-Technologien künftig bei der Standardisierung selbst einzusetzen, u.a. also die Dokumentanalyse, die Wissensrepräsentation und das maschinelle Lernen auch auf das Erstellen, Verteilen und Nutzen von Standards anzuwenden, um von maschinenlesbaren Standards hin zu maschinell interpretierbaren und überprüfbaren Standards zu kommen.

Diese Vorgehensweise bietet das Potenzial, die heute schon 17 Milliarden EUR jährlichen Einsparungen durch Standards in Deutschland nochmals signifikant zu steigern. Normen beschleunigen den Ergebnistransfer der exzellenten KI-Forschung in die deutsche Wirtschaft und öffnen internationale Märkte besonders auch für den Mittelstand und Start-up Unternehmen.

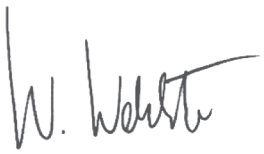
Ohne den unermüdlichen Einsatz unserer ehrenamtlich arbeitenden Expertinnen und Experten wäre die Erstellung dieser ersten KI-Normungsroadmap nicht möglich gewesen. Die mit zwanzig hochrangigen Persönlichkeiten besetzte Steuerungsgruppe hat seit ihrer Gründung sechs Sitzungen durchgeführt und zusätzlich in einer Klausursitzung zusammen mit der Leiterin und den Leitern der sieben Arbeitsgruppen die abschließenden Handlungsempfehlungen konsensual verabschiedet.

Auch im Namen der Steuerungsgruppe möchten wir uns an dieser Stelle bei allen 185 Autoren und 88 weiteren Beteiligten für das große Engagement bedanken, wobei wir Frau Filiz Elmas als exzellenter Koordinatorin des Gesamtprojektes unser besonderes Lob aussprechen möchten.

Nun gilt es, möglichst viele der Handlungsempfehlungen mit Unterstützung aller Bundesministerien, die für die aktuell mit fünf Milliarden EUR geförderte KI-Strategie zuständig sind, rasch umzusetzen und den Boden für die Fortschreibung unserer KI-Normungsroadmap bereits heute zu bereiten. Die Roadmap zeigt, dass noch ein erheblicher Forschungsbedarf u. a. zur Erstellung der erforderlichen Qualitätsmetriken und Prüfprofile für eine risiko-adaptierte Zertifizierung von KI-Komponenten besteht.

Wir wünschen allen Leserinnen und Lesern eine spannende Lektüre und bitten Sie um aktive Unterstützung bei der weiteren Ausgestaltung dieser Normungsroadmap.

Lassen Sie uns gemeinsam internationale Normen und Standards erarbeiten und einführen, die die sichere Anwendung von „KI made in Germany“ nach europäischen Wertmaßstäben unterstützen.



Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster
Leiter der Steuerungsgruppe, CEA DFKI

Christoph Winterhalter
Vorsitzender des Vorstandes, DIN

GRUSSWORT

Peter Altmaier
Bundesminister für Wirtschaft und Energie

Sehr geehrte Leserinnen und Leser,

bei Sprachassistenssystemen oder Bilderkennung ist der Einsatz von Künstlicher Intelligenz (KI) schon heute Realität. Und die Liste weiterer potentieller Anwendungsfelder ist lang: Sie reicht von autonomem Fahren und intelligenter Verkehrssteuerung über medizinische Diagnostik und Therapie bis hin zur industriellen Automation. Mit unserem Ansatz einer verantwortungsvollen, gemeinwohlorientierten und menschenzentrierten Entwicklung und Nutzung von KI-Technologien wollen wir Deutschland zu einem führenden KI-Standort machen und Akzente auf europäischer Ebene setzen. Damit wir uns künftig im internationalen Wettbewerb um die besten Ideen noch besser behaupten können, müssen wir heute die Weichen richtig stellen:

- → regulatorisch, in dem der Gesetzgeber einen Regelungsrahmen für die Entwicklung und den Einsatz von KI-Technologien schafft, der Innovationen fördert;
- → gesellschaftlich, in dem wir einen Dialog über die Chancen, Risiken und ethische Fragestellungen im Zusammenhang mit dem Einsatz von KI-Technologien führen und
- → technisch, in dem wir einheitliche Anforderungen mit Normen und Standards beschreiben, die die Umsetzung der rechtlichen Rahmenbedingungen und ethischen Werte unterstützen.

Der Standardisierung fällt dabei eine wesentliche Rolle zu: Normen und Standards sorgen für Interoperabilität, erhöhen die Nutzerfreundlichkeit und sind Grundlage für Vertrauen in technische Systeme und Prozesse. Gleichzeitig erleichtern sie unserer mittelständisch geprägten Wirtschaft den Zugang zu internationalen Märkten und steigern damit die Wettbewerbsfähigkeit. Vor zwei Jahren, im November 2018, hat die Bundesregierung ihre nationale Strategie Künstliche Intelligenz beschlossen und das Setzen von Standards als eines von zwölf zentralen Handlungsfelder identifiziert. Die nun vorliegende „Deutsche Normungsroadmap Künstliche Intelligenz“ ist ein erster Schritt zur Umsetzung der Maßnahmen dieses Handlungsfelds: Sie beschreibt das Umfeld, in dem KI-Standardisierung sich bewegt, identifiziert bereits bestehende Normen und Standards, die für das Feld der KI relevant sind und zeigt weitere Normungs- und Standardisierungsbedarfe auf. Darüber hinaus formuliert sie konkrete Handlungsempfehlungen, die sich überwiegend an Normungsakteure, aber auch an Stakeholder der Qualitätsinfrastruktur und die Politik richten.

Die Normungsroadmap zeigt uns, dass wir alle gefordert sind, um „KI made in Germany“ zum Erfolgsmodell zu machen. Sie ist nicht der Abschluss, sondern vielmehr der Auftakt zur Umsetzung der Maßnahmen aus dem Handlungsfeld „Standards setzen“. Wirtschaft, Zivilgesellschaft, Wissenschaft und die öffentliche Hand sind daher jetzt aufgefordert, gemeinsam daran zu arbeiten, die Handlungsempfehlungen der Normungsroadmap umzusetzen und die Spielregeln der zukünftigen digitalen Wirtschaft und Gesellschaft aktiv mitzugestalten. Ich bin überzeugt, dass Künstliche Intelligenz europäische Wertemaßstäbe braucht. Gehen wir es zusammen an!

Peter Altmaier
Bundesminister für Wirtschaft und Energie

Zusammenfassung

Rund ein Jahr haben DIN und DKE in einem gemeinsamen Projekt mit dem Bundesministerium für Wirtschaft und Energie und zusammen mit ca. 300 Fachleuten aus Wirtschaft, Wissenschaft, öffentlicher Hand und Zivilgesellschaft an der Normungsroadmap Künstliche Intelligenz gearbeitet. Eine hochrangige Steuerungsgruppe unter dem Vorsitz von Prof. Wolfgang Wahlster hat die Erarbeitung koordiniert und begleitet.

Ziel der Roadmap ist die frühzeitige Entwicklung eines Handlungsrahmens für die Normung und Standardisierung, der die internationale Wettbewerbsfähigkeit der deutschen Wirtschaft unterstützt und europäische Wertmaßstäbe auf die internationale Ebene hebt.

Mit der Normungsroadmap KI wird eine wesentliche Maßnahme der KI-Strategie der Bundesregierung umgesetzt, in der eines von zwölf Handlungsfeldern sich explizit dem Thema „Standards setzen“ widmet.

Gerade im Bereich der Künstlichen Intelligenz spielen Normen und Standards eine besondere Rolle: Sie fördern den schnellen Transfer von Technologien aus der Forschung in die Anwendung und öffnen internationale Märkte für Unternehmen und ihre Innovationen. Indem sie Anforderungen an Produkte, Dienstleistungen oder Prozesse festlegen, sorgen sie für Interoperabilität und Qualität. Normen und Standards tragen somit maßgeblich zur Erklärbarkeit und Sicherheit bei und unterstützen die Akzeptanz und das Vertrauen in KI-Anwendungen.

Die vorliegende Normungsroadmap KI wurde in einem breiten Beteiligungsprozess mit interdisziplinären Akteuren erarbeitet und skizziert die Arbeits- und Diskussionsergebnisse der Arbeitsgruppen. Sie liefert einen umfassenden Überblick über Status Quo, Anforderungen und Herausforderungen zu folgenden sieben Schwerpunktthemen:

→ Grundlagen
→ Ethik/Responsible AI
→ Qualität, Konformitätsbewertung und Zertifizierung
→ IT-Sicherheit bei KI-Systemen
→ Industrielle Automation
→ Mobilität und Logistik
→ KI in der Medizin

Für diese zentralen Themenbereiche wird das aktuelle Umfeld der KI-Standardisierung beschrieben und eine Übersicht über relevante Normen und Standards zu Aspekten der Künstlichen Intelligenz gegeben.

Mit über 70 identifizierten Normungs- und Standardisierungsbedarfen zeigt die Roadmap konkrete Potenziale auf und formuliert fünf zentrale und übergreifende Handlungsempfehlungen:

1. **Datenreferenzmodelle für die Interoperabilität von KI-Systemen umsetzen**
   In Wertschöpfungsketten kommen viele unterschiedliche Akteure zusammen. Damit auch die verschiedenen KI-Systeme dieser Akteure automatisiert zusammenarbeiten können, ist ein Datenreferenzmodell nötig, um Daten sicher, zuverlässig, flexibel und kompatibel auszutauschen. Standards für Datenreferenzmodelle aus unterschiedlichen Bereichen schaffen die Grundlage für einen übergreifenden Datenaustausch und stellen damit weltweit die Interoperabilität von KI-Systemen sicher.

2. **Horizontale KI-Basis-Sicherheitsnorm erstellen**
   KI-Systeme sind im Kern IT-Systeme – für letztere gibt es bereits viele Normen und Standards aus verschiedensten Anwendungsbereichen. Um ein einheitliches Vorgehen beim Thema IT-Sicherheit von KI-Anwendungen zu ermöglichen, ist eine übergreifende „Umbrella-Norm“ sinnvoll, die vorhandene Normen und Prüfverfahren für IT-Systeme bündelt und um KI-Aspekte ergänzt. Diese Basis-Sicherheitsnorm kann dann durch Sub-Normen zu weiteren Themen ergänzt werden.

3. **Praxisgerechte initiale Kritikalitätsprüfung von KI-Systemen ausgestalten**
   Wenn selbstlernende KI-Systeme über Menschen, deren Besitz oder Zugang zu knappen Ressourcen entscheiden, können ungeplante Probleme in der KI individuelle Grundrechte oder demokratische Werte gefährden. Damit sich KI-Systeme in ethisch unkritischen Anwendungsfeldern dennoch frei entwickeln lassen, sollte durch Normen und Standards eine initiale Kritikalitätsprüfung gestaltet werden – diese kann schnell und rechtssicher klären, ob ein KI-System solche Konflikte überhaupt auslösen kann.

4. **Nationales Umsetzungsprogramm „Trusted AI“ zur Ertüchtigung der europäischen Qualitätsinfrastruktur initiieren und durchführen**
   Bisher fehlen verlässliche Qualitätskriterien und Prüfverfahren für KI-Systeme – das gefährdet das wirtschaftliche Wachstum und die Wettbewerbsfähigkeit dieser Zukunftstechnologie. Es braucht ein nationales Umsetzungsprogramm „Trusted AI“, das die Basis für reproduzierbare und standardisierte Prüfverfahren legt, mit denen Eigenschaften von KI-Systemen wie Verlässlichkeit, Robustheit, Leistungsfähigkeit und funktionale Sicherheit geprüft und Aussagen über die Vertrauenswürdigkeit getroffen werden können. Normen und Standards beschreiben Anforderungen an diese und bilden so die Grundlage für die Zertifizierung und Konformitätsbewertung von KI-Systemen. Mit einer solchen Initiative hat Deutschland die Chance, ein weltweit erstes und international anerkanntes Zertifizierungsprogramm zu entwickeln.

5. **Use Cases auf Normungsbedarf analysieren und bewerten**
   Die KI-Forschung sowie die industrielle Entwicklung und Anwendung von KI-Systemen sind hoch dynamisch. Bereits heute gibt es viele Anwendungsfälle in den verschiedenen Einsatzfeldern von KI. Über anwendungstypische und branchenrelevante Use Cases lassen sich Standardisierungsbedarfe für industriereife KI-Anwendungen ableiten. Um Normen und Standards zu gestalten, ist es wichtig, wechselseitige Impulse aus Forschung, Industrie, Gesellschaft und Regulierung einzubinden. Im Zentrum dieses Ansatzes sollten die entwickelten Standards entlang von Use Cases erprobt und weiterentwickelt werden. So lassen sich anwendungsspezifische Bedarfe frühzeitig erkennen und marktfähige KI-Standards realisieren.

Die Ergebnisse der Normungsroadmap KI stellen den Auftakt der anstehenden Arbeiten dar und geben damit den Weg für die zukünftige Normung und Standardisierung im Bereich der Künstlichen Intelligenz vor. Ihre Umsetzung wird dazu beitragen, die deutsche Wirtschaft und Wissenschaft zu unterstützen und innovationsfreundliche Bedingungen für die Technologie der Zukunft schaffen. Insbesondere in der gesellschaftspolitischen Debatte auf europäischer Ebene über die künftige Rolle und den Einsatz von KI werden die Ergebnisse einen wichtigen Beitrag leisten.

Nur ein frühzeitiges und umfassendes Engagement der deutschen Stakeholder in der nationalen, aber vor allem auch der europäischen und internationalen Normung und Standardisierung wird die Position Deutschlands als Wirtschaftsnation und Exportland stärken und den Weg für „KI – Made in Germany“ ebnen.

Die Normungsroadmap KI wird dabei stetig aktualisiert und weiterentwickelt werden, um sich ändernde Anforderungen zu berücksichtigen.

Nun gilt es, konkrete Normungs- und Standardisierungsaktivitäten entlang der Handlungsempfehlungen auf den Weg zu bringen. Interessierte Fachleute sind ausdrücklich eingeladen, mitzuwirken und ihr Wissen in der Normung und Standardisierung einzubringen.

Vorwort ... **1**
Grußwort ... **3**
Zusammenfassung ... **4**

1 Einleitung ... **9**
1.1 Trends in der Künstlichen Intelligenz ... **11**
1.2 Standards für KI: vier praktische Beispiele ... **13**
1.3 Rolle der Normung und Standardisierung bei KI ... **15**
1.4 KI-Strategie der Bundesregierung ... **16**
1.5 Ziele und Inhalte der Normungsroadmap KI ... **18**
1.6 Hochrangige Steuerungsgruppe ... **20**
1.7 Methodisches Vorgehen ... **22**

2 Handlungsempfehlungen der Normungsroadmap KI ... **23**

3 Akteurs- und Normungsumfeld ... **27**
3.1 Gesellschaftspolitisches Umfeld ... **28**
3.2 Innovationspolitische Initiativen ... **29**
3.3 Normungs- und Standardisierungsumfeld ... **30**

4 Schwerpunktthemen ... **33**
4.1 Grundlagen ... **35**
4.1.1 Status quo ... 38
4.1.2 Anforderungen, Herausforderungen ... 39
4.1.3 Normungs- und Standardisierungsbedarfe ... 60
4.2 Ethik/Responsible AI ... **63**
4.2.1 Status quo ... 64
4.2.2 Anforderungen, Herausforderungen ... 65
4.2.3 Normungs- und Standardisierungsbedarfe ... 77
4.3 Qualität, Konformitätsbewertung und Zertifizierung ... **79**
4.3.1 Status quo ... 81
4.3.2 Anforderungen, Herausforderungen ... 85
4.3.3 Normungs- und Standardisierungsbedarfe ... 95
4.4 IT-Sicherheit bei KI-Systemen ... **97**
4.4.1 Status quo ... 99
4.4.2 Anforderungen, Herausforderungen ... 103
4.4.3 Normungs- und Standardisierungsbedarfe ... 114
4.5 Industrielle Automation ... **117**
4.5.1 Status quo ... 119
4.5.2 Anforderungen, Herausforderungen ... 120
4.5.3 Normungs- und Standardisierungsbedarfe ... 124
4.6 Mobilität und Logistik ... **127**
4.6.1 Status quo ... 128
4.6.2 Anforderungen, Herausforderungen ... 129
4.6.3 Normungs- und Standardisierungsbedarfe ... 132
4.7 KI in der Medizin ... **135**
4.7.1 Status quo ... 136
4.7.2 Anforderungen, Herausforderungen ... 136
4.7.3 Normungs- und Standardisierungsbedarfe ... 138

5 Anforderungen an die Erarbeitung und Nutzung von Normen und Standards ... 143
5.1 Überprüfung und Entwicklung von Normen und Standards im Bereich KI ... 144
5.1.1 Überprüfung bestehender Normen und Standards ... 144
5.1.2 Agile Entwicklung von Normen und Standards für KI ... 144
5.2 SMART Standards – Neugestaltung von Normen für KI-Anwendungsprozesse ... 144
5.2.1 Motivation ... 145
5.2.2 Status quo ... 145
5.2.3 SMART Standards – Stufenmodell ... 147
5.2.4 Normen und KI ... 148
5.2.5 Neugestaltung von Normen für KI-Anwendungsprozesse ... 149
5.2.6 Zusammenfassung und Ausblick ... 150

6 Übersicht über relevante Dokumente, Aktivitäten und Gremien zu KI ... 151
6.1 Veröffentlichte Normen und Standards zu KI ... 152
6.2 Veröffentlichte Normen und Standards mit Relevanz für KI ... 155
6.3 Laufende Normungs- und Standardisierungsaktivitäten zu KI ... 164
6.4 Gremien zu KI ... 171

7 Abkürzungsverzeichnis ... 175

8 Quellen- und Literaturverzeichnis ... 179

9 Autorenverzeichnis ... 199

10 Weitere Mitglieder der Arbeitsgruppen ... 205

11 Anhang ... 209
11.1 Glossar ... 210
11.2 Philosopische Grundlagen zur Ethik ... 214
11.3 SafeTRANS Roadmap ... 216
11.4 SMART Standards – Neugestaltung von Normen für KI-Anwendungsprozesse ... 217
11.4.1 Nutzung von granularen Inhalten mittels Technologieansatz ... 217
11.4.2 Bottom-up-Methode – Nachstrukturierung von Normen ... 223
11.4.3 Top-down-Methode – Entwicklung von SMART Standards ... 224

1 Einleitung

Künstliche Intelligenz (KI) ist bereits seit einigen Jahren allgegenwärtig und aus der heutigen digitalen Welt nicht mehr wegzudenken. Sie durchdringt immer mehr Bereiche des gesellschaftlichen und wirtschaftlichen Lebens und wird die Art und Weise, wie wir arbeiten, lernen, kommunizieren und konsumieren, verändern.

Bereits heute sind die Anwendungsfälle und existierenden Praxisbeispiele für KI zahlreich.

So spielen KI-basierte Systeme im Alltag eine sehr präsente Rolle, beispielsweise wenn Onlinehändler beim Einkauf im Internet weitere Produkte anpreisen, Streamingdienste neue Musik-Playlists oder Filme empfehlen, Social-Media-Plattformen auf Nachrichten hinweisen, Smartwatches Herzrhythmusstörungen erkennen oder Autofahrer in Echtzeit zu freien Parkplätzen geführt werden.

Auch in der industriellen Anwendung steigt die Bedeutung von KI rasant. Fachleute gehen davon aus, dass KI in Zukunft auf die industrielle Wertschöpfung einen so großen Einfluss haben wird, dass Unternehmen sich gegen den Einsatz von KI kaum werden verwehren können. Die Möglichkeiten sind fast grenzenlos: Ob Sprachassistenten und Chat-Bots, Programme zur Dokumentenrecherche, Systeme zur diagnostischen Bilderkennung bei Tumoren, mit Menschen interagierende Industrieroboter in der Fabrik oder autonom fahrende Autos.

Schon heute wird KI in Unternehmen vielfach zur Prozessoptimierung und Produktivitätssteigerung eingesetzt. Dabei handelt es sich vor allem um analytische Tätigkeiten, die bei Entscheidungsprozessen unterstützen. Der große Vorteil der KI: Sie lernt, bessere Ergebnisse zu produzieren als Verfahren, die nach starren Mustern vorgehen, und ermöglicht Produktivitäts- und Umsatzgewinne durch zunehmend personalisierte Angebote. Damit stellt sie eine Technologie dar, mit der der Fortschritt vorangetrieben und der Wirtschaftsstandort Deutschland und somit der Wohlstand einer ganzen Gesellschaft gesichert werden kann.

Die EU geht davon aus, dass die Wirtschaft innerhalb der nächsten zehn Jahre mithilfe von KI um 14 Prozent wachsen wird [1]. In Deutschland könnte KI Schätzungen zufolge bis zum Jahr 2030 das Bruttoinlandsprodukt um 11,3 Prozent steigern, was einer Wertschöpfung um 430 Milliarden Euro entspricht [2]. Nicht zuletzt deshalb haben die Europäische Kommission und die Bundesregierung diese Technologie zu einer Top-Priorität erklärt (siehe Kapitel 1.4).

Als wesentlicher Grund für den Aufstieg dieser Technologie wird der rasante Anstieg der verfügbaren Daten gesehen. Sowohl die produzierte Datenmenge als auch die verfügbare Rechenleistung steigen exponentiell. KI lebt von Daten, je mehr Daten verarbeitet werden, desto größer ist der potenzielle Lerneffekt und desto vielfältiger der gesellschaftliche Nutzen [1]. Unter den Top 10 der wertvollsten Unternehmen der Welt stehen sieben Firmen, die ihr Geld hauptsächlich mit Daten verdienen. Unter den Top 100 taucht lediglich ein deutsches Unternehmen auf, dessen Geschäftsmodell auf Daten basiert [3].

In der KI-Forschung genießt Deutschland einen ausgezeichneten Ruf. Viele Forschungseinrichtungen und -netzwerke[1] zählen zur globalen KI-Spitzenforschung und haben einen deutlichen Wissensvorsprung beispielsweise bei industriellen KI-Anwendungen in Produktionsbereichen. Doch wenn es darum geht, aus den Forschungsergebnissen innovative Produkte und Dienstleistungen zu entwickeln und diese schließlich zum kommerziellen Erfolg zu führen, sind andere Länder wie beispielsweise China oder die USA deutlich erfolgreicher.

Fest steht: Wenn es gelingt, den großen industriellen Erfahrungsschatz der deutschen Wirtschaft mit den Möglichkeiten der datengetriebenen KI-Methoden zu einer industriellen KI zu verbinden, könnte Deutschland ein Gewinner der neuen KI-Technologie werden und seine Wettbewerbsfähigkeit in den ohnehin dominierten industriellen Branchen sichern und sogar ausbauen.

Allerdings haben deutsche Unternehmen zum Teil sehr unterschiedliche Ausgangssituationen, gerade wenn es um den Einsatz von KI-Lösungen geht: Manche kennen KI bisher nur als Schlagwort, andere haben die Potenziale von KI-Technologien zwar erkannt, wissen aber nicht, wo sie ansetzen sollen. Wieder andere planen die Einführung von KI-Lösungen, tun sich aber noch schwer mit der Umsetzung. Über 99 Prozent aller Unternehmen in Deutschland sind kleine und mittlere Unternehmen und erwirtschaften über die Hälfte der gesamten Wertschöpfung. Somit sollte gerade der deutsche Mittelstand KI als Schlüsseltechnologie verstehen und die Potenziale für sich nutzen [4].

1 Zu nennen sind hier beispielsweise die europäischen Forschungsverbünde ELLIS und CLAIRE, mit denen Deutschland seine international starke Position in der Forschung und Entwicklung weiter gefestigt hat.

Wenn „KI – Made in Germany“ in Zukunft als Marke und Exportschlager etabliert werden soll, muss die KI-Technologie schon heute als integraler Bestandteil unserer Wirtschaft betrachtet werden.

Allerdings wird eine Technologie nur dann flächendeckend erfolgreich zum Einsatz kommen, wenn sie Akzeptanz in der Gesellschaft findet. Während in den Unternehmen die Chancen durch KI zunehmend erkannt werden, ist die öffentliche Diskussion in Deutschland sehr kontrovers. Aufgrund ethischer Bedenken stößt KI in der Bevölkerung teilweise auf Ablehnung.

Zwar sind KI-Methoden per se weder neutraler noch diskriminierender als Menschen, dennoch können sie problematische oder diskriminierende Entscheidungen hervorbringen. KI-Systeme werden mit Daten und Informationen trainiert, die in der Regel von Menschen erhoben und aufbereitet werden. Sind in diesen Daten gesellschaftliche Vorurteile oder Verzerrungen enthalten, übernimmt das KI-System diese Vorurteile oder verstärkt sie unter Umständen sogar, da ein KI-System kein moralisches Urteilsvermögen besitzt [4].

Es braucht daher einen klaren Handlungsrahmen, der sicherstellt, dass ethische Werte eingehalten werden. Genau hier können Normen und Standards ansetzen und helfen, die breite Akzeptanz von KI-Systemen zu erhöhen, indem sie beispielsweise Qualitätsmetriken definieren und dadurch die Zuverlässigkeit der Resultate der KI-Systeme besser beurteilbar machen.

Auch darüber hinaus gibt es noch viel Handlungsbedarf vor allem im Hinblick auf Sicherheit, Fairness, Robustheit, Transparenz und Angemessenheit der KI-Systeme und ihrer Entscheidungen. Was fehlt, ist ein definierter Handlungsspielraum, in dem KI-Systeme für den Menschen agieren und nachvollziehbare Entscheidungswege zugrunde liegen. Die Europäische Kommission hat hierfür eine High-Level Expert Group on Artificial Intelligence (HLEG-KI) eingesetzt (siehe Kapitel 3.1), die u. a. „Ethikleitlinien für eine vertrauenswürdige KI“ [5] als Orientierung für Unternehmen erarbeitet hat. Wenn Deutschland es schafft, die europäischen Wertmaßstäbe in KI-Anwendungen zu integrieren, können deutsche KI-Produkte weltweit eine höhere Akzeptanz finden als vergleichbare Produkte etwa aus den USA oder China. So kann die Vorreiterrolle der deutschen Wirtschaft gelingen.

Eine Beteiligung aller interessierten Kreise unter Einbindung interdisziplinärer Akteure – beispielsweise aus Informatik, Ingenieurwissenschaften, Philosophie, Psychologie, Soziologie, Rechtswissenschaften, Politik, Zivilgesellschaft und Endverbraucher – stellt eine solide Grundlage für eine menschenzentrierte Ausrichtung und Entwicklung von KI-Systemen dar.

Die Normungsroadmap KI legt für solch eine interdisziplinäre Zusammenarbeit einen signifikanten Meilenstein durch das Etablieren eines offenen, transparenten und nachhaltigen Austauschs. Daraus erwachsende Gremien, Plattformen, Foren oder Aktivitäten können als Katalysator in der Technologieentwicklung fungieren.

1.1 Trends in der Künstlichen Intelligenz

In den vergangenen Jahren haben verschiedene technologische Entwicklungen der Künstlichen Intelligenz einen enormen Schub gegeben und das Rennen um die weltweite Technologieführerschaft eröffnet. Mittlerweile hat sich der Einsatz von KI als globaler Trend etabliert, dem sich keine Volkswirtschaft und auch kaum noch ein Unternehmen entziehen kann. Die Facetten der aufkommenden Trends sind vielfältig. Gerade mit dem Fortschritt der KI-Technologien und zunehmenden Erfolgen in der Technologieentwicklung kommen fast täglich neue Einsatzfelder und Möglichkeiten hinzu. Einige dieser Trends werden im Folgenden aufgegriffen und exemplarisch dargestellt.

Seit jeher gibt es Bestrebungen nach menschlichen Interpretationen, Reaktionen und Verhaltensweisen durch KI-Systeme, welche in den vergangenen Jahren zunehmend an Bedeutung erlangt haben. Mit dem **Neuromorphic Computing** ist ein Meilenstein dahin gelegt. Bereits heute können mit Menschen interagierende und humanoide Roboter zunehmend auf traditionell weiche Faktoren, wie die Gefühlswelt des Menschen, eingehen. Auch können KI-Systeme kognitive Prozesse von Menschen nachahmen und in Arbeitsprozesse übernehmen, wodurch KI-Systeme in diversen Anwendungen nicht nur menschliche Entscheidungen vorbereiten, sondern sie mitunter schon selbst treffen können. Damit werden nicht nur einfache, kraftraubende oder routinemäßige Tätigkeiten durch Maschinen möglich, sondern auch kognitiv anspruchsvolle und kreative Tätigkeiten, die bisher Menschen vorbehalten waren.

Mit den Möglichkeiten steigt die Komplexität an Berechnungen, Entscheidungen, Interpretationen etc. durch ein KI-System, was immer größerer Datenmengen und höherer Verarbeitungsgeschwindigkeiten bedarf. Eine Antwort darauf

liefern Technologien wie **Quantencomputer**, durch die der Fortschritt in der Entwicklung der Künstlichen Intelligenz beschleunigt werden dürfte. Quantencomputer können nicht nur die Leistung der Informationsverarbeitung verbessern, sondern unter Umständen auch den Einsatz von KI-Methoden erst möglich machen. Ein Beispiel stellt das **Quantum Machine Learning** (QML) dar.

Parallel zur IT-Hardware entwickelt sich die Leistung von KI-Systemen auch aufseiten der IT-Software stetig weiter. Ein Beispiel hierfür stellen KI-Systeme dar, die mit nicht-hierarchischen Daten und Wissensstrukturen sowie Unsicherheit umgehen und auch unstrukturierte Daten in eine Antwortstruktur bringen können. Dabei sind neben statistischen Verfahren **Ontologien** ein zentrales Element, um aus Daten Bedeutungszusammenhänge zu erschließen, Umweltzustände zu erkennen und daraus automatisiert Handlungsempfehlungen oder Handlungen des KI-Systems abzuleiten.

Durch neue Technologien sowie neuartige Verarbeitungs-, Entscheidungs- und Handlungsprozesse (wie z. B. beim Neuromorphic Computing) bieten sich zunehmend weitere Möglichkeiten. So lässt sich auch bereits ohne die flächendeckende Anwendung der Technologie erkennen, dass sich unter Einhaltung ethischer Werte die Grenzen von automatisierten Systemen der Gegenwart zu hochgradig autonomen Systemen öffnen werden. Im Unterscheid zu automatisierten Systemen wählen autonome Systeme selbstständig ihre Mittel aus und entscheiden bis zu einem bestimmten Grad selbstständig, um ein vorgegebenes Ziel zu erreichen, und zwar ausgehend vom Erkennen der Situation, in der sie sich gerade befinden. In der Industrie finden niedrigstufige autonome Systeme bereits Anwendung, um Nützlichkeitsaspekte (Flexibilität, Ressourcen, Zeit, Qualität, Nachhaltigkeit) gegenüber Kosten, Sicherheitsaspekten und industriespezifischen Aspekten der Veränderung der Arbeitswelt (Personalbeschaffung, Qualifikation, Freisetzung etc.) abzuwägen. Im Bereich des Endkunden ist die Entwicklung bereits weiter fortgeschritten, wie z. B. die Technologien Smart Home und Service-Robotik zeigen.

Ein weiterer Trend, der sich abzeichnet und die wesentlichen Charakteristika erfolgreicher KI-Systeme ausmachen wird, ist die **Selbsterklärungsfähigkeit**. Sie wird durch Erklärungskomponenten realisiert, welche die Ergebnisse des KI-Systems und die ihnen zugrunde liegenden Verarbeitungsschritte für den jeweiligen Benutzer verständlich, kontextabhängig und auf unterschiedlichen Detaillierungsebenen in einem argumentativen Dialog erklären können. Die dynamisch erzeugten Erklärungen erfolgen meist sprachlich, vereinzelt aber auch grafisch oder multimodal. Schon Mitte der 1970er-Jahre wurden die ersten KI-Systeme (u. a. das System Mycin) implementiert, die ihre eigenen Inferenzprozesse einem Benutzer, der eine Warum-Frage stellt, erklären konnten. Dies ist besonders für medizinische Diagnosesysteme eine Voraussetzung für die Akzeptanz von Ärzten, die ja für die Verwendung eines Diagnose- oder Therapievorschlags gegenüber dem Patienten die Verantwortung tragen. Bei wissensbasierten Systemen ist eine solche Introspektion auf der symbolischen Ebene aufgrund der expliziten Modelle der Anwendungsdomäne erheblich einfacher zu realisieren als bei modellfreien Systemen, die auf statistischen oder neuronalen maschinellen Lernverfahren beruhen. Doch selbst für KI-Systeme, die auf neuronalem Deep Learning basieren, gibt es inzwischen erste Ansätze für elementare Erklärungskomponenten (vgl. [6]).

Ein weltweiter Trend in der KI ist die Entwicklung **hybrider kognitiver Systeme**, die wissensbasierte Methoden mit maschinellem Lernen kombinieren, sodass sich symbolische und subsymbolische Verfahren wechselseitig ergänzen und verstärken. Die Weltorganisation für KI (AAAI) hat in ihrer Roadmap für die nächsten 20 Jahre diesen Trend für die USA klar herausgearbeitet (vgl. [7]). Aus heutiger Sicht kann man vier Phasen der KI-Forschung unterscheiden (siehe Abbildung 1), wobei die hybriden kognitiven Systeme derzeit den höchsten Intelligenzgrad, die beste Robustheit, Transparenz sowie Anpassungsfähigkeit aufweisen.

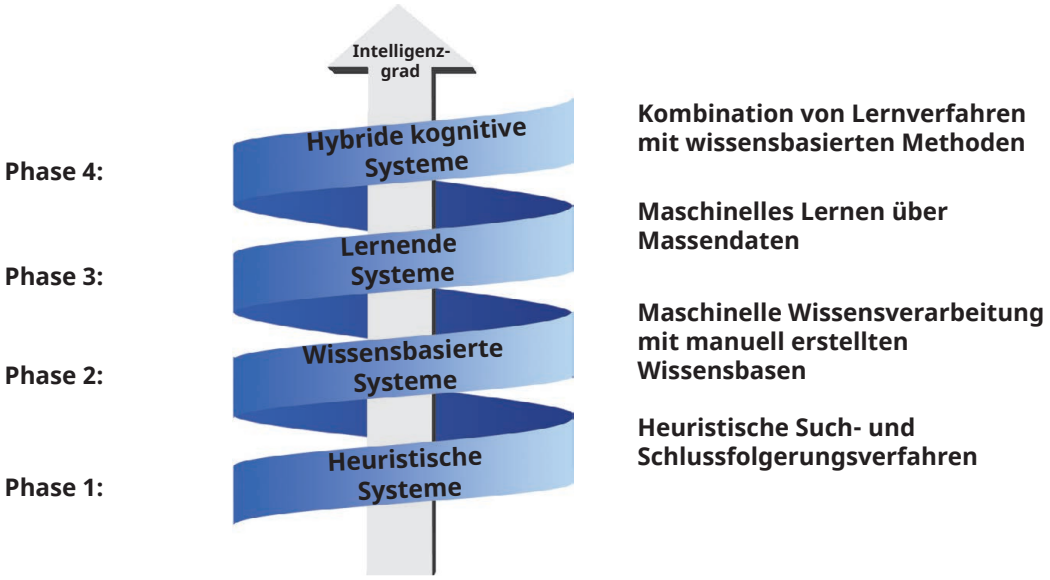
Bei der Entwicklung von Technologietrends ist zunehmend eine breite Beteiligung interessierter Kreise und interdisziplinärer Bereiche erkennbar. Neben der Einbindung neuer Akteure zeichnet sich mitunter auch ein direkter Austausch mit dem Endverbraucher ab, der dadurch eine stärkere Stimme bekommt und womit die Entwicklung von KI-Systemen eine menschenzentrierte Ausrichtung erhält.

Nachhaltigkeitspotenziale durch Künstliche Intelligenz

Neben den bereits genannten technologischen Trends eröffnen sich mittels KI-Anwendungen gerade auch bezüglich diverser Nachhaltigkeitsaspekte enorme Möglichkeiten. Nachfolgend werden exemplarische Anwendungsbereiche aufgezeigt.

→ In der Agrarwirtschaft kann KI in Kombination mit z. B. drohnen- oder sensorbasiertem Monitoring helfen, den Zustand von Pflanzen zu bewerten und infolgedessen Düngemittel und Pestizide gezielter und sparsamer einzusetzen („precision farming“).

Abbildung 1: Die vier Phasen der KI [8]



- → In der Produktion kann durch die Vernetzung und Robotik der Energieverbrauch reduziert werden.
- → In der Nutzungsphase kann durch vorausschauende Wartung die Produktlebensdauer verlängert werden.
- → In der Recycling- und Abfallwirtschaft kann KI die Erkennung und Sortierung von Abfällen verbessern und damit die Prozesseffizienz steigern und Kreislaufwirtschaft fördern.
- → Für Gebäudeeffizienz und Energiemanagement bietet KI die Möglichkeit einer verbesserten Systemsteuerung; sowohl was die Regelung von Heiz-, Kühl- und Lüftungssystemen anbelangt als auch die Handhabung miteinander vernetzter Produktionsmaschinen, insbesondere bei Einbeziehung von IoT-Aktivitäten [9].

Bei der Frage, wie nachhaltig KI und deren Anwendungen wirklich sind, muss jedoch nicht nur das Anwendungsfeld betrachtet werden, sondern auch die für Berechnungen benötigte Energie. Da manche Rechenleistungen sehr energieintensiv sind, muss sichergestellt werden, dass eine möglichst energieeffiziente Variante der Analyse gewählt wird.

Alles in allem kann somit KI stark zu mehr Nachhaltigkeit beitragen, wenn sie richtig eingesetzt wird und sowohl ökologische, ökonomische und soziale Aspekte berücksichtigt werden.

1.2 Standards für KI: vier praktische Beispiele

Beispiel 1: Standardisierter Qualitätsvergleich von KI-Systemen zur automatischen Übersetzung

Eines der ältesten in der KI bearbeiteten Anwendungsgebiete ist die maschinelle Übersetzung (MÜ) von Texten. Schon im Jahr 1954 hatte IBM zusammen mit der Georgetown University ein minimales Experimentalsystem für Russisch-Englisch mit nur 250 Wörtern und sechs Syntaxregeln programmiert. In den 1970er-Jahren wurden in Deutschland im Sonderforschungsbereich 100 in Saarbrücken mit dem MÜ-System SUSY wichtige Grundmodule für sprachwissenschaftlich fundierte automatische Übersetzer für das Deutsche entwickelt. Heute sind MÜ-Systeme weitverbreitet. Sie werden von Internet-Unternehmen wie Google, Microsoft, Facebook, Amazon, und Baidu, aber auch von einem sehr erfolgreichen deutschen KI-Start-up namens Deep angeboten und haben täglich Milliarden von Benutzern in Alltag und Beruf. Derzeit sind MÜ-Systeme auf der Basis neuronaler Übersetzungsalgorithmen am weitesten verbreitet. Diese werden über Millionen von Satzpaaren in Quell- und Zielsprache oft einfach über die Zeichenketten Ende-zu-Ende in sogenannten Transformer-Architekturen mit einer Sequenz von Encodierern für die Eingabe und einer weiteren Sequenz von Decodierern für die Ausgabe trainiert. Zuvor waren statistische Verfahren am erfolgreichsten, während rein symbolische Verfahren trotz ihrer höheren Präzision in Spezialdomänen (z. B. Wetterberichte) wegen ihrer zu geringen Abdeckung für beliebige Alltagssprache in den Hintergrund rückten.

Im Bereich der maschinellen Übersetzung ist der Goldstandard für die Messung der Richtigkeit einer Übersetzung der BLEU-Wert (**bil**ingual **e**valuation **u**nderstudy), der maschinell übersetzte Texte mit Varianten von von menschlichen Experten übersetzten Texten automatisch vergleicht. Obwohl der BLEU-Wert (auf einer Skala bis 100 Prozent) durch seine auf den Vergleich auf Wortsequenzen beschränkte, sehr vereinfachte Metrik die syntaktische und semantische Korrektheit nicht erfasst, hat er sich für eine grobe Qualitätseinschätzung und einen Leistungsvergleich zwischen MÜ-Systemen bewährt. Die Standardisierung von Qualitätsmetriken für KI-Systeme ist ein wichtiges Thema für die breite Akzeptanz solcher Systeme im praktischen Einsatz, da die Zuverlässigkeit der Resultate dadurch besser beurteilbar wird. NIST hat einen leicht modifizierten BLEU-Faktor entwickelt, der seit 2006 in den jährlichen Tests von MÜ-Systemen für unterschiedliche Übersetzungsdomänen in den jährlichen Workshops und Konferenzen (WMT) zur maschinellen Übersetzung zur Bewertung von Systemleistungen verwendet wird. Beispielsweise ist ein BLEU-Wert von 15 ein sehr schlechter Wert, der einen großen Aufwand in der Postedition der automatischen Übersetzung bedeutet, um überhaupt mit dem übersetzten Text weiterarbeiten zu können.

Spitzenwerte, die heute von den besten MÜ-Systemen erreicht werden, liegen derzeit bei etwas über 45. Allerdings kann mit der primitiven BLEU-Metrik nicht die Schwere von Übersetzungsfehlern bewertet werden, wenn z. B. durch eine Negation an der falschen Stelle die gesamte Aussage im Quelltext verfälscht wird.

Bei einer risikoadaptiven Zertifizierung von KI-Systemen müssen für bestimmte Anwendungsklassen Qualitätsschranken definiert werden, deren Unterschreitung eine Weiterverwendung der Ergebnisse des KI-Systems in einem kritischen Anwendungskontext ausschließt. Wenn z. B. eine Zeugenaussage in einer Fremdsprache vorliegt, kann deren maschinelle Übersetzung durch ein KI-System mit einer zu hohen Fehlerrate natürlich nicht vor Gericht verwendet werden, sondern muss von einem vereidigten menschlichen Übersetzer erstellt werden (roter Bereich in der Kritikalitätspyramide in Abbildung 2). Beim MÜ-Einsatz für Bedienungsanleitungen technischer Geräte sollten nur zertifizierte MÜ-Systeme zum Einsatz kommen und bei der MÜ für Vertragstexte und Arztbriefe muss die MÜ-Qualität kontinuierlich geprüft werden. Wenn ein systematischer Manipulationsverdacht durch inhaltsverfälschende Übersetzungen bei öffentlichen Tweets, Blogs oder Nachrichtenportalen vorliegt, sollte eine Ex-post-Kontrolle des benutzten MÜ-Systems möglich sein. Auf der anderen Seite kann auch ein KI-Übersetzungsdienst mit geringerem BLEU-Wert im Kontext einer Übersetzung privater Chats und Empfehlungen durchaus sinnvoll sein, weil hier ein sehr geringes Risiko für den Benutzer bei fehlerhafter Übersetzung besteht.

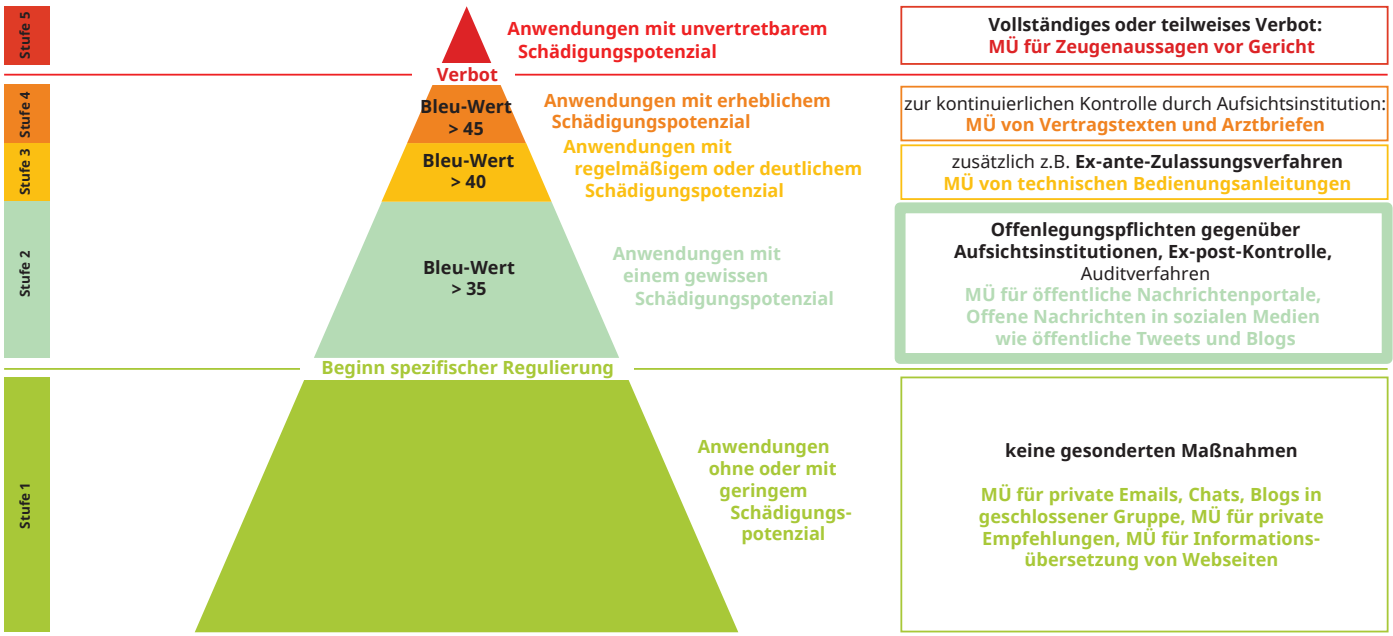


Abbildung 2: Die Kritikalitätspyramide und ein risikoadaptiertes Regulierungssystem für den Einsatz von KI-basierten maschinellen Übersetzungssystemen (nach [10])

Beispiel 2: Semantische KI-Technologien zur Gewährleistung der Interoperabilität von Systemen
In Industrie 4.0 kommunizieren Produktionsmaschinen verschiedener Hersteller untereinander. Dabei werden von den Werkzeugmaschinenbauern oftmals unterschiedliche Terminologien benutzt. Durch maschinenverstehbare Ontologien aus der KI wird es möglich, dass unterschiedliche Begriffssysteme automatisch ineinander überführt werden, sodass eine Verständigung von verschiedenartigen Maschinen untereinander im Internet der Dinge (IoT, Internet of Things) möglich wird. Dazu braucht man standardisierte Ontologiebeschreibungssprachen. Mit einer Konsortialnorm des W3C namens **OWL** (Ontology Web Language) wurde unter maßgeblicher Mitwirkung deutscher KI-Experten ein Standard geschaffen, der heute auch in der deutschen Industrie seinen Einsatz findet. Viele Firmen haben bereits eigene Begriffssysteme für ihre Produktionsmaschinen mithilfe von OWL maschinenverständlich spezifiziert.

Beispiel 3: KI für die Handlungsplanung
Die KI-basierte Handlungsplanung ist ein Gebiet, auf dem Deutschland in der internationalen Forschung als sehr erfolgreich gilt. Dabei geht es vereinfacht gesagt darum, eine Folge von einzelnen Handlungsschritten ausgehend von einem Anfangszustand zu finden, die zu einem gewünschten Zielzustand führt. Solche KI-Planungssysteme sind z. B. für die Planung der Aktionen autonomer Roboter, bei der Transportplanung in der Logistik oder der Produktionsplanung in der Smart Factory eine wichtige KI-Technologie. Hier hat sich ein De-facto-Standard namens **PDDL** (Planning Domain Definition Language) und für das hierarchische Planen neuerdings die Variante HDDL (Hierarchical Domain Definition Language) als Spezifikationssprache u. a. für die Vorbedingungen und Nachbedingungen eines Handlungsschrittes durchgesetzt, die z. B. für den weltweiten Wettbewerb und den Vergleich der besten Planungssysteme IPC 2020 verwendet wird. Deutsche KI-Planungssysteme haben in diesen weltweiten Wettbewerben, die seit 1998 ausgetragen werden, bereits mehrfach gewonnen. In Industrie 4.0 und beim autonomen Fahren gehören KI-Planungssysteme zu den erfolgskritischen Komponenten.

Beispiel 4: Standardisierungen im Bereich des datengetriebenen maschinellen Lernens (ML)
Mit KI-Verfahren des Deep Learning konnten im Bereich der automatischen Analyse von Bildern und Bildfolgen sehr gute Fortschritte auf der Basis mehrschichtiger neuronaler Netze gemacht werden. Allerdings hängt der Lernerfolg nicht nur von der Quantität der Trainingsdaten, sondern auch von deren Qualität ab. Besonders beim überwachten Lernen müssen umfangreiche Trainingsdaten oftmals zunächst durch menschliche Experten annotiert werden, z. B. welche Hunderasse auf einem von 100.000 Bildern im Trainingsdatensatz zu sehen ist. Um möglichst valide Datensätzen zu erhalten, werden dabei mehrere Annotatoren zur Bearbeitung überlappender Datensätze beauftragt. Hierbei muss deren Konsistenz in der Beurteilung geprüft werden. Dazu wurden als ein Goldstandard u. a. die sogenannten **Kappa-Statistiken** eingeführt. Damit wird als eines der Kriterien für die Qualität der annotierten Trainingsdaten die Reliabilität der Annotationen für eine Zertifizierung eines KI-Systems, das auf ML beruht, durch eine Standardmetrik operationalisierbar und vergleichbar.

1.3 Rolle der Normung und Standardisierung bei KI

Die Fähigkeit, neue Erkenntnisse und Ideen in Produkte und Dienstleistungen umzusetzen, ist entscheidend für die Wettbewerbsfähigkeit der deutschen Wirtschaft. Normung kann dabei als ein Katalysator für Innovationen dienen und helfen, Lösungen nachhaltig am Markt zu verankern.

In Normen und Standards werden Anforderungen an Produkte, Dienstleistungen oder Verfahren definiert und damit die Grundlage für die Technische Beschaffung und Produktentwicklung gelegt. Gleichzeitig sorgen Normen und Standards für Interoperabilität und dienen dem Schutz von Mensch, Umwelt und Sachen sowie der Qualitätsverbesserung in allen Lebensbereichen. Auf diese Weise stellen sie Transparenz und Vertrauen in die Anwendung von Technologien her und unterstützen zugleich die Kommunikation unter allen Beteiligten durch einheitliche Begriffe und Konzepte. Der gesamtwirtschaftliche Nutzen der Normung wird für Deutschland auf 17 Mrd. Euro im Jahr geschätzt [11].

Bei der Schaffung eines zukunftsfähigen Handlungsrahmens für Künstliche Intelligenz spielen Normen und Standards eine ganz besondere Rolle: Sie fördern den schnellen Transfer von Technologien aus der Forschung in die Anwendung und öffnen internationale Märkte für deutsche Unternehmen und ihre Innovationen. Gerade beim Thema KI kann ein frühzeitiges und umfassendes Engagement der deutschen Stakeholder in der Normung auf nationaler – vor allem aber auch auf europäischer und internationaler – Ebene dazu beitragen, die Position Deutschlands als Wirtschaftsnation und Exportland entscheidend zu stärken.

Insbesondere der deutsche Mittelstand kann davon profitieren. Hier zeigt sich ein wesentlicher Vorteil der Normung. Es gilt der Grundsatz: Nicht der Größere entscheidet, sondern der Konsens. Eine Mitarbeit eröffnet innovativen kleinen und mittelständigen Unternehmen so die Möglichkeit, auf Augenhöhe mit den großen nationalen und internationalen Konzernen an der Zukunft von KI zu arbeiten und eigene Vorstellungen in den Normungsprozess einzubringen. Über offene Schnittstellen und dank einheitlicher Anforderungen erhält er besseren Zugang zum globalen Markt und die Chance, dort seine Ideen zu positionieren.

Ein solches Engagement ist aus nationaler und europäischer Sicht enorm wichtig: Wer seine Standards international durchsetzt, hat einen Vorsprung, weil so die eigenen Regeln gelten und auf bestehenden Lösungen aufgebaut werden kann. Die Wettbewerber sind sich dessen bewusst, allen voran China und die USA. Diese verfolgen naturgemäß ihre ganz eigenen Interessen – und deren Vorstellungen können unseren europäischen Wertmaßstäben und ethischen Richtlinien widersprechen. Dass aber gerade im werteorientierten Deutschland und Europa der Frage nach technischer Souveränität und vor allem nach Datensouveränität nachgegangen wird, zeigen Leuchtturmprojekte wie GAIA-X, die den Mehrwert von „KI – Made in Germany“ im internationalen Kontext manifestieren sollen. Normen und Standards unterstützen die Souveränität, indem sie Transparenz fördern und Rahmenbedingungen setzen, die einen „moralischen Kompass“ geben. Zwar ist es Aufgabe von Gesellschaft und Politik, zu definieren, was ethisch ist; technische Standards können jedoch dazu beitragen, bestehende ethische Werte umzusetzen, und so von technischer Seite den Schutz beispielsweise gegen Verzerrungen, Diskriminierungen und Manipulationen sicherstellen.

In diesem Kontext tragen Normen und Standards maßgeblich zur Erklärbarkeit und Nachvollziehbarkeit bei – zwei essenzielle Bausteine, wenn es um die Akzeptanz von KI-Anwendungen geht. Gleichzeitig sorgen sie für Sicherheit und Vertrauen, insbesondere in einem sensiblen Themenfeld wie der KI. Auch die Bundesregierung weist Normen und Standards gerade beim Thema Künstliche Intelligenz eine zentrale Rolle zu. Nicht zuletzt deshalb sind die Normung und Standardisierung ein zentraler Baustein in der KI-Strategie der Bundesregierung.

1.4 KI-Strategie der Bundesregierung

Die Bundesregierung hat am 15. November 2018 die nationale Strategie „Künstliche Intelligenz“ verabschiedet [12] und forciert damit den Weg von „Künstliche Intelligenz – Made in Germany“ an die Weltspitze. Mit ihr will die Bundesregierung den exzellenten Forschungsstandort Deutschland sichern, die Wettbewerbsfähigkeit der deutschen Wirtschaft und von Europa ausbauen sowie die vielfältigen Anwendungsmöglichkeiten von KI in allen Bereichen der Gesellschaft fördern. Der Nutzen für Mensch und Umwelt soll dabei in den Mittelpunkt gestellt und der intensive Austausch zum Thema KI mit allen gesellschaftlichen Gruppen gestärkt werden. Um die ehrgeizigen Ziele zu erreichen, hat die Bundesregierung im Rahmen ihres jüngsten Konjunkturprogramms [13] beschlossen, die geplanten Investitionen zur KI-Förderung von drei Milliarden Euro auf fünf Milliarden Euro bis 2025 zu erhöhen. Im Fokus stehen die Bereiche Forschung, Transfer, gesellschaftlicher Dialog, Qualifikation und Datenverfügbarkeit. Damit wird ein wettbewerbsfähiges europäisches KI-Netzwerk unterstützt.

Wesentliche Ziele der KI-Strategie der Bundesregierung sind:

- → Stärkung der Wettbewerbsfähigkeit Deutschlands und Europas
- → Verantwortungsvolle und gemeinwohlorientierte Entwicklung und Nutzung von KI
- → Ethische, rechtliche, kulturelle und institutionelle Einbettung von KI in die Gesellschaft

Die Strategie beschreibt in insgesamt zwölf Handlungsfeldern (siehe Abbildung 3) konkrete Maßnahmen.

Normung und Standardisierung sind eines von insgesamt zwölf Handlungsfeldern. Beim Handlungsfeld 10 „Standards setzen“ heißt es:

„Die Bundesregierung wird (u. a.) in einem gemeinsamen Projekt mit DIN eine Roadmap zu Normen und Standards im Bereich KI entwickeln.“

Ferner wird die Überprüfung bestehender Normen und Standards auf „KI-Tauglichkeit“ sowie die Entwicklung maschinenlesbarer und von Maschinen interpretierbarer Normen und Standards (Smart Standards) für KI-Anwendungen angeregt (siehe Kapitel 5).

Abbildung 3: Die zwölf Handlungsfelder der KI-Strategie der Bundesregierung

KI-Strategien anderer Länder

Der Wettlauf um die weltweit führende Position bei der Künstlichen Intelligenz hat längst begonnen. Seither suchen viele Länder und Wirtschaftszonen nach Wegen, um die Forschung und Anwendung von KI in ihren Ländern zu fördern, und haben hierfür eigene nationale Strategien entwickelt. Im Folgenden werden exemplarisch KI-Strategien einzelner Länder mit ihren Besonderheiten vorgestellt [14]:

Weißbuch der Europäischen Kommission zu KI

Mit dem Weißbuch „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“ [15] hat die EU-Kommission ihre Vision für einen sicheren und verantwortungsvollen Umgang mit Künstlicher Intelligenz veröffentlicht. Es stellt einen ersten Versuch dar, klare Regeln aufzustellen, was KI darf und was nicht, und schlägt Ansätze vor, wie diese durchgesetzt werden können. Dabei liegt der Fokus darauf, KI für Wissenschaft, Wirtschaft und Gesellschaft nutzbar zu machen und sich gleichzeitig mit den damit einhergehenden Risiken zu beschäftigen. Die vorgeschlagenen Maßnahmen umfassen z. B. eine stärkere Zusammenarbeit mit und zwischen den Mitgliedstaaten sowie die Bündelung von Kompetenzen durch die Erleichterung der Einrichtung von Exzellenz- und Testzentren.

USA

Die Führungsposition der USA bei KI lässt sich anhand folgender Zahlen zusammenfassen: mit rund 22.000 KI-Publikationen publikationsstärktes Land weltweit, etwa 2.400 KI-Start-ups und damit die global größte KI-Startup-Landschaft, Platz 1 beim Einsatz von KI-Anwendungen in Unternehmen (25 Prozent der Unternehmen), sieben der weltweit zehn größten Technologiekonzerne und seit 40 Jahren gewachsene Kooperationsstrukturen zwischen Universitäten, Behörden und Unternehmen [4]. Angesichts dieser Faktoren verwundert es nicht, dass die Obama-Regierung auch die weltweit erste nationale KI-Strategie bereits im Jahr 2016 vorgelegt hat.

Darüber hinaus hat die DARPA (die Defense Advanced Research Project Agency der USA) jüngst eine neue Förderinitiative im Umfang von zwei Mrd. US-Dollar gestartet, die mit einer Laufzeit von fünf Jahren unter dem Namen „AI Next“ läuft, um die Grundlagen für die nächste Generation von KI-Systemen zu entwickeln [16]. DARPA will damit eine neue Welle von KI-Systemen fördern, die robuster und vertrauenswürdiger als die bisherigen Systeme sind, weil sie auf einer engen Integration von Komponenten zum Wahrnehmen, Lernen, Kontextverstehen, Inferieren und Planen sowie auf einer expliziten Repräsentation des bei der Problemlösung benutzten Wissens beruhen. DARPA will damit die zuletzt nach Meinung der US-Regierung in vielen Ländern zu stark auf maschinelles Lernen aus Daten fokussierte KI-Entwicklung überwinden und eine neue Generation von autonomen Systemen entwickeln, die auch in Teams mit Menschen zusammenarbeiten können. KI-Systeme sollen damit in einer „dritten Welle der KI“ von reinen Werkzeugen zu echten kollaborativen Partnern bei konkreten Problemlösungen werden.

China

In drei Schritten plant China, bis 2030 die führende KI-Nation der Welt zu werden, und legt hierfür volkswirtschaftliche Ziele fest. Über 700 Mio. chinesische Internetnutzer und leistungsfähige Hardware- und Technologiekonzerne schaffen gute Voraussetzungen. Zwar hinkt das Land den USA in der

Grundlagenforschung, der Ausbildung qualifizierter Fachkräfte, bei der Anzahl der KI-Start-ups und internationalen Patenten noch hinterher, die Entwicklungen der letzten Jahre lassen jedoch keinen Zweifel daran, dass China aufholt. Allein zur Förderung der Chipindustrie hat Peking 16,4 Mrd. Euro angekündigt, auf der subnationalen Ebene hat eine einzige Stadt (Tijian) einen Fonds von 12,8 Mrd. Euro für KI-Förderung aufgelegt. Mit dem „Thousand Talents"-Programm will Peking darüber hinaus hochqualifizierte Auslandschinesen zurückholen. Trotz des massiven Mitteleinsatzes lässt sich jedoch ein wissenschaftlicher Durchbruch, insbesondere in der schwachen Grundlagenforschung, nicht planen. Neben Kapital braucht es dafür vor allem förderliche akademische Rahmenbedingungen.

Großbritannien
Anfang 2018 einigten sich die britische Regierung und Privatwirtschaft, die Forschung und Kommerzialisierung von KI mit einer Milliarde Euro gemeinsam zu fördern. Wesentliche Stärken des Landes sind neben der international sehr einflussreichen KI-Forschung die KI-Start-up-Szene. Nirgendwo sonst in Europa konzentrieren sich mehr KI-Start-ups. Gleichzeitig hat die Regierung zur Entwicklung ethischer Richtlinien für KI Grundlagen geschaffen, etwa durch die Gründung eines Zentrums für Ethik. In den letzten Jahren hat Großbritannien die Technologiekooperation mit den USA ausgebaut. Dagegen weist das Land Schwächen in der Kommerzialisierung der Forschung auf, was sich u. a. durch die geringe Anzahl an Patenten manifestiert.

Frankreich
Frankreich formuliert einen Führungsanspruch auf einem auf europäischen Werten basierendem Mittelweg zwischen China und den USA. Bei der KI-Entwicklung setzt das Land auf eine zentralisierte Struktur und Organisation des Staatswesens. Die zuständigen Ministerien fokussieren mit ihren Strategien und Mitteln auf KI-Anwendungen in den Bereichen Gesundheit, Mobilität und Verteidigung. Strukturelle Schwächen zeigen sich in der geringen Anzahl der Institute und Lehrkörper, die aktiv in Bereichen mit direktem KI-Bezug forschen (Großbritannien hat fast achtmal mehr, Deutschland etwa viermal mehr) und der mangelnden Kooperation zwischen Universitäten und der Industrie. In geplanten KI-Exzellenzzentren werden daher einerseits Wissenschaftler gebündelt, um mit Anwendern in gewisser Autonomie zusammenarbeiten zu können. Zugleich setzt Frankreich neue Regeln fest, die es Forschern erlauben zugleich auch im Privatsektor tätig zu sein. Ein Netzwerk aus freiwilligen KI-Experten soll den Staat bei der Beschaffung von Technologien beraten und die Cyber-Sicherheit unterstützen.

1.5 Ziele und Inhalte der Normungsroadmap KI

Die frühzeitige Entwicklung eines Handlungsrahmens, der die Erfordernisse im Bereich der Normung und Standardisierung darlegt, ist essenziell und notwendig. Der daraus resultierende Anstoß entsprechender Arbeiten in der Normung und Standardisierung auf nationaler, aber vor allem auf europäischer und internationaler Ebene kann die Rolle Deutschlands als Wirtschaftsnation und Exportland entscheidend stärken.

Insbesondere in einem solch sensiblen Themenfeld wie der KI können dadurch entscheidende Schritte eingeleitet werden, die zu Vertrauen und Sicherheit beim Einsatz von KI führen. Das Thema KI besitzt unweigerlich auch einen erkennbaren Bezug zur Gesetzgebung. Die in Normen und Standards festgelegten Anforderungen können in diesem Prozess zugleich eine entlastende Wirkung auf die Gesetzgebung entfalten und damit zu einer Beschleunigung der Festschreibung z. B. von Rahmenbedingungen beitragen.

Um bei diesem Prozess eine federführende Rolle zu erlangen, ist es wichtig, sich entsprechend zu positionieren. Grundlage hierfür muss die abgestimmte Ausrichtung zu den relevanten Themenfeldern sowie ein funktionierendes Netzwerk sein. Die Erarbeitung entsprechender Empfehlungen in Form einer Normungsroadmap kann hier einen wesentlichen Beitrag leisten, um die nationale Position auf Basis eines breiten Abstimmungsprozesses auch auf europäischer und letztlich internationaler Ebene einzubringen. DIN und DKE bieten eine anerkannte und neutrale Plattform, um diese Arbeiten mithilfe ihrer langjährigen Expertise und Netzwerkkompetenz zu orchestrieren.

Um frühzeitig einen Handlungsrahmen für die Normung und Standardisierung im Bereich KI zu entwickeln, haben DIN und DKE im Auftrag des Bundesministeriums für Wirtschaft und Energie (BMWi) die Arbeiten an der Normungsroadmap „Künstliche Intelligenz" angestoßen. Mit diesem Schritt wird die KI-Strategie der Bundesregierung umgesetzt (siehe Kapitel 1.4). Am 16. Oktober 2019 wurde mit einer Auftaktveranstaltung unter Teilnahme von über 300 Teilnehmerinnen und Teilnehmern aus Wirtschaft, Zivilgesellschaft, Politik und Wissenschaft der Startschuss für die Normungsroadmap KI gegeben.















Abbildungen: Impressionen von der Auftaktveranstaltung

Die Normungsroadmap stellt ein „lebendes Dokument“ dar, das die bisherigen Arbeits- und Diskussionsergebnisse vorstellt und als zentrales Kommunikationsmedium zum Austausch zwischen Normungsgremien, Industrie, Verbänden, Forschungseinrichtungen und Politik dient.

Ziel der Normungsroadmap KI ist es, frühzeitig einen Handlungsrahmen zu beschreiben, der die deutsche Wirtschaft und Wissenschaft im internationalen Wettbewerb um die besten Lösungen und Produkte im Bereich der Künstlichen Intelligenz stärkt und innovationsfreundliche Rahmenbedingungen für die Technologie der Zukunft schafft.

Sie zeigt Bedarfe für Normen und Standards, insbesondere im Hinblick auf die Sicherheit, Zuverlässigkeit und Robustheit von KI-Systemen auf und trägt wesentlich dazu bei, die Qualität von KI-Lösungen sicherzustellen. Neben der Darstellung des Akteursumfeldes gibt sie eine Übersicht über bestehende Normen und Standards zu Aspekten der KI, skizziert die wesentlichen Normungs- und Standardisierungspotenziale und spricht Handlungsempfehlungen an die Politik, Forschung und Normung aus. Damit leistet sie einen wesentlichen Beitrag, um „KI – Made in Germany“ als starke Marke zu etablieren und neue Geschäftsmodelle, disruptive Innovationen und skalierbare Anwendungen zu entwickeln. Gleichzeitig bietet sie großes Potenzial, um europäische Wertmaßstäbe auf die internationale Ebene zu heben.

Die Normungsroadmap KI wird in einem offenen, transparenten und breit angelegten Beteiligungsprozess durch Vertreter aus Wirtschaft, Wissenschaft, öffentlicher Hand und Gesellschaft erarbeitet und regelmäßig fortgeschrieben.

Auf Basis der Ergebnisse der Normungsroadmap und der identifizierten Handlungsempfehlungen (siehe Kapitel 2) werden im nächsten Schritt konkrete Normungs- und Standardisierungsaktivitäten an die jeweiligen Gremien kommuniziert.

1.6 Hochrangige Steuerungsgruppe

Gesteuert wird die Erarbeitung der Normungsroadmap KI von einer Gruppe aus führenden Persönlichkeiten aus Wirtschaft, Politik, Wissenschaft und Zivilgesellschaft. Vorsitzender der Steuerungsgruppe ist Prof. Wolfgang Wahlster, Mitglied des Lenkungskreises der Plattform Lernende Systeme (PLS) und führender deutscher Wissenschaftler im Bereich KI.

Die 20-köpfige Gruppe ist verantwortlich für die inhaltliche und strategische Ausrichtung der Normungsroadmap KI und ebnet mit ihr den Weg für den Ausbau des KI-Standortes Deutschland. Die Mitglieder repräsentieren wichtige Themen, Disziplinen, Branchen und Unternehmen unterschiedlicher Größe im Bereich von KI und verstehen sich als Botschafter für den Transfer wissenschaftlicher Ergebnisse durch Normen und Standards in die Wirtschaft und wichtige Lebensbereiche.

Mit der Gründung der Steuerungsgruppe wurde ein wichtiger Schritt getan, den notwendigen Handlungsrahmen für KI zu schaffen.

MITGLIEDER DER STEUERUNGSGRUPPE:

Dr. Tarek R. Besold
neurocat GmbH

Jörg Bienert
Bundesverband künstliche Intelligenz e. V.

Dr. Julia Borggräfe
Bundesministerium für Arbeit uns Soziales

Dr. Joachim Bühler
Verband der TÜV e. V.

Susanne Dehmel
Bitkom e. V.

Dr. Dirk Hecker
Fraunhofer – Allianz Big Data und Künstliche Intelligenz

Thorsten Herrmann
Microsoft Deutschland GmbH

Stefan Heumann
Stiftung Neue Verantwortung

Dr. Wolfgang Hildesheim
IBM Deutschland

Prof. Jana Koehler
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Prof. Klaus Mainzer
Technische Universität München

Dr. Christoph Peylo
Robert Bosch GmbH

Alexander Rabe
eco Verband der Internetwirtschaft

Prof. Ina Schieferdecker
Bundesministerium für Bildung und Forschung

Stefan Schnorr
Bundesministerium für Wirtschaft und Energie

Andreas Steier, MdB
Christlich Demokratische Union Deutschland

Dr. Volker Treier
Deutscher Indusrie und Handelskammertag e. V.

Prof. Wolgang Wahlster
Lenkungskreis Plattform Lernende Systeme

Prof. Dieter Wegener
Siemens AG

Christoph Winterhalter
Deutsches Institut für Normung e. V.

STÄNDIGE GÄSTE DER STEUERUNGSGRUPPE:

Dr. Gerhard Schabhüser
Bundesamt für Sicherheit in der Informationstechnik

Dr. Johannes Winter
Plattform Lernende Systeme/acatech

1.7 Methodisches Vorgehen

Die Mitwirkung von Expertinnen und Experten aller relevanten Kreise stellt die wesentliche Grundlage bei der Erarbeitung der Normungsroadmap dar. Zu den einzubeziehenden Stakeholdern gehören Wirtschaftsvertreter der relevanten Branchen, Experten aus der Wissenschaft, Vertreter aus der Politik und der Zivilgesellschaft sowie Repräsentanten bereits konstituierter und mit dem Thema KI befasster Kreise. Hierbei ist die Berücksichtigung verschiedener Sichtweisen und damit verbundener Anforderungen von hoher Bedeutung, sodass sowohl technische als auch nicht-technische Aspekte gleichermaßen Eingang in den Entstehungsprozess der Normungsroadmap KI fanden.

Die Erarbeitung der Normungsroadmap KI umfasste die übergeordnete Koordination und Orchestrierung der relevanten Stakeholder und erfolgte in sieben Arbeitsgruppen zu verschiedenen Schwerpunktthemen (siehe Kapitel 4). Für die Leitung dieser Arbeitsgruppen konnten erfahrene Expertinnen und Experten gewonnen werden , die die inhaltlichen Arbeiten leiteten und an die Steuerungsgruppe berichteten:

1. Grundlagen (Leitung: Dr. Peter Deussen, Microsoft Deutschland GmbH, und Dr. Wolfgang Hildesheim, IBM Deutschland)
2. Ethik/Responsible AI (Leitung: Tobias Krafft, Technische Universität Kaiserslautern)
3. Qualität, Konformitätsbewertung und Zertifizierung (Leitung: Dr. Maximilian Poretschkin, Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIS, und Daniel Loevenich, Bundesamt für Sicherheit in der Informationstechnik)
4. IT-Sicherheit bei KI-Systemen (Leitung: Annegrit Seyerlein-Klug, secunet Security Networks AG)
5. Industrielle Automation (Leitung: Dr.-Ing. Christoph Legat, HEKUMA GmbH)
6. Mobilität und Logistik (Leitung: Dr. Reinhard Stolle und Bogdan Bereczki, beide Argo AI)
7. KI in der Medizin (Leitung: Prof. Dr. Johann Wilhelm Weidringer, Bayerische Landesärztekammer und Bundesärztekammer)

Rund 300 Expertinnen und Experten aus verschiedenen Branchen und mit unterschiedlichen Erfahrungshintergründen brachten ihr Fachwissen in den sieben Arbeitsgruppen ein. Die Zusammensetzung der Arbeitsgruppen zeigt die Abbildung 4.

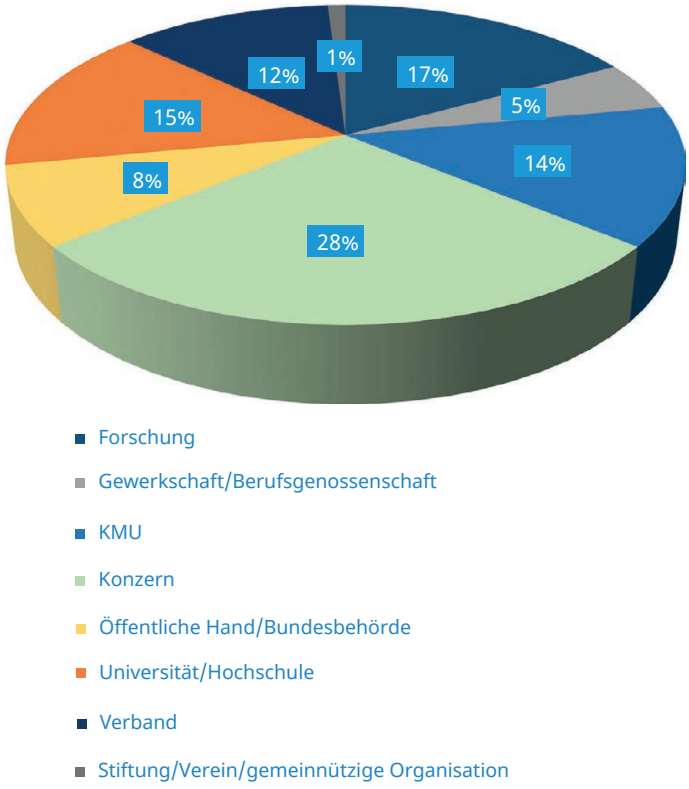


Abbildung 4: Zusammensetzung der sieben Arbeitsgruppen der Normungsroadmap KI

Die Inhalte wurden auf der digitalen Arbeitsplattform DIN.ONE (www.din.one/site/ki) erarbeitet.

Die Normungsroadmap KI wird im Rahmen des Digital-Gipfels 2020 vorgestellt und der Bundesregierung übergeben werden. Sie steht in deutscher und englischer Sprachfassung unter www.din.de/go/normungsroadmapki kostenlos zum Download bereit.

Der Veröffentlichung der Normungsroadmap KI schließt sich unmittelbar ihre Umsetzung an. Das bedeutet, dass auf der Basis der Ergebnisse der Roadmap und ihrer Handlungsempfehlungen (siehe Kapitel 2) im nächsten Schritt konkrete Normungs- und Standardisierungsaktivitäten angestoßen werden.

2 Handlungsempfehlungen der Normungsroadmap KI

Ziel der Normungsroadmap KI ist es, frühzeitig einen Handlungsrahmen zu beschreiben, der die deutsche Wirtschaft und Wissenschaft im internationalen Wettbewerb um die besten Lösungen und Produkte im Bereich der Künstlichen Intelligenz stärkt und innovationsfreundliche Rahmenbedingungen für die Technologie der Zukunft schafft. Damit leistet sie einen wesentlichen Beitrag, um „KI – Made in Germany" als starke Marke zu etablieren und neue Geschäftsmodelle, disruptive Innovationen und skalierbare Anwendungen zu entwickeln. Insbesondere der deutsche Mittelstand und die wachsende Start-up-Szene in Deutschland können davon profitieren, wenn es gelingt, den großen industriellen Erfahrungsschatz der deutschen Wirtschaft mit den Möglichkeiten der KI-Methoden zu verbinden. Normen und Standards bilden die Grundlage für technische Souveränität und schaffen einen Rahmen, der Transparenz fördert und Orientierung bietet. Somit sorgen sie für Sicherheit, Qualität und Zuverlässigkeit und tragen maßgeblich zur Erklärbarkeit von KI-Lösungen bei – essenzielle Bausteine, wenn es um die Akzeptanz von KI-Anwendungen geht. Damit bietet die Normungsroadmap KI großes Potenzial, um sowohl die Wettbewerbsfähigkeit Deutschlands zu sichern als auch europäische Wertmaßstäbe auf die internationale Ebene zu heben. Nicht zuletzt deshalb sollte ein besonderes Augenmerk auf die Umsetzung der vorliegenden Normungsroadmap KI und ihrer Handlungsempfehlungen gelegt werden.

Handlungsempfehlung 1: Datenreferenzmodelle für die Interoperabilität von KI-Systemen umsetzen

Ein großer Teil der deutschen Industrie ist geprägt von kleinen und mittelgroßen Unternehmen. Die Verwirklichung von übergreifenden Wertschöpfungsketten erfordert deshalb oft die Zusammenarbeit unterschiedlichster Akteure. Für die Anwendung von Methoden Künstlicher Intelligenz ist die Automatisierung der Zusammenarbeit von KI-Systemen verschiedener Akteure entlang der Wertschöpfungskette entscheidend. Dazu bedarf es eines Datenreferenzmodells, um den Datenaustausch sicher, zuverlässig, flexibel und kompatibel zwischen den Technologien ausführen zu können. In einem solchen Modell sollten grundlegende und für die Interoperabilität relevante Datenarten, ihre Strukturen und Beziehungen zueinander dargelegt werden.[2] Die Normungsroadmap KI empfiehlt daher ein Umsetzungsprogramm für die Standardisierung von Datenreferenzmodellen in unterschiedlichen Domänen. Durch eine solche Initiative kann Deutschland die Grundlage für einen übergreifenden Datenaustausch schaffen, indem es internationale Normen und Standards maßgeblich mitgestaltet und auf diese Weise weltweit die Interoperabilität von Systemen sicherstellt.[3]

Handlungsempfehlung 2: Horizontale KI-Basis-Sicherheitsnorm erstellen

Bei einem KI-System handelt es sich im Kern um ein IT-System, für dessen IT-Sicherheit bereits eine Vielzahl an Normen und Standards aus verschiedensten Branchen und Anwendungsfeldern existiert. Mit einer solchen Varietät steigt gleichzeitig die Komplexität und Unübersichtlichkeit, was schließlich zu einem inkonsistenten Vorgehen der beteiligten Akteure am Markt (z. B. Hersteller, Verbraucher, Regulierer) führen und die Technologieentwicklung von KI maßgeblich hemmen kann. Insbesondere die IT-Sicherheit für KI-Systeme leidet darunter, da sie über den gesamten Lebenszyklus in großem Maße etwa von Transparenz und Nachvollziehbarkeit, Security-by-Design, Security-by-Default und Privacy abhängt. Eine allumfassende „KI-Umbrella-Norm", die vorhandene Normen und Prüfverfahren für IT-Sicherheit (Security, Safety und Privacy) bündelt und um Aspekte speziell für KI-Systeme ergänzt, kann einerseits als Katalysator für die Technologieentwicklung dienen und andererseits zwischen den Akteuren vermitteln. Aus diesem Grund wird die Erstellung einer horizontalen Basis-Sicherheitsnorm für KI empfohlen, die in vertikalen Sub-Normen weitere Themen und Branchen mit ihren Spezifika berücksichtigt und ein IT-(Security-)Managementsystem für KI-Systeme integriert. Damit würden etablierte Vorgehensweisen beibehalten und ein prüf- und zertifizierbarer Standard geschaffen werden, der wirtschaftliche Aspekte berücksichtigt und die Akzeptanz steigert. Das wiederum baut Vertrauen auf und fördert den Einsatz von KI-Technologien.[4]

Handlungsempfehlung 3: Praxisgerechte initiale Kritikalitätsprüfung von KI-Systemen ausgestalten

Nicht intendierte ethische Probleme und Konflikte treten vor allem bei ADM-Systemen mit lernenden Komponenten auf, die über Menschen, deren Hab und Gut oder über den Zugang zu knappen Ressourcen entscheiden und dabei das Potenzial aufweisen, individuelle Grundrechte und/oder grundlegende demokratische Werte zu schädigen. Eine initiale Kritikalitätsprüfung, ob ein System solche Konflikte überhaupt auslösen

2 Siehe Beispiel 2: Semantische KI-Technologien zur Gewährleistung der Interoperabilität von Systemen

3 Grundzüge und Ziele eines solchen Umsetzungsprogramms sind in Kapitel 4.3.2.4 erläutert.

4 Für weitere Informationen siehe Kapitel 4.4.2.3

kann oder es sich um eine Anwendung fernab jeder ethischen Fragestellung handelt, muss durch Normung schnell und einfach gestaltet werden. Diese horizontal für alle Bereiche niedrigschwellige Überprüfung muss schnell und rechtsicher klären, ob das System überhaupt Transparenz- und Nachvollziehbarkeitsanforderungen erfüllen muss. Gerade im Hinblick auf die weiten Einsatzfelder von Künstlicher Intelligenz bietet eine solche risikoadaptive Kritikalitätsprüfung in kritischen Bereichen die Möglichkeit, adäquate Forderungen zu stellen und gleichzeitig dem Vorwurf des „ethical red taping“ zu begegnen, indem völlig unkritische Anwendungsfelder frei von zusätzlichen Anforderungen entwickelt werden können. Daher wird die Ausgestaltung einer praxisnahen und risikoadaptiven Kritikalitätsprüfung für KI-Systeme empfohlen.[5]

Handlungsempfehlung 4: Nationales Umsetzungsprogramm „Trusted AI“ zur Ertüchtigung der europäischen Qualitätsinfrastruktur initiieren und durchführen

Wirtschaft, Behörden und Zivilgesellschaft fordern verlässliche Qualitätskriterien und Prüfverfahren für die marktfähige Konformitätsbewertung und Zertifizierung von KI-Systemen. Das Fehlen solcher Prüfverfahren gefährdet das wirtschaftliche Wachstum und die Wettbewerbsfähigkeit dieser Zukunftstechnologie. Gleichzeitig sind Aussagen über die Vertrauenswürdigkeit von KI-Systemen ohne hochwertige Prüfmethoden nicht belastbar, wodurch der Nutzen von KI-Anwendungen in Wirtschaft und Gesellschaft wegen fehlender Akzeptanz unklar bleibt. Eine erfolgreiche Nutzung, die unseren ethisch-gesellschaftlichen Wertevorstellungen genügt, benötigt sachkundige, verlässliche und reproduzierbare Prüfungen. Die Grundlage hierfür kann ein nationales Umsetzungsprogramm für die Zertifizierung von KI-Systemen legen, das auf die hervorragende deutsche Prüfinfrastruktur aufbaut und Anforderungen beispielsweise an die Verlässlichkeit, Robustheit, Leistungsfähigkeit und funktionale Sicherheit definiert. Anhand konkreter Anwendungsfälle sollen dabei Prüfgrundlagen getestet, Pilotprüfungen durchgeführt und Standards abgeleitet werden, die die Grundlage für eine KI-Zertifizierung bilden und in die internationale Normung eingebracht werden. Die zu entwickelnden Prüfverfahren dienen einerseits der Bestätigung der zugesicherten Eigenschaften von KI-Systemen (Produktprüfung) und andererseits der Bewertung der Maßnahmen der Organisationen, die KI-Systeme bereitstellen (Managementsystemprüfung). Mit einer solchen Initiative hätte Deutschland die Chance, die Grundlage für ein weltweit erstes Zertifizierungsprogramm zu legen und damit führend bei der Entwicklung und Standardisierung eines international anerkannten KI-Zertifizierungsverfahrens zu sein. Die Normungsroadmap KI empfiehlt daher die schnellstmögliche Initiierung und Durchführung eines nationalen Umsetzungsprogramms „Trusted AI“ mit höchster Priorität.[6]

Handlungsempfehlung 5: Use Cases auf Normungsbedarf analysieren und bewerten

Use Cases beschreiben Anwendungsfälle, die im Kontext von KI-Technologien essenziell für das Verständnis von Funktion und Verhalten der KI-Systeme sind. Bereits heute existiert eine Vielzahl an Anwendungsfällen für verschiedene Einsatzfelder der KI. Durch die Betrachtung anwendungstypischer und branchenrelevanter Use Cases können für industriereife KI-Anwendungen Normungs- und Standardisierungsbedarfe abgeleitet werden. Allerdings ist die bewährte Vorgehensweise der klassischen Normung für KI-Anwendungen nicht immer zielführend. Der Grund ist: Viele Branchen verwenden abhängig vom Einsatzfeld der KI-Lösung unterschiedliche und auf den Anwendungsfall bezogene KI-Technologien. Hybride KI-Lösungen beruhen oftmals sogar auf einer Kombination von KI-Methoden. Die Anwendungsspezifika werden dabei in den allermeisten Fällen von modernsten Ansätzen aus KI-Teildisziplinen erfüllt, die individuell angepasst und verfeinert werden. Folglich ist die Dynamik an der Schnittstelle zwischen KI-Forschung und industrieller Entwicklung und Anwendung besonders hoch. So wird die angewandte KI ständig weiterentwickelt und industriell evaluiert. Die KI-Standardisierung muss diesem Spannungsbogen zwischen angewandter Forschung und industriereifer Entwicklung Rechnung tragen und pragmatische, bidirektionale Ansätze bei der Analyse der Standardisierungsbedarfe und der Entwicklung marktreifer Standards verfolgen. Hierfür braucht es einen iterativen Prozess, der bei der Gestaltung von Normen und Standards wechselseitig Impulse aus Forschung, Industrie, Gesellschaft und Regulierung einbezieht und ein kontinuierliches und gegenseitiges Lernen zwischen den Akteuren unterstützt. Im Zentrum dieses Ansatzes steht die Erprobung und sukzessive Verfeinerung der entwickelten Standards entlang von Use Cases. So können anwendungsspezifische Bedarfe frühzeitig erkannt und marktfähige KI-Standards realisiert werden. In der Folge wird damit die Akzeptanz der KI-Standards in Wirtschaft, Wissenschaft und Gesellschaft sichergestellt.[7]

5 Für weitere Informationen siehe Kapitel 4.1.2.1.5

6 Für weitere Informationen siehe Kapitel 4.3

7 Für weitere Informationen siehe Kapitel 4.5.2.2

3 Akteurs- und Normungsumfeld

Aktuell gibt es sowohl auf nationaler als auch auf europäischer und internationaler Ebene eine Vielzahl an Akteuren, Initiativen, Gremien sowie Normungs- bzw. Standardisierungsaktivitäten, die sich mit dem Thema KI auseinandersetzen. Im Folgenden wird das KI-Umfeld mit den wesentlichen Akteuren und Initiativen dargestellt[8].

3.1 Gesellschaftspolitisches Umfeld

Ethik-Kommission „Automatisiertes und vernetztes Fahren“
Die vom Bundesministerium für Verkehr und digitale Infrastruktur eingesetzte Ethik-Kommission „Automatisiertes und Vernetztes Fahren“ wurde im September 2016 eingerichtet. In der interdisziplinär besetzten Kommission wirkten hochrangige Experten aus Philosophie, Rechts- und Sozialwissenschaften, Technikfolgenabschätzung, Verbraucherschutz, Automobilindustrie sowie Digitalwirtschaft mit. Sie war das weltweit erste Gremium, das sich mit den wichtigen gesellschaftsrelevanten Fragen beim automatisierten und vernetzten Fahrzeugverkehr auseinandergesetzt hat. In ihrem Abschlussbericht hat die Ethik-Kommission insgesamt zwanzig ethische Regeln bzw. „Entwicklungsleitlinien“ erarbeitet [17].

Enquete-Kommission
Die Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale“ [18] wurde im Juni 2018 vom Deutschen Bundestag eingesetzt und soll den zukünftigen Einfluss von KI auf das gesellschaftliche Leben, die Wirtschaft und die Arbeitswelt untersuchen – allesamt Bereiche, auf die Normung und Standardisierung ebenfalls einen essenziellen Einfluss haben. Die Enquete-Kommission setzt sich je zur Hälfte aus Abgeordneten des Deutschen Bundestages (in prozentualer Vertretung der jeweiligen Fraktion im Parlament) sowie aus externen Sachverständigen zusammen.
Die Mitglieder arbeiten in sechs Projektgruppen, von denen jeweils drei parallel tagten bzw. tagen:

- → KI und Wirtschaft (Industrie/Produktion, Finanzen, Dienstleistungen, Innovationen)
- → KI und Staat (Verwaltung, Sicherheit, Infrastruktur)
- → KI und Gesundheit (Pflege, Sport)
- → KI und Arbeit, Bildung, Forschung
- → KI und Mobilität (Energie, Logistik, Umwelt)
- → KI und Medien (Social Media, Meinungsbildung, Demokratie)

Für die Projektgruppen KI und Wirtschaft, KI und Staat sowie KI und Gesundheit wurden im Dezember 2019 bereits Zusammenfassungen der vorläufigen Ergebnisse veröffentlicht [19]–[21]. Der Abschlussbericht der Enquete-Kommission wird für Herbst 2020 erwartet.

High-Level Expert Group on Artificial Intelligence
Die Hochrangige Expertengruppe für Künstliche Intelligenz (HLEG-KI, High-Level Expert Group on Artificial Intelligence) setzt sich aus 52 Experten aus Wissenschaft, Zivilgesellschaft und Industrie zusammen und stellt das zentrale Gremium der Europäischen Kommission im Bereich der KI dar. Ihre Aufgabe ist es, die Umsetzung der europäischen KI-Strategie zu unterstützen. Dazu gehört die Ausarbeitung von Empfehlungen zur zukunftsbezogenen Politikentwicklung und zu ethischen, rechtlichen und gesellschaftlichen Fragen im Zusammenhang mit KI, einschließlich sozioökonomischer Herausforderungen.

Die HLEG-KI hat 2018 und 2019 folgende Ergebnisse hervorgebracht:

- → Ethikrichtlinien zur Künstlichen Intelligenz [5]:
  Die Richtlinien legen einen menschenzentrierten Ansatz zu KI vor und listen sieben Schlüsselanforderungen auf, die KI-Systeme erfüllen sollten, um vertrauenswürdig zu sein. Sie umfassen Themen wie beispielsweise Fairness, Sicherheit, Transparenz, Zukunft der Arbeit, Demokratie, Privatsphäre und Schutz personenbezogener Daten.
- → Politik- und Investitionsempfehlungen:
  Aufbauend auf ihren ersten Ergebnissen hat die HLEG-KI 33 Empfehlungen zur Stärkung der Wettbewerbsfähigkeit Europas einschließlich Leitlinien für eine strategische Forschungsagenda zu KI und für den Aufbau eines Netzwerks von KI-Exzellenzzentren [22] vorgelegt. Die Empfehlungen sollen der Kommission und ihren Mitgliedstaaten helfen, den gemeinsam koordinierten Plan zu KI zu aktualisieren. Es wird erwartet, dass dieser eine Schlüsselrolle beim Aufbau der Zukunft der Künstlichen Intelligenz in Europa spielen wird.

8 Die Darstellung erhebt keinen Anspruch auf Vollständigkeit.

3.2 Innovationspolitische Initiativen

Plattform Lernende Systeme

Die Plattform Lernende Systeme (PLS) wurde 2017 durch das Bundesministerium für Bildung und Forschung initiiert mit dem Anspruch, KI zum Wohl der Gesellschaft zu gestalten. Sie vereint rund 200 KI-Expertinnen und Experten aus Wissenschaft, Wirtschaft, Politik und Zivilgesellschaft. In sieben Arbeitsgruppen (AG) entwickeln sie Handlungsoptionen und Empfehlungen für den verantwortungsvollen Einsatz von Lernenden Systemen, von denen fünf Parallelen zu den Themen dieser Normungsroadmap aufzeigen:

- → PLS AG 1 „Technologische Wegbereiter und Data Science“
- → PLS AG 2 „Arbeit/Qualifikation, Mensch-Maschine-Interaktion“
- → PLS AG 3 „IT-Sicherheit, Privacy, Recht und Ethik“
- → PLS AG 4 „Geschäftsmodellinnovationen“
- → PLS AG 5 „Mobilität, intelligente Verkehrssysteme“
- → PLS AG 6 „Gesundheit, Medizintechnik, Pflege“
- → PLS AG 7 „Lebensfeindliche Umgebungen“

In ihren veröffentlichten Publikationen analysiert die PLS technologische, wirtschaftliche, moralische und gesellschaftliche Voraussetzungen für den verantwortungsbewussten und selbstbestimmten Einsatz von KI-Systemen in verschiedenen Anwendungsbereichen (z. B. Medizin und Mobilität). Zudem beleuchtet sie Querschnittsfragen wie Diskriminierung, Zertifizierung oder die IT-Sicherheit von KI-Systemen. Anhand branchenspezifischer Anwendungsszenarien zeigt die PLS auf, was in wenigen Jahren mit KI technologisch möglich ist und welche Rahmenbedingungen dafür zu schaffen sind. Auf ihrer KI-Landkarte macht sie sichtbar, wo KI in Deutschland bereits zum Einsatz kommt und welche Institutionen zum Thema forschen. Die Verbindung aller Aktivitäten der PLS stellt eine Schnittmenge zur Normung und Standardisierung dar, woraus sich Normungs- und Standardisierungspotenziale ergeben.

Plattform Industrie 4.0

Die Plattform Industrie 4.0 (PI4.0) stellt das zentrale Netzwerk dar, um die digitale Transformation in der industriellen Wertschöpfung voranzutreiben. Gegründet im Jahr 2013 durch die Wirtschaftsverbände BITKOM, VDMA und ZVEI umfasst sie heute über 350 Akteure aus Unternehmen, Verbänden, Gewerkschaften, Wissenschaft und Politik. In aktuell sechs Arbeitsgruppen werden relevante Aspekte der Industrie 4.0 betrachtet. Ein Forschungsbeirat bringt Wissenschafts-, Forschungs- und Entwicklungsexpertise in die Arbeitsgruppen und gibt Impulse hinsichtlich künftiger Forschungsthemen.

KI wird in der PI4.0 als Querschnittsthema betrachtet. Demzufolge wurde innerhalb der Plattform eine Projektgruppe „Künstliche Intelligenz“ mit dem Ziel gegründet, dieses Thema hinsichtlich der Anwendung und thematischen Verankerung arbeitsgruppenübergreifend zu betrachten und entsprechende Impulse in den existierenden Arbeitsgruppen zu setzen [23], [15]. Die Projektgruppe hat ihre dedizierten Arbeiten im ersten Quartal 2020 abgeschlossen. Die weiteren Arbeiten werden nun in den sechs Arbeitsgruppen weitergeführt.

- → PI4.0 AG 1 „Referenzarchitekturen, Standards und Normung“ [25]
- → PI4.0 AG 2 „Technologie- und Anwendungsszenarien“
- → PI4.0 AG 3 „Sicherheit vernetzter Systeme“ [26], [27]
- → PI4.0 AG 4 „Rechtliche Rahmenbedingungen“ [28]
- → PI4.0 AG 5 „Arbeit, Aus- und Weiterbildung“ [29]
- → PI4.0 AG 6 „Digitale Geschäftsmodelle in der Industrie 4.0“

Plattform Zukunft der Mobilität

Die Nationale Plattform Zukunft der Mobilität (NPM) unterstützt die Bundesregierung als Initiative des Bundesministeriums für Verkehr und digitale Infrastruktur (BMVI) darin, ihre Ziele beispielsweise im Verkehrssektor und Klimaschutz zu erreichen. Im Detail hat die NPM die übergeordneten Ziele:

- → Entwicklung von verkehrsträgerübergreifenden und -verknüpfenden Lösungen für ein weitgehend treibhausgasneutrales und umweltfreundliches Verkehrssystem
- → Sicherstellung einer wettbewerbsfähigen Automobilindustrie und Förderung des Beschäftigungsstandorts Deutschlands
- → Ermöglichung einer effizienten, hochwertigen, flexiblen, sicheren, resilienten und bezahlbaren Mobilität für den Personen- und Güterverkehr.

Zur künftigen und vollumfänglichen Umsetzung der Ziele sind KI-Systeme essenziell. Konventionelle IT-Systeme sind wegen der Komplexität beispielsweise der Optimierungs- und Verarbeitungsprozesse bereits an ihre Grenzen geraten. Damit Normung und Standardisierung bei den Zielen der Bundesregierung und der NPM unterstützen kann, stellt insbesondere die AG 6 „Standardisierung, Normung, Zertifizierung und Typgenehmigung“ der NPM die Verbindung zwischen den Organisationen dar. Hierdurch existiert ein direkter Austausch zwischen technischen Regelsetzern, rechtlichen Regelsetzern sowie Industrie und Forschung.

3.3 Normungs- und Standardisierungsumfeld

Die wesentliche Normungsarbeit ist eine Gemeinschaftsaufgabe, die in Selbstverwaltung von den interessierten Kreisen (wie beispielsweise Wirtschaft, Wissenschaft, Forschung, Anwender, Verbraucherschutz, Arbeitsschutz, Gewerkschaften, öffentliche Hand und Umweltschutz) erfüllt wird. Ausgangspunkt ist stets ein Bedarf aus den Reihen der interessierten Kreise.

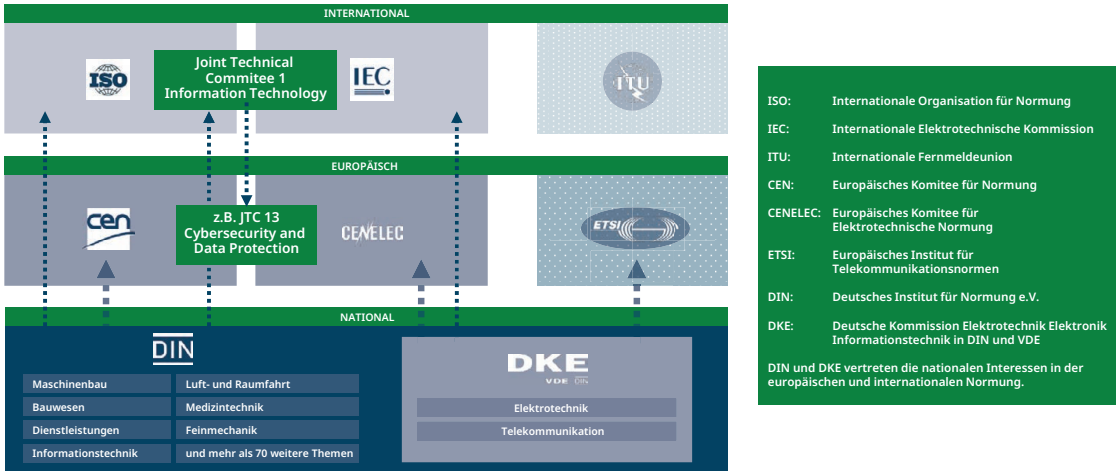
Die Entwicklung von Normen und Standards findet auf unterschiedlichen Ebenen (national, europäisch, international) statt (siehe Abbildung 5).

In Deutschland ist DIN[9] seit 1975 vertraglich die zuständige Normungsorganisation der Bundesrepublik Deutschland und vertritt die deutschen Interessen als Mitglied bei CEN[10] in der europäischen Normung sowie bei ISO[11] in der internationalen Normung. Die DKE[12] nimmt die Interessen der Elektrotechnik, Elektronik und Informationstechnik auf dem Gebiet der internationalen und regionalen elektrotechnischen Normungsarbeit wahr. Sie vertritt somit die deutschen Interessen sowohl bei der CENELEC[13] als auch in der IEC[14].

Heute ist die Normungsarbeit zu fast 90 Prozent europäisch und international ausgerichtet, wobei DIN und DKE den gesamten Prozess der Normung auf nationaler Ebene organisieren und über die entsprechenden nationalen Gremien die deutsche Beteiligung auf europäischer und internationaler Ebene sicherstellen.

Normen sind als Technische Regeln das Ergebnis nationaler, europäischer oder internationaler Normungsarbeit und werden von Ausschüssen nach festgelegten Grundsätzen, Verfahrens- und Gestaltungsregeln erarbeitet[15]. An der Ausschussarbeit können sich alle interessierten Kreise beteiligen, beispielsweise Hersteller, Verbraucher, Handel, Hochschulen, Forschungsinstitute, Behörden, Prüfinstitute etc. Normen entstehen im Konsens. Das bedeutet, die Experten verständigen sich unter Berücksichtigung des Standes der Technik auf gemeinsame Inhalte, welche die Interessen der Beteiligten berücksichtigen. Nach dieser Definition werden im Rahmen dieser Normungsroadmap KI alle Normungsdokumente der nationalen Normungsorganisationen (DIN/DKE), der europäischen Normungsorganisationen (CEN/CENELEC/ETSI) und der internationalen Normungsorganisationen (ISO/IEC/ITU) als „Normen“ bezeichnet.

Abbildung 5:
Nationale, europäische und internationale Normungsebenen



9 Deutsches Institut für Normung e. V., www.din.de

10 Comité Européen de Normalisation, Europäische Organisation für Normung, www.cen.eu

11 International Organization for Standardization, Internationale Organisation für Normung, www.iso.org

12 DKE Deutsche Kommission Elektrotechnik Elektronik Informationstechnik in DIN und VDE, www.dke.de

13 Comité Européen de Normalisation Électrotechnique, Europäisches Komitee für elektrotechnische Normung, www.cenelec.eu

14 International Electrotechnical Commission, Internationale Elektrotechnische Kommission, www.iec.ch

15 Die Anwendung von Normen ist grundsätzlich freiwillig. Erst wenn Normen zum Inhalt von Verträgen werden oder wenn der Gesetzgeber ihre Einhaltung zwingend vorschreibt, werden sie bindend.

Parallel dazu meint die allgemeine Bezeichnung „Standards“ alle weiteren Technischen Regeln wie Technische Reports (TR), Fachberichte, Vornormen, Spezifikationen (TS, DIN SPEC), Konsortialstandards, Anwendungsregeln (AR), Richtlinien, Expertenempfehlungen etc., für deren Erarbeitung und Herausgabe die zuvor genannten sowie auch andere Organisationen und technische Regelsetzer zuständig sein können. So werden beispielsweise Themen, die noch nicht vollkommen im Markt angekommen sind bzw. deren Markt noch nicht existiert, häufig in (Konsortial-)Standards behandelt. Dies kann auch mit dem geringen Reifegrad (oder „Technology Readiness Level“) zusammenhängen. Hierbei sind Konsens und die Einbeziehung aller interessierten Kreise nicht zwingend erforderlich.

Gegenwärtig werden auf allen Ebenen Normungs- bzw. Standardisierungsarbeiten zu KI verrichtet.

Die Normungsarbeiten im Bereich der KI finden auf nationaler Ebene im DIN-Normenausschuss „Informationstechnik und Anwendungen“ (NA 043-01-42 AA) statt. Dieser erarbeitet die deutsche Position in der KI-Normung und spiegelt zugleich Arbeiten auf internationaler und europäischer Ebene.[16]

Auf der europäischen Ebene ist die CEN/CENELEC Focus Group on Artificial Intelligence als relevantes Gremium zu nennen. Sie wurde 2019 von CEN und CENELEC als temporäre Arbeitsgruppe eingerichtet und hat die Aufgabe, einen Fahrplan für eine KI-Normung auf europäischer Ebene zu entwickeln.

Auf internationaler Ebene stellt ISO/IEC JTC 1/SC 42 „Artificial Intelligence“ das zentrale Gremium zur KI-Normung dar und ist damit für die Entwicklung und Veröffentlichung der internationalen Normen zu KI verantwortlich.

Abseits der klassischen Normung gibt es eine Reihe von Fachverbänden und Konsortien, die entsprechende Festlegungen oder Empfehlungen zu KI veröffentlichen. Eine Vielzahl der Konsortialarbeiten zur KI-Standardisierung findet innerhalb diverser Foren und Konsortien wie beispielsweise IETF, IEEE, CSA, OGC, OMG und W3C statt.

Kapitel 6 gibt eine umfassende Übersicht über die wesentlichen Dokumente, Aktivitäten sowie Gremien in der Normung und Standardisierung im Bereich der Künstlichen Intelligenz.

16 DIN und DKE haben im Auftrag des Bundesministeriums für Wirtschaft und Energie ein White Paper zu „Ethikaspekten in Normung und Standardisierung für KI in autonomen Maschinen und Fahrzeugen“ entwickelt, deren Ergebnisse in die Arbeiten der vorliegenden Normungsroadmap KI eingeflossen sind.

4 Schwerpunktthemen

Umfang und Komplexität lassen es sinnvoll erscheinen, das Thema KI nach Grundlagen, horizontalen Themen sowie betroffenen Wirtschafts- und Anwendungsbereichen zu strukturieren (siehe Abbildung 6).

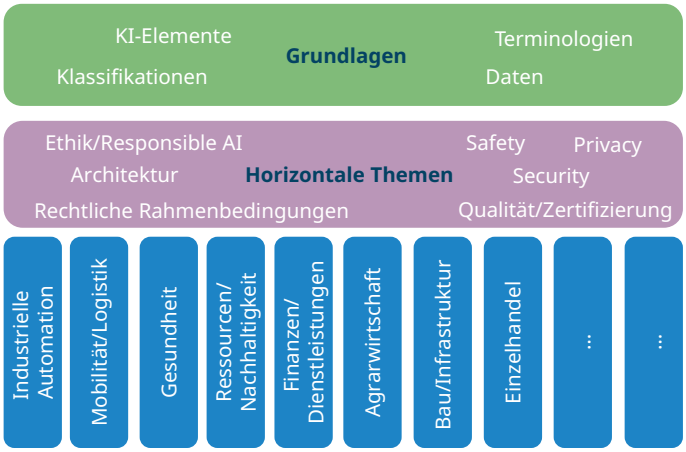


Abbildung 6: Organigramm von Grundlagen, horizontalen Themen und betroffenen Wirtschafts- und Anwendungsbereichen

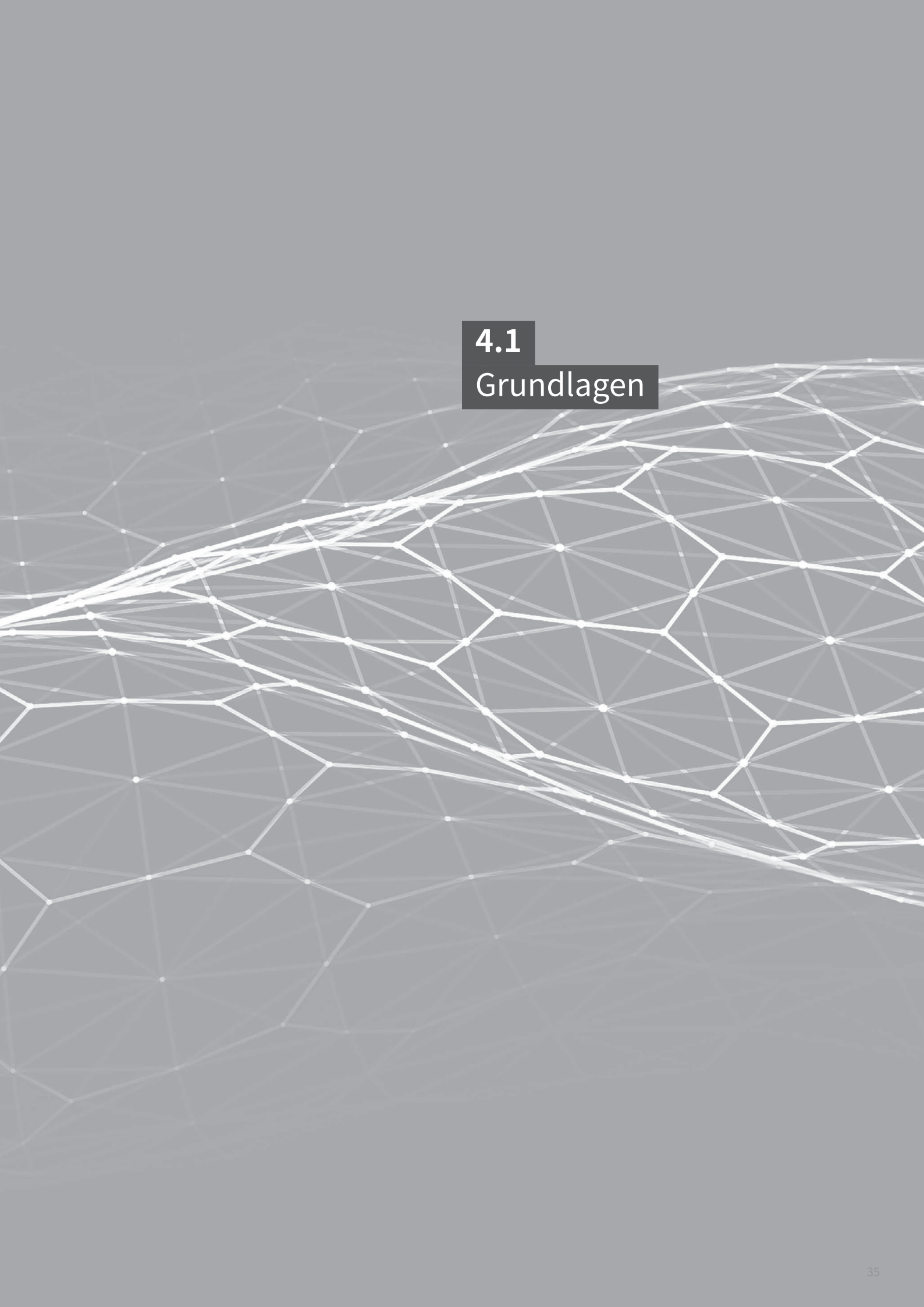
Die Basis für die Diskussion zu KI stellen die Grundlagenthemen dar. Hierzu zählen beispielsweise Terminologien (Begriffsbestimmungen), Klassifikationen, aber auch Themen wie Daten (Datenanalysen, Datenformate, Datengüte etc.).

Durch die neuen technischen Entwicklungen werden insbesondere in der Anwendung von KI neue Fragestellungen zu übergreifenden Themen wie IT-Sicherheit, Qualität, Ethik oder Rechtsrahmen aufgeworfen. Ethische Aspekte zur Verantwortung im Umgang mit KI-Technologien, aber auch Fragen beispielsweise zu Fairness, Sicherheit, sozialer Inklusion und Transparenz von Algorithmen sind hierbei zu berücksichtigen. Zudem müssen die Grundlagen für branchenübergreifende Qualitätskriterien entwickelt werden, die eine Analyse und Zertifizierung von KI-Systemen möglich machen. In welchem Rechtsverhältnis KI zukünftig stehen kann, stellt ein weiteres zu diskutierendes Querschnittsthema dar.

Die wirtschaftlichen Einsatzfelder für KI sind äußerst vielfältig. Nahezu für alle Wirtschaftsbereiche und auch sonstige Anwendungsbereiche außerhalb der Wirtschaft ist KI relevant und findet sich sowohl in Form von Komponenten in Endprodukten und Dienstleistungen als auch in den produktiven Kern- und Unterstützungsprozessen innerhalb der Unternehmen.

Die vorliegende Normungsroadmap KI fokussiert in ihrer ersten Version auf die Bereiche Grundlagen, Horizontale Themen (Ethik, Qualität/Konformitätsbewertungen/Zertifizierung, IT-Sicherheit) sowie die drei Anwendungsfelder Industrielle Automation, Mobilität/Logistik sowie Gesundheit.

In den nachfolgenden Kapiteln 4.1 bis 4.7 werden die Ausgangssituation, Herausforderungen und wesentlichen Normungs- und Standardisierungsbedarfe der sieben Schwerpunktthemen herausgearbeitet.



4.1 Grundlagen

Definitionen des Begriffs „Künstliche Intelligenz“

Die Bereitstellung einer präzisen Definition des Begriffs „Künstliche Intelligenz“ ist aufgrund einer Vielzahl unterschiedlicher Perspektiven und Meinungen zu diesem Thema ein schwieriges Unterfangen:

1. Bezieht sich der Begriff auf eine wissenschaftliche oder technische Disziplin oder ist eine Systemeigenschaft oder -fähigkeit gemeint?
2. Sollte sich der Begriff auf eine Umschreibung der Funktion von KI-Systemen beschränken oder auf ihre Implementierung referenzieren?
3. Sollen Begriffe, die gewöhnlich mit menschlicher Intelligenz assoziiert werden (wie „Wissen“, „Fertigkeiten“) verwendet werden, um KI zu erklären?

Beinahe jede Organisation, die sich mit KI befasst, definiert diesen Begriff in unterschiedlicher Weise. In Anbetracht der Schwierigkeiten, eine allgemein akzeptierte Definition zu finden, soll dies in diesem Dokument nicht geschehen. Kapitel 4.1.2.1 gibt einen Überblick über die verschiedenen Klassen von KI-Methoden und ihre Fähigkeiten und Anwendungsgebiete, der für die folgende Diskussion zur Eingrenzung des Begriffs verwendet werden wird. Allerdings soll die Bandbreite der möglichen Definition der KI anhand der folgenden Beispiele in Tabelle 1 illustriert werden:

Tabelle 1: Verschiedene Definitionen von KI

Beispiel	Deutsch	Englisch	Quelle
1	Künstliche Intelligenz beschreibt Systeme, die intelligentes Verhalten dadurch zeigen, dass sie – mit einem gewissen Grad an Autonomie – ihre Umgebung analysieren und entsprechend agieren, um spezifische Ziele zu erreichen.	Artificial intelligence (AI) refers to systems that display intelligent behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals.	[30]
2	<System> Fähigkeit, sich Wissen anzueignen, zu verarbeiten, zu kreieren und anzuwenden, das in einem Modell gespeichert wird, um eine oder mehrere vorgegebene Aufgaben zu erfüllen <Technische Disziplin> Disziplin zur Entwicklung und Erforschung von KI-Systemen <Künstliche Intelligenz> Informationen zu Objekten, Ereignissen, Konzepten oder Regeln, ihren Beziehungen und Eigenschaften, zur zielorientierten Nutzung organisiert Anmerkung 1 zum Begriff: Information kann in numerischer oder symbolischer Form existieren. Anmerkung 2 zum Begriff: Informationen sind kontextualisierte Daten, die damit interpretierbar werden. Daten werden durch Abstraktion oder durch Messungen der Umgebung kreiert.	<system> capability to acquire, process, create and apply knowledge, held in the form of a model, to conduct one or more given tasks <engineering discipline> discipline of developing and studying AI systems <artificial intelligence> information about objects, events, concepts or rules, their relationships and properties, organized for goal-oriented systematic use Note 1 to entry: Information may exist in numeric or symbolic form. Note 2 to entry: Information is data that has been contextualized, so that it is interpretable. Data is created through abstraction or measurement from the world.	ISO/CD 22989, fortlaufendes Projekt im ISO/IEC JTC 1/SC 42, zurzeit im Status Committee Draft (CD)
3	Das Design und die Konstruktion intelligenter Agenten, die Wahrnehmungen ihrer Umgebung erhalten und deren Handlungen ihre Umgebung beeinflussen.	The designing and building of intelligent agents that receive percepts from the environment and take actions that affect that environment.	[31]
4	Ein KI-System ist ein maschinenbasiertes System, das in der Lage ist, für eine vorgegebene Menge von durch den Menschen definierte Ziele vorhersagen, Empfehlung oder Entscheidungen, die reale oder virtuelle Umgebungen beeinflussen, vorzunehmen. KI-Systeme werden entwickelt, um mit verschiedenen Graden von Autonomie zu operieren.	An AI system is a machine-based system that can, for a given set of human defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.	[32]

Beispiel	Deutsch	Englisch	Quelle
5	Künstliche Intelligenz (KI) ist ein Teilgebiet der Informatik mit dem Ziel, intelligentes Verhalten und die zugrunde liegenden kognitiven Fähigkeiten auf digitalen Computern zu realisieren.	Artificial intelligence (AI) is a branch of computer science with the goal of realizing intelligent behavior and the underlying cognitive abilities on digital computers.	[33]

Autonome Systeme [33] können komplexe Aufgaben in einer bestimmten Anwendungsdomäne trotz variierender Zielvorgaben und Ausgangssituationen selbstständig lösen. Autonome Systeme müssen abhängig vom aktuellen Aufgabenkontext eigenständig einen Handlungsplan generieren, mit dem Gesamtziel, das vom Betreiber des autonomen Systems vorgegeben ist, ohne Fernsteuerung und möglichst ohne Eingriffe und Hilfe menschlicher Operateure im Rahmen der gesetzlichen und ethischen Vorgaben erreicht werden kann. Wenn einzelne Aktionen des autonomen Systems während der Planausführung scheitern, muss das System in der Lage sein, selbstständig eine Planrevision auszuführen, um durch Adaption des ursprünglichen Plans auf anderem Wege die vorgegebene Zielsetzung dennoch zu erreichen. Eine neue Generation von autonomen Systemen ist auch in der Lage, mit anderen autonomen Systemen und/oder einer Gruppe von Menschen gemeinsam eine Aufgabe verteilt zu lösen. Im Rahmen der Selbstregulation muss ein autonomes System auch über explizite Modelle der eigenen Leistungsgrenzen verfügen und bei Vorgaben oder Umgebungsbedingungen, die keine erfolgreiche autonome Zielerreichung erwarten lassen, den Systembetreiber auf diesen Umstand hinweisen (z. B. zu starke Scherwinde verhindern einen Drohnenflug, eine extrem steiler Streckenabschnitt übersteigt die maximale Steigfähigkeit eines autonomen Fahrzeuges).

Im Hightech-Forum der Bundesregierung wurde eine Referenzarchitektur für autonome Systeme erarbeitet (siehe Abbildung 7), deren Rahmen die Sensorik zur Beobachtung der Umgebung und die Aktorik zur Änderung von Umgebungszuständen im Hinblick auf die Zielerreichung des autonomen Systems bildet. Zusätzlich können sich durch die Kommunikation mit der vernetzten Umgebung des Systems und mit kooperierenden Menschen weitere wichtige Informationen für das Verhalten des autonomen Systems ergeben. Prinzipiell besteht das autonome System aus mehreren Modulen zur kognitiven Informationsverarbeitung, die durch verschiedene Mechanismen zur Selbstregulation gesteuert werden, sowie mehreren Wissensbasen, die durch maschinelles Lernen und Schlussfolgern ausgehend von einer initialen Konfiguration ständig adaptiert werden.

Bei den Wissensbasen dient ein episodisches Gedächtnis als Langzeitspeicher für Ereignisse, die das autonome System unmittelbar betroffen haben, um fallbasiertes Schließen und Lernen aus Erfahrung zu ermöglichen. Im Diskursgedächtnis wird der gesamte Verlauf der Kommunikation des Systems mit Menschen und technischen Systemen in der Umgebung gespeichert, um jederzeit Referenzen auf Vorerwähntes und Mehrdeutigkeiten im Kontext auflösen zu können. Eine Planbibliothek speichert erfolgreich ausgeführte Pläne für häufig auftretende Problemklassen, um durch Planrevision ohne Neuplanung Ziele effizienter erreichen zu können und durch Planerkennung aufgrund der Beobachtung von Aktionen anderer Agenten in der Umgebung deren Intention zu erkennen.

Domänenmodelle enthalten vernetzte Modelle aller relevanten Objekte, Relationen, Zustände und Ereignisse in einem Anwendungsfeld, die zur deren Erkennung über die Sensorik oder zur deren Transformation durch die Aktorik des autonomen Systems notwendig sind. In Aufgabenmodellen werden typische Aufgabenklassen für ein autonomes System schematisch erfasst, um eine durch den Systembetreiber neu gestellte Aufgabe rasch verstehen und einordnen zu können oder in eine Reihe bekannter Aufgaben zu dekomponieren. Benutzermodelle sind besonders beim Einsatz autonomer Systeme als Assistenzsysteme im Dienstleistungsbereich entscheidend, da diese u. a. Annahmen über die Präferenzen, Fähigkeiten und den Wissensstand eines Systemnutzers enthalten, die eine Personalisierung der Serviceleistung durch adaptives Verhalten ermöglichen.

Um das Vertrauen in die Nutzung autonomer Systeme zu stärken und das Risiko der Gefährdung von Menschen in der Umgebung bei einem technischem Komplettausfall der zentralen Steuerungsfunktionen zu minimieren, muss es gemäß der Referenzarchitektur eine technische Rückfallebene geben, die das autonome System im Notfall beispielsweise über eine redundante mechatronische Funktion oder eine funkbasierte Fernsteuerung in einen sicheren Betriebszustand versetzt und über Kommunikation mit der Umgebung eine Alarmmeldung generiert.

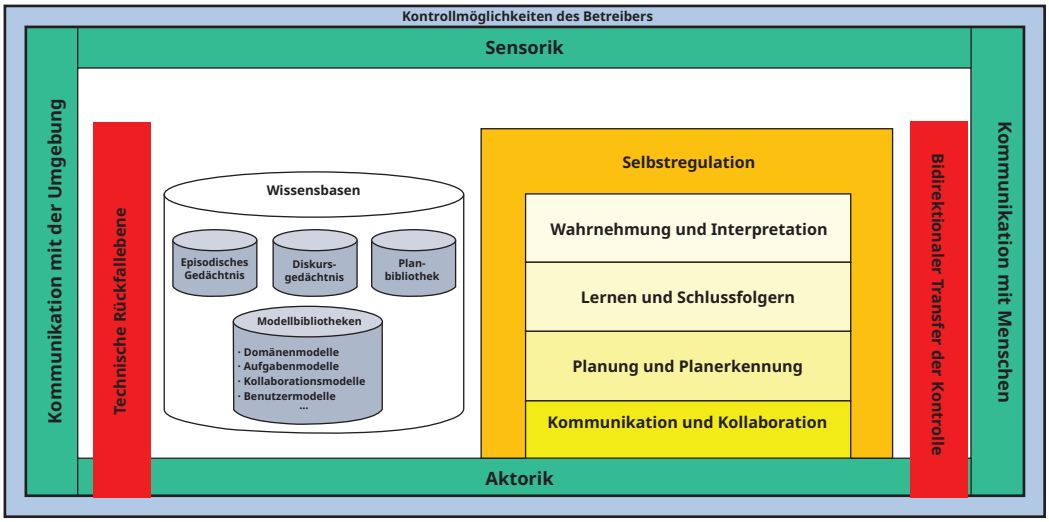


Abbildung 7: Referenzarchitektur für Autonome Systeme [33]

Häufiger wird es vorkommen, dass ein autonomes System die Grenzen seiner Fähigkeiten in anormalen Situationen erreicht und eine Kontrollübergabe an einen Menschen durchführen muss. Ein bidirektionaler Transfer der Kontrolle ist vorzusehen, damit der Mensch nach der Überwindung eines Hindernisses für die Zielerreichung, die für das autonome System allein nicht leistbar ist, die Kontrolle wieder vollständig an das System zurückgeben kann.

4.1.1 Status quo

Mit Blick auf KI-Grundlagen werden beim SC 42 Arbeiten an verschiedenen Dokumenten durchgeführt:

- → **ISO/IEC 22989**, **Artificial intelligence – Concepts and terminology** beschreibt Konzepte und Terminologie der Künstlichen Intelligenz. Diese Norm wird unter Leitung eines deutschen Editors entwickelt.
- → **ISO/IEC 23053**, **Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)** beschreibt ein begriffliches Rahmenwerk für maschinelles Lernen.
- → **ISO/IEC 23894**, **Information Technology – Artificial Intelligence – Risk Management** enthält Richtlinien für das Risikomanagement zur Entwicklung und Nutzung von KI-Systemen. Auch diese Norm wird unter Leitung eines deutschen Editors entwickelt.
- → **ISO/IEC 38507**, **Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations** behandelt organisatorische Governance im Zusammenhang mit KI.
- → **ISO/IEC 20546**, **Information technology – Big data – Overview and vocabulary** [34] behandelt Konzepte und Terminologie für den Bereich Big Data, der ebenfalls im SC 42 betrachtet wird.
- → **ISO/IEC 5059**, **Software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality Model for AI-based systems**

Verschiedene Technische Berichte geben einen Überblick über den augenblicklichen Stand der Technik. Unter anderem sind hier zu nennen:

- → **ISO/IEC TR 24027**, **Information technology – Artificial Intelligence (AI) – Bias in AI systems and AI aided decision making**
- → **ISO/IEC TR 24368,** **Information technology – Artificial intelligence – Overview of ethical and societal concerns**.

Projekte zu den folgenden Themen befinden sich zurzeit unter Abstimmung und werden voraussichtlich im Herbst 2020 die Arbeit aufnehmen:

- → Ein zertifizierbarer Managementstandard für KI, der Anforderungen und Organisationen zur verantwortlichen Entwicklung und Nutzung von KI-Systemen enthält.
- → Verschiedene Projekte zur Beschreibung von Methoden und Prozessen zur Datenqualität im Zusammenhang mit maschinellem Lernen. Eines dieser Projekte wird unter Leitung eines deutschen Editors durchgeführt werden.

4.1.2 Anforderungen, Herausforderungen

Die Beurteilung von KI-Anwendungen in Anbetracht einer angemessenen Eignung kann auf Grundlage von ethischen, rechtlichen sowie technischen Kriterien erfolgen. Angesichts des progressiv wachsenden KI-Marktes ist ein Überblick über Anwendungsszenarien sowie eingebettete Methoden (Kapitel 4.1.2.1) und Fähigkeiten (Kapitel 4.1.2.1.2) von KI unabdingbar. Darüber lassen sich Unzulänglichkeiten bei Entwicklung, Einsatz, einer Konformitätsbewertung sowie der Bestimmung von Qualitätsmerkmalen von KI vermeiden. Während Kapitel 4.1.2.1.3 einen Überblick über Anwendungen mit eingebetteten Methoden und Fähigkeiten von KI innerhalb von Softwaremärkten gibt, wird in Kapitel 4.1.2.1.4 eine Einordnung von KI-Anwendungen auf Grundlage unterschiedlicher Grade von Entschiedungsautonomie vorgestellt. Neben einer Beschreibung von KI durch Methoden, Fähigkeiten und Autonomiegrad können mittels Kritikalität (Kapitel 4.1.2.1.5) Aspekte wie „Recht auf Privatheit“, „Grundrecht auf Leben und körperliche Unversehrtheit“ widergespiegelt werden (siehe Abbildung 8).

In Anbetracht des weiten Fähigkeitsspektrums von KI-Anwendungen spielt das Schädigungspotenzial eine entscheidende Rolle bei der gesellschaftlichen Akzeptanz von KI. Am Beispiel der KI-basierten Erkennung von Verkehrszeichen kann das Schädigungspotenzial anwendungsabhängig variieren: Im Straßenverkehr kann bei selbstfahrenden Kraftfahrzeugen von einem erheblichen Schädigungspotenzial aufgrund der hohen Menge an Besorgnissen und Pflichten ausgegangen werden. Dagegen repräsentiert ein konventioneller Müllwagen mit gleicher KI-Technik zur Verkehrszeichenerkennung kein selbstfahrendes Fahrzeug, sodass von einem geringen Schädigungspotenzial ausgegangen werden kann (siehe Abbildung 9).

Ein einfaches Beispiel für eine KI-Anwendung im Fahrzeug sind Systeme zur videobasierten Verkehrszeichenerkennung. Hier gehört die Erkennung von Geschwindigkeitsbegrenzungen heute schon in vielen Autos zur Serienausstattung. Da viele Verkehrszeichen im Zusammenhang mit der zulässigen Höchstgeschwindigkeit nur eine temporäre Gültigkeit haben (z. B. Baustellen, Schilderbrücken zur dynamischen Verkehrs-

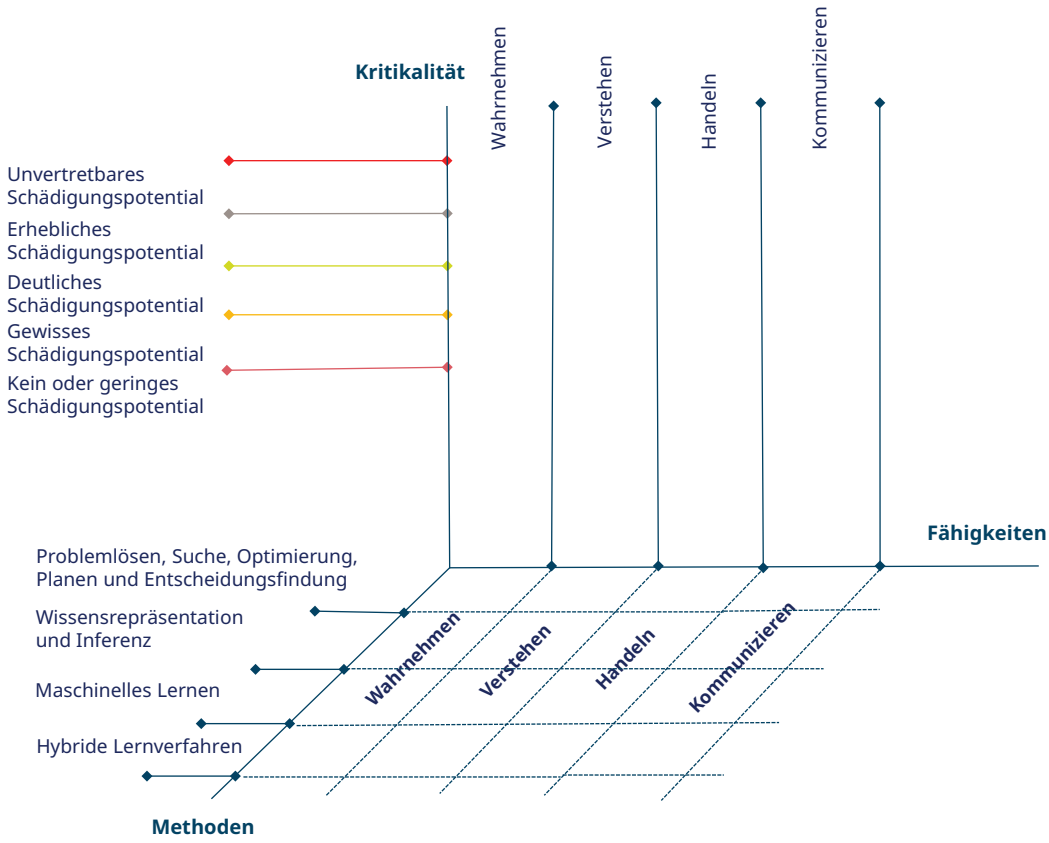


Abbildung 8: Dreidimensionales Klassifizierungsschema zur Bewertung eines KI-basierten Systems

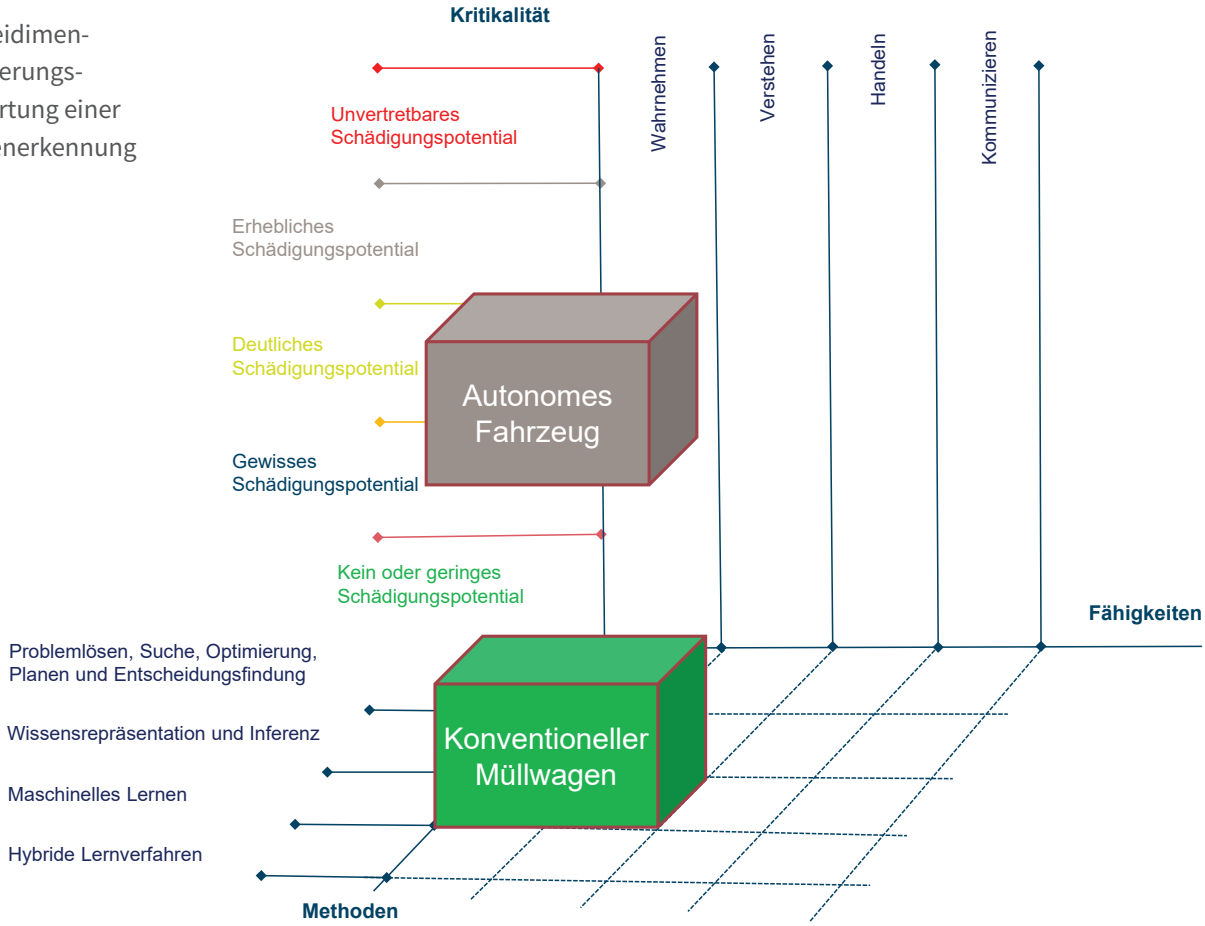


Abbildung 9: Dreidimensionales Klassifizierungsschema zur Bewertung einer KI-Verkehrszeichenerkennung

regelung), können die notwendigen Angaben nicht allein aus digitalen Karten entnommen werden, sondern werden über Mustererkennung über Bilder einer Kamera, zumeist am Innenspiegel, erkannt. Auf diese Weise werden auch erst kürzlich aufgestellte Schilder, beispielsweise an Baustellen, registriert. Damit aber nicht genug: Eine kamerabasierte Verkehrszeichenerkennung wertet nicht nur Daten anhand von Schildern aus. Vielmehr werden diese mit anderen Assistenzsystemen abgeglichen, etwa mit dem Navigationssystem, dem Regensensor und der Uhrzeit, um eingeschränkte Tempolimits korrekt zu interpretieren. Auf dem Markt erhältliche Fahrerassistenzsysteme zur Verkehrszeichenerkennung arbeiten aber nicht 100 Prozent korrekt, sondern ein Test ergab eine Erkennungsrate zwischen 32,5 Prozent beim schlechtesten System und 95 Prozent beim besten System [35] auf einem Parcours mit 40 Schildern zu Tempolimits, den zwölf Pkws durchlaufen haben. Als große Herausforderung erwiesen sich per Klebestreifen temporär ungültig gemachte Temposchilder sowie Tempoanzeigen in Tunneln und Leuchtschilder an Schilderbrücken, aber auch die Verwechslung bei Tempolimits für eine abbiegende Spur (siehe Abbildung 10).

An diesem einfachen Beispiel wird klar, dass Normen und Prüfverfahren für diese relativ einfache Teilaufgabe für das autonome Fahren nach Level 5 notwendig sind, um Konformität des Fahrens mit der Straßenverkehrsordnung sicherzustellen.

Hierzu muss ein standardisierter Trainingsdatensatz für die Verkehrszeichen festgelegt und Benchmark-Tests für eine Zertifizierung bereitgestellt werden. In einem risikoadaptierten Ansatz müssten hier für das autonome Fahren 99,9 Prozent Erkennungsrate erreicht werden, während auch bei reinen Assistenzfunktionen Erkennungsraten unter 80 Prozent erhebliche Risiken bei der Produkthaftung aufweisen.

Abbildung 10: KI-basierte Verkehrsschild-Erkennung

4.1.2.1 Klassifikation

In Anlehnung an das Positionspapier „A definition of AI: Main capabilities and scientific disciplines“ der HLEG-KI [30] wird im Folgenden zwischen Methoden und Fähigkeiten von KI unterschieden. In beiden Fällen orientieren sich die nachfolgenden Klassifizierungen am Standardwerk von Russell und Norvig [31] und integrieren den aktuellen Stand der Technik. Die Matrix in Abbildung 8 bildet ab, welche KI-Methoden eingesetzt werden, um bestimmte KI-Fähigkeiten zu realisieren. Um darüber hinaus auch den Ist-Zustand der aktuellen industriellen KI-Märkte angemessen abzubilden, erfolgt außerdem eine Klassifizierung von KI-Anwendungen, die sich aus den KI-Methoden und KI-Fähigkeiten ergeben. Detaillierte Informationen befinden sich in den Tabellen 2-5 sowie im neu erschienenen Beuth Pocket [36].

4.1.2.1.1 Klassifikation von KI-Methoden

Die Methoden der KI bewegen sich ganz allgemein innerhalb einer Art Spektrum zwischen symbolischen und subsymbolischen – manchmal auch numerisch genannt – Methoden (siehe Tabelle 2). Auf der Seite der symbolischen Methoden stehen insbesondere Techniken der Wissensrepräsentation und des logischen Schließens, während die Seite der subsymbolischen Methoden primär durch Techniken des maschinellen Lernens vertreten wird. Dazwischen sind Methoden des Problemlösens/Optimierens/Planens/Entscheidens sowie hybride Lernverfahren, die sowohl symbolische als auch subsymbolische Techniken verwenden, einzuordnen.

Symbolische KI zeichnet sich insbesondere durch eine deduktive Verfahrensweise aus, d. h. aus dem (algorithmischen) Anwenden logischer Regeln oder Zusammenhänge auf Einzelfälle. Unterschieden wird dabei zwischen Methoden zur Repräsentation von Wissen einerseits sowie Methoden zur Anwendung dieses Wissens andererseits. Wissen kann dabei entweder als sicher oder unsicher repräsentiert werden. In der Wissensanwendung eignen sich die klassischen Methoden des logischen Schließens für sicheres Wissen. Für das Schließen auf Basis von unsicherem Wissen sind probabilistische Ansätze weitverbreitet, es existiert aber auch eine Reihe nicht-probabilistischer Ansätze.

Subsymbolische KI zeichnet sich insbesondere durch eine induktive Verfahrensweise aus, d. h. durch das (algorithmische) Ableiten genereller Regeln oder Zusammenhänge aus Einzelfällen. Hierbei wird zumeist zwischen überwachtem Lernen zum Erreichen eines vorgegebenen Ziels und unüberwachtem Lernen ohne vergleichbare Zielvorgabe unterschieden. Werden beide Ansätze kombiniert, spricht man von teilüberwachtem Lernen. Daneben ist noch das ohne feste Zielparameter auskommende bestärkende Lernen bekannt,

das zwar keinen festen Zielwert, dafür jedoch qualitative Vorgaben (richtig/falsch) benötigt.

Der Methodenkomplex des Problemlösens/Optimierens/Planens/Entscheidens umfasst Algorithmen und Verfahren, die sich auf diese Teilbereiche konzentrieren. Beispiele sind intelligente Agenten, Methoden der Spieltheorie und evolutionäre Algorithmen.

Hybride Verfahren zeichnen sich häufig dadurch aus, dass sie subsymbolische mit anderen KI-Techniken kombinieren, um z. B. sowohl induktiv als auch deduktiv arbeiten zu können. Im Gegensatz zu klassischen subsymbolischen Verfahren wird dabei oft zusätzlich eine Form von Wissensrepräsentation verwendet. Im Unterschied zu klassischen symbolischen Verfahren werden solche Wissensrepräsentationen jedoch häufig in Abhängigkeit von Eingabedaten algorithmisch modifiziert.

Tabelle 2: Klassifikation nach KI-Methoden

KLASSIFIKATION DER METHODEN NACH THEMENBEREICH			Beispiele
PROBLEMLÖSEN, SUCHE, OPTIMIERUNG, PLANEN, ENTSCHEIDUNGSFINDUNG	Problemlösen	Problemlösende Agenten, Problemlösen durch Suche, Suchstrategien	Uninformierte und informierte Suchstrategien
			Adversiale Suche (Spieletheorie)
			Suche mit Rand- und Nebenbedingungen (Constraint-Solving)
	Optimierung	Statistische Optimierungsverfahren	Lokale Suche zur Optimierung
			Suche in stetigen Räumen
			Mit partieller Beobachtung suchen
			Suche in unbekannten Umgebungen
			Dynamisches Programmieren
		Bio-inspirierte Optimierungsverfahren	Evolutionäre Algorithmen
			Genetische Algorithmen/Genetische Programmierung
			Schwarmintelligenz
	Planen und Planerkennung	Autonome und semi-automatische Planungsverfahren	Zustandsraumsuche
			Planungsgraphen
			Hierarchisches Planen
			Planen in nichtdeterministischen Domänen
			Zeit- und Ressourcenplanungsverfahren
			Generierung von Plänen
		Planerkennungsverfahren	Planerkennung durch abduktives Schließen
			Deduktive Planerkennung
			Erkennung durch Planbibliotheken
			Erkennung durch Plansynthese
	Entscheidungsfindung	Ansätze zur Entscheidungsfindung	Modelle
			Nutzen/Wert von Informationen
			Entscheidungsnetze
			Entscheidungstheoretische Expertensysteme
			Sequenzielle Entscheidungsprobleme
			Iterationsmodelle

KLASSIFIKATION DER METHODEN NACH THEMENBEREICH			Beispiele
WISSENSREPRÄSENTATION UND INFERENZ	Repräsentation von Wissen	Wissensrepräsentationssprachen und Modelle	RDF
			RDFS
			OWL
			KIF
			Strukturiertheit und Formalität
		Ontologisches Engineering	Taxonomie
			Ontologie
			Interpretation
			Kalküle
			Deduktion
			Abduktion
			Ontologie Mapping
		Wissensgraphen und Semantische Netzwerke	Wissensnetze/-graphen
			Existenzgraph
			Graphtraversierungsalgorithmen
			Mapping
			Semantic Web
		Modellierung in formaler Logik	Aussagenlogik und Prädikatenlogik
			Logiken höherer Stufen, nicht-monotone Logiken
			Temporal- und Modallogik
	Logisches Schließen	Automatische Beweisverfahren	Resolutions- und Konnektionsbeweiser
			SAT und SMT solver
			Model checking
		Interaktive Beweisverfahren	Taktisches Theorembeweisen
	Unsicheres Wissen	Unsicherheit Quantifizieren	Bayes’sche Regel
		Repräsentation von unsicherem Wissen	Bayes’sches Netz
	Probabilistisches Schließen	Inferenz in Bayes'schen Netzen	Exakte Inferenz
			Annähernde Inferenz
			Markov-Ketten-Simulation
		Relationale Wahrscheinlichkeitsmodelle	Relationale Wahrscheinlichkeitsmodelle in geschlossenen/offenen Universen
		Zeit und Unsicherheit beim probabilistischen Schließen	Hidden-Markov-Modelle
			Kalman-Filter
			Dynamische Bayes’sche Netze
	Nicht-probabilistische Ansätze	Qualitative Ansätze	Schließen mit Defaultinformation
			Begründungsverwaltungssysteme (TMS)
		Regelbasierte Ansätze	Regelbasiertes Schließen mit „Sicherheitsfaktor“
		Schließen mit Vagheit	Fuzzy-Mengen und -Logik
		Schließen mit Glaubensfunktion	Dempster-Shafer-Theorie
	Weitere Ansätze zum unsicheren Schließen		Räumliches Schließen
			Fallbasiertes Schließen (CBR)
			Qualitative Physik
			Psychologisches Schließen

KLASSIFIKATION DER METHODEN NACH THEMENBEREICH			Beispiele
MASCHINELLES LERNEN	Überwachtes Lernen	Neuronale Netze	Multi-layer Perceptron
			Learning Vector Quantisation (LVQ)
			Radiale-Basisfunktionen-Netze (RBF)
			Adaptive Resonance Theory (ART)
			Convolutional Neuronal Networks (CNN)
			Rekurrente Neuronale Netze (RNN)
			Time-Delay-Netze (TDNN)
			Long-Short Term Memory (LSTM)
			Hopfield-Netze
			Boltzmann-Maschine
		Statistisches Lernen	Entscheidungsbäume
			Random Forest
			Support Vector Machine (SVM)
		Probabilistische Verfahren	Naive-Bayes
			Fuzzy Classifier
	Unüberwachtes Lernen	Clustering	kMeans
			Hierarchisches Clustering
			DBSCAN
			Fuzzy Clustering
			Self-Organizing Map
		Dimensionsreduktion	Autoencoder
			Principal Component Analysis
		Probabilistische Verfahren	Fuzzy k-Means
	Teilüberwachtes Lernen	Statistische Ansätze	Erwartungswertmaximierung (EM) mit generativen Mixmodellen
			Transduktive Support-Vector-Maschinen
		Modifizierte Lernstrategien	Self-training
			Co-training
		Graphenbasierte Ansätze	Graphenbasierte Ansätze
	Bestärkendes Lernen	Temporal Difference Learning	Q-Learning
			SARSA
		Monte-Carlo-Methoden	Markov Chain Monte Carlo
		Adaptive dynamische Programmierung	Aktive und passive adaptive dynamische Programmierung
HYBRIDE LERNVERFAHREN	Hybride Neuronale Systeme	Unified Neural Architectures	Constructivist Machine Learning
		Transformation Architectures	Regelextraktion für neuronale Netze, Neuro-Fuzzy-Expertensysteme
		Hybrid Modular Architectures	
	Lernen über Wissensstrukturen	Logisches Lernen	Current-Best-Learning
		Induktive logische Programmierung	Sequential Covering-Algorithmus, konstruktive Induktionsalgorithmen
		Erklärungsbasiertes Lernen	
		Lernen mit Relevanzinformation	
	Konversationales Lernen	Aktives, dialogbasiertes Lernen	Dialogbasiertes überwachtes Lernen
			Dialogbasiertes bestärkendes Lernen

4.1.2.1.2 Klassifikation von KI-Fähigkeiten

KI als wissenschaftliche Disziplin ist von kognitiven Fähigkeiten des Menschen inspiriert [31]. Solche Fähigkeiten werden innerhalb von Didaktik und Pädagogik bereits seit Mitte des vergangenen Jahrhunderts auf Basis sogenannter Lernziele klassifiziert. Das heute am weitesten verbreitete Klassifikationssystem unterscheidet menschliche Fähigkeiten sowohl hinsichtlich sechs Kognitionslevel als auch hinsichtlich vier grundsätzlichen kognitiven Domänen [37], wodurch sich bis zu 24 menschliche kognitive Fähigkeiten unterscheiden lassen.

Vor diesem Hintergrund betrachtet, repräsentieren alle aktuell existierenden KI-basierten Systeme jeweils nur einen Teilbereich des kognitiven Fähigkeitenspektrums des Menschen. Folgt man der Annahme, dass KI-Fähigkeiten menschliche Fähigkeiten nachahmen, können diese grob in die Kernbereiche Wahrnehmen, Verstehen, Handeln und Kommunizieren unterteilt werden (siehe Tabelle 3). Die meisten dieser Fähigkeiten werden durch den Zusammenschluss aus mechatronischen und softwaretechnischen Komponenten realisiert. Die vorgeschlagene Einteilung hilft, die Diskussion zu strukturieren, ist aber nicht trennscharf.

KI-Fähigkeiten aus dem Bereich Wahrnehmen umfassen eine Informationsverarbeitung durch die sensorischen Fähigkeiten Bildverstehen, Geräuschinterpretation, Haptik, Geruchs- und Geschmacksverarbeitung bis hin zu dem komplexen Feld der Erkennung und Interpretation von sozialen Signalen.

Mittels der Fähigkeit Verstehen wird die Informationsverarbeitung mit Blick auf Bewertung, Vorhersage und Entscheidung beschrieben. Das Spektrum umfasst die Unterpunkte Fusion von Wahrnehmungen, episodisches Gedächtnis, Erklärung und Selbstregulation.

Die KI-Fähigkeit Handeln beschreibt insbesondere mechanisch bzw. physisch ausgeführte Tätigkeiten wie Roboterwahrnehmung, Bewegungsplanung, Sensorik und Manipulatoren, Kinematik und Dynamik sowie das Gebiet der Mensch-Roboter-Interaktion, da bei dieser Interaktionsform eine physische Mensch-Maschine-Interaktion im Vordergrund steht.

Unter Kommunizieren ist dabei die Informationsübertragung zur Verarbeitung natürlicher Sprache sowie während der Mensch-Maschine-Interaktion zu verstehen. Die Verarbeitung natürlicher Sprache korrespondiert in der Computerlinguistik mit den Fähigkeiten Textgenerierung, maschinelle Übersetzung, Textanalyse, Informations- und Wissensextraktion, Information Retrieval, Dokumentanalyse und Sprachdialogsysteme. Bei der Mensch-Maschine-Interaktion finden sich kognitive Systeme sowie Interaktionsparadigmen und Modalitäten.

Tabelle 3: Klassifikation nach KI-Fähigkeiten

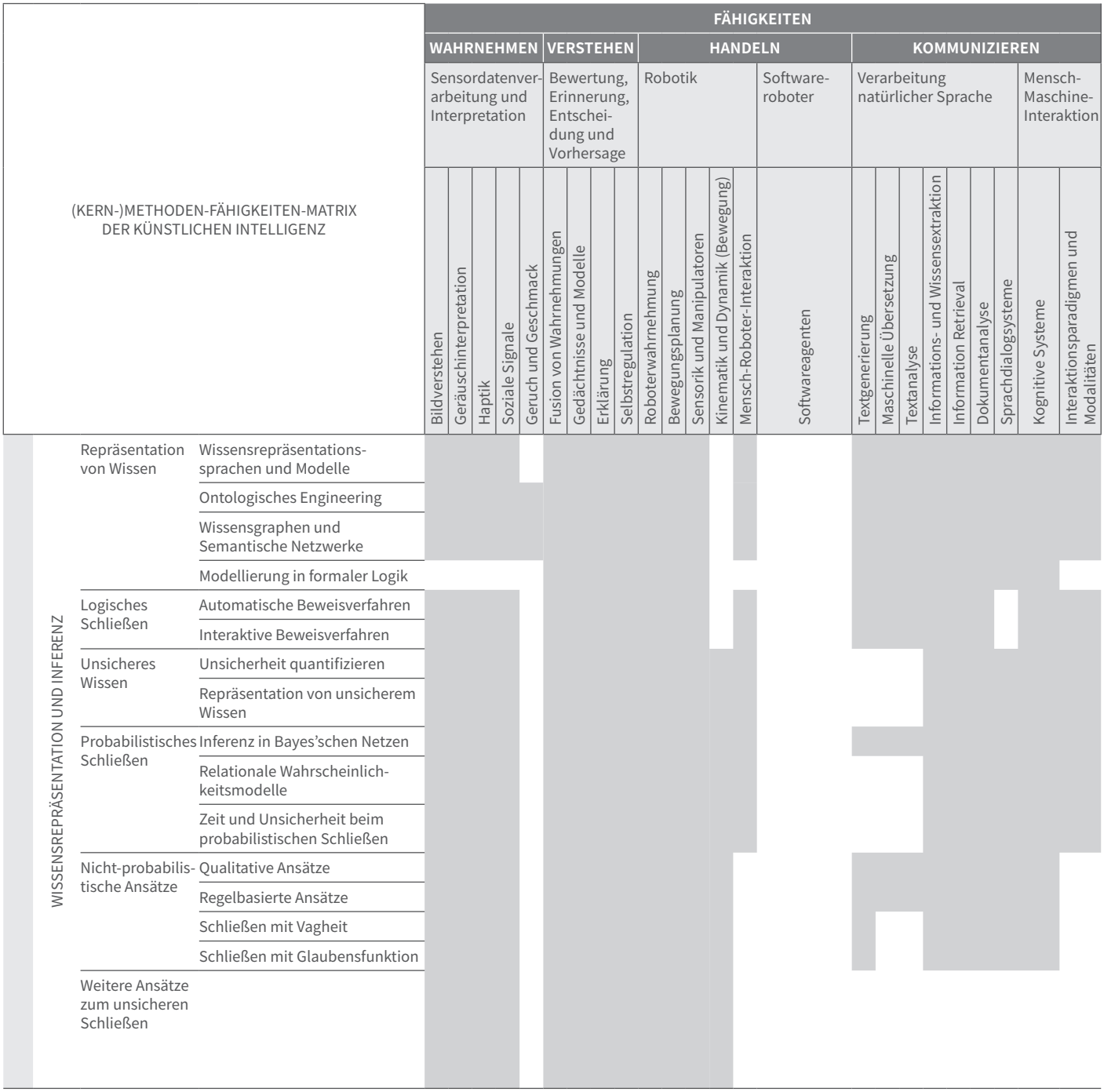
FÄHIGKEITEN VON KÜNSTLICHER INTELLIGENZ			BEISPIELE
WAHRNEHMEN	Sensordaten-verarbeitung- und Interpretation	Bildverstehen	Bildanalyse, Objekterkennung, Videoanalyse, Perceptual Reasoning, Szenenanalyse, Photometrie, physische Attribute, 3D-Modellierung, Simulation, virtuelle Realität
		Geräuschinterpretation	Spracherkennung und -synthese, Geräuscherkennung und -synthese, Anomalieerkennung
		Haptik	Sensornahe Technologien und Methoden der Wahrnehmung zur taktilen Ein-und Ausgabe (Sinneswahrnehmungen wie Struktur, Kitzel, Berührung, Bewegung, Vibration, Temperatur, Druck und Spannung)
		Soziale Signale	Erkennung und Interpretation von Gestik, Mimik, Körperhaltung, Affekt und Stimmung, Emotionen
		Geruch und Geschmack	Sensornahe Technologien und Methoden der Wahrnehmung zur Erkennung und Synthese von Gerüchen, Anomalieerkennung bei Gerüchen und Erkennung vom Geschmack

FÄHIGKEITEN VON KÜNSTLICHER INTELLIGENZ			BEISPIELE
VERSTEHEN	Bewertung, Erinnerung, Entscheidung und Vorhersage	Fusion von Wahrnehmungen	Sensordatenfusion und Interpretation auf der semantischen Ebene, Datenassoziation, Entscheidungsfusion, Statusabschätzung, ML-basierte/modellgestützte/Faktorgraphbasierte/probabilistische Sensordatenfusionsverfahren
		Gedächtnisse und Modelle	Episodisches und semantisches Gedächtnis, Aufgaben- und Prozessmodellierung, Umgebungsmodellierung, Prozessgedächtnis, Diskursgedächtnis, Planbibliothek
		Erklärung	Erklärungsableitung und -generierung, Rationalisierung, hybride Modelle, integrierte Vorhersage- und Erklärungsmodelle, Erklärung durch Architekturmodifikation, Modellagnostische Erklärung
		Selbstregulation	Modellierung eigener Leistungsgrenzen, ressourcenadaptive Handlungsplanung, Methoden zur Selbstoptimierung, dynamische „Weltmodellierung“
HANDELN	Robotik Softwareroboter	Roboterwahrnehmung	Sensornahe Technologien und Methoden der Wahrnehmung in Robotersystemen, 2D und 3D-Wahrnehmungsverfahren, Lokalisierung
		Bewegungsplanung	Methoden zur Planung unsicherer Bewegungen, Steuerungsmethoden
		Sensorik und Manipulatoren	„Passive und aktive Sensoren, Effektoren, Manipulatoren, kooperierende Manipulation“
		Kinematik und Dynamik (Bewegung)	Kinematiksysteme, räumliche Kinematik, Vorwärtskinematik, inverse Kinematik, dynamische Bewegungssysteme
		Mensch-Roboter-Interaktion	Soft-Robotik, Mensch-Roboter-Kollaboration, multimodale Teleoperation
		Softwareagenten	„autonome Softwaresysteme, Prozessautomatisierung, (Chat-)Bots die Transaktionen durchführen, handelnde Assistenzsysteme“
KOMMUNIZIEREN	Verarbeitung natürlicher Sprache	Textgenerierung	Paraphrasing, Markov Textgenerierung, Bedeutung-Text-Modell, Generierung von Zusammenfassungen, Berichte, künstlerische Texte
		Maschinelle Übersetzung	Transfer-, Interlingua-Methode, beispielbasierte, statistische, neuronale und semiautomatische Ansätze
		Textanalyse	Parsing (syntaktische Analyse), Shallow- und Deep-Analyse (semantische Interpretation)
		Informations- und Wissensextraktion	Text und Web Mining, Entitätenextraktion, Disambiguierung, Relationsextraktion, Ereignisextraktion
		Information Retrieval	Vektorraummodell, LSA, pLSA, semantische Suche, Faktensuche, Frage-Antwort-Systeme, Autovervollständigung
		Dokumentanalyse	OCR, ICSR, Dokumentklassifikation, Segmentierung, Bereichserkennung
		Sprachdialogsysteme	Sprechakterkennung, Referenzauflösung, Klärungsdialoge, Diskursmodellierung, Dialogmanagement, Sprechwechselstrategien
	Mensch-Maschine-Interaktion	Kognitive Systeme	Human Factors, Human Processor Modelle, Benutzermodellierung, Kognitionstheorie (Kognition, mentale Modelle, Gedächtnis, Lerntypen, Kognitive Belastung)
		Interaktionsparadigmen und Modalitäten	Interaktionsdesign, Pattern, multimodale Interaktion, User Experience, Fusion- und Fission von Modalitäten

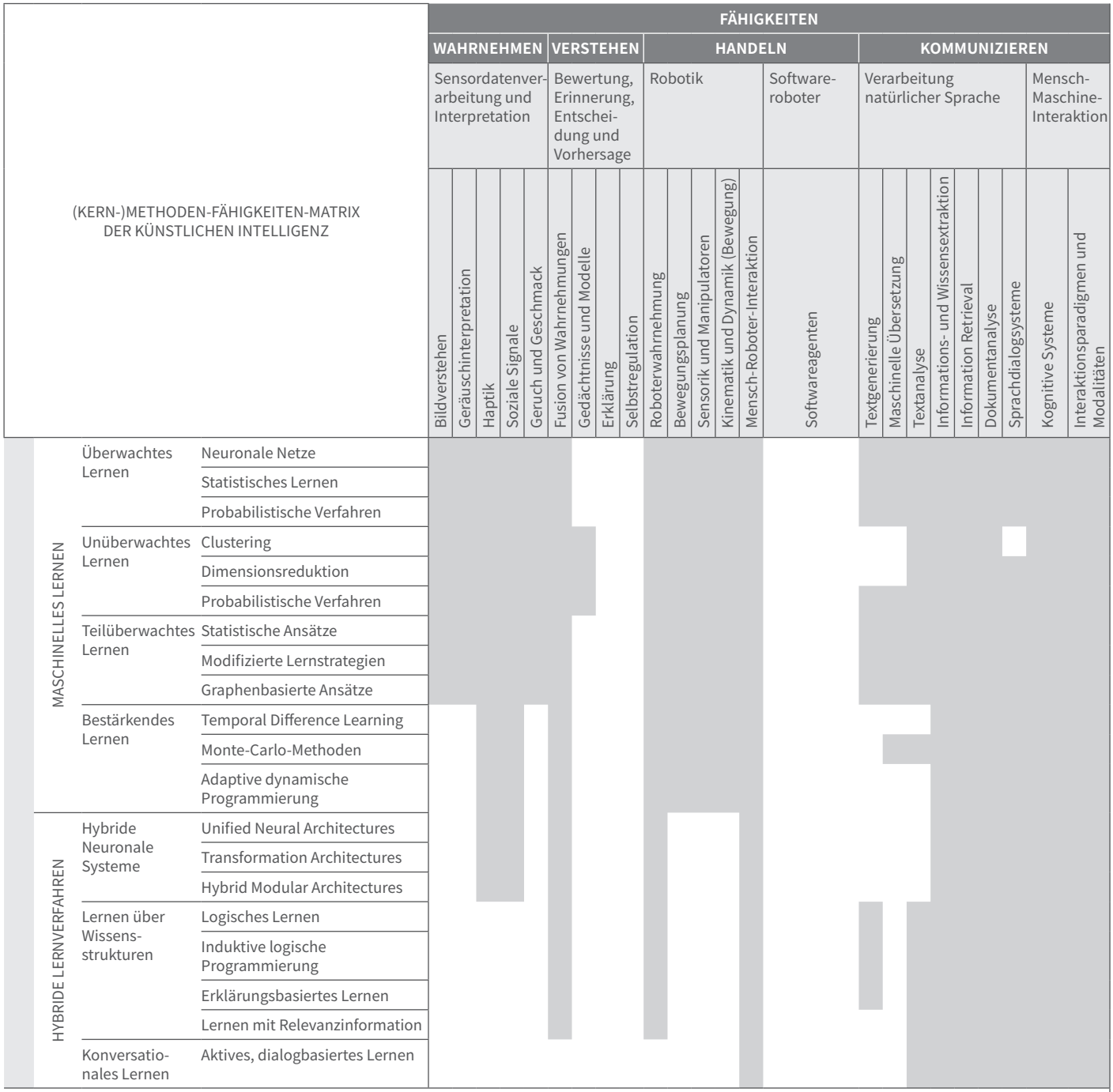
Unter Einbezug der Klassifizierungsmatrix für Methoden und Fähigkeiten (siehe Tabelle 4) kann für KI-Anwendungen eine Kennzeichnungsanforderung für umgesetzte Methoden und Fähigkeiten etabliert werden. Kapitel 4.3 bietet einen Überblick über Anforderungen und Herausforderungen bezüglich der Konformitätsbewertung sowie Qualitätsbeurteilung von Systemen auf KI-Basis.

Tabelle 4: Methoden-Fähigkeiten-Matrix

(KERN-)METHODEN-FÄHIGKEITEN-MATRIX DER KÜNSTLICHEN INTELLIGENZ			FÄHIGKEITEN																								
			WAHRNEHMEN					VERSTEHEN				HANDELN						KOMMUNIZIEREN									
			Sensordatenverarbeitung und Interpretation					Bewertung, Erinnerung, Entscheidung und Vorhersage				Robotik				Softwareroboter		Verarbeitung natürlicher Sprache							Mensch-Maschine-Interaktion	
			Bildverstehen	Geräuschinterpretation	Haptik	Soziale Signale	Geruch und Geschmack	Fusion von Wahrnehmungen	Gedächtnisse und Modelle	Erklärung	Selbstregulation	Roboterwahrnehmung	Bewegungsplanung	Sensorik und Manipulatoren	Kinematik und Dynamik (Bewegung)	Mensch-Roboter-Interaktion	Softwareagenten	Textgenerierung	Maschinelle Übersetzung	Textanalyse	Informations- und Wissensextraktion	Information Retrieval	Dokumentanalyse	Sprachdialogsysteme	Kognitive Systeme	Interaktionsparadigmen und Modalitäten
METHODEN	PROBLEMLÖSEN, SUCHE, OPTIMIERUNG, PLANEN, ENTSCHEIDUNGSFINDUNG	Problemlösen	Problemlösende Agenten, Problemlösen durch Suche, Suchstrategien														Dem Anwendungsfall entsprechend aus den vorigen Spalten zu entnehmen										
		Optimierung	Statistische Optimierungsverfahren																								
			Bio-inspirierte Optimierungsverfahren																								
		Planen und Planerkennung	Autonome und semi-automatische Planungsverfahren																								
			Planerkennungsverfahren																								
		Entscheidungsfindung	Ansätze zur Entscheidungsfindung																								


(KERN-)METHODEN-FÄHIGKEITEN-MATRIX DER KÜNSTLICHEN INTELLIGENZ
FÄHIGKEITEN
WAHRNEHMEN
VERSTEHEN
HANDELN
KOMMUNIZIEREN
Sensordatenverarbeitung und Interpretation
Bewertung, Erinnerung, Entscheidung und Vorhersage
Robotik
Softwareroboter
Verarbeitung natürlicher Sprache
Mensch-Maschine-Interaktion
Bildverstehen
Geräuschinterpretation
Haptik
Soziale Signale
Geruch und Geschmack
Fusion von Wahrnehmungen
Gedächtnisse und Modelle
Erklärung
Selbstregulation
Roboterwahrnehmung
Bewegungsplanung
Sensorik und Manipulatoren
Kinematik und Dynamik (Bewegung)
Mensch-Roboter-Interaktion
Softwareagenten
Textgenerierung
Maschinelle Übersetzung
Textanalyse
Informations- und Wissensextraktion
Information Retrieval
Dokumentanalyse
Sprachdialogsysteme
Kognitive Systeme
Interaktionsparadigmen und Modalitäten
WISSENSREPRÄSENTATION UND INFERENZ
Repräsentation von Wissen
Wissensrepräsentationssprachen und Modelle
Ontologisches Engineering
Wissensgraphen und Semantische Netzwerke
Modellierung in formaler Logik
Logisches Schließen
Automatische Beweisverfahren
Interaktive Beweisverfahren
Unsicheres Wissen
Unsicherheit quantifizieren
Repräsentation von unsicherem Wissen
Probabilistisches Schließen
Inferenz in Bayes'schen Netzen
Relationale Wahrscheinlichkeitsmodelle
Zeit und Unsicherheit beim probabilistischen Schließen
Nicht-probabilistische Ansätze
Qualitative Ansätze
Regelbasierte Ansätze
Schließen mit Vagheit
Schließen mit Glaubensfunktion
Weitere Ansätze zum unsicheren Schließen

(KERN-)METHODEN-FÄHIGKEITEN-MATRIX DER KÜNSTLICHEN INTELLIGENZ



LEGENDE: Methodenklasse kommt zur Erreichung der Fähigkeit vermehrt zum Einsatz Methodenklasse kommt selten oder gar nicht zum Einsatz

4.1.2.1.3 Klassifikation von KI-Anwendungen

Die Klassifikation von KI-Anwendungen orientiert sich häufig an den oben beschriebenen KI-Methoden und KI-Fähigkeiten. Ziel der KI-Anwendung ist es, die mathematischen Methoden und abstrakten Fähigkeiten mittels Software konkret zu implementieren. Auf diese Weise sind spezialisierte Softwaremärkte entstanden, die diese typischen KI-Produkte vermarkten. Diese können von Unternehmen und Anwendern gekauft oder gemietet werden, um die Produktivität der Geschäftsprozesse zu steigern oder Innovationen der Geschäftsmodelle möglich zu machen. Auch sind die typischen Softwaremärkte (siehe Tabelle 5) weltweit einheitlich bezeichnet und werden von unabhängigen Marktanalysten (z. B. IDC, Gartner, Forrester etc.) regelmäßig beobachtet, sodass potenzielle Anwender, Projekte und Investoren sich gut über den Stand der Fähigkeiten informieren können.

Die Softwaremärkte können grob in die Bereiche Business Intelligence & Decision Support, AI based Customer Interaction, AI based Services und AI Development Environment & Tools eingeteilt werden.

Bei Business Intelligence & Decision Support steht das zeit- und themengerechte Erstellen von Reporten im Mittelpunkt. Diese haben das Ziel, einen quantitativen und qualitativen Überblick über das Geschäft zu gewährleisten, und sind schon seit vielen Jahren in allen Bereichen – z. B. Finanz, Human Resources (HR), Entwicklung, Marketing und Vertrieb – kommerziell verfügbar. Auf diese Weise werden Entscheidungen unterstützt und komplette Planungsprozesse in komplexen Umgebungen ermöglicht. Diese Fähigkeiten beinhalten auch Analytics, da sie typischerweise die Analyse vieldimensionaler Datenräume bedingen. Wesentliche Produkte in diesem Bereich sind Software-Umgebungen zur mathematischen und KI-basierten Optimierung sowie Berechnung von Vorhersagen. Ein weiterer Bereich ist die Verarbeitung von Sprache typischerweise zur Suche, Navigation und Exploration in großen Textkörpern. Setzt man mehrere dieser Funktionen zusammen, können ganze Geschäftsprozesse automatisiert werden, was häufig als Robotic Process Automation (RPA) bezeichnet wird.

Seit 2012 hat sich der KI-Trend deutlich beschleunigt, dadurch dass die verfügbaren CPUs und GPUs (Central und Graphics processing units) immer leistungsfähiger werden und KI-Methoden auf der Basis von künstlichen neuronalen Netzen dadurch schneller und kostengünstiger realisiert werden können. Dies erlaubt neue Möglichkeiten für die Mensch-Maschine-Schnittstelle auf der Basis von KI-Anwendungen, welche SMS, Chats, Sprache und physische Bewegungen simulieren und entsprechende Prozesse, z. B. einfache Dialoge in Call- und Servicecentern, automatisieren.

Um die Nutzung von KI-Anwendungen zu vereinfachen, werden typische KI-Anwendungen aus Public- oder Private-Cloud-Umgebungen angeboten. Dies erlaubt es dem Anwender, sofort mit der Anpassung der Anwendung an die eigenen Bedürfnisse anzufangen und nicht erst hohe Aufwände für den Aufbau von Hard- und Software zu haben. Typische KI-Services, welche out-of-the-box angeboten werden, sind: Bilderkennung, Videoanalyse, Sprache-zu-Text-Umwandlung, Text-zu-Sprache-Umwandlung, Übersetzung, Textanalyse, intelligente Suche und maschinelles Lernen. In allen wird die eigentliche Nutzung des künstlichen neuronalen Netzes gekapselt und durch eine einfache grafische Nutzeroberfläche oder durch simple Funktionsaufrufe aus Standardsprachen (z. B. Java, C, Python, etc.) erleichtert.

Für die Entwicklung von KI-Anwendungen braucht man entsprechende KI-Entwicklungsumgebungen und -werkzeuge. Diese tragen den typischen Phasen eines KI-Projekts Rechnung: Build, Train und Run. In allen Phasen kommen häufig Open-Source-Technologien und Software-Bibliotheken zum Einsatz, welche zum einen die KI-Methoden anbieten und zum anderen professionelle Softwareentwicklung, z. B. methodengestützt und in verteilten Teams.

Mittels Regulierung von Systemen auf KI-Basis können mögliche Unzulänglichkeiten von KI-Anwendungen sowie wettbewerbsverzerrende Konstellationen vermieden werden. In Anlehnung an das Weißbuch der Europäischen Kommission „Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“ sind mit Blick auf Regulierung folgende Aspekte von Bedeutung: Haftung, Transparenz und Zuständigkeiten sowie Trainingsdaten, Aufbewahrung von Daten und Aufzeichnungen, vorzulegende Informationen, Robustheit, Genauigkeit, Menschliche Aufsicht und besondere Anforderungen an bestimmte KI-Anwendungen, z. B. Anwendungen für die biometrische Fernidentifikation.

Die ethischen Aspekte der Entwicklung, des Nutzens und der Normung von KI werden aktuell besonders diskutiert. Folgende Eigenschaften spielen hier eine wichtige Rolle, welche methodisch und technisch für jede KI-Anwendung durchdacht und sichergestellt werden sollten: Autonomie & Kontrolle, Transparenz, Stabilität gegenüber Störungen, Sicherheit und alle Fragen des Datenschutzes.

Tabelle 5: Überblick der Softwaremärkte & typische KI-Anwendungen

Softwaremärkte & typische KI-Anwendungen		
Softwaremarkt	**Typische Softwareprodukte**	**Grundsätze**
Business Intelligence & Decision Support Systems	Business Intelligence	Autonomy & Control
	Decision Support	
	Workflow-Systems	
	Planning Analytics	
	Constraint Based Optimization	
	Prediction Capability	Fairness
	Text Processing Platforms & Search Engines	
	Robotic Process Automation (Rule-Based)	
	Cognitive Automation (Training-Based)	
	Real-Time Processing	
AI based Customer Interaction	Chatbots	Transparency
	Voicebots	
	Avatars	
	Virtual & Augmented Reality	
AI based Services consumed from Public- or Private-Cloud	Image Recognition	
	Video Analytics	Robustness
	Speech To Text	
	Text To Speech	
	Translation	
	Deep Learning as a Service	
	Knowledge Navigation	Security
	Knowledge Exploration	
	Intelligent Search	
	Natural Language Processing	
	Automatical Annotation	
AI development environment & tools	Build & Develop AI	Data Governance
	Train & Optimize AI	
	Run & Manage AI	
	Ethic Support Tools	

4.1.2.1.4 Klassifikation von KI-Autonomie

KI-Anwendungen und die Computersysteme, die diese realisieren, können unterschiedliche Grade von Entscheidungsautonomie aufweisen [33]. So unterscheidet beispielsweise die Datenethikkommission der Bundesregierung [10] drei Autonomieklassen: algorithmisch-basierte, algorithmisch-getriebene und algorithmisch-determinierte Systeme.

Algorithmisch-basierte KI-Anwendungen arbeiten als reine Assistenzsysteme ohne autonome Entscheidungsbefugnis. Die von ihnen berechneten (Teil-)Ergebnisse und (Teil-)Informationen sind jedoch Grundlage menschlicher Entscheidungen.

Algorithmisch-getriebene KI-Anwendungen nehmen dem Menschen Teilentscheidungen ab oder prägen durch die von ihnen berechneten Ergebnisse menschliche Entscheidungen. Dadurch schrumpft der tatsächliche Entscheidungsspielraum des Menschen und folglich dessen Selbstbestimmungsmöglichkeiten.

Algorithmisch-determinierte KI-Anwendungen führen selbstständig Entscheidungen herbei und weisen damit ein hohes Maß an Autonomie auf. Durch den hohen Automatisierungsgrad ist im Einzelfall keine menschliche Entscheidung mehr vorgesehen, insbesondere keine menschliche Überprüfung von automatisiert getroffenen Entscheidungen.

4.1.2.1.5 Risikoadaptierte Beurteilung von Anwendungen

Mit Blick auf Vielfalt, Komplexität und Dynamik von Anwendungen sieht die Datenethikkommission eine Notwendigkeit in der risikoadaptierten Beurteilung. Darüber soll zu einer menschenzentrierten sowie werteorientierten Gestaltung und Nutzung von Systemen beigetragen werden. Vor diesem Hintergrund sind auf Basis eines ethisch-rechtlichen Ordnungsrahmens Vorgaben für Transparenz, Erklärbarkeit und Nachvollziehbarkeit vorgesehen. Hierbei soll insbesondere Wert gelegt werden auf die Aspekte Reichweite von Informationsrechten und -pflichten sowie Haftung durch menschliche Entscheidungsträger.

Die Beurteilung ist auf Grundlage einer Kritikalitätspyramide beabsichtigt. Nach der Pyramide soll ein möglicher Schadenseintritt (z. B. menschlich verursacht und/oder algorithmendeterminiert) mit dessen Ausmaß (z. B. „Recht auf Privatheit“, „Grundrecht auf Leben und körperliche Unversehrtheit“ sowie „Diskriminierungsverbot“) für ein sozio-technisches System beurteilt werden. Zur Beurteilung wird die Einbeziehung aller technischen Komponenten (u. a. Hardware, Software und Trainingsdaten), menschlichen Akteure (u. a. Entwickler, Hersteller, Prüfer und Nutzer) sowie Lebenszyklusphasen (u. a. Entwicklung, Implementierung, Konformitätsbewertung und Anwendung) angestrebt. Neben dem Gesetzgeber soll die Kritikalität eines Systems anhand der Pyramide auch von Entwicklern, Prüfern und Anwendern eingeschätzt werden können.

Die Kritikalitätspyramide (siehe Abbildung 11) weist fünf Kritikalitätsstufen (bzw. Kritikalitätsgrade) auf. Mit zunehmender Kritikalitätsstufe wachsen die Anforderungen an ein zu bewertendes, sozio-technisches System. Systeme der Stufe 1: „Anwendungen ohne oder mit geringem Schädigungspotenzial“ werden auf Qualitätsanforderungen überprüft und unterliegen keiner risikoadaptierten Beurteilung (Anwendungsbeispiel: automatische Kaufempfehlung; Anomalie-Erkennung in der industriellen Produktion). Für Systeme der Stufe 2 bis 5 sollte eine Risikofolgenabschätzung vollzogen werden. Zu der Stufe 2 zugehörige Systeme „Anwendungen mit einem gewissen Schädigungspotenzial“ sollten Offenlegungspflichten zu Transparenz haben. Darüber hinaus sind Untersuchungen auf Fehlverhalten, beispielsweise über die Analyse des Ein- und Ausgabeverhaltens, erforderlich (Anwendungsbeispiel: nicht personalisierte, dynamische Preissetzung; automatische Abwicklung der Schadensregulierung). Für Systeme der Stufe 3 „Anwendungen mit regelmäßigem oder deutlichem Schädigungspotenzial“ sollten zusätzlich zu den Maßnahmen der Stufe 2 Zulassungsverfahren eingesetzt werden (Anwendungsbeispiel: automatische Kreditvergabe; vollautomatisierte Logistik). Systeme der Stufe 4 „Anwendungen mit erheblichem Schädigungspotenzial“ sollten zusätzlich zu den Maßnahmen der Stufen 2 und 3 weitere Pflichten zu Kontrolle und Transparenz erfüllen, wie beispielsweise Veröffentlichung von Algorithmen und Berechnungsparametern sowie Schaffung einer Schnittstelle zur direkten Beeinflussung des Systems (Anwendungsbeispiel: KI-basierte Diagnostik in der Medizin; automatisiertes Fahren). Für Systeme der Stufe 5 „Anwendungen mit unvertretbarem Schädigungspotenzial“ soll ein anteiliges oder vollständiges Einsatzverbot gelten (Anwendungsbeispiel: Systeme, die die Unschuldsvermutung außer Kraft setzen, oder Systeme, welche ohne menschliche Einflussnahme billigend letal wirken).

Die Anwendung der Kritikalitätspyramide hat mit Blick auf KI einen weiteren, tiefergehenden Diskussionsbedarf aufgezeigt.

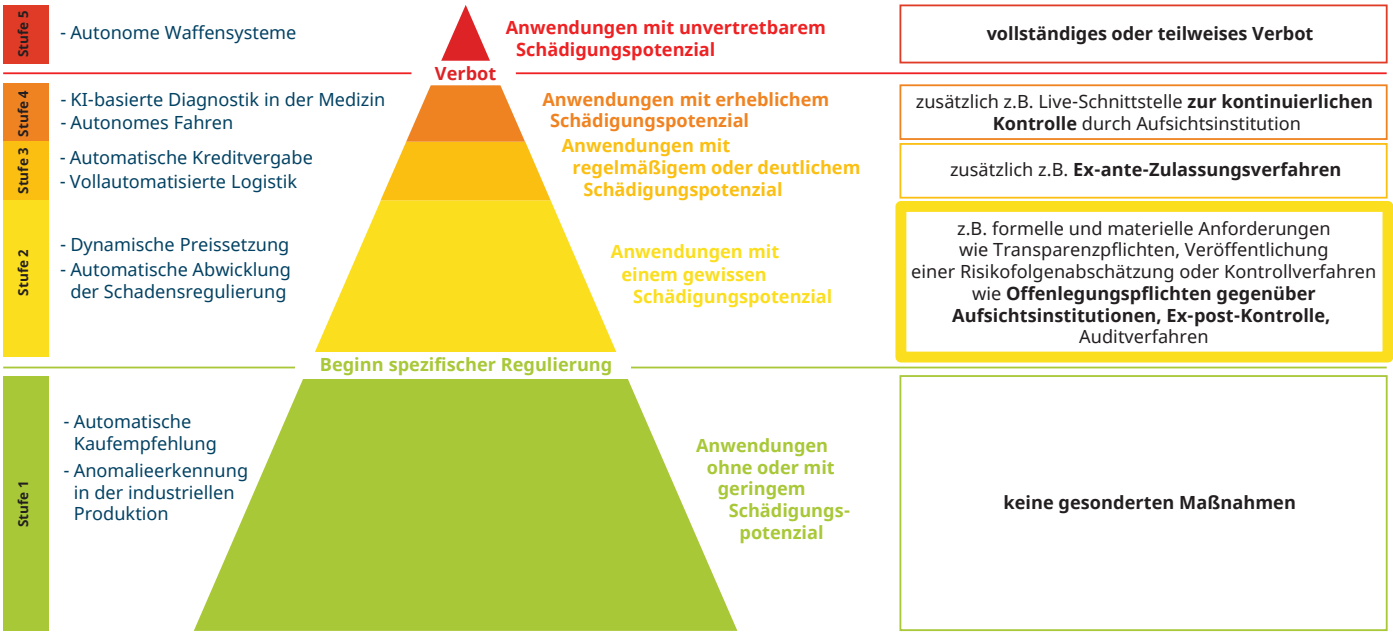


Abbildung 11: Die Kritikalitätspyramide [10] und risiko-adaptiertes Regulierungssystem für den Einsatz algorithmischer Systeme

Im Zuge dessen sollte sich eine Vorgehensweise für die rechtliche Beurteilung sowie ethische Bewertung von KI-Anwendungen herauskristallisieren. Dadurch ließe sich beispielsweise der Anwendungsbereich von Grund- und Haftungsrechten für eine KI-Anwendung festlegen. Darüber hinaus könnte die Aussagekraft der Kritikalitätspyramide über den Einbezug mehrerer zusätzlicher Dimensionen erhöht werden, sodass ein mögliches Schadensausmaß konkreter beschrieben werden kann. Des Weiteren sollte mittels Zertifizierung im Zuge einer Konformitätsbewertung die Erfüllung von Anforderungen mit Blick auf das Schädigungspotenzial von KI-Anwendungen innerhalb der Stufen 1 bis 4 ausgewiesen werden können. Für die Stufe 5 soll die Darlegung einer Konformität verboten werden, da beispielsweise mittels Zertifizierung die Abwendung einer hohen Schadensschwere nicht sichergestellt werden kann. Schlussfolgernd besteht mit Blick auf Regulierung und im Zuge einer Konformitätsbewertung durchgeführten Zertifizierung für Systeme der Stufen 2 bis 4 die größte Vielfalt an Pflichten, Anforderungen, Vorbehalten, Bedenklichkeiten sowie ethischen und rechtlichen Implikationen.

Zur Beurteilung von KI-relevanten Kriterien können standardisierte Konformitätsbewertungsverfahren von akkreditierten Prüflaboren eingesetzt werden, beispielsweise auf Grundlage der ISO/IEC 17000er Normenreihe [38]–[44]. Im Zuge der Konformitätsbewertung können Produkte, Systeme und Prozesse einer Prüfung, Kalibrierung, Inspektion oder Zertifizierung sowie Personen einer Zertifizierung unterzogen werden. Dazu soll der Sachverstand von bereits etablierten, akkreditierten Zertifizierungsstellen mit Blick auf Methoden und Fähigkeiten der KI ausgeweitet werden. Einen Einblick in relevante Aspekte der Konformitätsbewertung mit dem Fokus auf KI bietet Kapitel 4.3.

4.1.2.2 Vertrauenswürdigkeit

Der Begriff „Vertrauenswürdigkeit“ kann sich grundsätzlich sowohl auf Organisationen wie auch auf technische Systeme beziehen. Einem technischen System (d. h. einem Produkt oder einer elektronisch bereitgestellten Dienstleistung) kann bzgl. gewisser Eigenschaften wie Sicherheit oder Zuverlässigkeit vertraut werden, wenn ein Beleg (z. B. in Form eines Prüfberichts oder eines Zertifikats) dafür vorliegt, dass das System solche Eigenschaften erfüllt.[1] Die Vertrauenswürdigkeit einer Organisation ist weiter gefasst: Sie bezieht sich darauf, dass einer Organisation zugetraut wird, geeignete Maßnahmen durchzuführen und Managementstrukturen – ein sogenanntes Managementsystem – zu unterhalten, um

die Erwartungen ihrer Teilhaber und anderer interessierter Parteien zu erfüllen. Neben einem entsprechenden Prüfbericht kann auch die Reputation einer Organisation oder ihre Akzeptanz am Markt zu ihrer Vertrauenswürdigkeit beitragen[17].

Im Rahmen dieses Papiers sollen technische Systeme betrachtet werden, die KI-Funktionen implementieren (man spricht dann von KI-Systemen), bzw. Organisationen, die solche Systeme implementieren, anbieten oder betreiben.

4.1.2.2.1 Anforderungen an Vertrauenswürdigkeit

Die „**Hochrangige Expertengruppe für Künstliche Intelligenz der Europäischen Kommission**“ (HLEG-KI) hat in ihren Ethik-Leitlinien [5] eine Reihe von Anforderungen an KI-Systeme im Hinblick auf ihre Vertrauenswürdigkeit beschrieben. Es handelt sich hierbei in den allermeisten Fällen um hybride Anwendungen, d. h. sie bestehen aus KI-Komponenten und nicht KI-basierter Software und Hardware und werden grundsätzlich als spezielle IT verstanden. In diesem Kapitel sollen diese Anforderungen stellvertretend für eine Anzahl ähnlicher Ansätze verwendet werden, um Standardisierungsbedarfe abzuleiten. Abbildung 12 gibt einen Überblick über die in den Leitlinien genannten Anforderungen, die im Folgenden weiter diskutiert werden:

1. **Vorrang menschlichen Handelns und menschliche Aufsicht**, zudem wird die Einhaltung und Sicherstellung von Grundrechten genannt. Es wird gefordert, dass im Zusammenhang mit KI-Systemen Auskunfts-, Aufsichts- und Kontrollmechanismen zur Verfügung stehen sollen, um negative Auswirkungen z. B. auf Grundrechte, aber auch den Missbrauch von KI-Systemen zu vermeiden. Diese Fragestellungen haben einerseits technische Implikationen, die sich auf die Entwicklung von KI-Systemen beziehen, nämlich die Implementierung von effektiven Monitoring- und Kontrollfunktionen. Die Verwendung solcher Funktionen muss aber in die Managementprozesse der betreibenden Organisation eingebettet werden, damit sie wirksam werden können. Schließlich bezieht sich die Frage nach dem Vorgang menschlichen Handelns und der Kontrolle technischer Systeme durch den Menschen auf die Zielsetzungen, den Auftrag und die Risikobereitschaft einer Organisation, die KI-Systeme betreibt (Governance). Im Zusammenhang z. B. mit öffentlicher Sicherheit werden andere Abwägungen für den Einsatz von KI eine Rolle spielen als beim Einsatz durch ein Wirtschaftsunternehmen. Die HLEG-KI fordert, dass in Bereichen, in denen der Einsatz von KI Grundrechte beeinträchtigen kann, eine Folgeabschätzung durchgeführt wird.
2. **Technische Robustheit und Sicherheit**, z. B. die Widerstandsfähigkeit gegen Angriffe und Sicherheitsverletzungen, Auffangplan und allgemeine Sicherheit, Präzision,

Abbildung 12:
Anforderungen an eine vertrauenswürdige KI
[In Anlehnung an [45]]

17 Letztlich wird hier die Vertrauenswürdigkeit eines technischen Systems auf die Vertrauenswürdigkeit einer Organisation zurückgeführt, nämlich die der prüfenden Stelle. Da sich die Prüfung aber auf das System und nicht auf seinen Hersteller oder Anbieter bezieht, soll diese Unterscheidung zwischen Systemvertrauenswürdigkeit und Organisationsvertrauenswürdigkeit zur besseren Stukturierung der Diskussion beibehalten werden.

Zuverlässigkeit und Reproduzierbarkeit. Aus Sicht der Standardisierung ergeben sich eine ganze Reihe relevanter Fragestellungen:

→ Sind gängige Ansätze zum Management IT- bzw. Cybersicherheit ausreichend für den Einsatz von KI? Welche spezifischen Verletzlichkeiten weisen KI-Systeme auf? Sind neue Kontrollen oder Managementprozesse notwendig?

→ Welchen Beschränkungen muss ein KI-System unterliegen? Wann muss die KI durch klassische Systeme oder durch den Menschen eingeschränkt bzw. überstimmt werden, um Schaden an Personen oder Objekten zu vermeiden?

→ Wie können die Präzision von KI-Systemen und ihre Zuverlässigkeit bemessen bzw. sichergestellt werden? Existieren allgemein akzeptierte Metriken und Maßeinheiten? Welche Rolle spielen Entwicklungs- und Qualitätssicherungsprozesse?

1. **Schutz der Privatsphäre und Datenqualitätsmanagement**, z. B. die Achtung der Privatsphäre, Qualität und Integrität der Daten sowie Datenzugriff. Fragestellungen, die Standardisierungsaktiviäten betreffen, sind Datenschutzmanagement im Zusammenhang mit KI, aber auch, wie Datenqualität insgesamt sichergestellt werden kann. Dies betrifft insbesondere auch den Fall, in dem Daten für maschinelles Lernen durch externe Anbieter zur Verfügung gestellt werden.
2. **Transparenz**, z. B. Nachverfolgbarkeit, Erklärbarkeit und Kommunikation. Die HLEG-KI fordert einerseits, dass Datensätze und Prozesse, die zu der Entscheidung des KI-Systems geführt haben, dokumentiert werden sollen. Andererseits bezieht sich der Begriff „Erklärbarkeit“ auf die Nachvollziehbarkeit der internen Funktion von KI-Systemen (z. B. die Frage, mithilfe welcher Kriterien eine automatische Entscheidung durch ein KI-System getroffen wurde).
3. **Vielfalt, Nichtdiskriminierung und Fairness**, z. B. Vermeidung unfairer Verzerrungen, Zugänglichkeit und universeller Entwurf sowie Beteiligung der Interessenträger.
4. **Gesellschaftliches und ökologisches Wohlergehen**, z. B. Nachhaltigkeit und Umweltschutz, soziale Auswirkungen, Gesellschaft und Demokratie.
5. **Rechenschaftspflicht**, z. B. Nachprüfbarkeit, Minimierung und Meldung von negativen Auswirkungen, Kompromisse und Rechtsbehelfe.

Zusammenfassend lässt sich sagen, dass die Empfehlungen der HLEG-KI eine Reihe von wichtigen Themen aufgreifen. Allerdings kann die Veröffentlichung nicht unmittelbar zur Ableitung von Aufträgen an die Standardisierungsgremien verwendet werden:

1. Standards sind grundsätzlich technischer Natur, d. h., sie beziehen sich auf Anforderungen und Empfehlungen technisch-organisatorischer Art sowie darauf, wie solche im Rahmen einer Organisation angewendet werden können. Gesellschaftliche, rechtliche und politische Anforderungen können nicht in Standards kodifiziert werden, lediglich technisch-organisatorische Implikationen, die sich aus solchen Anforderungen ergeben, können zum Gegenstand eines Standards werden. Nicht alle Themen, die von der HLEG-KI genannt werden, eignen sich dementsprechend schon zur Standardisierung.
2. Die HLEG-KI unterscheidet nicht zwischen dem Vertrauen in das KI-Produkt oder den KI-Dienst (im Sinne eines Produkt oder Dienstes, das oder der KI-Funktionen verwendet) und dem Vertrauen in die Organisation, die einen solchen Dienst bereitstellt oder ein solches Produkt verwendet, herstellt oder vertreibt.
3. Wird Standardisierung auf internationaler Ebene, d. h. im ISO, IEC oder in der ITU als Ziel gesehen, muss von einer ethischen Grundlegung solcher Arbeiten abgesehen werden, sofern diese nicht in der internationalen Staatengemeinschaft allgemein akzeptiert ist. So ist das Vorhaben, ein nicht international anerkanntes Wertegerüst mithilfe einer internationalen Norm zu propagieren, durch die für diese drei Organisationen verbindlichen Prinzipien der Welthandelsorganisation ausgeschlossen [46].

4.1.2.2.2 Vertrauen in Produkte und Dienste

Common Criteria (CC)

Die sogenannten Common Criteria (CC) [47] beschreiben eine Methodik zur Prüfung von Produkten und Diensten mit Fokus auf deren Sicherheit, die als Begriffsgerüst für entsprechende Prüfungen von KI-Systemen verwendet werden können. Die CC liegen ebenfalls als internationaler Standard ISO/IEC 15408 [48]–[50] vor. Unterstützend wird eine abgestimmte Methodik für die Evaluierung auf Grundlage der CC im internationalen Standard ISO/IEC 18045 [51] beschrieben. Diese Dokumente stellen die technische Basis des Common Criteria Recognition Arrangement (CCRA) [52] dar, das von einer Vielzahl von Staaten, so auch von Deutschland, unterzeichnet wurde. Weitere Informationen zu den CC finden sich u. a. auf der Website des BSI [53].

Anforderungen an eine Prüfung nach den CC werden in sogenannten Evaluation Assurance Levels (EALs) zusammengefasst:

EAL1 funktionell getestet
EAL2 strukturell getestet
EAL3 methodisch getestet und überprüft
EAL4 methodisch entwickelt, getestet und durchgesehen
EAL5 semiformal entworfen und getestet
EAL6 semiformal verifizierter Entwurf und getestet
EAL7 formal verifizierter Entwurf und getestet

International ist eine Zertifizierung bis zum EAL4 anerkannt.

4.1.2.2.3 Vertrauen in Organisationen

Das Verhältnis von Governance, Management und technisch-organisatorischen Maßnahmen – Managementsysteme

Zur weiteren Untersuchung der Anforderungen der HLEG-KI zur Vertrauenswürdigkeit von KI soll ein begrifflicher Exkurs zur Unterscheidung der Begriffe „Governance" und „Management" unternommen werden, wie sie im ISO/IEC zurzeit etwa in ISO/IEC 38500 [54] vorgenommen wird (vgl. Abbildung 13). Es ist hierbei zu beachten, dass sich der Begriff „Managementsystem" auf alle drei im Folgenden diskutierten Ebenen, nämlich das Leitungsgremium, das Management und konkrete technisch-organisatorische Maßnahmen, bezieht.

Governance

Governance bezieht sich auf die allgemeinen Aufgaben und Zielsetzung einer Organisation, ihres Selbstverständnisses und auf die sich daraus ergebenden Werte und die Kultur der Organisation, die ihr Handeln bestimmt. Ein zentraler Begriff ist der der Risikobereitschaft. Nach dem Begriffsgerüst der ISO/IEC 38500 [54] ist das Leitungsgremium (**governance body**) einer Organisation verantwortlich für die Umsetzung ihrer Rechenschafts- und Sorgfaltspflichten. Gerade auch Fragen der Haftung bekommen in Verbindung mit KI besondere Relevanz, da durch den möglichen Autonomiegrad der KI die Frage, wer haftet bei Fehlern und Schäden, wichtig ist. Dies sollte die Governance berücksichtigen, da der rechtliche Rahmen sich in diesem Feld dynamisch entwickelt. Das Leitungsgremium leistet hierzu Vorgaben und erstellt Richtlinien, die innerhalb der Organisation umgesetzt werden müssen. Weiterhin ist das Leitungsgremium für die Etablierung von Managementstrukturen (Prozesse, Rollen, Verantwortlichkeiten) und die Bereitstellung adäquater Ressourcen verantwortlich.

Management

Das Management einer Organisation setzt die Vorgaben und Richtlinien des Leitungsgremiums in konkrete Prozesse, Rollen und Verantwortlichkeiten um. Beispiele für Managementaufgaben sind u. a

→ die Identifikation und Analyse potenzieller Risiken und die Etablierung von Handlungsoptionen basierend auf der Risikobereitschaft der Organisation.
→ die Etablierung eines Datenschutzmanagements sowie von Prozessen zur Sicherstellung ausreichender Datenqualität.
→ die Einführung eines Sicherheitsmanagements für KI-basierte IT-Systeme.
→ effektives Management der Entwicklung und des Betriebs von KI-Systemen.

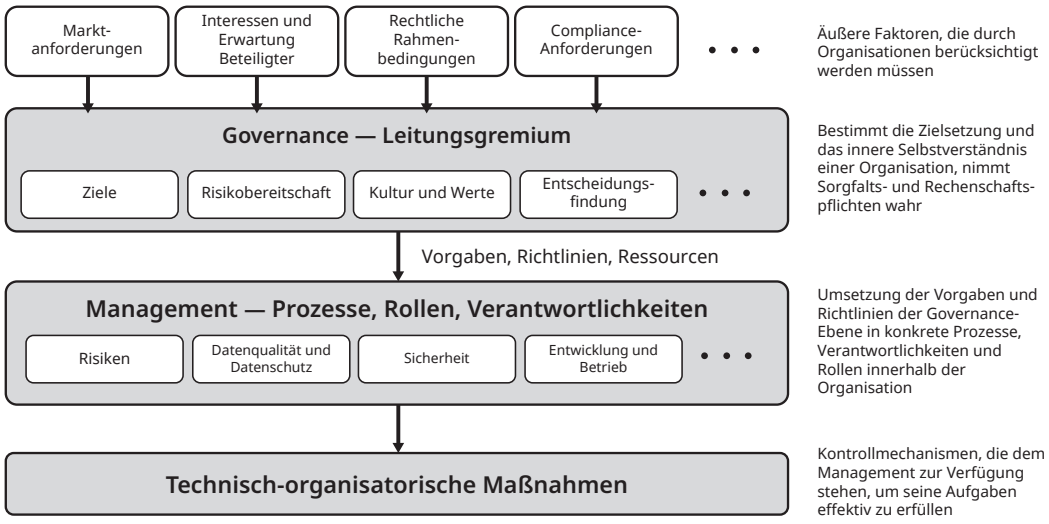


Abbildung 13: Managementsystem: Governance, Management und technisch-organisatorische Maßnahmen

Technisch-organisatorische Maßnahmen
Dieser Begriff umfasst alle technischen und organisatorischen Hilfsmittel, die dem Management zur Verfügung stehen, um seine Aufgaben effektiv und nachprüfbar zu erfüllen. Technisch-organisatorische Maßnahmen reichen von der Verfügbarkeit von Verschlüsselungsfunktionen zur Erhöhung der Datensicherheit über die Anwendung statistischer Methoden zur Identifikation unfairer Verzerrungen bzw. Kontaminierung in Datensätzen bis hin zur Verfügbarkeit von Test- und Validationswerkzeugen.

Anforderungen an das Managementsystem
Im Kontext der internationalen Standardisierung spielt der Begriff des Managementsystemstandards (MSS) eine zentrale Rolle. Ein MSS definiert Anforderungen an Organisationen zur Durchführung eines effektiven und verantwortungsvollen Managements. Zum Teil werden auch Anforderungen an das Leitungsgremium einer Organisation gestellt, und viele MSS enthalten weiterhin konkrete Kontrollen im Sinne von technisch-organisatorischen Maßnahmen. Der Begriff „Managementsystem“ bezieht sich damit auf das Gesamtbild aus Abbildung 13. Mindestanforderungen an das Managementsystem sind in den Richtlinien der Internationalen Standardisierung des ISO/IEC, in der der sogenannten: High Level Structure (HLS) [55] beschrieben:

1. **Kontext der Organisation**, hierzu zählen u. a. rechtliche Rahmenbedingungen, gesellschaftliche Erwartungen, Bedürfnisse und Erwartungen interessierter Parteien, Ziele und Werte der Organisation sowie der eigentliche Geltungsbereich des Managementsystems.
2. **Leitung,** das Leitungsgremium muss verbindliche Bereitschaften der Organisation definieren und in Form von Vorgaben niederlegen. Weiterhin muss es Prozesse, Rollen, Verantwortlichkeiten für ein effektives Management bestimmen.
3. **Planung,** es muss Aktivitäten beschreiben, um mit Risiken und Chancen umzugehen.
4. **Unterstützung,** dies umfasst die Bereitstellung von Ressourcen, die Bestimmung notwendiger Kompetenzen, die Sicherstellung von notwendiger Achtsamkeit, die Kommunikation und Dokumentation.
5. **Betrieb**, das umfasst die operative Umsetzung von Managementanforderungen.
6. **Leistungsevaluation,** diese umfasst das Monitoring, die Analyse und Evaluation, die interne Auditierung und Begutachtung durch das Management.
7. **Verbesserung,** diese befasst sich mit der Identifikation von Nonkonformität bzgl. der Anforderungen des MSS, korrektiven Maßnahmen und der kontinuierlichen Verbesserung des Managementsystems.

Organisationen können Konformität mit MSS nachweisen (z. B. durch eine Selbstbewertung oder Zertifizierung durch eine unabhängige dritte Partei) und damit die Vertrauenswürdigkeit der Organisation bezüglich der spezifischen Aspekte des MSS erhöhen. Betrachtet man den Einsatz einer Klasse von Technologien wie die der KI, muss das Managementsystem einer Organisation deshalb auf die besonderen Charakteristiken und Wirkungsreichweiten der KI Bezug nehmen. Dies kann geschehen, indem existierende MSS um KI-spezifische Anforderungen erweitert werden. Da jedoch die verschiedenen MSS durch unterschiedliche Gremien im ISO und IEC publiziert und gewartet werden, die weder über ein gemeinsames Begriffsgerüst noch über eine synchronisierte Arbeitsweise verfügen, und es darüber hinaus nicht klar ist, ob existierende MSS überhaupt ausreichend sind, um alle Aspekte der KI zu berücksichtigen, ist es erfolgversprechender, einen neuen MSS zu entwerfen, der sich auf KI-spezifische Anforderungen konzentriert.

Unterstützende Standards
MSS umfassen lediglich Anforderungen an ein Managementsystem, beschreiben jedoch nicht seine Implementierung. Dies erlaubt es Organisationen, ihre eigenen Managementstrukturen in der für sie angepassten Weise zu definieren, solange ein Nachweis erfolgen kann, dass die Anforderungen des MSS erfüllt sind. Solche Strukturen, aber auch unterliegende technische und organisatorische Maßnahmen, werden in der Regel in ergänzenden Standards beschrieben, die nun keine Anforderungen, sondern lediglich Richtlinien enthalten.

4.1.2.3 Entwicklung von KI-Systemen

Mit Software erhalten Maschinen einen immer größer werdenden Funktionsumfang. Hardware und Software bilden dabei eine Symbiose und es gibt Verfahren, wie das V-Modell® XT [56], [57] – mit und ohne agile Methoden (z. B. Scrum) –, welche helfen, die Qualität des Gesamtergebnisses bei dessen Entwicklung sicherzustellen. Für Software mit einem vorbestimmten Funktionsablauf gibt es allgemein akzeptierte Entwicklungs- und Qualitätssicherungsverfahren, wie z. B. Code Reading, Modul- und Applikationtests auf verschiedenen Integrationsstufen, Verifikation und Validierung. Diese Methoden und Verfahren wirken auch bei der Software mit regelbasierten KI-Systemen. Neben der Qualität des Softwarecodes und der verwendeten Compiler kommt bei der Entwicklung von KI-Systemen der Softwarearchitektur, der Qualität der verwendeten Daten und der Lernphase eine besondere Bedeutung zu.

Lernende KI-Systeme erhalten wesentliche Funktionalitäten durch die Lernphase. Diese Lernphase kann statisch oder dynamisch erfolgen, überwacht (supervised) oder unbewacht (unsupervised). Wie beim Menschen auch stellt die Prüfung dessen, was gelernt wurde, eine große und für die Softwareentwicklung neue Herausforderung dar. Dieses ist insbesondere dadurch kritisch, dass KI-Systeme besonders dort ihre Stärke zeigen, wo Entscheidungen oder Entscheidungsempfehlungen auf Basis vieler Daten sehr zeitnah getroffen werden sollen.

Werden KI-Systeme zur automatisierten oder autonomen Entscheidungsfindung im sicherheitskritischen Bereich eingesetzt, so werden darauf bezogene Verfahren zur Nachweisführung und Konformitätsbewertung auch durch Dritte erforderlich. Dies gilt insbesondere für Nachweise im Rahmen der Nachweisführung zur funktionalen Sicherheit bei der Produkthaftung.

Ein zweckmäßiger Ansatz bei der Entwicklung von KI-Systemen ist ein risikoadaptiertes[18] Vorgehen unter Betrachtung des gesamten Lebenszyklus eines KI-Systems in dessen Anwendungsumfeld sowie die Sicherstellung der Datenqualität in der Lern- und Anwendungsphase.

Eine weitere Betrachtung müssen KI-Systeme finden, deren Source Code und/oder deren Lerninhalte selber oder durch andere KI-Systeme generiert wurden. Dadurch entwickelt ein bestehendes KI-System ein neues oder verändert dessen Lerninhalte, sodass eine Art Evolution der Maschinen stattfindet.

18 entspricht Englisch „risk based“

4.1.2.3.1 Der Lebenszyklus eines KI-Systems

Analog zur traditionellen Softwareentwicklung bestehen die Lebenszyklusphasen eines KI-Systems aus: **Concept**, **Development**, **Deployment**, **Operations** und **Retirement**, wobei sich insbesondere für Systeme, die auf maschinellem Lernen beruhen, das in verschiedenen Phasen des Lebenszyklus vom Development bis hin zur Operation angewendet werden kann, eine deutlich engere Verzahnung der Phasen ergibt, als es bei klassischen Softwaresystemen der Fall ist.

Im Rahmen der Konzeptphase ist zu definieren, ob die zu erstellende Applikation als regelbasiertes, statisches oder dynamisches KI-Modul angelegt wird und welche Anforderungen sich aus dem Kontext des Anwendungsbereichs sowie der notwendigen Datenqualität ergeben. Für regelbasierte KI-Systeme kann der etablierte Software-Lebenszyklus nach ISO/IEC/IEEE 12207 [58], oder für sicherheitskritische Systeme auch nach ISO 26262 [59]–[70], ISO/IEC 27034 [71]–[78] oder IEC 61508 [79]–[86], Anwendung finden. Für statische und dynamische KI-Systeme ist ein risikoadaptierter Ansatz notwendig.

Darauf basierend ist eine Risikoanalyse durchzuführen, z. B. auf Basis einer FMECA (Failure Mode and Effects and Critically Analysis), welche den gesamten Lebenszyklus des KI-Systems berücksichtigen muss. Im Rahmen der Risikobewertung kann eine erste einfache Klassifizierung, wie in der DIN SPEC 92001-1 [87] vorgestellt, erfolgen (siehe Abbildung 14). Dies kann in Low Risk und High Risk ausreichend sein, jedoch scheint ein feinteiligeres Stufenmodell zielführender, zumal hierbei auch auf Aspekte dynamischer Modelle vertieft eingegangen werden kann.

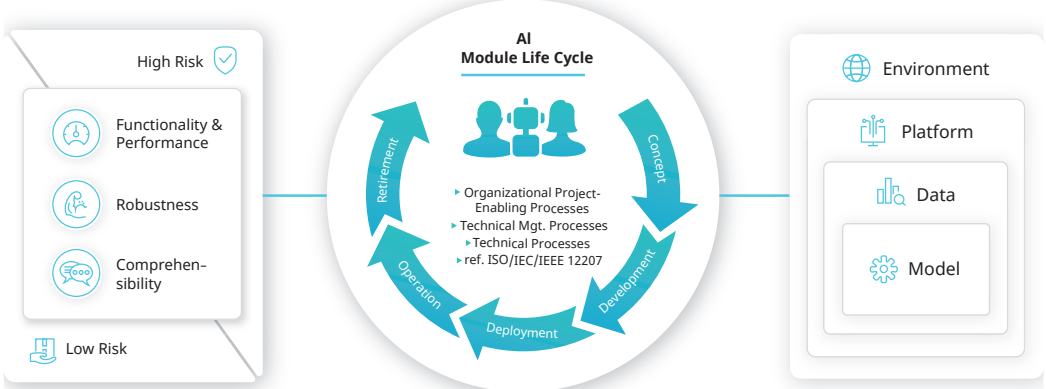
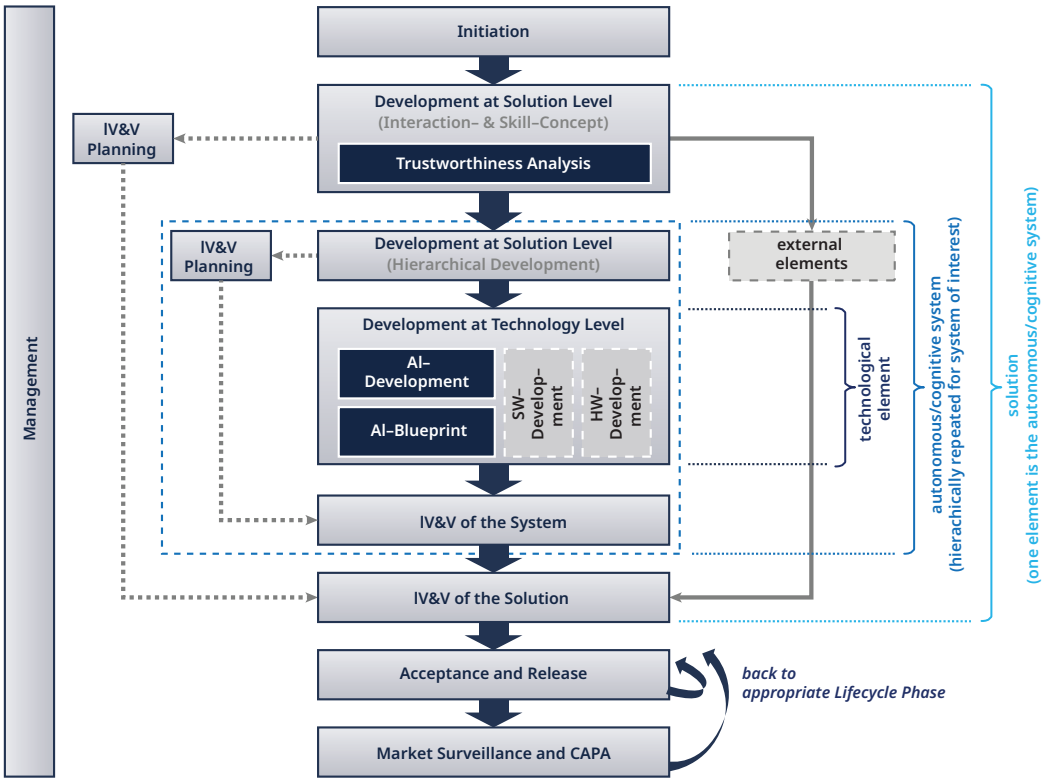


Abbildung 14: AI quality metamodel der DIN SPEC 92001-1 [87]

Abbildung 15: VDE DKE Referenzmodell KI [88]



Alternativ zur zuvor dargestellten DIN SPEC 92001-1 [87] wurde vom VDE DKE ein „Referenzmodell KI“ [88] vorgestellt (siehe Abbildung 15), welches einen Entwicklungsprozess für KI-Systeme beschreibt, in Anlehnung an das V-Modell® XT. Im Rahmen der Normungsroadmap KI ist ein Konsensmodell für den KI-Lebenszyklus zu entwickeln.

4.1.2.3.2 Grundsätze zur Datenqualität für KI-Module

Die Qualität der Daten für das Lernen, Prüfen und die anschließende Anwendung ist ein wesentlicher Faktor für die erfolgreiche Entwicklung und in der Anwendungsphase für die Nutzung von KI-Systemen. Eine allgemeine Definition der Datenqualität im Bereich der Softwareentwicklung wird in der ISO/IEC 25012:2008 [89] beschrieben und besteht aus inhärenten und systemabhängigen Merkmalen. Inwieweit dieser Standard auch für die Entwicklung von KI-Anwendungen geeignet ist oder andere bzw. weitere Qualitätsmerkmale Bedeutung finden, ist zu überprüfen und ggf. für KI-Anwendungen spezifisch zu standardisieren. Es wird sinnvoll sein, für den jeweiligen Anwendungsfall ein Tailoring und/oder eine Priorisierung der Dimensionen durchzuführen. Werden z. B. Simulationsdaten („synthetische Daten“) zum Lernen und/oder Testen verwendet, muss deren Nutzbarkeit/Beispielhaftigkeit sichergestellt sein. Werden zu Lern-, Test- und Prüfzwecken bewusst fehlerhafte Daten bereitgestellt, so sind diese entsprechend zu kennzeichnen und von den nicht fehlerhaften Daten geeignet zu trennen, sodass keine ungewollte Durchmischung erfolgt.

Der Fraunhofer Leitfaden für qualitativ hochwertige Daten und Metadaten (NQDM) von 2019 [90] führt die folgenden Dimensionen der Datenqualität auf:

1. **Aktualität**
   Daten beschreiben die aktuelle Wirklichkeit. Daher wird empfohlen, bei der Erfassung und Benennung der Daten auf einen Zeitstempel und ggf. eine Versionsnummer zu achten. Daten sollen in angemessenen Intervallen auf ihre Aktualität überprüft werden, damit sie repräsentativ sind.
2. **Fehlerfreiheit**
   Die Daten sollen korrekte Werte beinhalten und möglichst fehlerfrei sein. Hierbei ist ein Datum dann fehlerhaft, wenn es nicht mit dessen Klassifizierung übereinstimmt. Somit ist ein fehlerhaftes Datum, das dem KI-System zum Training auch als fehlerhaft mitgeteilt wurde, in diesem Sinne nicht fehlerhaft. Zum Training von KI-Systemen werden bewusst auch fehlerhafte Daten genutzt, diese aber auch als fehlerhaft klassifiziert.

3. **Genauigkeit**
   Je nach Anwendungsfall ist die Genauigkeit der Daten von hoher Relevanz, sodass beispielsweise ein Runden von Werten vermieden werden sollte. Auch inhaltliche Beschreibungen der Daten sollten möglichst präzise sein, um die Relevanz von Daten schnell einschätzen zu können.
4. **Konformität**
   Bei der Bereitstellung von Daten ist auf die Erwartungskonformität der enthaltenen Information in einem bestimmten Nutzungskontext und -format zu achten, beispielsweise bei der Benennung von Attributen und Vokabeln. Für eine möglichst universelle Nutzung der Daten sind nach Möglichkeit entsprechende Standards zu verwenden, wie z. B. die ISO 8601 [91] für Datumsangaben.
5. **Konsistenz**
   Daten sollten widerspruchsfrei sein, sowohl in sich selbst als auch datensatzübergreifend. Diese Dimension wird ggf. bereits durch Fehlerfreiheit abgedeckt.
6. **Transparenz und Vertrauenswürdigkeit**
   Ursprung, Originalität und Veränderungen der Daten sollten nachvollziehbar gemacht werden, damit die Transparenz und Glaubwürdigkeit der Daten gestärkt und somit das Vertrauen der Nutzerinnen und Nutzer gewonnen und auch ethische Anforderungen erfüllt werden können.
7. **Verlässlichkeit**
   Zur Einschätzung der Verlässlichkeit oder auch des Reifegrads einer Information kann ihr ein Status zugewiesen werden (siehe auch DCAT-AP.de).
8. **Verständlichkeit**
   Die Datenstruktur, die Benennung der Daten sowie Datenschnittstellen sollten leicht verständlich sein.
9. **Vollständigkeit**
   Ein Datensatz sollte vollständig sein: Attribute, die zwingend für die Weiternutzung des Datensatzes erforderlich sind, müssen demnach einen Wert enthalten.
10. **Zugänglichkeit und Verfügbarkeit**
    Die Ressourcen sollten leicht zugänglich sein. Dazu gehören eine einfache Auffindbarkeit, langlebige Verlinkungen und Referenzen sowie verständliche Beschreibungen.

Um ein hohes Maß an Datenqualität zu erreichen, ist eine genaue Spezifizierung der Anforderungen an die Daten und Datenschnittstellen notwendig. Ergebnisse der Plattform Lernende Systeme zeigen, dass Datenmanagement als Fundament für lernende Systeme zu sehen ist [92]. Für die Vertrauenswürdigkeit und Nachvollziehbarkeit von Anwendungen sowie für die Beurteilung ihrer Qualität ist ein tiefes Verständnis aller einzelnen Komponenten des Data-Science-Prozesses nötig sowie für den Prozess als Ganzes. Zu den Komponenten des Prozesses gehören: Datenerfassung, Datenbereinigung, Datenintegration, Datenexploration, Datenanalyse, Modellbildung, Datenvisualisierung und Dateninterpretation sowie interaktive Abläufe bzw. Feedback-Schleifen innerhalb der gesamten Prozesskette (z. B. Monitoring, Evaluation).

4.1.3 Normungs- und Standardisierungsbedarfe

BEDARF 1:
Unterstützung der internationalen Standardisierungsarbeiten zu einem MSS für KI
Ein Projekt zur Entwicklung eines MSS für KI wurde kürzlich im ISO/IEC/JTC 1/SC 42 „Artificial Intelligence“ initiiert [Anmerkung: Tatsächlich wird über den Vorschlag zurzeit abgestimmt, ein positives Votum der Nationalvertretungen im SC 42 kann jedoch als sicher angesehen werden]. Da ein solcher Standard letztlich vollumfänglich zertifizierbar ist und damit einen internationalen Standard für Anforderungen und Prozesse für Organisationen, die KI entwickeln oder nutzen, darstellen wird, wird eine Mitwirkung interessierter Kreise in Deutschland an diesem Projekt ausdrücklich empfohlen. Umsetzungsaktivitäten im Rahmen der KI-Normungsroadmap sollten insbesondere die Bereitstellung von Mitteln und Ressourcen für eine solche Mitwirkung in Erwägung ziehen.

BEDARF 2:
Erstellung einer Technologie-Roadmap für KI
Ergänzend zur dem in Kapitel 4.1.2.1 skizzierten Klassifikationsmethodik der KI wird empfohlen, Arbeiten zur Erstellung einer Technologie-Roadmap zu fördern, die augenblickliche Technologietrends in der KI zusammenfasst und Empfehlungen für eine perspektivische Weiterentwicklung des Standorts Deutschlands gibt.

BEDARF 3:
Risikomanagement für KI
Basierend auf der ISO 31000 **Risk management** [93] wird zurzeit im ISO/IEC JTC 1/SC 42 unter der Nummer ISO/IEC 23894 ein Projekt zum Risikomanagement für KI durchgeführt. In seiner augenblicklichen Fassung beschreibt das Dokument Erweiterungen der generischen Richtlinien aus der ISO 31000 für KI-spezifische Aspekte. Das Risikomanagement muss weiterhin durch Richtlinien zur Folgeabschätzung für den Einsatz von KI-Systemen ergänzt werden.

BEDARF 4:
Datenqualitätsmanagement für KI
Datenqualitätsmanagement ist im Zusammenhang mit maschinellem Lernen ein vorrangiges Thema. Eine Reihe von Projekten zum Datenqualitätsmanagement werden zurzeit im ISO/IEC JTC 1/SC 42 initiiert und werden voraussichtlich im Herbst 2020 starten, die von deutscher Seite aus kritisch beobachtet und ggf. durch Beiträge unterstützt werden sollten.

BEDARF 5:
Management von Transparenz und Vermeidung von Diskriminierung
Wie in Kapitel 4.1.2.2 dargestellt wurde, ist die Erklärbarkeit und Nachvollziehbarkeit von KI-Systemen ein weiteres Thema im Zusammenhang mit KI, das Gegenstand der Standardisierung werden sollte. Dies sollte durch die Definition von technischen und organisatorischen Maßnahmen zur Vermeidung von Diskriminierung ergänzt werden.

BEDARF 6:
Designprinzipien für KI-Systeme
Arbeiten zur Definition eines Lebenszyklus-Modells für KI-Systeme werden zurzeit bereits national im Rahmen der DIN SPEC 92001 sowie international im ISO/IEC JTC 1/SC 42 durchgeführt. Diese Aktivitäten sollten harmonisiert und im Rahmen von internationalen Standards weitergeführt werden.

4.2
Ethik/Responsible AI

Ethik ist ein Spezialgebiet der Philosophie und die Grundlage für den verantwortungsvollen Umgang mit Technologie im Allgemeinen und KI im Speziellen. Ein kleiner Exkurs zu den Grundlagen der Philosophie und damit ihrem Spezialgebiet der Ethik in unserem Kulturkreis, den Begriffen des ethischen Dilemmas und der KI-Ethik ist im Anhang 11.2 gegeben.

Bei verantwortungsvoller KI (Responsible AI) geht es um die Schaffung von Rahmenbedingungen für die Bewertung, den Einsatz und die Überwachung der KI, um neue Möglichkeiten für bessere Dienstleistungen für Bürger und Institutionen zu schaffen. Es bedeutet, Lösungen zu entwerfen und umzusetzen, die den Menschen in den Mittelpunkt stellen. Mithilfe des design orientierten Denkens untersuchen Organisationen ethische Kernfragen im Kontext, bewerten die Angemessenheit von Richtlinien und Programmen und schaffen eine Reihe von wertorientierten Anforderungen an KI-Lösungen.

Algorithmische Entscheidungssysteme, insbesondere solche, die ihre Entscheidungsregeln mit Verfahren des maschinellen Lernens aus historischen Trainingsdaten ableiten, entstehen in einer langen Kette, entlang der sich die Verantwortlichkeiten verteilen. In diesem Systembegriff sind also explizit die Menschen und Prozesse mit inbegriffen. Die Verantwortlichkeiten gilt es, im Zuge von Normungsbestrebungen im Bereich Künstlicher Intelligenz in den Fokus der Arbeit zu rücken. Normung kann dabei helfen, die Übergabepunkte in dieser Verantwortungskette transparent zu machen und dadurch eine Modularisierung zu ermöglichen, die Spezialisten im Wettbewerb die besten Lösungen finden lassen.

Es ist zu beachten, dass generierende KIs mit ethisch relevanten Aspekten, wie z. B. die Deep-Fake-Technologie (Imitation von Personen und ihrem Verhalten in Bild, auch Video, und Ton), aus der Betrachtung herausfallen. Sie haben zwar durch ihre Möglichkeiten durchaus weitreichende ethische Fragestellungen, diese beziehen sich aber einzig auf die Anwendung dieser Systeme und weniger auf den Entwicklungs- und Entstehungsprozess, den die Normungsarbeit in den nächsten Jahren zu bewältigen hat.

Allgemein wird festgestellt, dass die ausschließliche Betrachtung der technischen Komponente nicht ausreicht. Der mögliche Einsatz der gleichen technischen Komponente, also des gleichen Entscheidungssystems in verschiedenen Anwendungsfeldern, zeigt deutlich, dass eine Zertifizierung der rein technischen Teile der Komplexität des Problems nicht gerecht werden kann. Eine sozioinformatische Gesamtbetrachtung ist deshalb angebracht, wobei alle sozialen Akteure sowie die Einbettung des Automated-Decision-Making-Systems (ADM-Systems) in dem sozialen Prozess Beachtung finden. Da ADM-Systeme aber stets dem Rechtsrahmen unterliegen und dieser sich aktuell stark ändert und anpasst (siehe Richtlinie 2006/42/EG (Maschinenrichtlinie) [94], Verordnung (EU) 2016/679 (Datenschutz-Grundverordnung DSGVO) etc.) [95], wird der Brückenschlag zwischen Normungsbestreben und Rechtskonformität eine aufwendige Aufgabe der zukünftigen Normungsarbeit sein (beispielsweise Fragen der Gefahrenabwehr und Haftungsthemen). Dabei muss stets berücksichtigt werden, dass die Verletzung von Rechtsgütern bei der Nutzung von KI oft schwer zu erkennen und nachzuweisen ist.

4.2.1 Status quo

In diesem Sinne sind parallel zu immer stärker in alle Lebensbereiche eindringenden KI-Systemen, insbesondere durch die immer elaborierteren Systeme des Machine Learnings (ML), große Initiativen, politische Statements und Expertenkommissionen, die sich mit Ethikgrundsätzen, Werten und Kriterien beschäftigen und diese in Grundsatz- bzw. Positionspapieren festgehalten haben, Berichte sowie Studien entstanden. Diese sind wiederum aufgehängt und initiiert durch politische Strategien der EU sowie der Länder. Gerade im europäischen Kontext wird im Zuge der KI-Entwicklung der Bedarf nach einem souveränen Umgang mit der voranschreitenden Digitalisierung adressiert. Unter dem Stichwort digitale Souveränität wird u. a. die Befähigung der Bürgerinnen und Bürger, aber auch die Gestaltung menschenzentrierter Angebote formuliert.

Die meisten dieser Initiativen haben dabei einen interdisziplinären Blickwinkel, betrachten mehrere Anwendungsgebiete und/oder erläutern das breite Hintergrundwissen der beteiligten Experten. Daraus ergibt sich eine weitgefasste Sammlung an Aussagen, Werten und Kriterien.

Zudem wird das Thema aktiv in der Forschung durch Ausschreibungen von Stiftungen, dem Bund und/oder der EU mittels einer Vielzahl an Projekten vertieft (z. B. vom BMAS, BMBF und dem BMI).

Gerade die Interdisziplinarität erfordert die Bedeutung eines gemeinsam zu entwickelnden Begriffsverständnisses inkl. gemeinsamer Einigung auf verwendetes Vokabular. Dies ist hier umgesetzt im Glossar (siehe Anhang 11.1).

4.2.2 Anforderungen, Herausforderungen

Der aktuelle Diskurs um Ethik und KI wird von zwei Gedanken dominiert, einerseits werden Chancen und Potenziale diskutiert, auf der anderen Seite werden aufgestellte ethische Anforderungen als red-taping verstanden, die Wirtschaft und Gesellschaft ausbremsen und daran hindern, KI-Systeme (im weltweiten Markt) ungehindert/wirtschaftlich einzusetzen. Diese Sorge ist nicht unbegründet, da der Einsatz von KI-basierten ADM-Systemen im Vergleich zu sonstigen Algorithmen einige zusätzliche ethische Herausforderungen mit sich bringt. Hier ist es jedoch durch Normung möglich, die Einhaltung ethischer Mindeststandards sicherzustellen.

Systeme, die bereits im Einsatz sind oder waren, haben gezeigt, dass sich die Folgen des Einsatzes teilweise nur schwer oder gar nicht vorab einschätzen lassen. Dieser Umstand zeigt sich sehr plakativ am Umgang der amerikanischen Bürgerrechtsorganisation ACLU (American Civil Liberties Union) mit dem System COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), mit dem im amerikanischen Justizsystem die Rückfälligkeitswahrscheinlichkeit von Straftätern vorhergesagt wird [96]. 2011 forderte die ACLU den Einsatz von algorithmenbasierten Entscheidungssystemen in allen Phasen des Strafvollzugs in Amerika [97] mit dem Argument einer objektiveren Entscheidung für Angeklagte und Verurteilte. Infolge mehrerer Studien bezüglich mangelnder Fairness bzw. Diskriminierungsfreiheit solcher Systeme schloss sich die ACLU im Jahr 2018 einer gegenteiligen Forderung an und sprach sich für ein Verbot von gelernten ADM-Systemen im gerichtlichen Kontext aus [98].

Gerade aus dem Bereich der Technikfolgenabschätzung wissen wir, dass die Konsequenzen eines Systems erst dann so ausreichend detailliert abschätzbar sind, dass man es entsprechend anpassen kann, wenn die Technologie von einer ausreichend großen Menge an Personen über einen ausreichend langen Zeitraum hinweg benutzt wird. Dann jedoch ist das System schon so etabliert, dass sich die Durchführung grundlegender Änderungen als sehr schwierig, vielleicht sogar unmöglich gestaltet. Diese Problematik ist als Collingridge-Dilemma bekannt [99]. Aus diesem Grund wird der Gesetzgeber durch die Entwicklung neuer Technologien und durch neue Anwendungen stets mit neuen Herausforderungen konfrontiert, denen er sich meist unter hohem Zeitdruck stellen muss. Hier kann Normung unterstützen, indem sie Schnittstellen und Überprüfungsstellen festlegt und beschreibt und somit eine überprüfende Instanz dahingehend unterstützt, dass sie den in diesem Zusammenhang möglichen weitreichenden ethischen Folgen mit ausreichend Flexibilität entgegensteht. Dies bietet der entwickelnden Industrie einen gewissen Grad an Anpassungsfähigkeit an einen sich stetig wandelnden Anforderungskatalog.

Im Kern sind KI-Systeme mathematische Komponenten, die der Optimierung und/oder Klassifikation dienen und durch weitere Programmelemente in für Menschen relevante Entscheidungssituationen oder (Teil-)Automatisierungen eingebettet sind. Dennoch ist in KI-Zusammenhängen, also auch im Rahmen der Beschäftigung mit KI-Ethik, häufig eine „Anthropomorphisierung von KI“ festzustellen [100]. Der KI werden hierbei menschliche Eigenschaften und Verhaltensweisen zugeschrieben – wie Denken, Lernen, Entscheiden etc. Dabei kann die Anthropomorphisierung von KI der Kommunikation über das Thema schaden, da sie technische Tatsachen verschleiern kann und Problembezüge unangemessen emotionalisiert. Aus diesem Grund ist es von größter Notwendigkeit, bei der Beschäftigung mit KI klar die Systemgrenzen herauszuarbeiten und vor allem auch das Verhalten technischer KI-Komponenten von menschlichem Verhalten zu trennen. Erschwerend kommt hinzu, dass manche Systeme im Einsatz weiter lernen, weshalb eine einmalige Prüfung zu einem bestimmten Zeitpunkt für die Lebensdauer eines Systems potenziell nicht ausreichend sein kann.

Im Folgenden werden die im Hauptfokus dieser NRM liegenden Ansätze zur „entscheidungsfindenden KI“ von Ansätzen zur „generierenden KI“, die nicht primär im Fokus des Dokuments liegen, unterschieden. Dabei ist eine scharfe, grundsätzliche Trennung unmöglich; die Praxis zeigt aber, dass Ansätze der Künstlichen Intelligenz zumeist in eine von zwei Kategorien eingeteilt werden können: Ansätze, deren wesentliche Funktion darin besteht, aus Eingabedaten eine abstrahierende bzw. komplexitätsreduzierende „Entscheidung“ oder „Bewertung“ vorzunehmen („entscheidungsfindende KI“), und Ansätze, die grundsätzlich neue Daten erzeugen und dabei entweder gar nicht von Eingabedaten abhängen oder die Komplexität der Eingabedaten durch synthetische Ergänzungen erheblich erhöhen sollen („generierende KI“).

Dabei ist zunächst festzustellen, dass diese Unterscheidung in keinem Zusammenhang zur Komplexität der KI steht: Es gibt einfache Entscheidungsverfahren (z. B. Bayes-Klassifizierer) und einfache Generierungsverfahren (z. B. zelluläre Automaten wie Conways Game of Life), aber ebenso komplexe Entscheidungsverfahren (z. B. Objekterkennung mittels neuronaler Netze) und Generierungsverfahren (z. B. Erzeugung fotorealistischer Gesichter mittels generativer neuro-

naler Netze). Ebenso gibt es Verfahren, die keiner Kategorie zugeordnet werden können, wie etwa Sprachübersetzung, oder Verfahren, bei denen auch auf Basis generierter Daten Entscheidungen getroffen werden können (beispielsweise durch synthetisch erzeugte Phantombilder in polizeilichen Ermittlungen).

Dennoch wurde die Unterscheidung und die Schwerpunktsetzung auf „entscheidungsfindende KI“ für die AG Ethik der Roadmap als zweckdienlich befunden: In „entscheidungsfindenden KIs“ (denen auch die beschriebenen Herausforderungen [97], [98] zuzuordnen sind) entstehen die ethischen Implikationen bereits in der Konzeptions- und Entwicklungsphase der Systeme. Hier kann Normung dazu beitragen, den Entwicklungsprozess zu begleiten und ethische Risiken in der Anwendung zu minimieren. Bei „generierenden KIs“ hingegen tritt in der Regel stark das benannte Collingridge-Dilemma auf: Durch die Komplexitätssteigerung ergeben sich aus einem vergleichsweise einfachen Verfahren unzählige denkbare Anwendungen, anhand derer erst die ethische Bewertung möglich ist. Damit treten für „generierende KIs“ die ethischen Implikationen nicht hauptsächlich im Entwicklungsprozess zutage, sondern erst mittelbar in den jeweiligen, oft schwer abschätzbaren Anwendungen; diese werden deshalb im Rahmen dieses Kapitels zur Ethik nicht in den Fokus gestellt.

4.2.2.1 Ethisch relevante Problemstellungen beim Entstehungsprozess von KI-Systemen

Die verschiedenen am Entwicklungs-, Umsetzungs- und Evaluierungsprozess beteiligten Akteure haben unterschiedliche Anforderungen, wenn es um die Anwendbarkeit von Rahmenwerken für ethische Überlegungen geht. Anbieter von KI-Systemen benötigen Ansätze, die die Umsetzung solcher Prinzipien so einfach wie möglich machen. Dies wird durch den Umstand erschwert, dass große Systeme häufig in abgegrenzten funktionalen Subsystemen entwickelt werden. Die Komplexität, die durch das Zusammenspiel dieser Subsysteme entsteht, erschwert eine Übersicht, inwiefern das Gesamtsystem ethischen Anforderungen gerecht wird. Um diese Komplexität beherrschbar zu machen, erfordert es einen Prozess, der sowohl die Subsysteme als auch das Gesamtsystem evaluiert. Unabhängig von der konkreten Umsetzung muss berücksichtigt werden, dass Systeme, die nach dem Deployment weiter lernen, auf jeden Fall zusätzliche Betreuung und regelmäßige Evaluationen benötigen. Diese Erkenntnis zeigt sich bereits z. B. in der Weiterentwicklung des Industriestandards CRISP-DM (Cross-industry standard process for data mining) zu ASUM-DM (Analytics Solutions Unified Method for Data Mining/Predictive Analytics), welche eine Betreuung nach Deployment berücksichtigt, vgl. [101] oder im Referenzmodell für eine vertrauenswürdige KI der DKE [87]. Sich ändernde Bedingungen im Umfeld der Anwendung können zu unerwarteten Resultaten führen, da die Trainingsphase nicht auf diese Bedingungen ausgerichtet war, weshalb eine Betrachtung des Anwendungskontextes zwingend erforderlich ist. Zudem wäre es zwingend erforderlich, mit jeder Veränderung oder Erweiterung des Anwendungskontextes eine Reevaluation durchzuführen.

In den letzten Jahren erweiterte die Verfügbarkeit von großen Datenmengen gepaart mit den technischen Möglichkeiten die Anwendungsfelder von KI enorm. Welche Aus- und Wechselwirkung ihr Einsatz in gesellschaftlich relevanten Prozessen hat, ist noch kaum erforscht. In verschiedenen Beiträgen wurde auf die Beteiligung von Interessengruppen an der Gestaltung von ADM-Systemen hingewiesen, um deren Akzeptanz sicherzustellen [102]–[104]. Gerade im Hinblick auf Systeme mit weitreichenden ethischen Fragestellungen ist somit eine Integration vieler möglichst verschiedener Stakeholder am und um den Entwicklungsprozess wünschenswert. Hierbei müssen die Belange der „interessierten Parteien“ [105] ausreichend berücksichtigt werden. Um die Komplexität der daraus resultierenden Entscheidungen und Entscheidungsvorbereitungen zu erfassen und darauf aufbauend Handlungsempfehlungen aussprechen zu können, ist es notwendig, ihre Entwicklung bis hin zur Integration in den sozialen Prozess zu betrachten, hier bildet sich eine lange Kette der Verantwortlichkeiten (siehe Abbildung 16).

Algorithmendesign und Implementierung: Die Entwicklung und Implementierung von ADM-Systemen ist äußerst aufwendig und von vielen Designentscheidungen durchsetzt. Deshalb greift man häufig auf von Firmen oder Gemeinschaften von Programmierern bereitgestellte Softwarepakete zurück, welche fertige Implementierungen wichtiger Bausteine bereitstellen. Dabei können Designentscheidungen wie z. B. die Wahl einiger Hyperparameter[19] oft nicht nachvollziehbar oder gar nicht erst einsehbar sein. Eine Prüfung, ob das verwendete Softwarepaket für den eigenen Anwendungskontext geeignet ist, findet oft nicht statt.

19 Parameter die vor dem eigentlichen Training einer KI festgelegt werden, wie z. B. die Struktur eines künstlichen neuronalen Netzes

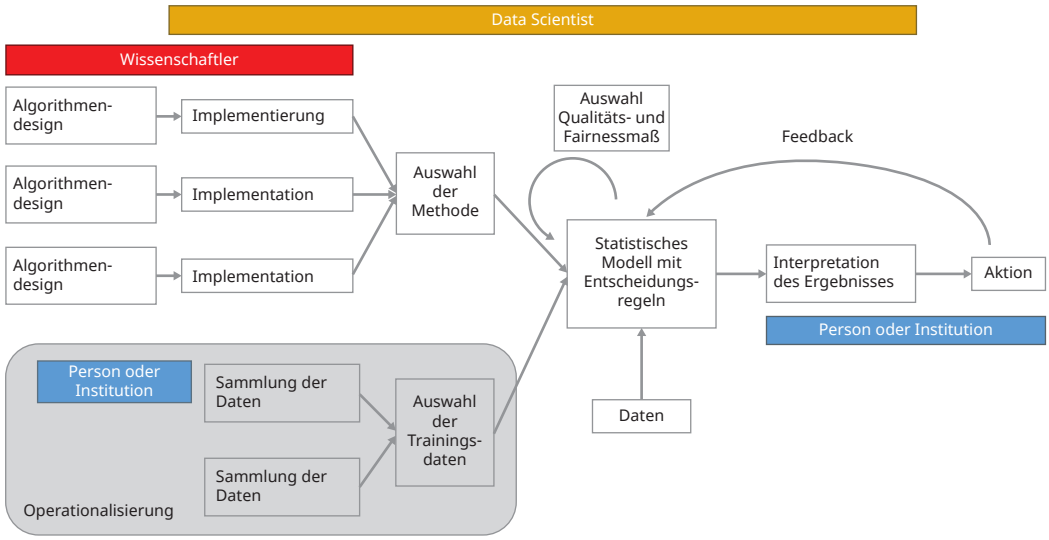


Abbildung 16: Lange Kette der Verantwortlichkeit (nach [106])

Auswahl der Methode: Es gibt eine Vielzahl von (Teil-)Methoden des maschinellen Lernens, die sich zu großen Teilen miteinander kombinieren lassen; so besteht ein ADM-System mindestens aus zwei großen Komponenten, einer auf Basis der Trainingsdaten lernenden und einer auf Basis des von der ersten Komponente erstellten Modells eine Entscheidung fällenden Komponente. Die Unterschiede wirken sich dabei nicht nur auf funktionale Aspekte (Trainingsdauer, Fehleranfälligkeit, …), sondern auch auf nicht-funktionale Anforderungen aus, die im ethischen Kontext besonders relevant sind (z. B. Erklär- und Nachvollziehbarkeit). Jede Methode bringt Modellannahmen mit, die durch die Daten, das Setting und die Trainingsart sichergestellt werden müssen, damit die Methode zielführend arbeiten und etwaige Zusicherungen an die erreichte Qualität halten kann. Data Scientists fehlt in der Regel die notwendige Ausbildung, den Anwendungskontext zum Erkennen potenzieller Risiken beim Einsatz der jeweiligen Technologie ohne explizite Hinweise in Betracht zu ziehen. Dieser Bereich sollte in den Normungsbestrebungen der nächsten Jahre eine wichtige Rolle spielen.

Datensammlung und -auswahl: Werden Daten aus externen Quellen bezogen, wie z. B. von staatlichen, wirtschaftlichen oder wissenschaftlichen Institutionen oder von Datenhändlern, besteht unter Umständen die Gefahr, dass diese durch ein mangelhaftes Sammelverfahren oder eine fehlerhafte Aufbereitung verfälscht sind oder werden. Der Bereich der Datensammlung über die Datenverarbeitung bis hin zur Datenspeicherung muss im Kontext von Künstlicher Intelligenz ausführlich betrachtet werden. Dabei können rechtliche Rahmenbedingungen wie beispielsweise die DSGVO einen groben Rahmen bieten, aber dieser muss durch Normungsarbeit weiter geklärt und präzisiert bzw. für den Anwendungskontext angepasst gestaltet werden. Dabei ist ferner zu berücksichtigen, dass auch das Zusammenspiel und Verwobensein von maschinen- und -personenbezogenen Daten zu neuen Herausforderungen führen wird. Eine Normung sollte unter Beachtung des jeweiligen Anwendungsfelds eine gegebene Zweckbindung der Daten berücksichtigen, welche sicherstellt, dass eine Dokumentation vorhanden ist, zu welchem Zweck die Daten gesammelt wurden. Eine Regulierung ermöglicht die Etablierung von Bedingungen, unter welchen die Daten einem anderen Zweck zugeführt werden dürfen oder nicht. Daten- und Methodenauswahl folgen in keiner festgelegten Reihenfolge, vielmehr werden verschiedene Daten und Methoden in Kombination getestet, sofern der dafür notwendige Aufwand vertretbar ist. Ziel ist die Operationalisierung einer abstrakten Größe, die messbar gemacht werden soll (z. B. Kreditwürdigkeit oder Relevanz einer Nachricht). Die Ergebnisse werden insbesondere durch die Datengrundlage (insbesondere Auswahl und Qualität) massiv geprägt.

Konstruktion des Entscheidungssystems: Bei der Konstruktion von Entscheidungssystemen werden Trainingsdaten und Methoden des maschinellen Lernens zusammengebracht. Ob die Vorhersagefähigkeit des Entscheidungssystems den gegebenen Anforderungen genügt, wird auf Grundlage eines selbst gewählten Qualitätskriteriums ermittelt. Die Wahl des Qualitätskriteriums (es können über zwei Dutzend infrage kommen) liegt oft ebenfalls bei Data Scientists und ist daher höchst subjektiv. Die Auswahl des Qualitätskriteriums kann weitreichende Folgen haben, z. B. wenn ein ungeeignetes Maß gewählt wurde, weshalb dieser Prozess von einer Normung profitieren würde. Das System wird so lange optimiert

(es gibt eine Vielzahl an Parameter, die festgelegt werden können) und/oder trainiert, bis die Anforderungen an das gewählte Qualitätskriterium (z. B.: Unterschreiten einer bestimmten Fehlerquote) erfüllt sind.

Einbettung in den gesellschaftlichen Prozess: Bei der Einbettung des Entscheidungssystems in den Anwendungskontext wird festgelegt, wie Ergebnisse interpretiert werden können und wie mit ihnen umzugehen ist. Die Anwender werden in irgendeiner Form eingewiesen, wie das System zu verwenden ist, und können Ausgaben auf Basis eigener Eingaben erhalten, wodurch eine Qualitätskontrolle im Einsatz des Systems ermöglicht wird. Häufig ist ein Data Scientist mit der Kontrolle der Qualität des Systems im konkreten Einsatz betraut, z. B. wenn es um die grafische Aufbereitung geht. Er kann damit, zusammen mit den Anwendern, einen Beitrag zur Interpretation der Ergebnisse leisten, auf deren Basis entschieden wird, wie mit den Ergebnissen zu verfahren ist. Ein automatisches Entscheidungssystem kann basierend auf den Ergebnissen selbstständig Aktionen auswählen und in Gang setzen.

Re-Evaluation: Wenn der Entwicklungs- und Integrationsprozess abgeschlossen ist, wird das Gesamtsystem in der Regel noch einmal evaluiert, entweder vom Data Scientist oder von den Anwendern. Je nach Evaluationsergebnis können beliebige Teilkomponenten des technischen Systems noch einmal geändert oder sogar ganz ausgetauscht werden (vgl. Rückkopplungspfeile in Abbildung 16). So sind Fälle in Amerika bekannt, in denen eine KI im Rahmen des Rekrutierungsprozesses einer Firma die Distanz zum Firmensitz als relevanten Indikator für die Kündigungswahrscheinlichkeit identifiziert hat. Da damit Bewerber, die sich keine Wohnung im teuren Umfeld des Firmensitzes leisten können, impliziert diskriminiert wurden, haben die Entwickler das Kriterium aus dem Entscheidungsprozess ausgeschlossen [107].

Zu erwähnen ist, dass korrekt durchgeführte agile Entwicklungsprozesse die Probleme im Allgemeinen reduzieren könnten, da jeder Entwickler potenziell auch als Data Scientist fungiert, (abwechselnd) jede Rolle in der Kette der Verantwortlichkeiten übernehmen kann und deutlich mehr Kommunikation zwischen den Entwicklern stattfindet, wodurch sich auch einzelne Expertisen zu konkreten Aspekten besser verbreiten können. Dennoch handelt es sich nur um eine Reduktion des Problems. Je nach System können mehrere 100 Personen an der Entwicklung beteiligt sein (z. B. automatisiertes Fahren, hier entsteht das Gesamtsystem aus einer Vielzahl von Modulen die teilweise unabhängig voneinander entwickelt werden, in ihrem Verhalten bezüglich des Gesamtsystems jedoch einander beeinflussen). Die dadurch entstehende Komplexität kann von agilen Entwicklungsmethoden nicht mehr kompensiert werden. Darüber hinaus befindet sich die Einbettung in den sozialen Prozess außerhalb der Reichweite der Softwareentwickler, weshalb diese nur begrenzt Berücksichtigung finden kann.

Diese Beispiele zeigen, dass es wichtig ist, das Produkt auch in seinem Einsatz, dem sogenannten Feldeinsatz, zu beobachten. Wie schon in den Grundlagen beschrieben, verbleiben trotz Evaluierung bei lernenden KI-Systemen Ungewissheiten über das Verhalten des KI-Systems nach Auslieferung, das sogenannte Restrisiko, aufgrund unzureichender Qualitätsmethoden zur vollständigen Verifizierung des Gelernten. Die Produktbeobachtung im Feldeinsatz beinhaltet auch die Interaktion mit fremden Produkten, welche mit dem KI-System in Verbindung gebracht werden können. Ziel ist, bis dato unerkannt gebliebenes, ungewolltes und unethisches Verhalten während der Anwendung systematisch zu identifizieren und korrektiv gegenzusteuern, um unnötige Gefährdungen oder unethisches Verhalten einzugrenzen oder besser zu verhindern. Eine solche Produktbeobachtung mit Rückkopplung, als Teil des KI Life Cycle, wird in Kapitel 4.3.2.3.2.4 näher beschrieben.

Verantwortungs-/Haftungsübernahme: In der Entwicklung von Anwendungen mit höherer Komplexität ist die frühe Beteiligung nicht-technischer Disziplinen in die Arbeitsschritte der Data Scientists und dabei die Berücksichtigung der ethischen Perspektive sowie der interessierten Kreise sehr ratsam. So können Risiken, fehlerhafte oder systematisch verzerrte Urteile und unerwünschte Effekte frühzeitig erkannt und minimiert werden. In Kontext der Verantwortungsübernahme sollten bereits bei der Projektierung mehrere evaluative Prozessschritte vorgesehen werden. Ethisch gute KI kann nachhaltig und großflächig vermarktet werden. Eine systematisch ethische Reflexion bereits in die Investitionskosten einzuplanen lohnt sich mit zunehmender Komplexität der KI. Das komplexe Zusammenspiel vieler Akteure wirft die Frage auf, wer im Falle eines entstandenen Schadens die Verantwortung trägt. Das kann in einer langen Kette der Verantwortlichkeiten der Data-Scientist, der Verwender des Systems (z. B. derjenige, der ADM-Systeme in seine Produkte und Dienstleistungen integriert) oder der Betroffene (z. B. derjenige, der die Produkte und Dienstleistungen nutzt oder mit ihnen in Berührung kommt) sein. Wann die Verantwortlichkeit eines Akteurs endet und die eines anderen beginnen kann, wird als problematisch angesehen. Hier könnten „Übergabepunkte“, wie man sie an anderer Stelle unseres

Rechtssystems kennt, für mehr Transparenz sorgen. Normung könnte aufgrund der Sachnähe zu den jeweiligen Anwendungen bei der Festlegung und Ausgestaltung solcher „Übergabepunkte“ einen wichtigen Beitrag leisten. Ein dabei nutzbarer Unterschied in der Perspektive besteht zwischen dem Recht, das Fragen der Haftung und Schadensregulierung im engeren Sinne und innerhalb strenger Grenzen klärt, und der Ethik mit dem Begriff der Verantwortung, der über Haftungsfragen von einzelnen Beteiligten weit hinausgeht und am anderen Punkt des Prozesses ansetzt. Mittels einer Klärung der Verantwortlichkeiten für die Qualität der KI im Sinne der Einhaltung der nutzerrelevanten ethischen Werte können weit im Vorfeld einer Haftungsfrage oder von wirtschaftlichen Einbußen Risiken erkannt und minimiert werden. Normen können eine frühe Beteiligung interessierter bzw. relevanter Kreise fordern und auch Prozessketten mit Verantwortungspunkten definieren und damit potenziell eine präventive Aufgabe zur Vermeidung bzw. Verringerung potenzieller Risiken und Schäden leisten. Mit der Sicherung hoher ethischer Qualität kann Normung zudem ein breitflächig hohes Marketingpotenzial für „gute Produkte“ anbieten und Märkte mit anderem Rechtsrahmen zu erschließen helfen.

4.2.2.2 Fairness und Freiheit von Diskriminierung

Obwohl automatisierte Klassifikations- oder Entscheidungssysteme den Anspruch erheben können, menschliche Vorurteile und Unzulänglichkeiten durch Kalkulation zu vermeiden, sind sie ihrerseits nicht davor gefeit, Diskriminierung und Ungerechtigkeit zu produzieren („systematische Diskriminierungen“, vgl. [108]). Dies liegt neben Hardwareproblemen sowie dem Zusammenspiel von Hard- und Softwaren in erster Linie an der technischen Logik der Systeme [109]: Intelligente Maschinen lernen aus verfügbaren Trainingsdaten, die ihrerseits ein Abbild des bisherigen Verhaltens von Personen und Institutionen sowie der Art und Weise der Erhebung sind. Die Trainingsdaten reflektieren dabei auch die Fehler oder subjektiven Einschätzungen von Menschen und führen mitunter zur Wahrnehmung der Systeme als „unfair“ oder „diskriminierend“. Dabei ist die Wahrnehmung selbst eine subjektive, die weder durch den Gesetzgeber noch durch Normung verändert werden kann. Allerdings – und hier kann Normung durchaus einen wichtigen Beitrag leisten – können die Erhebungs- und Verarbeitungsprozesse, also die Entscheidungsprozesse bis hin zum ADM-System, d. h. die Verfahrensregeln, „fair“ und „diskriminierungsfrei“ ausgestaltet werden (z. B. [108]). Genannt werden dabei zumeist fünf Kriterien, nach denen die Fairness von Verfahrensregeln ausgestaltet sein und auch bewertet werden sollte: (1) Konsistenz, (2) Neutralität, (3) Genauigkeit, (4) Revidierbarkeit und (5) Repräsentativität (u. a. [110]).

- → Konsistenz: Die Entscheidungsregeln sollten konsequent angewendet werden, unabhängig vom Entscheidungsträger, von den Betroffenen sowie vom Zeitpunkt der Entscheidung.
- → Neutralität: Die persönlichen (Prozess-)Präferenzen der Entscheidungsträger sollten eine Entscheidung nicht beeinflussen können. Neutralität bezieht sich somit auf die vermeintliche Unvoreingenommenheit automatisierter Entscheidungssysteme.
- → Genauigkeit: Faire Entscheidungen sollten auf möglichst vollständigen und korrekten Informationen beruhen. Damit ist im Falle von automatisierten Entscheidungen die Zuverlässigkeit und Gültigkeit des Dateninputs angesprochen.
- → Revidierbarkeit: Im fairen Entscheidungsprozess ist sichergestellt, dass fehlerhafte oder unangemessene Entscheidungen rückgängig gemacht werden können.
- → Repräsentativität bedeutet, dass für verschiedene Identitäten, Kulturen, Ethnien und Sprachen, die Gegenstand des Verfahrens sind, aussagekräftige Daten verfügbar sind und berücksichtigt werden.

Normung kann hier wichtige Impulse liefern, etwa durch Vorgaben zu Entscheidungsabläufen (Ablaufdiagramme) für den Erhebungs- und Verarbeitungsprozess, Verfahrensabläufe, die beim Revidieren von Fehlern zu durchlaufen sind, Ausgestaltung etwaiger Produktbeobachtungspflichten o. Ä.

4.2.2.3 Wertekanon

In einem Wertekanon, der in der Entwicklung eines maschinellen Systems beachtet werden soll, stehen die Werte nicht beziehungslos nebeneinander, sondern er enthält in der Regel komplementäre und konkurrierende Werte und auch solche, die relativ unabhängig voneinander stehen. Hierbei ist es zwingend notwendig, den konkreten Anwendungskontext des ADM-Systems zu berücksichtigen, da oft auch offene ethische Fragen im gesellschaftlichen Bereich tangiert werden. Insbesondere für die miteinander konkurrierenden Werte lässt sich feststellen:

Das grundsätzliche Bewerten von Risikosituationen führt in der Praxis häufig zu einer Quantifizierung des Risikos unter

ökonomischen Gesichtspunkten. Im konkreten Anwendungsfall ist kein Wertekanon frei von einer auch ökonomischen Bewertung. Dies zeigt sich mit besonderer Deutlichkeit in der (aktuellen) Corona-Krise. Obwohl unstreitig sein sollte, dass das Leben von Menschen das höchste Gut ist, an dem sich andere Werte messen lassen, ist doch zu beobachten, dass an vielen Stellen und in unterschiedlichen Ländern eine Diskussion geführt wird, inwieweit der ökonomische Schaden durch die Aufrechterhaltung von Quarantänebestimmungen nicht etwa größer ist als der Schaden durch Aufhebung der Quarantänebestimmungen. Was in der Konsequenz nichts anderes ist, als Menschenleben gegen ökonomischen Erfolg aufzurechnen. Es wird ein Risiko gegen ein anderes abgewägt.

4.2.2.3.1 Hinführung

Ethische Leitlinien für algorithmische Entscheidungssysteme werden in unterschiedlichen Kontexten diskutiert (z. B. [5], [32], [111], [112]). So gibt es etwa ethische Richtlinien für die statistische Praxis der American Statistical Association, der „Code of ethics and professional conduct“ der Association for Computing Machinery (ACM) und die ethischen Richtlinien der Gesellschaft für Informatik (eine Diskussion der hier genannten Rahmenwerke findet sich in Garzcaraek und Steuer [113]. Eine umfassende Darstellung der unterschiedlichen AI Guidelines findet sich bei Fjeld et al. [114], Jobin et al. [115] oder Hagendorff [116]).

Auch das Gutachten der Datenethikkommission [10] ist hierbei zu nennen, welches in Abschnitt B allgemeine ethische und rechtliche Grundsätze behandelt (Menschenwürde, Selbstbestimmung, Privatheit, Sicherheit, Demokratie, Gerechtigkeit und Solidarität sowie Nachhaltigkeit). Diese werden dort jedoch in einen allgemeineren Rahmen und in Bezug zu Ethik und Recht gesetzt.

Ebenfalls zu erwähnen ist das Ethik-Briefing [117], ein Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI von Expertinnen und Experten der Plattform Lernende Systeme. Dieser enthält Empfehlungen für einen ethisch reflektierten Entwicklungs- und Anwendungsprozess von KI-Systemen. Diese Empfehlungen lassen sich in drei grundlegende Werte (Selbstbestimmung, Gerechtigkeit und Schutz der Privatheit und der Persönlichkeit) sowie weiterführende Prinzipien (Förderung der Autonomie, Verantwortungswahrnehmung, Gleichheit, Diskriminierungsfreiheit, Diversität und Vielfalt, fairer Zugang zu den Vorteilen von KI, Nachhaltigkeit, Privatsphäre als Rückzug aus der Öffentlichkeit, Anonymität als Schutz der Privatheit, informationelle Selbstbestimmung und Integrität der persönlichen Identität) und notwendige Voraussetzungen zur Realisierung untergliedern. Diesen Überlegungen vorgelagert sind die drei Basisannahmen Vermeidung von Schäden, Rechtskonformität und technische Robustheit [117]. Darüber hinaus werden in einem weiteren White Paper Kriterien für eine gelingende Mensch-Maschine-Interaktion vorgestellt. Diese lassen sich in vier Cluster unterteilen: Schutz des Einzelnen, Vertrauenswürdigkeit, sinnvolle Arbeitsteilung und förderliche Arbeitsbedingungen [118].

4.2.2.3.2 Werte

Es gibt unterschiedliche Ansätze, einen Wertekanon zur Auswahl von Handlungsoptionen oder für die Bewertung und damit auch für die Bemessung des Risikos von ADM-Systemen festzulegen. Implizit läuft es jedoch immer darauf hinaus, zum einen die (Ziel-)Werte eines Wertekanons, zum zweiten, in welcher Beziehung diese Werte zueinander stehen (Beziehungsrelationen), und zum dritten, wie sich Werte und Beziehungsrelationen für das jeweilige Einsatzgebiet des ADM-Systems ändern, festzulegen. Beispiele für einen bestimmten Ansatz sind die im bereits erwähnten Gutachten der Datenethikkommission [10] genannten Grundwerte, das Ethik-Briefing der Plattform Lernende Systeme [117] und die Kriterien für eine gelingende Mensch-Maschine-Interaktion der Plattform Lernende Systeme [118] sowie diejenigen aus dem White Paper „Ethikaspekte in Normung und Standardisierung für KI”.

Die Festlegung auf einen konkreten Wertekanon kann im gegebenen Anwendungskontext schwierig werden. Daher ist eine methodisch systematische Abwägung der Einzelwerte und von deren (Beziehungs-)Zusammenhängen unabdingbar (vgl. Operationalisierung von Werten). So können in einem spezifischen Anwendungskontext einzelne der festgelegten Werte nur marginal zum Tragen kommen oder durch bestehende Gesetze bereits hinreichend berücksichtigt sein. In jedem Fall ist es zielführend, operationalisierbare Werte zu definieren, anhand derer eine konkrete Messung des jeweiligen Gefährdungspotenzials möglich ist, das ein konkretes ADM-System für den Menschen im jeweiligen Kontext darstellt. Die Festlegung von operationalisierbaren Werten begünstigt zudem deren Vertretbarkeit im internationalen Kontext, selbst wenn sich konkrete Wertvorstellungen unterscheiden. Eine differenzierte Berücksichtigung von Werten in Abhängigkeit des konkreten Anwendungsfeldes

(z. B. Medizin, Mobilität, ...) kann dabei zielführend sein. Die betrachteten Werte bestehen dabei häufig nicht unabhängig voneinander, sondern stehen jeweils miteinander in Beziehung. Abhängig vom jeweiligen Anwendungskontext können diese an Gewicht gewinnen oder auch verlieren. Folgende Beispiele aus dem Umfeld autonomer[20] Maschinen sollen dies illustrieren:

1 Beispiele für softwareseitige Beschränkung: Ein Fahrzeug nimmt in Bereichen mit erhöhtem Gefährdungspotenzial (z. B. vor Kindergärten) automatisch eine zusätzliche Geschwindigkeitsreduzierung vor. Eine Maschine im Produktionsprozess reagiert auf die Annäherung eines Menschen mit sofortigem Stopp.

2 Beispiele für physische Beschränkungen: Produktionsmaschinen, bei denen durch eine Absperrung verhindert wird, dass Menschen zu Schaden kommen. Bei autonomen Fahrzeugen wäre es die Höchstgeschwindigkeit, die bauartbedingt nicht überschritten werden kann.

An diesen Beispielen wird deutlich, dass die Werte „Autonomie des Menschen“ und „Sicherheit“ in Abhängigkeit von den Umständen zum einen unterschiedlich gewichtet sind, zum anderen die Wichtigkeit anderer Werte gegen null gehen lassen, z. B. die im Ergebnispapier zur Ethik-Roadmap genannte Transparenz der Kommunikation.

Es zeigt sich also, dass in einem konkret vorliegenden Anwendungsfall die Gewichtung unterschiedlicher ethischer Anforderungen nicht unbedingt konstant sein muss, vielmehr es in der Regel nicht sein wird. Der Grund dafür liegt im Anwendungsfall und nicht in den ethischen Werten. Dieser Wert hängt immer vom konkreten Kontext und den jeweiligen Gegebenheiten ab.

Eine Möglichkeit, derartige Werte aufzustellen, ist es, exemplarisch Anwendungskontexte (wie die hier genannten Fallbeispiele) zu benennen und für den jeweiligen konkreten Anwendungskontext die gegebenen Attribute ins Verhältnis zueinander zu setzen. Dies kann zum einen mithilfe von Fachleuten im jeweiligen Anwendungsgebiet geschehen, z. B. in der Medizin, im Flugzeugbau usw., zum anderen mittels Bürgerbeteiligung. Zusätzlich wäre damit einhergehend eine Steigerung der fachspezifischen und gesellschaftlichen Akzeptanz zu erwarten (vgl. hierzu [119]).

Insgesamt zeigt sich, dass ein fest vorgegebener Wertekanon im konkreten Anwendungsfall unzureichend sein kann. Um eine konkrete Situation zu bewerten, bietet sich eine Methode an, die in der Industrie und auch im Bereich der ISO-Normen weitverbreitet ist, die Methode des risikoadaptierten[21] Ansatzes.

Für den Entwurf eines spezifischen Wertekanons für ein KI-Systems wird das Risiko einer Handlungsoption nicht immer hinreichend genau abschätzbar sein und damit eben auch nicht die Auswirkung auf einen spezifischen Wert im Wertekanon. Die Bandbreite der Risiken der einzelnen Handlungsoptionen kann so groß sein, dass eine differenzierte Handlungspräferenz nicht möglich ist, sei es wegen unzureichender Menge an Daten, qualitativ schlechter Daten oder der Komplexität des Gesamtsystems.

Zugleich sollte für die Zertifizierung von KI-Systemen die Erweiterung von Normen um ethische Aspekte mitgedacht werden (vgl. zu ersten Ansätzen [45]). Dabei sollte Normung nicht versuchen, einen einzigen Wertekanon für alle KI-Anwendungen festzulegen, da eine solche Festlegung die bestehenden Beziehungsrelationen in den verschiedenen Anwendungsfeldern nicht hinreichend abbilden könnte.

4.2.2.3.3 Privatsphäre

Als besonderer Wert ist der Grundsatz des Schutzes der Privatsphäre Ausdruck der Menschenwürde, Autonomie und der individuellen Freiheit. Der Schutz der Privatsphäre durch Technikgestaltung und datenschutzrechtliche Voreinstellungen (Art. 25 DSGVO [95]) hat demnach auch eine ethische Dimension, da Wertvorstellungen und funktional-kognitive Aspekte die zu schützende Persönlichkeit maßgeblich prägen. Normung sollte daher eine an ethischen Kriterien orientierte Technikgestaltung sowohl beim Design als auch der Begleitung von KI-Anwendungen befördern, um die Persönlichkeitsinteressen der Anwender sowie der durch die Systeme Betroffenen zu wahren (im Sinne eines „privacy ethical design“).

20 Unter dem Begriff „autonom“ wird hier – in Übereinstimmung mit der Definition in Kapitel 4.1.2.1 – bei Fahrzeugen/Maschinen verstanden, dass sie ihre Umwelt erfassen und sich ohne menschlichen Eingriff im gleichen Verkehrsraum bewegen wie Menschen. Sie unterliegen dabei selbstverständlich sowohl hardware- als auch softwareseitig Beschränkungen.

21 entspricht Englisch „risk based“

Bislang gibt es diesbezüglich jedoch keine einheitliche Strategie. Normung und Standardisierung von KI-Anwendungen beschränken sich im Wesentlichen auf Terminologien und Begriffsdefinitionen, die Interoperabilität von KI-Systemen sowie sicherheitstechnische Voreinstellungen. Normen, welche neben der Absicherung von Interoperabilität und der technischen Zuverlässigkeit die Berücksichtigung ethischer Aspekte und Werte bei der Produkt- oder Prozessgestaltung in den Vordergrund rücken oder den verantwortungsvollen Umgang mit KI-Anwendungen abfordern, fehlen weitgehend. Lediglich im medizinischen und arbeitssicherheitsrechtlichen Bereich werden über produkt- und/oder personenbezogene Organisations- und Dokumentationspflichten diese Aspekte mit abgedeckt. **Bereichsübergreifend** sind wertorientierte Anforderungen jedoch nicht vorhanden, wenngleich es Anknüpfungsmöglichkeiten gäbe: So könnte z. B. im Rahmen der ISO 9000 ff. [105], [120] im Sinne einer „erklärbaren KI“ sichergestellt werden, dass die Belange der „interessierten Parteien“ ausreichend berücksichtigt werden; ebenso könnten beim Umgang mit „Risiken“ [120] die bislang eher technischen Risiken um ethische erweitert werden. Diese Lücke sollte Normung künftig unter Berücksichtigung operationalisierbarer Werte schließen.

Normung, die im Sinne eines „privacy ethical design“ um diese ethischen Aspekte bereichert wäre, böte dabei nicht nur das Potenzial, über eine Sammlung entsprechender Standards zum europaweiten Benchmark zu werden, sondern könnte zugleich maßgeblich dazu beitragen, die Akzeptanz sowie das Vertrauen in die KI-Systeme ob des sodann integrierten Persönlichkeitsschutzes zu erhöhen.

Der Grundsatz des Schutzes der Privatsphäre durch Technikgestaltung und datenschutzrechtliche Voreinstellungen (Art. 25 DSGVO [95]) ist Ausdruck der Menschenwürde, Autonomie und individuellen Freiheit und hat demnach eine ethische Dimension, da Wertvorstellungen und funktional-kognitive Aspekte die zu schützende Persönlichkeit maßgeblich prägen. Normung sollte daher eine an ethischen Kriterien orientierte Technikgestaltung sowohl beim Design als auch der Begleitung von KI-Anwendungen befördern, um die Persönlichkeitsinteressen der Anwender sowie der durch die Systeme Betroffenen zu wahren (im Sinne eines „privacy ethical design“).

4.2.2.4 Kritikalitätsmatrix zur Risikoabschätzung

KI wirft durch ihren breiten Einsatz eine Vielzahl von rechtlichen wie auch gesamtgesellschaftlichen Fragen auf. Die Verwendung von KI gestützten Systemen im HR Bereich z. B. macht die Diskussion über arbeits- und datenschutzrechtliche Themen unter einer neuen Perspektive notwendig. Diese Bestrebungen sind zu unterstützen. Da ADM-Systeme durch ihre Entscheidungen und deren Fehler weitreichende Folgen auslösen können wie beispielsweise Diskriminierungen (siehe [120]), ist es ein wichtiger Punkt, diese ausreichend transparent und nachvollziehbar zu gestalten.

Wie in Kapitel 4.1.2.1.4 beschrieben, erhalten lernende KI-Systeme ihre wesentliche Funktionalität durch die Lernphase. Wie beim Menschen auch stellt die Prüfung dessen, was gelernt wurde, eine große und für die Softwareentwicklung neue Herausforderung dar. Ein diesbezüglicher Nachweis ist bei lernenden KI-Systemen heute nicht möglich, da keine Methode bekannt ist, das Gelernte vollständig zu verifizieren/validieren – somit bleibt eine Ungewissheit, ob das System im operativen Umfeld alle Anforderungen und Erwartungen erfüllt. Nach DIN EN ISO 9000 [105] werden die Auswirkungen von Ungewissheit als Risiko bezeichnet. Mit der DIN EN ISO 9001:2015 [120] wurde für Qualitätsmanagementsysteme (QM-Systeme) der risikoadaptierte[22] Ansatz für das Denken und Handeln hervorgehoben. Hierbei gilt es, fortlaufend diejenigen Faktoren zu bestimmen, die bewirken könnten, dass ihre Prozesse, Produkte und Dienstleistungen von den geplanten Ergebnissen abweichen, und vorbeugende Maßnahmen zur Steuerung umzusetzen. Der systematische Ansatz für ein Risikomanagementsystem wird in der DIN ISO 31000 [93] vorgestellt, in dem, in Übereinstimmung mit ISO/IEC Guide 51 [121], Risiken üblicherweise anhand der Risikoursachen, der potenziellen Ereignisse, ihrer Auswirkungen und ihrer Wahrscheinlichkeit dargestellt werden. Zur Vermeidung bzw. Begrenzung des Schadens, welcher bei ethischen Kriterien nicht nur monetär sein kann, gilt es, mögliche Risikoszenarien bereits in der Auslegungs- und Entwicklungsphase zu ermitteln und zu bewerten. Die Bewertung muss entlang des Entstehungsprozesses (siehe Kapitel 4.2.2.1) erfolgen, hierbei sollte immer gegen zuvor festgelegten Kriterien (Zielen), d. h. auch gegenüber ethischen Kriterien geprüft werden. Neben dem Fehlen von Informationen kann auch die Interpretierbarkeit von Information eine Rolle spielen.

22 entspricht Englisch „risk based“

Bei sicherheitskritischen Systemen wird bereits bei nicht-lernenden Systemen standardmäßig die Qualitätsmethode FMECA angewendet. Diese könnte auch auf lernende Systeme und zugehörige ethische Kriterien Anwendung finden (AI-FMECA) und dahingehend genutzt werden, um Faktoren zu identifizieren, welche bewirken können, dass das Gesamtsystem unvorhergesehenen Schaden verursacht, der über eine Kosten-Nutzen-Betrachtung hinausgeht, und um notwendige Transparenz- und Nachvollziehbarkeitspflichten bezüglich der Entscheidungslogik festzulegen. Dazu ist es notwendig, das ADM-System im Rahmen dieser Risikobetrachtung in seiner Gesamtheit einer Kritikalitätsstufe zuzuordnen.

Für die oben beschriebene Gefahrenvorsorge ist es unerlässlich, klare Kriterien zu haben, anhand derer der jeweilige ethische Wert bzw. das entsprechende Attribut messbar gemacht werden kann. In erster Näherung kann dies mit einer vergleichsweise groben Einteilung, wie oben schon erwähnt, geschehen. Hierzu könnte etwa eine Ordinalskala mit Werten wie gering, mittel und hoch herangezogen werden. Ein ähnliches Vorgehen, wie es auch in den entsprechenden Normen zur IT-Sicherheit (etwa ISO/IEC 27001 [122]) angewendet wird. Das Ergebnis einer solchen Risikoanalyse muss dann mit den festgelegten Risikokriterien verglichen werden. Dabei darf es nicht zulässig sein, ein niedriges Risiko in Bezug auf einen normativen Wert mit einem hohen oder mittleren Risiko in Bezug auf einen anderen zu verrechnen und insgesamt zu einem mittleren Risikolevel zu kommen. In Bezug auf ethische Werte und die Bewertung von Algorithmen nach ethischen Kriterien kann es hier nur ein Maximum-Prinzip geben, d. h. dass ein bestimmtes System oder ein Algorithmus nicht mehr zulassungsfähig ist, wenn auch nur ein Risiko in Bezug auf einen ethischen Grundsatz einen gewissen Grenzwert überschreitet.

Zur wirksamen Anwendung dieser ethischen Normen auf den Erstellungsprozess eines automatischen Entscheidungssystems wie auch für dessen Betrieb ist es erforderlich, die entsprechenden ethischen Anforderungen operationalisierbar zu machen. Dies erfordert die Definition von Kriterien, anhand derer sich der jeweilige Erfüllungsgrad (oder der Risikograd) eines jeden einzelnen Kriteriums messen lässt. Derartige Kriterien können ggf. aus den zuvor (exemplarisch) erwähnten, noch zu definierenden Anwendungsfällen kommen. Es gibt aktuelle Bestrebungen, ethische Werte bei KI-Anwendungen überprüfbar zu machen, hierzu wurde unlängst eine Methodik vorgestellt [123]. Um die ausgewählten Werte letztendlich zu operationalisieren, muss es das Ziel sein, Observable zu definieren, anhand derer überprüft werden kann, ob ein ADM-System einer Anforderung gerecht wird. Eine Möglichkeit dazu bietet das WKIO-Modell (**W**erte, **K**riterien, **I**ndikatoren, **O**bservablen), welches eine systematische Basis vorgibt, um allgemeine Werte durch Aufschlüsselung in Kriterien, Indikatoren und schlussendlich messbare Observable zu konkretisieren und damit überprüfbar zu machen, ob ein ADM-System eine Anforderung erfüllt. Ein solcher Prozess der Operationalisierung von Werten ist in den nächsten Jahren auch in der Normung zu begleiten. Wie in der Einleitung dieses Unterkapitels beschrieben ist es für die Vermeidung von Diskriminierungen und uneinschätzbaren Langzeitfolgen notwendig, je nach Potenzial des Gesamtschadens unterschiedliche Transparenz- und Nachvollziehbarkeitsforderungen an die Entscheidungslogik eines ADM-Systems zu stellen.

Angesichts der Vielfalt der Möglichkeiten, wie ADM-Systeme um- und eingesetzt werden können, erscheint ein differenzierter Regulierungsansatz notwendig. IT-Systeme mit sicherheitstechnischer Relevanz folgen dem ISO/IEC Guide 51 [121] und sind z. B. nach ISO 12100 [124], [125], ISO 13849 [126], [127], ISO 14971 [128] oder IEC 62061 [129] auszulegen. Für andere IT-Systeme kann die Betrachtung analog zu ISO 31000 [93] auf Basis einer matrixbasierten Risikoabschätzung geschehen[23], indem regulatorische Bestimmungen an verschiedene Risiken angepasst werden, z. B. im Finanzsektor (insbesondere das Arrow-II-Modell, [131], [132]) oder im Hinblick auf Umweltrisiken [133].

Der Zweck einer Risikomatrix besteht nicht darin, konkrete und harte Schwellenwerte zwischen den Kategorien zu identifizieren [133], da die konzeptionelle Unterscheidung zwischen Risikoklassen eine gründliche und detaillierte Bewertung konkreter Fälle durch eine Regulierungsbehörde nicht ersetzen kann. Auch der Umgang mit einer bestimmten Risikoquelle ist letztlich eine Frage der gesellschaftlichen Werte und der Risikotoleranz, die eine gewisse Formbarkeit und Mehrdeutigkeit mit sich bringen. Deshalb ist es kaum angebracht, starre Grenzen zwischen den Risikokategorien zu ziehen. Zudem erfordert ein praktikabler Ansatz für den Umgang mit der KI-Ethik die Berücksichtigung des jeweiligen Anwendungskontextes, da dieser unabhängig von der eingesetzten Technologie einen enormen Einfluss auf das potenzielle Risiko hat (vgl. den Einsatz eines Empfehlungssystems im Rahmen von Online-Recherchen und anschließender gezielter Werbung und Empfehlungen für die Wahl eines geeigneten Medikaments). In Anbetracht der unterschiedlichen

23 Andere Regierungen verfolgen den gleichen Ansatz, vgl. [130].

Anwendungsbereiche und gesellschaftlichen Kontexte, in denen ADM-Systeme eingesetzt werden können, ist es wichtig, dass sich eine Lösung allen Bedürfnissen anpassen kann, wenn es um die Beherrschung der Risiken algorithmischer Systeme geht [134]–[136].

Die Klassifikation müsste den gesamten potenziellen Schaden berücksichtigen, den ein KI-System in seinem jeweiligen Anwendungskontext verursachen kann. Entscheidende Faktoren bei der Beurteilung dieses Potenzials sind das Ausmaß der möglichen Verletzung von Rechtsgütern und Menschenleben durch das ADM-System und die Einschränkung der Handlungsfreiheiten des Individuums. Hier zeigt sich, dass die Verwendung einer zweidimensionalen Risikomatrix, auf der diese Faktoren die Achsen beschreiben, den Klassifizierungsprozess vereinfacht, ohne zu sehr von der gegebenen Komplexität eines KI-Systems zu abstrahieren [136].

4.2.2.4.1 Aktivitäten unabhängiger Dritter

Prüfungen im Bereich der KI durch Dritte können insbesondere durch Konformitätsbewertungsstellen (KBS) erfolgen, die außerhalb staatlicher Stellen und nicht in Verbindung mit den Anwendern und Herstellern des Systems stehen. Im Bereich der ISO-Normen gilt hier grundsätzlich, dass die KBS erforderliche Kompetenzkriterien an das Personal definieren müssen, sodass diese in der Lage sind, Konformität zu bestätigen. Sollten darüber hinaus gesetzliche Rahmenbedingungen gelten, haben diese stets Vorrang (Grundsatz „Gesetz bricht Norm“).

Dies soll beispielhaft an der Zertifizierung eines Maschinenbaubetriebs im Vergleich zur Zertifizierung einer Arztpraxis dargelegt werden. Als zugrunde liegende Norm wird die ISO 9001:2015 [120] herangezogen.

Fall 1: Zertifizierung eines Maschinenbaubetriebs
Auf der Basis der für die KBS geltenden ISO-Regeln (in diesem Fall ISO/IEC 17021-1:2015, 7.1.2 [40]) ist die KBS verpflichtet, für alle am Zertifizierungsprozess beteiligten Personen einen Mechanismus vorzuhalten, der die Kompetenzkriterien für die jeweilige Prüfperson (Fachexperte, Auditor) festlegt. Im vorliegenden Fall könnte die KBS also voraussetzen, dass zur Prüfung eines Maschinenbaubetriebs der Prüfer die Kompetenz eines Ingenieurs oder einen vergleichbaren Kenntnisstand besitzt. Derartige Regelungen besitzen in aller Regel eine Öffnungsklausel für Personen, die etwa Maschinenbau studiert, jedoch nicht abgeschlossen haben, oder auch für Personen, die keine Ingenieure im Maschinenbaubereich, sondern etwa Verfahrenstechniker sind. Wenn die KBS dies schlüssig darlegt, können auch Personen zur Prüfung eines solchen Betriebs herangezogen werden, die nicht über eine Ausbildung auf den Qualifikationsniveau eines Ingenieurs verfügen.

Fall 2: Zertifizierung einer Arztpraxis
Unabhängig davon, wie eine KBS hier ihre eigenen Kompetenzkriterien definiert, gilt nach deutschem Recht, dass Ärzte einer Berufsverschwiegenheit (MBO Ä 1997 § 9 [137]) unterliegen (vergleichbare Regelungen gelten auch etwa für Steuerberater oder Rechtsanwälte). In diesem Fall ist es so, dass bei der Prüfung der Prozesse in einer Arztpraxis, ob diese normkonform eingeführt sind, als Prüfer der KBS **grundsätzlich nur ein Arzt** oder eine Person mit einer noch höherwertigen Ausbildung infrage kommt. Allen anderen Personen, und seien sie auch medizinisch vorgebildet, darf der Arzt aufgrund seiner Berufsverschwiegenheit keinen Einblick in Prozesse der Praxis gewähren. Eine Prüfung kann daher von Rechts wegen gar nicht stattfinden.

Insofern ist es wichtig bei der Regulierung, innerhalb des Bereichs der KI in Schlüsselbereichen klare gesetzliche Anforderungen bzw. harmonisierte Normen zu haben, die die **Kompetenz der prüfenden Personen** sicherstellen. Andernfalls ist es einem unabhängigen Dritten (einer KBS) erlaubt, Kompetenzkriterien in einer Art und Weise zu definieren, die unter Umständen nicht problemadäquat, obwohl normkonform, sind.

Ein unbedingt zu beachtender Sachverhalt hierbei ist, dass wenn ein unabhängiger Dritter, also eine KBS, tätig wird, unbedingt gefordert werden muss, dass dieses ausschließlich im **akkreditierten Bereich** erfolgt. Andernfalls gibt es keinerlei Handhabe hinsichtlich der Anwendung internationaler Normen, die zuvor dargestellten Voraussetzungen auch wirklich einzufordern. Dies ist nur möglich, wenn die KBS selbst regelmäßig durch eine unabhängige Organisation geprüft wird, wie es in Deutschland auf Basis des Akkreditierungsstellengesetzes von der Deutschen Akkreditierungsstelle (DAkkS) durchgeführt wird. Hier wäre somit auch eine europaweit gültige Regelung existent, die dann greifen würde, da die entsprechenden Akkreditierungsstellengesetze in jedem EU-Land gleichermaßen gelten und durchgesetzt werden.

4.2.2.4.2 Kritikalitätsmodell

Eine Unterteilung in die zwei Bereiche „hohes Risiko“ und „kein hohes Risiko“, wie sie die Europäische Kommission gefordert hat (Weißbuch EU-Kommission [15]) schafft es nicht, dieser komplexen Problematik gerecht zu werden. Es wird sich daher eine differenzierte Betrachtung durchsetzen, die es durch Normung zu begleiten und zu gestalten gilt. Das von Krafft und Zweig entwickelte differenzierte Regulierungskonzept [136] dagegen ordnet ADM-Systeme mit ihrem jeweiligen Anwendungskontext anhand zweier Kriterien in eine von fünf möglichen Klassen ein und ermöglicht so eine grobe Einschätzung des Handlungsbedarfs (vgl. Abbildung 17). Eine visuelle Aufbereitung dieses Konzepts wurde durch die Datenethikkommission vorgenommen, Details hierzu sind in Kapitel 4.1.2.2 ausgeführt.

4.2.2.4.3 Prüfung der Notwendigkeit einer ausführlichen Kritikalitätsprüfung

Die wirtschaftliche Tragfähigkeit dieses Konzepts ist nur dann gegeben, wenn der horizontal gelagerte, für alle ADM-Systeme geltende Ordnungsrahmen minimal gehalten wird. Zusätzlich müssen sektorspezifisch in allen (!) Anwendungsfeldern von Künstlicher Intelligenz bestehende Normen und Richtlinien überprüft/überarbeitet/angepasst und ergänzt werden.

Aktuelle Forschungsergebnisse weisen darauf hin, dass die ADM-Systeme, welche aktuell in Deutschland in Anwendung sind, lediglich zu einem geringen Anteil das Potenzial aufweisen, in persönlichem oder gesellschaftlichem Schaden zu resultieren. Zusätzlich gibt es Untersuchungen, die darauf hinweisen, dass die meisten Schäden nur die Privatwirtschaft betreffen [138]. Dennoch ist es wichtig, gerade die Systeme mit einem solchen Potenzial zu identifizieren. Dies kann durch eine noch zu gestaltende Kritikalitätsprüfung geschehen.

Ein ADM-System wird anhand von zwei Achsen in eine von fünf Kategorien einsortiert. Je höher die Kategorie, desto mehr Transparenz- und Nachvollziehbarkeitsauflagen gibt es an die Entscheidungslogik des Systems.

Ausmaß der möglichen Verletzung von Rechtsgütern und Menschenleben (x-Achse)
Für die x-Achse ist der kritische Aspekt das Ausmaß der möglichen Verletzung von Rechtsgütern und Menschenleben durch ein KI-System. Um dies zu beurteilen, ist mindestens Folgendes zu beachten:

→ **Auswirkungen auf Grundrechte, Gleichheit oder soziale Gerechtigkeit:** Hat eine KI negative Auswirkungen auf die Grundrechte einer natürlichen oder juristischen Person, sind die Mechanismen der sozialen Gerechtigkeit (z. B. Rente, Krankenversicherung) für eine demografische Gruppe gefährdet oder können die

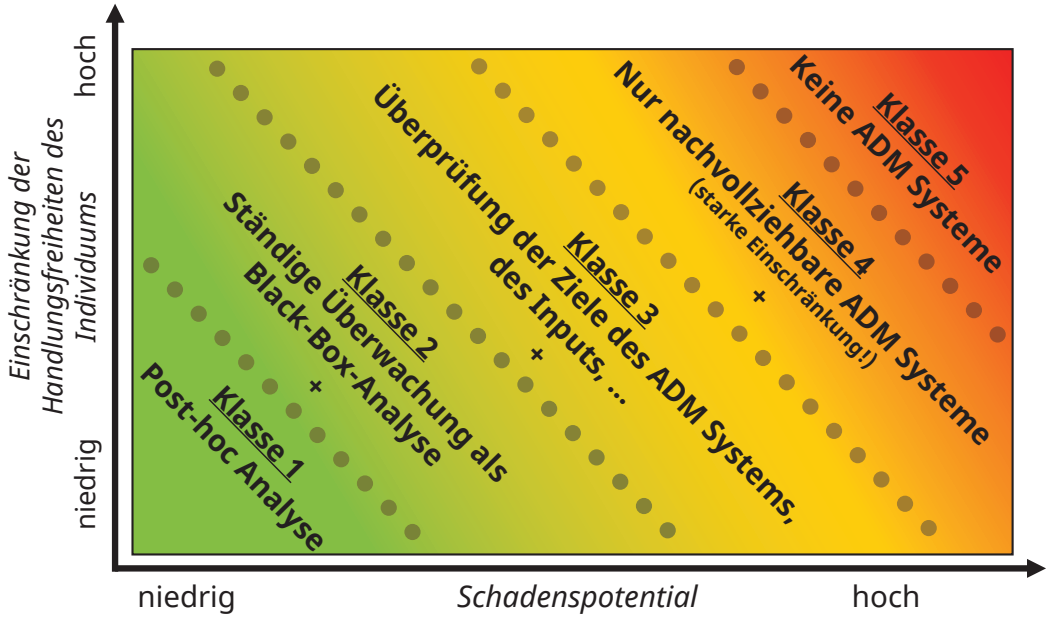


Abbildung 17: Kritikalitätsmodell (nach [136]).

Auswirkungen sogar katastrophal sein und zum Verlust von Leben führen (z. B. die Behandlung von Intensivpatienten)?

→ **Anzahl der betroffenen Menschen:** Ist eine große Anzahl von Menschen betroffen (z. B. faire Bewertung bei einer Bewerbung)?

→ **Auswirkungen auf die Gesellschaft:** Trägt das System das Risiko, die Gesellschaft als Ganzes zu beeinträchtigen (z. B. personalisierte Auswahl von politischen Nachrichten), unabhängig von direkt wahrnehmbaren Schäden?

In jedem Fall ist es unmöglich, die Intensität des potenziellen Schadens zu bewerten, indem man einfach die Schadenshöhe mit der Eintrittswahrscheinlichkeit multipliziert. Dies würde bedeuten, das Risiko, dass jemand bei einem drohenden Sturm ohne Schirm das Haus verlässt (hohe Eintrittswahrscheinlichkeit, geringes Schadenspotenzial), mit dem Risiko eines nuklearen Unfalls (geringe Eintrittswahrscheinlichkeit, hohes Schadenspotenzial) gleichzusetzen. Mit zunehmendem Schadenspotenzial können Makrorisiken entstehen, die unsere Handlungsfähigkeit als Ganzes bedrohen und daher nicht akzeptabel sind.

Einschränkung der Handlungsfreiheiten des Individuums (y-Achse)

Die y-Achse zeigt die Einschränkung der Handlungsfreiheiten der potenziell betroffenen Individuen bezüglich der algorithmischen Entscheidung und geht damit auf die Optionen zur Vermeidung des auf der x-Achse angegebenen potenziellen Schadens ein. Je besser die Chancen stehen, die möglichen negativen Folgen einer Entscheidung oder den durch sie verursachten Schaden zu vermeiden, desto weiter unten auf der y-Achse würde das ADM-System stehen. Die drei Hauptfaktoren, die bei der Beurteilung der Abhängigkeit von der Entscheidung eine Rolle spielen, sind Kontrolle, Auswahl und Korrektur [123].

→ Entscheidungen und Handlungen eines KI-Systems, die zusätzlich durch menschliche Interaktion gefiltert werden (z. B. der Kauf von empfohlenen Artikeln in einem Online-Shop), implizieren einen geringeren Regulierungsbedarf als Maschinen, die ohne menschliche Mittler agieren (z. B. die Notabschaltung eines Kernkraftwerks). Dieser Aspekt wird unter **Kontrolle** zusammengefasst.

→ Die Fähigkeit, das KI-System gegen ein anderes auszutauschen (z. B. durch Wechsel eines Anbieters) oder zu vermeiden, dass man einer algorithmischen Entscheidung überhaupt ausgesetzt ist, wird als **Auswahl** bezeichnet. Ein einseitiges Abhängigkeitsverhältnis zwischen Produzenten oder Betreibern und Nutzern sowie monopolistische Strukturen führen zu einer Abhängigkeit von einem oder wenigen Systemen. Im schlimmsten Fall hat der Nutzer nicht die Möglichkeit, sich von der Nutzung bestimmter Dienste abzuwenden, ohne mit persönlichen oder gesellschaftlichen Auswirkungen (z. B. durch fehlenden Zugang zum Gesundheitswesen, Finanzmarkt) konfrontiert zu werden.

→ Die Bedeutung der Möglichkeit, eine automatisch generierte Entscheidung anfechten oder korrigieren lassen zu können, sowie die Zeit, die für eine angemessene Weiterverfolgung des entsprechenden Antrags benötigt wird, sollte nicht unterschätzt werden. Dies wird in dem Begriff **Korrektur** zusammengefasst. Maschinelle Entscheidungen, die überhaupt nicht angefochten werden können, erhöhen die Abhängigkeit von der Entscheidung. Die Behebung eines signifikanten individuellen Schadens erfordert mehr Zeit und Mühe als viele Fälle mit geringerem Schaden. Dieser Aspekt betrifft den Schadensausgleich/die Haftung, der in der Abhängigkeit von der Entscheidung (y-Achse) angesprochen wird.

4.2.2.4.4 Risikoklassen

Für Systeme, die in Klasse 1 fallen, würden keine Transparenzpflichten gefordert und keine Kontrollprozesse dauerhaft installiert werden. In Verdachtsfällen könnte in einer Post-hoc-Analyse auf relevanten Schaden geprüft werden. Sollte sich der Verdacht bestätigen, wäre eine erneute Bewertung in eine höhere Klasse denkbar.

In Klasse 2 würden erste Transparenzpflichten gefordert werden. Um sogenannte Black-Box-Analysen [139] zu ermöglichen, muss eine entsprechende Schnittstelle für das System bereitgestellt werden, sodass eine kontrollierende Instanz das Eingabe-Ausgabe-Verhalten des Systems überprüfen kann. Zudem wäre eine Beschreibung der Einbettung des Systems in den sozialen Entscheidungsprozess erforderlich.

Für Systeme in Klasse 3 sollten die Eingangsdaten gegenüber einer kontrollierenden Instanz vollständig beschrieben werden. Die angegebene Qualität (im Sinne von die Qualität beschreibenden numerischen Werten) des Entscheidungssystems müsste überprüfbar sein.

In Klasse 4 müssten sämtliche Angaben über und Entscheidungen von der Software zumindest für eine kontrollierende Instanz in angemessener Zeit nachvollziehbar und überprüf-

bar sein. Die Forderung nach Nachvollziehbarkeit schließt viele Lernverfahren generell aus (z. B. künstliche neuronale Netze), da sie diese zum aktuellen Stand der Forschung nicht erfüllen können. Alle notwendigen Interfaces müssten bereitgestellt werden.

Systeme, die in Klasse 5 eingeordnet werden, sollten nicht umgesetzt werden. Diese Klasse wird durch Systeme gerechtfertigt, die mit den Grundsätzen einer Demokratie nicht vereinbar sind, z. B. Bewertungssysteme, die auf einer kontinuierlichen Überwachung der Bevölkerung basieren, Systeme, die die Unschuldsvermutung außer Kraft setzen, oder Systeme, welche ohne menschliche Einflussnahme billigend letal wirken. Darüber hinaus wären Systeme, die ein gewisses Schadenspotenzial überschreiten und aufgrund der schwierigen Datenlage (z. B. unvollständig oder fehlerhaft) nur mit hoher Fehlerrate umgesetzt werden könnten, in dieser Klasse verortet (z. B. Identifikationssysteme für Terroristen). Die Klasse schließt statistische Methoden, die Muster in großen Datenmengen suchen, nicht aus, jedoch dürfte das Finden solcher Muster nicht unreflektiert in Entscheidungen münden.

4.2.3 Normungs- und Standardisierungsbedarfe

BEDARF 1:
Initiale Kritikalitätsprüfung von KI-Systemen schnell und einfach gestalten
Nicht intendierte ethische Probleme und Konflikte treten vor allem bei ADM-Systemen mit lernenden Komponenten auf, die über Menschen, deren Hab und Gut oder über den Zugang zu knappen Ressourcen entscheiden und dabei das Potenzial aufweisen, individuelle Grundrechte und/oder grundlegende demokratische Werte zu schädigen. Eine initiale Kritikalitätsprüfung, ob ein System solche Konflikte überhaupt auslösen kann oder es sich um eine Anwendung fernab jeder ethischen Fragestellung handelt, muss durch Normung schnell und einfach gestaltet werden. Diese horizontal für alle Bereiche niedrigschwellige Überprüfung muss schnell und rechtsicher klären, ob das System überhaupt Transparenz und Nachvollziehbarkeitsanforderungen erfüllen muss. Gerade im Hinblick auf die weiten Einsatzfelder von Künstlicher Intelligenz bietet eine solche Kritikalitätsprüfung in kritischen Bereichen die Möglichkeit, adäquate Forderungen zu stellen und gleichzeitig dem Vorwurf des „ethical red taping“ zu begegnen, indem völlig unkritische Anwendungsfelder frei von zusätzlichen Anforderungen entwickelt werden können.

BEDARF 2:
Operationalisierung ethischer Werte
Aktuell ist unklar, wie Organisationen, die KI-Systeme entwickeln und einsetzen, abstrakte ethische Werte messen und operationalisieren können. Es gibt eine Reihe von vielversprechenden Ansätzen, die das Potenzial haben, der Herausforderung zu begegnen (wie z. B. das WKIO-Modell), jedoch steckt die konkrete praktische Anwendung solcher Ansätze noch in den Kinderschuhen. Offene Fragen, Probleme und Herausforderungen können aktuell nur begrenzt adressiert werden, weshalb Normen die Möglichkeit bieten, den Prozess, theoretische Konzepte zur Operationalisierung von Ethik in die Praxis zu überführen, zu begleiten und im Dialog mit den Firmen konsensual zu gestalten.

BEDARF 3:
Normung eines Konzepts für privacy ethical design
Der Grundsatz des Schutzes der Privatsphäre ist Ausdruck der Menschenwürde, Autonomie und individuellen Freiheit und ein wesentliches Kriterium für die Akzeptanz neuer Systeme. Daher sollte Normung eine Technikgestaltung fördern, die die Persönlichkeitsinteressen von Anwendern und Betroffenen im Sinne eines „privacy ethical designs“ wahrt. Dies sollte die bisherigen Ansätze aus den Bereichen Medizin und Arbeitssicherheit in einem bereichsübergreifenden Konzept aufgreifen und ausgestalten. Dies kann im Rahmen des aktuell im ISO/IEC JTC 1/SC 42 initiierten Projekts zu einen MSS für KI (Kapitel 4.1.3, Bedarf 1 „Unterstützung der internationalen Standardisierungsarbeiten zu einem MSS für KI“) erfolgen, indem die Erklärbarkeit von KI-Systemen in den Anforderungskatalog des entstehenden Dokuments aufgenommen wird, sowie durch eine Ausweitung des Risikobegriffs auf ethische Risiken, wie sie bereits im Projekt ISO/IEC 23894 **Risk Management** vorgenommen wurde.

BEDARF 4:
Gestaltung des Wertesystems
Intelligentes Entscheiden auf der Basis allgemeiner Ethikgrundsätze benötigt eine Auseinandersetzung mit ethischen Werten. Wenn die Maschine den Bedeutungszusammenhang verschiedener Werte und Objekte mittels einer Ontologie kennt, ist das hilfreich. Autonome Systeme müssen auch ungeplante Situationen verarbeiten können. Wenn beispielsweise von autonomen Maschinen bei der Erkennung der Umgebung die interne Darstellung der Objekte durch Wissen aus einer Ontologie angereichert wird, ist dies eine Möglichkeit, den Maschinen ein Wertesystem zugänglich zu machen. Ontologien lassen die Maschine Bedeutungszusammenhänge herstellen, ohne dass zuvor Fallmuster spezifiziert sein müs-

sen (wie im W3C [140]). Die Erforschung und anschließende Normung der Schnittstellen von Ontologien zur Berücksichtigung ethischer Prinzipien in konkreten Szenarien verspricht, das Potenzial der Herausforderung zu begegnen.

BEDARF 5:
Zweckbindung von Daten gestalten
Eine Normung sollte die bestehende Zweckbindung von Daten weiter gestalten. Diese kann sicherstellen, dass eine Dokumentation vorhanden ist. Zu welchem Zweck die Daten gesammelt wurden, und eine Regulierung ermöglichen, unter welchen Bedingungen die Daten einem anderen Zweck zugeführt werden dürfen oder nicht.

BEDARF 6:
Schnittstellen des Entwicklungsprozesses von KI gestalten
Der lange Entwicklungsprozess von KI-Systemen sollte durch standardisierte Schnittstellen gestaltet werden. Hier kann Normung einen wichtigen Beitrag leisten. Diese Schnittstellen könnten beispielsweise den Zugriff zu relevanten Trainingsdatensätzen und -modellen eines KI-Systems als Grundlage für eine externe Überprüfung beinhalten. Vornehmlich internationale Standards würden sowohl eine Austauschbarkeit von Komponenten fördern als auch überprüfenden Instanzen den Zugang ermöglichen und ohne große Aufwände sicherstellen, dass Vorgaben erfüllt sind, um dadurch das Vertrauen in das System zu erhöhen.

BEDARF 7:
Quality Backward Chain in den KI Life-Cycle aufnehmen
Es wird empfohlen, einen Quality Backward Chain mit Felddatengewinnung in den KI Life-Cycle aufzunehmen, um unethisches Verhalten während der Anwendung zu identifizieren und dem korrektiv gegenzusteuern (siehe Kapitel 4.3.2.3.2.4 Prozessprüfungen: Qualitätssicherung nach Auslieferung durch Produktbeobachtung).

BEDARF 8:
Re-Evaluierung von KI-Systemen gestalten
KI-Systeme sollen im komplexen gesellschaftlichen Kontext breiten Einsatz finden. In der KI-Entwicklung sollte daher ein systematischer Prozess mit ethischer Reflexion und Beteiligung gestartet werden. Abhängig von der Komplexität der KI und potenziellen Risiken sind mehrere Evaluationsschritte und eine kontinuierliche Beteiligung interessierter Kreise sowie von Experten aus dem Bereich Ethik und ethisch geschulter Mitarbeiter empfehlenswert.

4.3 Qualität, Konformitätsbewertung und Zertifizierung

KI findet zunehmend Anwendung in unterschiedlichen Bereichen des Alltags (siehe auch die Kapitel 4.5 bis 4.7). Davon ausgehend, dass KI nur dann ihr volles Anwendungspotenzial entfalten kann, wenn ihr Einsatz nach hohen Gütekriterien erfolgt, beschäftigt sich das folgende Kapitel mit dem sich hieraus ergebenden Standardisierungsbedarf hinsichtlich von Qualitätskriterien und ihrem Nachweis durch eine entsprechende Konformitätsbewertung (auf Grundlage der ISO/IEC 17000er Normenreihe [38]–[44]). „Einige Gedanken, die in diesem Kapitel diskutiert werden, finden sich auch im Impulspapier und Whitepaper „Zertifizierung von KI-Systemen“ der Plattform Lernende Systeme.“

Bei der Prüfung von KI-Systemen lassen sich zwei Ebenen unterscheiden (siehe Abbildung 18): Zum einen können durch eine technische Prüfung zugesicherte Eigenschaften eines KI-Systems bestätigt werden, beispielsweise können für eine Klassifizierung die Genauigkeit durch Precision und Recall ermittelt werden (Technische Ebene der Prüfung). Die zweite Ebene stellt die Bewertungsebene dar, welche prüft, ob ein System für einen bestimmten Einsatzzweck geeignet ist (reicht die geprüfte Genauigkeit für den Einsatzzweck aus?) bzw. bestimmten ethischen, rechtlichen oder gesellschaftlichen Anforderungen genügt. Für ethische Betrachtungen ist ein Gütesiegel [123] vorgeschlagen worden, welches einen interessanten Ansatz für die ethische Bewertung von KI-Systemen darstellt und auf einem Wert-Analyseverfahren aus einer Kombination von Zielkriterien, Indikatoren und messbaren Größen beruht. Sämtliche Prüfungen der zweiten Art sollten sich immer auf technische Prüfungen stützen können. Es ist zu erwarten, dass Normen und Standards vor allem für die erste Prüfebene formuliert werden können, Fragestellungen der zweiten Prüfungsebene aber oft Gegenstand der Regulatorik oder gesellschaftlicher Diskurse sind.

Solche Konformitätsbewertungen können durch den Hersteller selbst, den Käufer oder eine akkreditierte Drittstelle durchgeführt werden. Im Zuge einer Konformitätsbewertung können Produkte, Systeme und Prozesse einer Prüfung, Kalibrierung, Validierung, Verifizierung und Zertifizierung oder Inspektion unterzogen werden. In bestimmten Bereichen (wie etwa gemäß der EU-Medizinprodukteverordnung [141]) ist die Zertifizierung durch eine benannte Stelle sogar verpflichtend vor dem Inverkehrbringen.

Eine Zertifizierung erfolgt im Rahmen der Konformitätsbewertung durch eine Drittstelle nach dem angewendeten Konformitätsbewertungsprogramm.

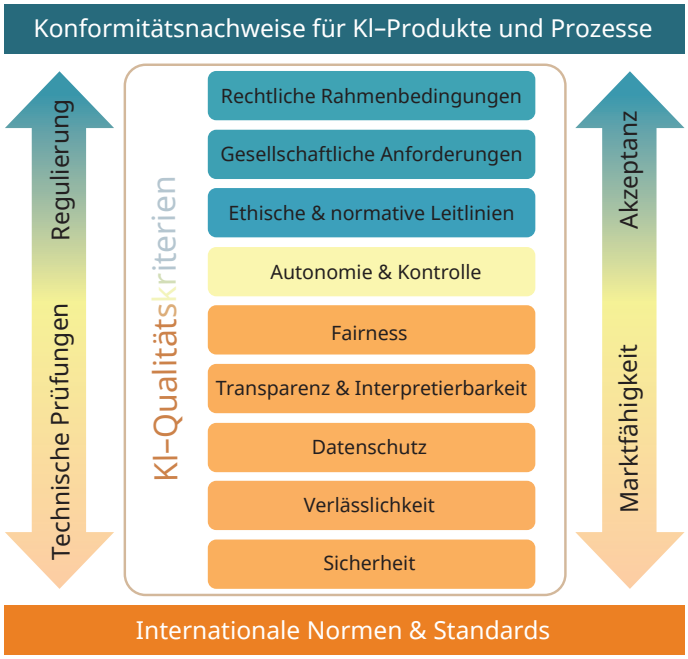


Abbildung 18: Einordnung der Kategorien der KI-Qualitätskriterien in die Konformitätsprüfung in Anlehnung an [45]

Konformitätsnachweise für KI-Produkte und Prozesse basieren nach Auffassung von internationalen Expertenkommissionen, wie z. B. der HLEG-KI, auf folgenden normativen, rechtlichen und technischen Qualitätskriterien, vergleiche auch [45].

Recht, Gesellschaft und Ethik

KI-Anwendungen besitzen disruptives Potenzial. Die Konformität mit gesellschaftlichen, ethischen und rechtlichen Rahmenbedingungen dient hauptsächlich dem Schutz rechtlich bzw. ethisch grundlegender Interessen von Personen (Kapitel 4.2.2.3). Die KI-Konformitätsprüfungen in diesen Kategorien sollen Beeinträchtigungen von Gruppen und Einzelpersonen, Unrecht bzw. ethisch nicht gerechtfertigte Zustände von der Gesellschaft verhindern und vermeiden helfen.

Autonomie und Kontrolle

KI-Anwendungen arbeiten in zunehmendem Maße autonom, d. h. sie verfolgen ein vorgegebenes Ziel bei freier Auswahl der dorthin führenden Mittel. Dem KI-System wird die Wahl der Mittel, nicht aber die eigentliche Zielsetzung freigestellt. Man spricht in diesem Kontext irreführend von der „Handlungsautonomie“ des Systems, obwohl die Zielsetzung nicht verändert wird. Aus dieser Analogie ergibt sich ein Spannungsfeld zur Autonomie des Menschen, da solche KI-Anwendungen den Menschen ihrerseits in der Wahl seiner Ziele und Mittel

beeinflussen können. KI-Konformitätsprüfungen müssen an der Schnittstelle zum technischen KI-System Aussagen zur Autonomie und Kontrolle treffen können, wenn die KI-Anwendung mit menschlicher Entscheidungsfindung interagiert, indem sie z. B. Entscheidungsvorschläge generiert, Steuerbefehle erzeugt und ggf. ausführt, mit dem Menschen kommuniziert oder in Arbeitsprozesse integriert ist.

Die folgenden Qualitätskategorien sind Bestandteil der technischen Prüfung von KI-Systemen.

Fairness und Nichtdiskriminierung
KI-Anwendungen lernen aus historischen Daten, die nicht notwendigerweise vorurteilsfrei sind. Damit ungerechtfertigte Ungleichbehandlung in einer KI-Anwendung vermieden und unzulässige Diskriminierungen ausgeschlossen werden, müssen die KI-Anwendungen darauf hin überprüfbar sein, dass Individuen nicht aufgrund ihrer Zugehörigkeit zu einer marginalisierten oder diskriminierten Gruppe wiederum im sozialen Ergebnis diskriminiert werden dürfen (siehe Kapitel 4.2.2.2).

Transparenz und Interpretierbarkeit
Die Transparenz einer KI-Anwendung kann erheblich zu ihrer Akzeptanz beitragen. Dazu müssen Informationen zum richtigen Umgang mit der KI-Anwendung verfügbar sein. Im Wesentlichen müssen Anforderungen an die Interpretierbarkeit, Nachverfolgbarkeit und Reproduzierbarkeit von Ergebnissen geprüft werden, die Einsichten in die inneren Prozesse der KI-Anwendung erfordern. Für die damit umgangssprachlich verbundene Forderung nach der Erklärbarkeit einer KI-Anwendung besteht, selbst wenn sich die Erklärbarkeit auf Wirkungen KI-spezifischer Technologiemerkmale beschränkt, noch erheblicher Forschungsbedarf.

Datenschutz
Die technische Prüfung der datenschutzrechtlichen Vorgaben, insbesondere der DSGVO [95], des BDSG [142] und die Anforderungen der Hambacher Erklärung [143], sind für KI-Konformitätsprüfungen zu beachten.

Verlässlichkeit
Aus technischer Sicht umfasst die Prüfung der Verlässlichkeit eines KI-Systems Anforderungen an die Korrektheit, der Nachvollziehbarkeit der Einschätzung der Unsicherheit der Ergebnisse und der Robustheit gegenüber Angriffen, Fehlern und unerwarteten Situationen und überschneidet sich damit mit dem Begriff der Sicherheit im engeren Sinn. Prüfungen der Verlässlichkeit und Sicherheit von KI-Anwendungen sind essenzielle Grundvoraussetzungen, um Aussagen über deren Vertrauenswürdigkeit zu machen.

Sicherheit
Die Sicherheit von KI-Anwendungen umfasst die Sicherheit vor Gefährdungen und Angriffen und die Funktionssicherheit im weiteren Sinne. Die Sicherheit von KI-Systemen wird im Kapitel 4.4 ausführlich betrachtet. Dabei werden auch Verlässlichkeit, Datenschutz und Datensicherheit in den Blick genommen. Prüfmethodisch ist festzuhalten, dass die technischen Prüfgrundlagen für KI-Systeme entwickelt und mit bestehenden Prüfverfahren in Beziehung gesetzt werden müssen.

4.3.1 Status quo

Im Folgenden werden die wesentlichen Begriffe von Gegenständen und Aktivitäten der Konformitätsbewertung aufgeführt.

4.3.1.1 Konformitätsbewertung

Darlegung, dass festgelegte Anforderungen erfüllt sind (ISO/IEC 17000 [38]). Festgelegte Anforderungen (d. h. Erfordernisse oder Erwartungen) können detailliert (z. B. konkrete technische Spezifikationen) oder allgemein (z. B. sicher, robust, transparent, fair) sein.

Zur Unterscheidung der Gegenstände der Konformitätsbewertung:
1. Produkt (z. B. Hard-/Software)
2. Prozess
3. System
4. Dienstleistung
5. Managementsystem
6. Person
7. Information (z. B. Deklarationen, Behauptungen, Vorhersagen)

Gegenstand einer Konformitätsbewertung können auch Kombinationen dieser einzelnen Objekte sein (z. B. Entwicklungsprozess + Produkt, Produkt + Dienstleistung, System + Behauptung). Die festgelegten Anforderungen sind dem Gegenstand eindeutig zuzuordnen (z. B. technische Spezifikation für die Hardware, Fairness-Kriterien für den Prozess, Robustheit für ein System, Kompetenzanforderungen für eine Person, Plausibilitätsbedingungen für eine Behauptung).

Zur Unterscheidung der Tätigkeiten:
Durch eindeutige Zuordnung der festgelegten Anforderungen zu definierten Gegenständen (s. o.) lassen sich die Tätigkeiten zur „Auswahl“ und „Ermittlung“ (s. Prozess der Konformitätsbewertung) bestimmen. Deren Ergebnisse können für die gegebene Situation (z. B. zur Analyse oder Charakterisierung) ausreichen oder anschließend im Hinblick auf eine „Entscheidung“ über Konformität des Gegenstandes einer „Bewertung“ unterzogen werden.

4.3.1.1.1 Arten der Konformitätsbewertung

Im folgenden werden die Arten der Konformitätsbewertung (siehe Abbildung 19) beschrieben.

Prüfung
Unter Prüfung wird die Ermittlung eines oder mehrerer Merkmale an einem Gegenstand der Konformitätsbewertung nach einem Verfahren verstanden. Das Verfahren kann dazu gedacht sein, Variablen innerhalb der Prüfung als Beitrag zur Genauigkeit oder Zuverlässigkeit der Ergebnisse zu kontrollieren. Die Ergebnisse der Prüfung können in Form spezifizierter Einheiten oder objektiver Vergleiche mit vereinbarten Referenzen dargestellt werden. Das Ergebnis der Prüfung kann Bemerkungen (z. B. Meinungen und Interpretationen) über die Prüfergebnisse und Erfüllung der festgelegten Anforderungen einschließen.

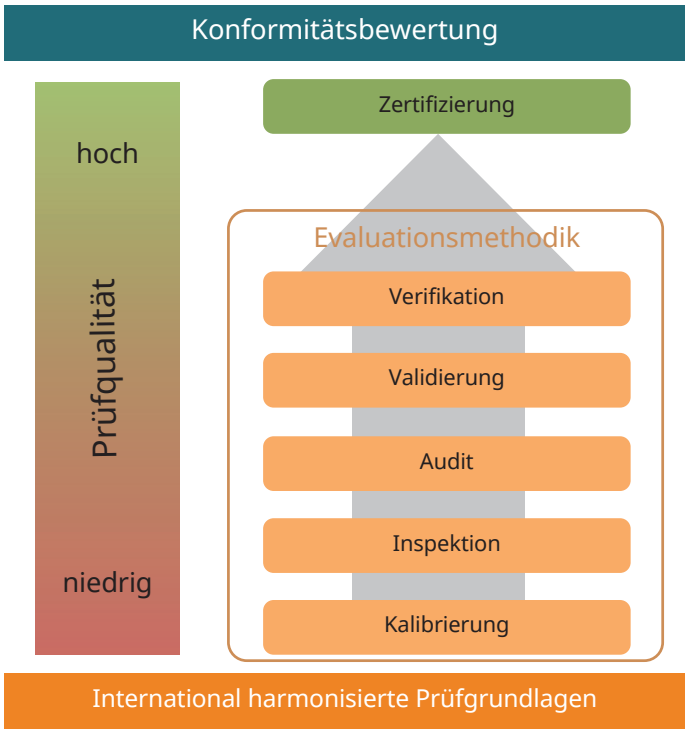


Abbildung 19: Evaluationsmethodik und Prüfqualität

Kalibrierung
Kalibrierung bezeichnet eine Tätigkeit, die unter festgelegten Bedingungen in einem ersten Schritt eine Beziehung zwischen den durch Normale zur Verfügung gestellten Größenwerten mit ihren Messunsicherheiten und den entsprechenden Anzeigen mit ihren beigeordneten Messunsicherheiten herstellt und in einem zweiten Schritt diese Informationen verwendet wird, um eine Beziehung herzustellen, mit deren Hilfe ein Messergebnis aus einer Anzeige erhalten wird.

Das Ergebnis einer Kalibrierung kann in Form einer Angabe, einer Kalibrierfunktion, eines Kalibrierdiagramms, einer Kalibrierkurve oder einer Kalibriertabelle ausgedrückt werden. In einigen Fällen kann sie aus einer additiven oder multiplikativen Korrektion der Anzeige mit der beigeordneten Messunsicherheit bestehen. Kalibrierung sollte nicht mit Justierung eines Messsystems verwechselt werden, das oft fälschlicherweise „Selbst-Kalibrierung“ genannt wird, und auch nicht mit Verifizierung der Kalibrierung. Oft wird nur der erste Schritt in dieser Definition als Kalibrierung angesehen [144].

Inspektion
Inspektion ist die Untersuchung eines Gegenstands der Konformitätsbewertung und Ermittlung seiner Konformität mit detaillierten Anforderungen oder, auf der Grundlage einer sachverständigen Beurteilung, mit allgemeinen Anforderungen. Eine Untersuchung kann direkte oder indirekte Beobachtungen einschließen, die Messungen oder das Ablesen von Messgeräten umfassen können. Inspektionen können in Konformitätsbewertungsprogrammen oder Verträgen auf Untersuchungen beschränkt werden.

Audit
Die Überprüfung, ob die Prozesse, Praktiken und Verfahrensweisen einer Organisation bestimmte Anforderungen erfüllen, die in einem Standard (z. B. einem MSS, vgl. Kapitel 4.1.2.2.3) formuliert sind, wird als Audit bezeichnet. Diese Überprüfung erfolgt in der Regel aufgrund eines aus der zugrunde liegenden Norm abgeleiteten Kriterienkatalogs, der beschreibt, wie Anforderungen überprüft werden.

Prüfungen umfassen die Einsichtname in Dokumentationen, die von der zu auditierenden Organisation zur Verfügung gestellt werden, Interviews durch den Prüfer, aber auch Vor-Ort-Inspektionen.

ISO unterscheidet drei Ebenen des Audits:
- → Audit durch die Organisation, auf die sich das Audit bezieht (Selbstauskunft);
- → Audit durch einen Kunden, Lieferanten oder Partner der zu auditierenden Organisation;
- → Audit durch eine unabhängige dritte Stelle. Ein solches Audit kann zu einer Zertifizierung führen.

ISO 19011 [145] stellt Leitlinien zur Auditplanung, Auditdurchführung und Auditnachbereitung zur Verfügung.

Validierung

Bestätigung der Plausibilität eines bestimmten Verwendungs- oder eines bestimmten Anwendungszwecks durch Bereitstellung eines objektiven Nachweises, dass festgelegte Anforderungen erfüllt wurden. Die Validierung kann auf Behauptungen angewendet werden, um die durch eine Behauptung angegebenen Informationen in Bezug auf die beabsichtigte zukünftige Verwendung zu bestätigen.

Verifizierung

Bestätigung der Wahrheitsmäßigkeit durch Bereitstellung eines objektiven Nachweises, dass festgelegte Anforderungen erfüllt wurden. Die Verifizierung kann auf Behauptungen angewendet werden, um die durch eine Behauptung angegebenen Informationen zu bestätigen, die sich auf Ereignisse beziehen, die bereits eingetreten sind oder die sich auf Ergebnisse beziehen, die bereits erlangt wurden.

Zertifizierung

Bestätigung durch eine dritte Seite bezogen auf einen Gegenstand der Konformitätsbewertung (Akkreditierung ausgenommen). Eine „dritte Seite“ (third party) ist unabhängig vom Anbieter des Gegenstands der Konformitätsbewertungstätigkeit und hat kein Interesse als Anwender. Prüfungen, Inspektions- und Validierungs-/Verifizierungstätigkeiten können auch vom Anbieter (first party) des zu bewertenden Gegenstandes oder von einer Person/Organisation mit Interesse als Anwender dieses Gegenstands (second party) durchgeführt werden. Zertifizierungen werden allein von unabhängigen Stellen angeboten.

4.3.1.1.2 Prozess der Konformitätsbewertung

Die Konformitätsbewertung gliedert sich in fünf Phasen:
- → **Auswahl** (selection) = Auswahl der anzuwendenden Anforderungen, Wahl der Methoden, Planung, Probenahme
- → **Ermittlung** (determination) = Tätigkeiten zur Sammlung der Belege für Konformität im Hinblick auf die festgelegten Anforderungen, d. h. Analysen, Tests, Evaluierungen, Untersuchungen, Audits, Prüfungen, Inspektionen, Validierungen, Verifizierungen etc.
- → **Bewertung** (review) = Schlussfolgerung bzgl. Eignung, Angemessenheit und ausreichender Anzahl der gesammelten Belege
- → **Entscheidung** (decision) = Entscheidung, ob Konformität des bewerteten Gegenstands mit den festgelegten Anforderungen dargelegt wurde oder nicht
- → **Bestätigung** (attestation) = formelle Ausstellung der Konformitätsaussage, z. B. Prüfbericht (Test bestanden/nicht bestanden) oder Zertifikat

4.3.1.1.3 Arten der Konformitätsbewertungsstellen

Je nach Art der Konformitätsbewertung unterscheidet die ISO/IEC 17000er-Reihe verschiedene Arten von Bewertungsstellen, die gemäß der oben aufgeführten Tätigkeiten Produktsicherheit und Qualität sowie Schutzgüter prüfen, analysieren, testen bzw. messen:
- → **Prüflaboratorium** (ISO/IEC 17025 [42])
- → **Inspektionsstelle** (ISO/IEC 17020 [39])
- → **Validierungs-/Verifizierungsstelle** (ISO/IEC 17029 [43])
- → **Zertifizierungsstelle** (ISO/IEC 17021-1 [40] für Managementsysteme, ISO/IEC 17024 [41] für Personen und ISO/IEC 17065 [44] für Produkte, Prozesse und Dienstleistungen)

4.3.1.2 Bestehende Standards und Normen aus anderen Bereichen mit Relevanz für KI-Qualität und Konformitätsbewertung

KI-Anwendungen sind meist als Komponenten in größere IT-Systeme eingebettet. Diese KI-Komponenten können wiederum durch eine Vielzahl an unterschiedlichen Technologien realisiert sein. Diese KI-Anwendungen kommen in vielen industriellen und alltäglichen Anwendungen zum Einsatz, wobei oft eine Interaktion der eigentlichen KI-Komponente

mit weiteren softwaretechnischen, informationstechnischen, mechanischen und elektronischen Modulen des Gesamtsystems vorliegt.

In einem ersten Schritt ist es daher Aufgabe der Standardisierung, bestehende Normen und Standards zu identifizieren, welche für die Qualität (und deren Überprüfung) dieser Systeme relevant sind. Hierbei sind insbesondere Normen aus den Bereichen Software (KI-Komponente), IT-Sicherheit (IT-Gesamtsystem), Datenqualität sowie Funktionale Sicherheit (Anwendungskontext) in Betracht zu ziehen.

Die Tabelle 13 im Kapitel 6.4 zeigt nationale sowie weltweit arbeitende Standardisierungsgremien. Für Qualität, Konformitätsbewertung und Zertifizierung relevante Arbeitsgruppen sind in der Spalte „Relevanz für Qualität, Konformitätsbewertung und Zertifizierung (4.3)“ gekennzeichnet.

Grundsätzlich ist, technologieunabhängig, jede Norm, die Anforderungen an einen Softwareeinsatz formuliert, auch für KI-Komponenten als eine spezielle Softwarekomponente von Relevanz. Es muss zunächst überprüft werden, welche Normen die KI-spezifischen Eigenschaften bereits ausreichend abdecken und ob Ergänzungen oder Änderungen notwendig werden.

Im Folgenden werden einige prominente Beispiele aus den Bereichen Softwareentwicklung und Funktionale Sicherheit genannt. Diese Aufzählung erhebt jedoch keinen Anspruch auf Vollständigkeit. Zusätzlich gibt es sehr viele relevante Normen zur IT-Sicherheit, die im Kapitel 4.4 umfassend behandelt werden. Darüber hinaus werden in diesem wie in weiteren Bereichen Normen mit Fokus auf KI überarbeitet, um KI-relevante Aspekte aufzugreifen. Tabelle 11 in Kapitel 6.2 gibt einen Überblick über Normen und Standards unterschiedlicher thematischer Bereiche, die noch keine detaillierten Aussagen zur Anwendung von KI-Komponenten machen. Darin sind die Normen, die relevante Anforderungen und Qualitätskriterien für Software formulieren, in der Spalte „Relevanz für Qualität, Konformitätsbewertung und Zertifizierung (4.3)“ gekennzeichnet.

4.3.1.2.1 Softwareentwicklung

KI-Prozesse können in bereits bestehende, für die Softwareentwicklung erstellte Normen integriert werden, beispielsweise in die ISO/IEC/IEEE 12207 [58] (Software life cycle processes), ISO/IEC 27034 [71]–[78] (Application Security) sowie die ISO/IEC 25010 [146] (System and software quality models). Beispielsweise kann in Anlehnung an ISO/IEC 25010 eine Prüfung von austrainierten KI-basierten Softwaresystemen auf „Funktionale Sicherheit, Effizienz, Übertragbarkeit, Wartbarkeit und Zuverlässigkeit“ erfolgen [147].

4.3.1.2.2 Funktionale Sicherheit

Die IEC 61508 [79]–[86] Normenreihe definiert Anforderungen an die verschiedenen Lebenszyklusphasen von elektrischen, elektronischen und programmierbaren elektronischen (E/E/PE) Systemen, die sicherheitsrelevante Funktionen ausführen. Ein spezieller Fokus ist in der IEC 61508-3 auf die Anforderungen für die Entwicklung von sicherheitsrelevanter Software gelegt. Dies beinhaltet auch Anforderungen an die verwendeten Tools im Entwicklungsprozess. Es werden vier Sicherheitsintegritätsstufen (Safety Integrity Level, SIL) definiert als Maß für die notwendige risikomindernde Wirksamkeit von Sicherheitsfunktionen und die daraus resultierenden Anforderungen an das sicherheitsrelevante System. Bisher wird die Verwendung von KI-Funktionalität von der IEC 61508 zwar nicht empfohlen, aber auch nicht ausgeschlossen. Das zuständige Kommittee IEC/SC 65A betrachtet aber die Thematik KI für ein Update der IEC 61508 und arbeitet dazu mit ISO/IEC JTC1 SC42 zusammen. Dort ist ein technischer Report zu funktionaler Sicherheit und KI-Systeme im Entstehen. Die IEC 61508 hat eine breite Akzeptanz und Anwendung in der Industrie und ist die Basis für mehrere Anwendungsbereiche für spezifische Standards, z. B. für die Prozessindustrie, den Maschinenbau, die Leittechnik in Kernkraftwerken und Bahnsignaltechnik. Der Standard ISO/PAS 21448 [148] beschreibt die Sicherheit der Soll-Funktion und bezieht Performance-Einschränkungen, die ihren Ursprung im Umgebungseinfluss oder der Kommunikation haben, mit ein. Die Norm ISO 12100 [124], [125] definiert als Sicherheitsgrundnorm generelle Prinzipien und Methoden der Maschinensicherheit, stellt jedoch keine Norm der Funktionalen Sicherheit im engeren Sinne dar.

4.3.1.2.3 Datenqualität

Da die Qualität einer KI-Komponente eng mit der Datenqualität verknüpft ist, werden in Tabelle 11 auch Normen zum Thema Datenqualität und Big Data gelistet. Die DIN ISO/IEC 25012 [88] führt ein Modell der Datenqualität ein. Die Standards ISO/IEC 20546 [34] und ISO/IEC TR 20547-2 [149] und -5 [150] befassen sich mit Big Data, deren Terminologie und Referenzarchitekturen.

4.3.1.3 Bestehende Standards und Normen im Bereich KI-Qualität und Konformitätsbewertung

In Tabelle 10 in Kapitel 6.1 werden bestehende Normen und Standards, die KI-Anwendungen explizit behandeln, aufgelistet. Darin sind die Normen, die Anforderungen formulieren, in der Spalte „Relevanz für Qualität, Konformitätsbewertung und Zertifizierung (4.3)“ gekennzeichnet. Diese Auflistung ist nicht abschließend, stellt jedoch aus heutiger Sicht den Gros der relevanten Normen und Standards dar.

Auf deutscher Seite wurden von DIN zwei DIN SPEC veröffentlicht, in welchen ein Qualitätsmetamodell für KI (DIN SPEC 92001 [87]) sowie ein Leitfaden von Deep-Learning-Bilderkennungssystemen (DIN SPEC 13266 [151]) vorgestellt sind. Auf europäischer Ebene wird bei ETSI Künstliche Intelligenz in technischen Spezifikationen in Bezug auf Emotionserkennung (ETSI TS 103 296 [152]) und autonome Netzwerke (ETSI TS 103 195-2 [153]) thematisiert. Auf internationaler Ebene fokussiert sich die ITU-T innerhalb der veröffentlichten Standards auf Anforderungen (Y.3170 [154]) und KI-Fähigkeiten (Y.3173 [155]) mit Blick auf KI in Zukunftsnetzen. Die Konsortien IEEE und UL befassen sich innerhalb der veröffentlichten Dokumente mit der Beurteilung von autonomen Systemen (IEEE 7010-2020 [156]und UL 4600 [157]).

Außer der DIN SPEC 92001-1 [87] behandeln alle genannten und in der Tabelle gelisteten Standards KI-Komponenten bezogen auf eine konkrete Anwendung. In weiterführenden Normungsaktivitäten müssten die dort genannten Qualitätskriterien mit den Qualitätskriterien der ISO/IEC 25010 [146] abgeglichen und die KI-spezifischen Anforderungen herausgehoben werden.

4.3.1.4 Standardisierungsaktivitäten mit Relevanz für KI-Qualität und Konformitätsbewertung

In der Tabelle 12 in Kapitel 6.3 werden Standardisierungsaktivitäten aufgelistet, wobei solche mit Relevanz für KI-Qualität und Konformitätsbewertung in der Spalte „Relevanz für Qualität, Konformitätsbewertung und Zertifizierung (4.3)“ gekennzeichnet sind. Diese Auflistung ist nicht abschließend, stellt jedoch aus heutiger Sicht den Gros der relevanten Normungs- und Standardisierungsprojekte dar.

Aktuell finden auf allen Standardisierungsebenen zahlreiche Aktivitäten zur KI-Standardisierung statt. Insbesondere die Arbeit im ISO/IEC NP 5059 ist hervorzuheben, da hier an Qualitätsanforderungen für KI in Anlehnung an die Softwarequalitätsanforderungen nach ISO/IEC 25010 [146] gearbeitet wird.

4.3.2 Anforderungen, Herausforderungen

4.3.2.1 Prüfungsbedarf und Marktfähigkeit

Im April 2019 veröffentlichte die HLEG-KI ethische, rechtliche und technische Key Requirements für vertrauenswürdige KI-basierte Systeme [22]. Es handelt sich dabei in den allermeisten Fällen um hybride Anwendungen, d. h. sie bestehen aus KI-Komponenten und nicht KI-basierter Software und Hardware und werden grundsätzlich als spezielle IT verstanden. Die Anwenderindustrie in Europa erwartet die marktgerechte Entwicklung von Kriterien und Methoden zur technischen Prüfung von KI-Systemen. Im Folgenden wird der Umfang einer solchen Prüfung diskutiert.

4.3.2.2 Scope einer Prüfung

In diesem Kapitel wird diskutiert, welche Aspekte im Rahmen einer Prüfung eines KI-Systems betrachtet werden sollten. Dies schließt die Komponenten eines KI-Systems ein sowie KI-spezifische Herausforderungen, welche sich bei der Prüfung dieser Systeme ergeben.

Weitere Qualitätsanforderungen ergeben sich aus der Tatsache, dass KI-Systeme oftmals auch Bestandteile eines größeren Produkts (z. B. der Platform Economy) sind, für dessen Interoperabilität ebenfalls Standards gesetzt werden müssen, um eine zusätzliche Anschlussfähigkeit und Austauschbarkeit im Endprodukt zu gewährleisten. Ohne solche Garantien wird ein globales Zusammenspiel nahezu unmöglich, was schlussendlich auch die Skalierbarkeit von Lösungen verhindert.

Hinsichtlich des Aspekts der Überprüfung von Qualitätskriterien besteht teilweise große Affinität zu Prüfverfahren der funktionalen Sicherheit, der Softwareentwicklung und IT-Sicherheit, welche sich darauf zurückführen lässt, dass es sich bei KI-Anwendungen um hybride IT-Systeme handelt. Die Marktfähigkeit eines potenziellen Prüfverfahrens, welches die oben genannten Aspekte adressiert, erfordert daher ein integriertes Vorgehen, welches bestehende Prüfverfahren auf

KI-spezifische Kriterien erweitert. Im Folgenden werden die Komponenten eines KI-Systems analysiert, welche bei einer entsprechenden Prüfung zu berücksichtigen sind. Zudem werden die KI-spezifischen Herausforderungen dargestellt, welche adressiert werden müssen, um eben dargestellte Lücke zu schließen.

4.3.2.2.1 Komponenten eines KI-Systems

Zu Komponenten eines KI-Systems gehören Algorithmen, Datenbasen und Schnittstellen zum Gesamtsystem. Grundsätzlich fußen Systemkomponenten auf KI-Basis auf symbolischen sowie subsymbolischen Methoden der Künstlichen Intelligenz. Dazu zählen Techniken zur Entscheidungsfindung (z. B. entscheidungstheoretische Expertensysteme), Wissensrepräsentation (z. B. Ontologien und Wissensgraphen), Verfahren zur Anwendung von Wissen (z. B. logisches Schließen und probabilistische Verfahren) sowie maschinelle Lernverfahren (z. B. überwachtes Lernen und unüberwachtes Lernen). Eine ausführliche Darstellung zur Klassifikation von KI-Komponenten findet sich in Kapitel 4.1.2.

Hierbei können die Methoden der Künstlichen Intelligenz in einer KI-Anwendung mittels Software realisiert werden. In Abhängigkeit des Fähigkeitsspektrums einer KI-Anwendung können auch hybride Methoden (z. B. hybride neuronale Netzwerk-Modelle) zum Einsatz kommen, in welchen symbolische und subsymbolische Techniken miteinander kombiniert sind. In KI-Anwendungen besteht eine Anpassungsfähigkeit (Dynamik) der Teilekomponenten von Methoden der Künstlichen Intelligenz. Beispielsweise bestimmen bei maschinellen Lernverfahren Aktivierungs-, Übertragungs- und Summationsfunktion die Dynamik eines neuronalen Netzes [158]. Andererseits kann sich eine Dynamik in der Veränderbarkeit von Wissen durch KI-Methoden äußern, beispielsweise auf Grundlage der AGM-Theorie [159], [160].

Unabhängig von der konkreten Realisierung der KI-Anwendung sollte eine Beurteilung der Qualität folgende Aspekte umfassen:

→ Die Qualität der verwendeten Daten: Dies umfasst u. a. einen möglichen Bias in den Daten, der die Fairness des Gesamtsystems negativ beeinträchtigen kann und die Integrität der Daten, da diese maßgeblich das Verhalten einer KI-Komponente bestimmen und somit eine Absicherung der Trainingsdatensätze gegen indirekte Angriffe durch deren Manipulation notwendig machen. Dies betrifft insbesondere kontinuierlich lernende (selbstlernende) Systeme, die im Feld weitertrainiert werden und deren Eingangsdaten für das kontinuierliche Lernen nicht unter unmittelbarer Kontrolle des Herstellers stehen. Daher ist auch eine Qualitätssicherung der Datenlieferkette selbst notwendig, da diese bezüglich der Qualitätsaspekte der Daten eine wesentliche Rolle spielt. Auch müssen die für das Training eines Modells verwendeten Daten und deren Verteilung (beispielsweise Bildauflösung, statistische Verteilung) der operativen Einsatzumgebung entsprechen.
Bei der Entwicklung von KI-Systemen kommen zunehmend auch synthetisch generierte Daten zum Einsatz. Hierbei wird eine künstliche Repräsentation eines Originaldatensatzes erzeugt, welche die wichtigsten statistischen Eigenschaften des Originaldatensatzes besitzt. Solche synthetisch generierten Datensätze sind besonders dann hilfreich, wenn entweder die Menge der Originaldaten zu klein ist (ein Beispiel ist das Trainieren von ML-Modellen für das autonome Fahren) oder wenn die Originaldaten sensible personenbezogene Merkmale enthalten. Die Qualität solcher synthetisch generierter Daten ist messbar und sollte insbesondere denselben Qualitätsanforderungen genügen wie realiter erzeugte Datensätze.

→ Die Auswahl der Methode/der Algorithmen, deren Hyperparameter und die Beurteilung eines gelernten Modells. Hier bieten sich im Allgemeinen empirische Verfahren des Testens sowie Verifikationsverfahren zur Qualitätssicherung eines trainierten Systems an. Beide unterliegen den im nachfolgenden Kapitel beschriebenen Herausforderungen. Zur Prüfung der Qualität ist eine Betrachtung alternativer Hyperparameter und deren Einfluss auf die Qualität notwendig. Für die Parameterauswahl und anderer Bereiche des Engineerings befinden sich Verfahren des automatisierten maschinellen Lernens unter dem Begriff AutoML in der Entwicklung und ersten Verwendung, u. a. als Dienstleistung.

→ Außerdem die Beurteilung des Gesamt-IT-Systems, in welche die KI-Komponente eingebettet ist. Hierbei ergeben sich insbesondere Schnittstellen zu weiteren technischen IT-Umgebungen wie etwa Cloud-Architekturen, Serverfarmen, Data Repositories und Data Supply Chains, statistische Analysepakete etc.

→ Die Mensch vs. Maschine – Schnittstelle, hierbei sind Human Factors und Maschine Factors zu berücksichtigen. Aber auch Maschine vs. Maschine, bzw. KI-System vs. KI-System ist zu validieren. Die Prüfung eines Human vs. Machine Interface kann erleichtert werden, durch eine

Rückmeldung des KI Systems an den Menschen, mit der Erläuterung, was es „verstanden“ hat.

→ Das Verhalten der KI-Anwendung nach Auslieferung während seiner Anwendung im operativen Umfeld (Produktbeobachtung) bis zum Abschluss seines Lebenszyklus (siehe Kapitel 4.3.2.3.2.1 und 4.3.2.3.2.4).

4.3.2.2.2 KI-spezifische Herausforderungen

Im Gegensatz zu herkömmlichen IT-Systemen weisen KI-Anwendungen einige Besonderheiten auf, für welche Qualitätskriterien und Prüfverfahren etabliert werden müssen und die für bestehende und zu entwickelnde Prüfverfahren substanzielle Herausforderungen darstellen. Dazu zählen:

→ Korrektheitsbegriff für KI-Systeme: Bei regelbasierten Algorithmen liegt ein klarer Source Code vor, welcher mit klassischen Methoden geprüft werden kann. Als Verifikationsverfahren bieten sich z. B. das klassische automatisierte Beweisen und Beweisassistenten an. Auch etwaige bestimmte Parameter lassen sich angemessen prüfen. Bei lernenden Systemen kommt neben der Softwarearchitektur, z. B. Auswahl des NN-Modells, und der Source-Code Qualität noch der gelernte Anteil hinzu. Im Gegensatz zu klassischen Systemen arbeiten KI-basierte Systeme zudem oftmals statistisch und werden daher nicht eine Genauigkeit bezüglich des spezifizierten Verhaltens von 100 % erreichen. Daher muss eine Prüfung von KI-basierten Systemen eine hinreichende Anforderung an die Genauigkeit definieren und zum Ziel haben, zum einen diese Anforderung und für die übrigen Fälle eine Absicherung des Systems durch weitere Maßnahmen, bspw. Safeguards, zu argumentieren. Unbenommen bleibt, dass als Teil des anwendungsspezifischen Risikomanagements bestimmte Restrisiken toleriert werden können.

→ Dynamik von KI-Systemen: KI-Systeme, welche auf Verfahren des maschinellen Lernens beruhen, unterliegen im Betrieb oftmals einer Dynamik, welche zwei Ursachen hat: Zum einen kann sich die Betriebsumgebung ändern, sodass das ursprünglich gelernte Modell die Realität nur noch unzureichend abbildet (Concept-drift). Zum anderen kann das Modell im Betrieb weiterlernen, beispielsweise durch Nutzerfeedback. Dies bezeichnet man als Model-Drift. Für eine potenzielle KI-Prüfung bedeutet dies, dass das Ergebnis über die zugesicherten Eigenschaften des KI-Systems zu einem späteren Zeitpunkt nicht mehr gültig sein muss. Dies stellt einen weiteren zentralen Unterschied zur Verifizierung von herkömmlicher Software dar. Folgende Maßnahmen sind denkbar, um dieser Problematik zu begegnen: 1) Model-Drift lässt sich durch die Einführung von strukturierten Model-Updates vermeiden. Für solche Updates lassen sich potenziell Qualitätsanforderungen definieren, sodass die zugesicherten und geprüften Eigenschaften auch nach dem Update erhalten bleiben. 2) Ähnlich zu Cloud-Zertifizierungen ist eine kontinuierliche Prüfung des KI-Systems durch das Monitoring geeigneter KPIs denkbar. Eine geeignete Auswahl solcher KPIs ist aktuell allerdings noch Gegenstand von Forschung und Entwicklung. Alternativ kann durch geeignete Maßnahmen, z. B. Safeguards, dafür gesorgt werden, dass das System keine kritischen Zustände einnehmen kann. 3) Mögliches, nicht-beherrschbares Verhalten von KI-Systemen kann durch das Einbinden einer Aufsichtsperson (human-in-the loop) unterbunden werden.

→ Unsicherheit: Unsicherheit bezüglich der Korrektheit einer Ausgabe (Uncertainty) ist eine intrinsische Eigenschaft von datengetriebenen KI-Anwendungen. Neben der einfachen Beobachtung, dass die Anwendung eines durch ein maschinelles Lernverfahren erstellten Modells auf eine neue, bislang unbekannte Eingabe ein korrektes oder auch nicht korrektes Ergebnis erzielen kann, beschäftigt sich die Forschung bzgl. der Unsicherheit von Modellen im engeren Sinn mit der Sicht, dass ein gelerntes Modell als probabilistische Funktion betrachtet werden kann und somit jede von dem Modell getroffene Aussage zugleich im Prinzip mit einer Konfidenz versehen ist, deren Kenntnis wiederum verschiedene Schlüsse bgzl. der Verwendung des Modells im gegebenen Fall erlaubt. Unglücklicherweise sind jedoch bei komplexen gelernten Modellen die tatsächlich gültigen Konfidenzen nicht direkt ersichtlich. Verkompliziert wird der Sachverhalt dadurch, dass die bei der Anwendung des Modells auftretenden Unsicherheiten durch verschiedene, aber auch in ihrer Wirkung interagierende Aspekte verursacht werden können: Unzureichende oder unpräzise Datenlage, Beschränkungen in der Ausdrucksfähigkeit der gewählten Modellklasse oder auch ein immanentes, nicht-deterministisches Verhalten der modellierten Zielfunktion (z. B. langfristige Wetterprognose). Dementsprechend ließe eine genaue Kenntnis der Modellunsicherheit wiederum Rückschlüsse über die Datenlage, die Modellkomplexität und die Vorhersagequalität in der Anwendung zu. Letzteres ist wiederum ein zentrales Element schichtweise aufgebauter Sicherheitsarchitekturen, auf denen auf oberen Ebenen alternative Mechanismen angewendet werden (z. B. Fahrer übernimmt das Steuer), wenn die KI-Anwendung auf der unteren Ebene eine zu große

Unsicherheit signalisiert (Monitoring-Ansatz). Es gibt ein breites Spektrum von Forschungsansätzen, um die mit einem gelernten Modell verbundenen Unsicherheiten unter einschränkenden Voraussetzungen zu erfassen, die von einer einfachen nachträglichen „Modellkalibrierung“ über gezielte Eingriffe in das eigentliche Lernverfahren bis zu aufwendigen Redundanzverfahren und mehr oder weniger ganzheitlich ansetzenden mathematischen Analysen reichen. Angesichts der stets wachsenden Modellkomplexität und Breite der Anwendungen von gelernten Modellen und auch sicherheitskritischen Bereichen ist die Entwicklung von effizienten, exakt wirksamen und allgemein anwendbaren Verfahren zur Bestimmung und zum Testen der Unsicherheit von Modellen dringend geboten.

→ Transparenz/Nachvollziehbarkeit: Ein KI-System ist transparent, wenn seine Genese und Wirkweise offen, vollständig und verständlich dargestellt wird. Dies umfasst insbesondere die Datengrundlage und die algorithmische Komponente. Die Entscheidungen/Vorschläge eines KI-Systems sind nachvollziehbar, wenn die Faktoren, die zu ihrem Zustandekommen geführt haben, durch einen Menschen verstanden werden können.
Bei der Transparenz spielen insbesondere die folgenden Aspekte eine wichtige Rolle: Transparenz der zum Training verwendeten Daten, der Annotation der Daten (z. B. Inter-Annotator-Agreement mittels Cohens-Kappa oder Fleiss-Kappa). Transparenz bei der Auswahl der Methoden. Transparenz und Nachvollziehbarkeit der Ergebnisse (Einflussgewichtung der eingegebenen Variablen). Transparenz in der Herangehensweise (z. B. durch eine Historie der überprüften Hypothesen bei der Parameteroptimierung bzw. Modellerstellung). Transparenz in der gesicherten Anwendung (also wann ein Modell solide Entscheidungen treffen kann oder wann es außerhalb oder in Randgebieten der Eingabedaten operiert). Generell ist zwischen der Transparenz für den Endanwender und Interpretierbarkeit zu unterscheiden.

Aus technischer Sicht ist die Frage der grundsätzlichen Transparenz nicht einfach zu beantworten und das Spannungsfeld zwischen höherer Genauigkeit bzw. Robustheit und der Erklärbarkeit von Modellen ist in der KI-Welt ein altbekanntes Dilemma. »Blackbox«-Modelle sind zwar in vielen Fällen genauer bzw. robuster als beispielsweise regelbasierte Modelle, jedoch sind sie nur bedingt interpretierbar. Teilweise kann diese Erklärbarkeit auch durch nachgeschaltete Verfahren, wie z. B. durch das Trainieren von Erklärmodellen oder eine Analyse des Eingabe/Ausgabe-Verhaltens von Modellen, sogenannte Local Interpretable Model-agnostic Explanations (LIME) Analyse, erreicht werden. Zurzeit ist die Interpretierbarkeit von Modellen ein aktives Forschungsfeld und es werden viele Anstrengungen unternommen, die Lernprozesse von »Blackbox«-Modellen besser zu verstehen sowie ihre internen Prozesse zu visualisieren und die resultierenden Entscheidungen erklären zu können.

→ IT-Sicherheit: Bei KI-Komponenten und KI-basierten Systemen treten neue IT-Sicherheitsrisiken wie Adversarial Attacks auf. Da diese oftmals statistisch arbeiten und deren Funktionsweise noch nicht gänzlich verstanden ist, stellt die Qualitätssicherung die IT-Sicherheit von KI-Komponenten vor große Probleme. Für Menschen nichtwahrnehmbare Modifikationen von Daten, beispielsweise in Bildern, führen beispielsweise bei Anwendung von Adversarial Samples zu Fehlklassifikationen, beispielsweise durch subtile Manipulation von Verkehrszeichen auf der Straße oder durch Hinzufügen von gezieltem Rauschen in bereits vorliegenden Bildern. Daneben unterliegen KI-Systeme selbst und die darin enthaltenen Modelle IT-Sicherheitsrisiken. Das austrainierte Modell stellt einen zu schützenden Geschäftswert dar und muss daher gegen Reverse Engineering sowie dessen Trainingsdaten geschützt werden. Entsprechende Angriffe können auch Einfluss auf den Datenschutz haben, da bereits Techniken existieren, die es erlauben, einzelne Trainingsdatensätze zu extrahieren. Detaillierte Ausführungen hierzu sind im Kapitel 4.4 zu finden.

→ Hyperparameter: Neben der ausgewählten KI-Methode bzw. dem ausgewählten KI-Algorithmus und den zum Training und Testen verwendeten Daten bestimmen die zugehörigen Hyperparameter wesentlich dessen Qualität und können zu Effekten wie Overfitting führen, bei denen das System eine besonders hohe Genauigkeit für die Trainigsdaten erreicht, allerdings eine nur niedrige Genauigkeit im operativen Betrieb. Zu den Hyperparametern zählen neben Eigenschaften des Modells bezüglich dessen Größe (beispielsweise Anzahl der Schichten eines tiefen neuronalen Netzes) auch Lernparameter wie die Anzahl der Epochen und die Lernrate.

4.3.2.3 Prüfverfahren

4.3.2.3.1 Nachweisführung von KI-Systemen

Eine angemessene Nachweisführung bildet die Basis jeder Konformitätsbewertung im Rahmen der Entwicklung von Systemen. Die in der Einführung erwähnten spezifischen

Herausforderungen lernender Systeme stellen auch Anforderungen an die entwicklungsbegleitende Nachweisführung. Neben einer entsprechenden Dokumentation mit Konfigurationsmanagement liegt hier die Betrachtung auf Test (Prüfung gegenüber Kriterien), Verifikation (Formale Prüfung des KI-Moduls gegenüber der Spezifikation) und Validation (Formale Prüfung der Anwendung im Einsatzumfeld).

Das „Product Quality Model“ der ISO/IEC 25010 [146] zeigt hierbei zwei Themenfelder auf:

1. Functional Testing: „what the system does“
2. Non-Functional Testing: „how the system does it“

Für eine Testumgebung sollte ein anwendungsspezifischer und von den Randbedingungen abhängiger Handlungsrahmen geschaffen werden, innerhalb von welchem Prüfverfahren festgelegt werden können. Hierbei richten sich Prüfverfahren und Prüftiefen u. a. nach der Identifikation von relevanten Nutzergruppen wie Entwickler- und Anwendergruppen sowie Anwendungsszenarien, Datenschutz und Schädigungspotenzial (siehe Abbildung 8). Die Randbedingungen, Strukturen und Schnittstellen eines möglichen Einsatzorts des Systems auf KI-Basis sollen innerhalb der Testumgebung nachgebildet werden. Die Testumgebung soll von der äußeren Umgebung getrennt sein, sodass Verzerrungen in Ergebnissen vermieden werden können. In der Testumgebung sollen Ablauffolgen für Prüfungen unterschiedlicher Tiefen gewährleistet sein. Die Prüftiefe kann in Anlehnung an Einsatzrisiko, Komplexität der KI-Anwendung sowie Aufwand und Kosten festgelegt werden. Um die Konformität eines Systems nachvollziehbar darlegen zu können (Konformitätsbewertung), ist es erforderlich, die zugrunde liegenden Anforderungen unmissverständlich festzulegen. Für Systeme auf KI-Basis sollte ein Anforderungskatalog entwickelt werden, in welchem Aspekte wie Systemanforderungen, Systemarchitektur, Softwareanforderungen, Softwarearchitektur, Struktur des Quelltexts, Modulaufbau, Softwareintegration, Qualität der Software, Qualität der Trainings- und Testdaten, Systemintegration und Systemqualität dokumentiert sind [161]. Auf Grundlage eines Handlungsrahmens können KI-Methoden und -Fähigkeiten in Abhängigkeit von Qualitäts- und Validierungsanforderungen einer Konformitätsbewertung mit Blick auf eine angemessene Eignung unter Einbezug von ethischen, rechtlichen sowie sozialen Bewertungsschemata unterzogen werden. Hierbei können KI-Anwendungen auf Grundlage der Klassifikationsmatrix aus Kapitel 4.1.2 beschrieben werden. Ähnlich wie bei der Feststellung der Messfähigkeiten bei der Kalibrierung/Prüfung von Messgeräten anhand rückgeführter, validierter metrologischer Normale und Referenzen können Referenzdaten, Benchmarks und Referenzverfahren in bestimmten Bereichen ein wichtiger Teil der Prüfung sein. Beispielsweise können in der EKG-Analyse validierte Benchmarks mit der KI-Methode mit vorher nicht bekannten Testdaten durchgeführt und mit dem Ergebnis von Referenzverfahren verglichen werden.

Bei der Prüfung von KI-Systemen lassen sich zwei Ansätze verfolgen: Durch Prozessprüfungen lassen sich Qualitätsstandards für Betrieb und Entwicklung des KI-Systems prüfen, während Produktprüfungen zugesicherte Eigenschaften von KI-Systemen verifizieren. Beide Prüfungsansätze müssen sich in ein übergeordnetes Prüfungsframework einfügen, welches die Vergleichbarkeit der Prüfungen von unterschiedlichen KI-Systemen sicherstellt. Dieses Prüfframework sollte offen in Hinblick auf die Auswahl sich anschließender Prüfverfahren sein, jedoch für etablierte Prüfverfahren anschlussfähig und kompatibel sein. Beispiele für etablierte Prüfverfahren sind die Konformitätsbewertung des New Legislative Frameworks (NLF) oder CC [47].

4.3.2.3.2 Prozessprüfungen

Das Produkt KI bringt einige neue Herausforderungen mit sich. Unter anderem ist, in Abhängigkeit vom eingesetzten Verfahren, Transparenz oder Nachvollziehbarkeit in Bezug auf eine Entscheidung, die von einer KI getroffen wurde, nur eingeschränkt möglich. Daher ist die Transparenz in Bezug auf den Entwicklungsprozess der KI in Form einer Prozessprüfung umso wichtiger [162]. Zudem ist zu beachten, dass KI-Produkte oftmals in Form von internetbasierten Diensten zur Verfügung gestellt werden bzw. auf Internet- und Clouddienste zugreifen. Solche Dienste werden häufig fortlaufend aktualisiert: Eine Prüfung von KI-Produkten, insbesondere dienstbasierten KI-Produkten, sollte deshalb durch ein Audit der Prozesse der Organisation ergänzt werden, die diese Produkte zur Verfügung stellen. Zudem sollten auch Organisationen, die KI-Produkte nutzen, in der Lage sein, einen Nachweis ihres verantwortungsvollen Umgangs mit solchen Technologien beispielsweise in Form eines Audit-Berichts oder eines entsprechenden Zertifikats zu erhalten.

4.3.2.3.2.1 Abschätzung der Folgen des Einsatzes von KI

Am Anfang dieser Prozesse steht neben den an die KI-Herausforderungen angepassten Anforderungen auch eine Betrachtung der Erwartungen, Ansprüche und Befürchtungen weiterer betroffener Parteien, also z. B. Kunden und Partner einer Organisation, Endnutzer von KI-Produkten usw. Organisationen sollten in der Lage sein, die Auswirkungen und Folgen der Nutzung solcher Produkte zu verstehen und ggf. mit den eigenen Zielsetzungen in Einklang zu bringen: Ein solches erweitertes Management von Risiken des Einsatzes von KI, das neben Risiken für eine Organisation insbesondere auch Auswirkungen auf Dritte betrachtet, sollte durch entsprechende Management-Funktionen, -Rollen und -Verantwortlichkeiten implementiert und überprüfbar dokumentiert werden.

4.3.2.3.2.2 Entwicklungsprozess von KI-Systemen

Transparenz in Bezug auf den Entwicklungsprozess der KI in Form einer Prozessprüfung sollte die Dokumentation wichtiger Entscheidungen in Bezug auf die Auswahl bestimmter Kriterien und Indikatoren (beispielsweise Metriken, Accuracy, Precision, Recall, Specificity und Sensitivity) beinhalten. Weiterhin sollten Anforderungen an kontinuierlich lernende KI-Systeme angemessen gestaltet (goal alignment) und dokumentiert werden. Da der Trainingsprozess einen wesentlichen Einfluss auf die Qualität einer KI hat, ist ein Trainingsfortschritt zu gewährleisten. Hierzu ist eine Versionierung der Software inklusive der zum Training verwendeten Daten unumgänglich. Neben der Versionierung der Software ist die Dokumentation zu zentralen und systemrelevanten Entscheidungen wichtig, wie z. B. Entscheidungen und Entscheidungsänderungen in Bezug auf die Modellwahl, die Datenaufbereitung (Feature Engineering) sowie die Einteilung in Trainings- und Testdaten. Vor der Validation einer KI ist die Dokumentation zu Tests und Verifikation abzuschließen.

4.3.2.3.2.3 Nutzung von KI-Systemen und ihre Bereitstellung als Dienstleistungen

Prozesse, die in der Nutzung von KI-Systemen insbesondere auch bei ihrer Bereitstellung als Dienstleistungen zum Tragen kommen, umfassen die kontinuierliche Prüfung und Bewertung von Leistungs- und Sicherheitsmetriken, die Bestimmung angemessener Reaktionen auf Zwischenfälle und die Etablierung geeigneter Gegenmaßnahmen. Neben diesen generischen Prozessen ist beispielsweise im Zusammenhang mit KI noch zu betrachten und durch entsprechende Managementprozesse zu unterlegen:

- → Die Auswirkungen automatischer Entscheidungen, die durch KI-Systeme getroffen werden, und der damit einhergehende Kontrollverlust.
- → Der Verlust organisatorischen Wissens, der durch den Einsatz automatisierter Entscheidungssysteme verursacht werden kann, und damit einhergehend eine starke Bindung an solche Systeme („blindes Vertrauen“).
- → Die Möglichkeit, dass Dienste von Dritten zu Zwecken verwendet werden, die innerhalb des ethischen Selbstverständnisses einer Organisation fragwürdig sind.
- → Der Umgang mit eingeschränkter Transparenz und Erklärbarkeit von KI-Systemen.

Prozessprüfungen sollten auf etablierten MSS beruhen (z. B. ISO 9001 [120], ISO/IEC 27001 [122], ISO/IEC 27701 [163], usw.); allerdings decken solche Standards, soweit sie zurzeit veröffentlicht sind, lediglich Teilbereiche der Entwicklung und Nutzung von KI ab (Qualität, Sicherheit, Datenschutz usw.). Die Entwicklung eines eigenständigen MSS für KI, der bereits im Kapitel 4.1 diskutiert wurde, wird deshalb empfohlen. Dieser kann ergänzend zu Produktprüfungen (vgl. Kapitel 4.3.2.3.3) zur Konformitätsbewertung und Zertifizierung als Ergebnis eines Audits verwendet werden.

4.3.2.3.2.4 Qualitätssicherung nach Auslieferung durch Produktbeobachtung

Als qualitätssichernde Maßnahme während der operativen Nutzung im KI Life Cycle (siehe Kapitel 4.1.2.3.1) soll bei KI-Systemen eine aktive Produktbeobachtung mit Auswertung gewonnener Felddaten normativ definiert werden, wie sie bereits für sicherheitskritische Systeme in der Automotive [164] sowie Luft-, Raumfahrt und Verteidigung [165] praktiziert wird.

Um sicherzustellen, dass erkannte Probleme und Risiken bei der Anwendung eines KI-Systems, z. B. unethisches oder unnötig gefährdendes Verhalten im operativen Umfeld, zu entsprechenden Korrektur- und nachhaltigen Verbesserungsmaßnahmen führt, ist eine Rückkopplung der durch die Produktbeobachtung gewonnenen qualitätsrelevanten Information an die entsprechende Wirkstelle im KI Life Cycle notwendig. Diese Rückkoppelung muss die Möglichkeit von Warnmeldungen an Kunden und Behörden wie auch einen Produktrückruf einschließen.

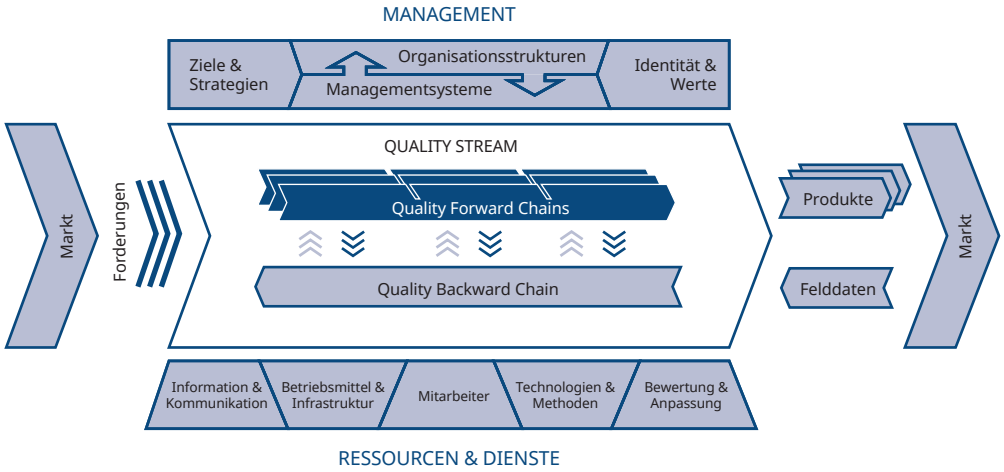


Abbildung 20: Das Aachener QM-Modell mit „Felddaten“ und „Quality Backward Chain“ [168]

Eine systematische Rückkopplung zur Produkt-, System- und Prozessverbesserung durch interne [166] und externe [167] qualitätsrelevante Informationen ist z. B. im „Aachener QM Modell (AQMM)“ [168] als „Quality Backward Chain“ beschrieben (siehe Abbildung 20).

Für KI-Systeme kann diese Produktbeobachtung dadurch erfolgen, dass Bewertungsergebnisse zusammen mit den dazugehörigen Quelldaten (z. B. Sensordaten) gespeichert, an den KI-Entwicklungsbetrieb übertragen und dort ausgewertet werden. Bei Verkehrsflugzeugen erfolgt dies heute zum Teil während des Fluges oder beim elektrischen Koppeln des Flugzeugs mit dem Gate am Flughafen, mit rund 1 GByte Daten pro Flugstunde. Auch aktuelle Kraftfahrzeuge verfügen über integrierte Mobilfunkschnittstellen zur Übertragung rudimentärer Felddaten an den Fahrzeughersteller und zum Teil auch über digitale Datenlogger, um zumindest einen bestimmten Zeitraum vor einem Schadensfall aufzuzeichnen. Bei KI-Systemen kann die benötigte Datenmenge sehr umfassend sein, sodass eine praktikable Lösung für jede KI-Anwendung während der Entwicklung bestimmt werden muss.

4.3.2.3.3 Produktprüfungen

Zusätzlich zu Prozessprüfungen, welche die Compliance mit guten Standards für die Entwicklung und den Betrieb von KI-Anwendungen sicherstellen, bedarf es Produktprüfungen, welche die Eigenschaften eines KI-Systems selbst prüfen. Dabei kann die Prüfung einerseits die durch den Entwickler zugesicherten Produkteigenschaften umfassen, andererseits die Einhaltung bestimmter branchen- oder produktspezifischer Standards bestätigen.

Für eine derartige Produktprüfung wird ein Framework benötigt, nach welchem die Funktionalitäten der KI-Anwendung einheitlich spezifiziert werden können. Zusätzlich werden Bewertungsgrundlagen benötigt, aus welchen hervorgeht, wann die zugesicherten Funktionalitäten als eingehalten gelten können. Diese Bewertungsgrundlagen sollten insbesondere auch eine Übersicht zu gängigen Metriken umfassen, welche die Performanz hinsichtlich unterschiedlicher technischer Eigenschaften (wie z. B. Robustheit gegenüber Adversarial Attacks, Verlässlichkeit etc.) messbar machen. Hierbei stellt sich die Herausforderung, dass die Angemessenheit der verwendeten Metriken eine starke Abhängigkeit vom Einsatzkontext oder dem Use Case aufweisen kann. Einige KI-Modelle (z. B. Word Embeddings) besitzen kein eigenes Gütemaß, sondern können nur in der Anwendung durch weiterführende Verfahren gegenübergestellt werden. Auch dafür sollte eine Lösung gefunden werden.

Beispiel für einen Use Case aus dem Bereich Mobilität Autonomes Fahren

Diese Fragestellungen werden aktuell in ersten Pilotprojekten wie beispielsweise dem BMWi-Projekt KI-Absicherung untersucht und lassen wertvolle Erkenntnisse für die Normung und Standardisierung von KI-Systemen erwarten:

Ziel des Projekts KI-Absicherung ist die Entwicklung und Untersuchung von Methoden und Maßnahmen für die Absicherung KI-basierter Fahrfunktionen für den Use Case „Fußgängererkennung“. Die gewonnenen Erkenntnisse sollen es ermöglichen, die Technologie besser bestimmbar und abschätzbar werden zu lassen. Zudem soll

damit eine stringente Argumentationskette geschaffen werden, die aus Expertensicht eine Absicherbarkeit von KI-Funktionen begründet. Letztlich soll durch Kommunikation mit normativen Gremien und Zertifizierungstellen ein Industriekonsens bezüglich einer KI-Teststrategie unterstützt werden.

Weiterhin bedarf es der Spezifikation von unterschiedlichen Stufen an Vertrauenswürdigkeit (vgl. Kritikalitätspyramide, siehe Kapitel 4.1.2.1.5, bzw. Risikopyramide der Datenethikkomission, siehe Kapitel 4.4.1.2), die durch eine Prüfung gemäß Umfang und Tiefe bestätigt werden. Hierzu ist es erforderlich, ein geeignetes Framework an unterschiedlichen Prüftiefen festzulegen. Das Methodenspektrum umfasst hierbei Dokumentenprüfungen, Audits, Black- und Whitebox-Tests sowie Validierung bzw. Verifikation.

Zur Durchführung solcher Produktprüfungen werden zudem geeignete Werkzeuge benötigt, mit denen sich die Erfüllung von Funktionalitäten sowie die Performanz hinsichtlich einer geeigneten Metrik messbar machen lässt. Diese Prüfwerkzeuge müssen entwickelt werden, zudem bedarf es Kriterien für ihre Bewertung und Zulassung. Darüber hinaus kann für KI-Anwendungen eine Kennzeichnungsanforderung für umgesetzte Methoden und Fähigkeiten etabliert werden, beispielsweise unter Einbezug der Klassifizierungsmatrix für Methoden und Fähigkeiten aus Kapitel 4.1.2.

4.3.2.3.4 Prüfbarkeit-by-Design (Testbarkeit)

Analog zu bestehenden Konzepten wie „Privacy-by-design“ oder „Safety-by-design“ sollen auch Qualitätsanforderungen an KI-Systeme bereits in der Konzipierung der Anwendung berücksichtigt werden.

Hierbei muss der gesamte Lebenszyklus eines KI-Systems von der Spezifikation der Eingangsdaten über die Aufbereitung von Rohdaten zu Trainingsdaten und die repräsentative Modellierung eines zweckbestimmten, domänenspezifischen Wissens bis hin zu den Einsatzszenarien vollständig berücksichtigt werden. Dies gilt insbesondere auch für die Transparenzanforderungen.

Die Konzepte und Standards einer Prüfbarkeit-by-Design für KI-Anwendungen sind ein mittelfristiger Forschungsgegenstand. Mit Blick auf die eingangs genannten Qualitätseigenschaften ergeben sich beim Einsatz von KI-Systemen zumindest die folgenden grundsätzlichen Forschungsfragen:

- → Wie kann eine KI-spezifische FMECA (Failure Mode and Effects and Criticality Analysis) durchgeführt werden?
- → Wie muss diese (FMECA) über die Entwicklungszeit hinweg aktualisiert werden?
- → Zu welchen Zwecken werden ein KI-System und dessen KI-Komponenten eingesetzt? Welche Anforderungen entstehen daraus für das prüfbare Design?
- → Welche KI-Modelle werden für die verwendeten KI-Komponenten genutzt? Gibt es standardisierte Designs, die prüfbar sind?
- → Sind an der Entscheidungsfindung bzw. Prognose durch eine KI-Komponente Menschen beteiligt und, wenn ja, in welcher Form? Welches sind die Verantwortlichkeiten im Hinblick auf die Eingangs- und Ausgangsgrößen der KI-Anwendung?
- → Wie erfolgt die Auswahl der KI-Modelle, der Implementation und der Trainingsmethoden? Welche Anforderungen ergeben sich für ein prüfbares Design der Anwendung?
- → Welche Testmethoden sind relevant, wie kann die KI-Komponente getestet werden, ob sie die zweckgemäßen Eigenschaften aufweist,
- → Und wie wird der laufende Betrieb dieser Komponente im Hinblick auf die Einhaltung des Zwecks überwacht? Welche Schlussfolgerungen sind für ein prüfbares Design zu ziehen, um die Prüfungen zu vereinfachen?

Es ist ferner zu erwarten, dass gewisse Qualitätseigenschaften von KI-Systemen einfacher zu überprüfen sind bzw. ihre Überprüfung überhaupt erst ermöglicht wird, wenn entsprechende Anforderungen hierfür bereits bei der Konzeption und weiteren Entwicklung der KI-Systeme berücksichtigt werden. Mögliche Ansatzpunkte sind beispielsweise Dokumentationen des Entwicklungsprozesses, Logging von (Zwischen-) Ergebnissen oder Schnittstellen für entsprechende Prüfwerkzeuge (siehe 4.3.2.3.6).

4.3.2.3.5 Prüfinfrastruktur für Konformitätsbewertung und Zertifizierung

Um die hier formulierten Qualitätsanforderungsbedarfe prüfen zu können, bedarf es einer Prüfinfrastruktur bestehend aus Prüfstellen, technischen Prüfern sowie den dafür benötigten Akkreditierungsmechanismen und -stellen. Die Akkreditierungsmechanismen sollten insbesondere sicherstellen, dass Prüfstellen und Prüfer über ein fundiertes

technologisches Verständnis verfügen, um diese Prüfungen durchzuführen. Beim Aufbau der Prüfinfrastruktur sollte so weit wie möglich auf die bereits existierende technische IT-Prüfinfrastruktur zurückgegriffen werden, um marktfähige Prüfungen zu entwickeln und Anschlussfähigkeit an bestehende Prüfverfahren herzustellen. Zur Zertifizierung von Personen kann die Kompetenz von bereits etablierten, akkreditierten Zertifizierungsstellen mit Blick auf Methoden und Fähigkeiten von Künstlicher Intelligenz ausgeweitet werden.

Die Zertifizierung kann auf Grundlage einer potenziell aktualisierten Variante der Norm ISO/IEC 17024 [41] erfolgen. Zum Nachweis spezifischer Kompetenzen sollten weitere Dokumente wie Empfehlungen, Verordnungen und weiterführende Normen mit Blick auf KI erstellt beziehungsweise erweitert und hinzugezogen werden.

Der Einsatz von innovativen, KI-gestützten Prüfdienstleistungen erfordert einen Nachweis über eine vorhandene Fachkompetenz von Prüfern, Fachexperten, Begutachtern und Auditoren, um eine Qualitätssicherung gewährleisten zu können. Abgesehen von der Validierung technischer Aspekte soll das Schädigungspotenzial einer KI-Anwendung auf Grundlage ethisch-rechtlicher Prinzipien durch qualifizierte Personen eingeschätzt werden können.

4.3.2.3.6 Neue Prüfmethoden und neue Prüfwerkzeuge

Methodische Ansätze

Nach einer Spezifikation des zu prüfenden Systems bietet sich die Prüfung beispielsweise mit Blick auf maschinelle Lernverfahren trainingsbegleitend sowie an einem austrainierten System an. Dies kann über die Analyse des Eingabe- und Ausgabeverhaltens von Modellen zur Bewertung von Invarianz, Regularität sowie Äquivalenz erfolgen. Hierfür bieten sich beispielsweise Sensitivitätsanalysen an. Bei trainingsbegleitetem Lernen kann darüber hinaus eine Lernkurve nachverfolgt und intentionell hinsichtlich Deklaration, Fehlerwahrscheinlichkeit und Adaptivität bewertet werden. Bei austrainierten Systemen lassen sich Key-Performance-Indizes einbeziehen, welche Kriterien für Eignung und Ausschluss der KI für Forschungszwecke oder einen Markt bewerten. Damit soll über interpretierbare Qualitätsmerkmale festgelegt werden können, in welchem Umfeld einzelne Methoden und Fähigkeiten der KI nutzbar sind.

Am Beispiel des Ansatzes von LIME liegt das Ziel in der Erklärbarkeit von Systemen auf Basis des maschinellen Lernens [44]. Des Weiteren können für einzelne KI-Methoden Modelle für die Interpretation der Lernmechanismen einbezogen werden. Für mehrschichtige neuronale Netzwerke bieten sich hierbei die Verfahren „Activation Maximization“ und „Deep Taylor Decomposition“ an [169].

Methoden wie LIME, Shapley [170], DeepLIFT [171] und QII [172] können oft nur auf strukturierte Daten angewendet werden. Methoden für unstrukturierte Datensätze von Datentypen wie Texte, Bilder und Audio sind zurzeit im frühen Entwicklungsstadium.

Eine Verifikation des Quelltextes von KI-basierten Systemen ist über konventionelle Software-Testverfahren nur beschränkt möglich. Darunter fallen die statistische Code-Analyse (Grammatech’s CODESURFER), Runtime Verification (Java Pathfinder) oder Model Checking (SPIN model checker) [173].

Für unterschiedliche Klassen neuronaler Netze lassen sich unterschiedliche Verifikationen angeben, die aus verschiedenen Theorien der Logik und Mathematik ableitbar sind. Dazu gehören Verifikationsverfahren, die auf der Erfüllbarkeit von Formeln der Booleschen Aussagenlogik (Satisfiability Theories, SAT), Erfüllbarkeit von Formeln der Prädikatenlogik 1. Stufe (Satisfiability Modulo Theories, SMT), Reduktion auf lineare Probleme (Mixed Integer Linear Programming, MIP) und Robustheit von mehrschichtigen Perzeptron-Netzen (Multi-Layer Perceptron, MLP) beruhen. Bei SAT- und SMT-Verifikationen wird die klassische KI (symbolische KI) des automatischen Beweisens (automated reasoning) mit dem ML verbunden. MIP beruht auf der Logik und Algebra linearen Programmierens. Robustheitsuntersuchungen von MLP wenden Erkenntnisse aus der Theorie komplexer dynamischer Systeme im ML an [174]. Verifikationsverfahren für subsymbolische KI-Systeme und ML bedürfen neuer Techniken, die wegen ihrer Parameterexplosion (z. B. neuronale Netze beim autonomen Fahren) äußerst rechenintensiv sind.

IT-Sicherheitstests für KI-basierte Systeme

Einen wesentlichen Aspekt der Prüfungen für Konformitätsbewertungen und Zertifizierungen stellen Sicherheitsprüfungen dar, die in statische und dynamische Prüfungen unterteilt werden. Hier spielen dynamische Sicherheitstests eine zentrale Rolle, die ein großes Spektrum an Methoden und Techniken bieten. Einen knappen Überblick bieten Übersichten wie beispielsweise das Dokument ETSI TR 101 583 [175]. Hier findet sich eine Aufzählung und

Erläuterung relevanter Methoden und Ansätze für Sicherheitstests, wie beispielsweise Risikoanalyse und risikoadaptiertes[24] Sicherheitstesten, funktionales Testen von Sicherheitsfunktionen, Performancetesten, Robustheitstesten und Penetrationstesten.

Es existiert eine Vielzahl an Techniken, die für die Sicherheitstests traditioneller Softwaresysteme entwickelt wurden. Diese lassen sich, wenn überhaupt, nur bedingt auf KI-basierte Systeme anwenden. Sicherheitstests für klassische Systeme können teilweise für KI-basierte Systeme adaptiert werden, beispielsweise lässt sich das weitverbreitete Fuzzing auch für KI-basierte Systeme in modifizierter Form einsetzen (vgl. beispielsweise [176], [177]). Um die für KI-basierte Systeme spezifischen Sicherheitsrisiken und -angriffe abzudecken, sind also neue Techniken und Ansätze nötig, die KI-spezifischen Aspekten Rechnung tragen, beispielsweise der Relevanz von Trainingsdaten. Die Techniken, die auf KI-spezifische Sicherheitsaspekte eingehen, werden unter dem Begriff Adversarial ML (AML) zusammengefasst [178]. Eine besondere Hürde stellen bei Sicherheitstests noch Abdeckungskriterien dar. Hierzu gibt es zwar eine Reihe von veröffentlichten Metriken. Jedoch wurde in einer Metastudie nur eine geringe Korrelation zwischen den bestehenden, für KI-Systeme entwickelten Metriken und der Robustheit gegen Angriffe festgestellt, wenn diese Metriken bei Tests berücksichtigt wurden [179]. Ein Überblick über bestehende Techniken und Metriken inklusive Anwendungshinweisen ist derzeit Gegenstand des laufenden Projekts ETSI DGS SAI003 „Security Testing of AI“.

4.3.2.4 Nationales Umsetzungsprogramm zur Normungsroadmap KI

Durch die rasante Verbreitung und hohe Komplexität von KI-Systemen entstehen branchenübergreifend neue technologische Herausforderungen. Die in KI-Entwicklungen vorherrschende Dynamik setzt einen stabilen Handlungsrahmen für alle Akteure in Forschung, Wirtschaft und Gesellschaft voraus, um die vorhandene Innovationskraft gemeinsam richtungsweisend einzusetzen und den wirtschaftlichen und gesellschaftlichen Nutzen des Einsatzes von KI-Systemen konvergierend zu fördern. Zur Operationalisierung der Handlungsempfehlungen der Normungsroadmap KI, welche die technischen Anforderungen an KI-Systeme betreffen, wurde beschlossen, ein nationales Umsetzungsprogramm vorzuschlagen. Die Mission dieses Umsetzungsprogramms ist die zeitnahe und bedarfsgerechte Entwicklung solcher Prüf- und Qualitätssicherungsstandards als zentrale technische Bestandteile des erforderlichen Handlungsrahmens und ihrer zukunftssicheren Fortschreibung auf der Grundlage des wirtschaftlichen und technischen Fortschritts.

Mit Blick auf den Erfolg der Mission wird das Programm die kurz- und mittelfristige Nachfrage der Wirtschaft nach einer Operationalisierung der technischen Aspekte der Normungsroadmap KI und den langfristigen Forschungsbedarf zu Fragestellungen der KI-Absicherung ausbalancieren.

In Anlehnung an die oben zusammengefassten Ergebnisse der NRM KI verfolgt das Programm folgende Ziele:

- → Entwicklung der **technologischen Grundlagen** für eine **neue Generation von KI-Systemen**, die **resillient & vertrauenswürdig „by-Design“** sind.
- → Entwicklung von **erweiterbaren Prüfkriterien** auf der Basis etablierter und zu entwickelnder **Prüftechnologien** auf Basis einer **einheitlichen Terminologie**.
- → **Evaluation dieser Prüfgrundlagen** in Pilotprojekten mit industriereifen, hybriden KI-Lösungen im Zuge eines **kontinuierlichen Verbesserungsprozesses mit breiter Beteiligung**.
- → Ableitung und Entwicklung von **Referenzarchitekturen und Prüfprofilen** für Use Cases und KI-Technologien mit dem Ziel der **Reduzierung von Prüfaufwänden**.
- → Entwurf und Etablierung einer **Prüfinfrastruktur** für die Testierung (Konformitätsprüfung) und für die **Zertifizierung auf Bundesebene** auf **Basis bestehender Prüfinfrastrukturen**.
- → **Standardisierung und Normung** der Prüfgrundlagen und ihre Einordnung auf Basis bestehender Normen und Kriterienwerke. **Etablierung** des Prüfstandards auf **europäischer Ebene**.

Die Basis für das Umsetzungsprogramm bildet das gemeinsame Programm CERTIFIED AI des Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS und des Bundesamtes für Sicherheit in der Informationstechnik (BSI), existierende internationale Normen und Standards und ihre Umsetzung im Bereich der IT-Sicherheit durch das BSI sowie die Forschungsaktivitäten des Deutschen Forschungszentrums für Künstliche Intelligenz (DFKI).

24 entspricht Englisch „risk based“

4.3.3 Normungs- und Standardisierungsbedarfe

Die folgenden Hinweise und Bedarfe betreffen die Umsetzung der Normungsroadmap KI auf der Grundlage des in Kapitel 4.3.2.4 genannten Konzepts.

BEDARF 1:
Umsetzungsprogramm
Im Zentrum der gesamten Prüfsystematik steht die technische Prüfung von Anforderungen an KI-Algorithmen, Modellen, Methoden und Daten nach international gültigen Standards. Die Entwicklung solcher Normen und Standards muss im Fokus der Umsetzung der Normungsroadmap KI stehen. Zu diesem Zweck sollte ein nationales Umsetzungsprogramm auf der Basis der in Kapitel 4.3.2.4 genannten Initiative ins Leben gerufen werden.

BEDARF 2:
Beziehung zwischen technischen Anforderungen einerseits und rechtlichen und ethischen Anforderungen andererseits
Technische Prüfungen sind für das Vertrauen und die Akzeptanz des Einsatzes von KI entscheidend. Technische Produkteigenschaften müssen erkannt und hinreichend evaluiert werden, bevor ihre rechtliche oder normative Zulässigkeit für den Einsatz festgestellt werden kann. Das Umsetzungsprogramm sollte zunächst die Zertifizierung technischer Anforderungen industriereifer KI-Anwendungen mit vertretbarem Prüfaufwand ermöglichen. Die Konformität nichttechnischer Anforderungen an KI-Systeme kann dann methodisch getrennt mit eigenen Kriterienkatalogen überprüft werden.

BEDARF 3:
Einbettung in existierende Prüfschemata und Prüfinfrastrukturen
Um eine zeitnahe Umsetzung der NRM KI in Kooperation mit Forschung, Industrie, gesellschaftlichen Gremien und staatlichen Stellen zu realisieren, sollten existierende Prüfverfahren sowie Prüfinfrastrukturen genutzt werden. Die Entwicklung von Kriterien und Methoden muss auf der Grundlage existierender, international gültiger Prüfstandards und Prüfverfahren erfolgen.

BEDARF 4:
Entwicklung von KI-Standards als Beteiligungsprozess
Innerhalb des Umsetzungsprogramms sollten die erforderlichen Prüfkriterien und Prüfmethoden für technische Prüfungen von KI-Lösungen in einem breiten Beteiligungsprozess entwickelt und getestet werden, um Vertrauen und Akzeptanz in Wirtschaft und Gesellschaft zu erhalten. Die bewährte Vorgehensweise der Normungsroadmap KI sollte gespiegelt werden, indem unter Beteiligung der Wirtschaft Pilotprojekte für Use Cases durchgeführt werden, Prüfverfahren verbessert und angepasst werden und schrittweise Normungsreife erreicht wird.

BEDARF 5:
Notwendigkeit eines Managementsystem-Standards
Die Ganzheitlichkeit der Umsetzung und die Interoperabilität auf der Ebene der Prüfkonzepte sollte normungstechnisch durch ihre Einbettung in ein umfassendes Managementsystem gewährleistet werden, das organisatorische, technische und prozessbezogene Prüfverfahren ebenso wie Prüfschemata für unterschiedliche rechtliche und ethische Fragestellungen über den gesamten Lebenszyklus von KI-Systemen berücksichtigt.

BEDARF 6:
KI-Normen: Smarte Assistenten für Behörden und öffentliche Ämter
Die demokratische Grundordnung ist eine weitere Stärke unserer Gesellschaft. Sie baut u. a. auf eine Vielzahl von historisch gewachsenen Behörden, Ämtern und Verwaltungsprozessen auf, die unsere Werte und Normen in die gelebte Praxis umsetzen. Diese soll nun für das Internet geöffnet werden. Das Onlinezugangsgesetz verpflichtet die öffentliche Hand, viele der Bürgerprozesse zu digitalisieren. Im Herzen dieser Anstrengung werden smarte Assistenten entstehen, die auf der Basis von KI eine neue Form der Mensch-Maschine-Schnittstelle bzw. Bürger-Amt-Schnittstelle etablieren. Smarte Assistenten können erhebliche Vorteile gegenüber dem „klassischen Ämtergang“ haben: Bürgernähe, 24/7-Verfügbarkeit, Schnelligkeit, einheitliche Qualität, direkte Verarbeitung der digitalen Unterlagen, Automatisierung, Kosteneinsparungen, Barrierefreiheit, leichtere Bedienbarkeit für ältere Menschen und Menschen mit Behinderung und vieles mehr. Diese gilt es zu entwickeln, am besten auf der Basis einheitlicher KI-Standards. Viele davon werden auch die deutsche Sprache betreffen, was besonders wichtig ist. Warum sollen z. B. nicht alle Rathäuser eines Bundeslandes den gleichen KI-Standard haben, um einen Chatbot produktiv zu setzen, der die Öffnungszeiten, Verantwortlichkeiten und Termine erklärt? In diesem Feld gibt es umgehenden Handlungsbedarf. Arbeitsteilung zwischen den Behörden, klare Verantwortlichkeiten und eine KI-Roadmap der „kleinen schnellen Schritte“ sind für den Erfolg entscheidend. KI-Normen der öffentlichen Hand machen unser Gemeinwesen fit

für die Zukunft und helfen, unsere Werte und Prozesse digital auszubauen.

BEDARF 7:
Forschungsbedarf
Die Entwicklung hochwertiger Prüfstandards und Prüfmethoden erfordert die massive Unterstützung durch die relevante KI-Forschung in Deutschland. Dabei besteht für eine vollständige Operationalisierung der Handlungsempfehlungen in manchen Bereichen noch erheblicher Forschungsbedarf, z. B. im Bereich der Verifikation von neuronalen Netzen. Die Forschungsschwerpunkte müssen parallel zur Umsetzung bearbeitet und wechselseitige Synergien aus Forschung, Umsetzung und Standardisierung bestmöglich genutzt werden.

4.4 IT-Sicherheit bei KI-Systemen

Sicherheit und insbesondere IT-Sicherheit gelten häufig als Spielverderber oder sogar Verhinderer von Innovation; die Historie zeigt allerdings, dass Sicherheit immer ein Begleiter und Förderer von Innovationen war. Nicht umsonst bestehen so vielfältige Normen, Standards und Gesetze für Sicherheit, die erst eine umfassende und vertrauensvolle wirtschaftliche Nutzung ermöglicht haben. Ohne umfängliche Sicherheit und Risikominimierung fährt heute kein Auto, fliegt kein Flugzeug, findet keine Operation statt und wird kein Haus, keine Brücke oder Straße gebaut. Auch in mehr virtuellen Bereichen wie beispielsweise Elektrizität wurde die Sicherheit umfänglich geregelt. Der Mensch vertraut auf sicherheitsgeprüfte Kaffeemaschinen oder sichere Komponenten in Atomkraftwerken sowie geschultes Personal. Innovationen werden nicht verhindert, sondern ihre wirtschaftliche Nutzung möglich gemacht, indem für Sicherheit im Einsatz gesorgt wird. Für Anwendung von Informationstechnik und insbesondere von KI gilt dies ebenso. Natürlich bedeutet die Berücksichtigung von Sicherheit immer einen zusätzlichen Aufwand und zusätzliche Kosten. Ohne Sicherheit sind die Kosten und der Schaden im Zweifel jedoch deutlich höher, siehe den jüngsten Fall von Stilllegung der Produktion bei Honda [180]. Hinzu kommen mögliche politische Dimensionen wie die massiven Cyberangriffe auf Behörden und Unternehmen in Australien, die im Juni 2020 bekannt wurden [181].

Wie viel Sicherheit erforderlich ist, ist eine Abwägung des Aufwands zur Akzeptanz eines möglichen Schadens und kann erst entschieden werden, wenn die Abwägung (Risikofolgeabschätzung) stattgefunden hat. Dies gilt auch und besonders für den Einsatz von KI, da hier zusätzliche Unwägbarkeiten durch stochastische Ergebnisse und Dual-Use-Möglichkeiten hinzukommen. Um Innovationen also breit im Markt wirtschaftlich umsetzen zu können, muss Vertrauen geschaffen werden, z. B. durch Nachweis der IT-Sicherheit.

Grundprinzipien der IT-Sicherheit

Der Begriff IT-Sicherheit wird ambivalent verwendet. Daher ist es wichtig, bei der Betrachtung des Themas zunächst die drei wichtigsten Teilaspekte zu klären.

Safety bezieht sich auf die Erwartung, dass ein System unter bestimmten Umständen nicht zu einem Zustand führt, in dem menschliches Leben, Gesundheit, Eigentum oder die Umwelt gefährdet sind.

Ein IT-System ist funktionssicher (safe), wenn seine tatsächliche Ist-Funktionalität mit der gewünschten, spezifizierten Soll-Funktionalität übereinstimmt und das System keine unerlaubten Zustände annimmt.

Ein funktionssicheres System ist informationssicher (secure), wenn es nur solche Zustände annimmt, die zu keiner unerlaubten Informationsgewinnung oder Informationsveränderung führen. Ein funktionssicheres System ist datensicher, wenn es nur solche Zustände annimmt, die kein unautorisiertes Erzeugen, Löschen, Lesen oder Verändern von Datenobjekten zulassen. IT-Systeme, die informationssicher und datensicher sind, werden als zuverlässig (reliable) bezeichnet.

Security hat zum Ziel, negative Auswirkungen, die ein Mensch oder eine andere Maschine auf das KI-Modul haben kann, zu verhindern. Vertraulichkeit, Integrität und Verfügbarkeit sind die wichtigsten Sicherheitsziele.

Für solche Systeme gelten neben Anforderungen an die Funktionssicherheit im Allgemeinen spezielle Anforderungen an die Vertraulichkeit von Informationen. Die Integrität von Daten (Datenschutz) im Zusammenhang mit IT-Systemen beschreibt die Kontrolle einer natürlichen Person – als sozio-technisches Systemsubjekt – über die Weitergabe personenbezogener Informationen und die Verfügbarkeit von Objekten und Subjekten.

Privacy/Datenschutz (Datensicherheit) bezieht sich auf die Erhebung und Verarbeitung personenbezogener Daten gemäß den einschlägigen Vorschriften wie der EU-Datenschutz-Grundverordnung. Beispielsweise haben betroffene Personen in Europa das Recht, dass ihre privaten Daten ausreichend vor IT-Angriffen geschützt werden.

Diese drei Aspekte – Saftey, Security, Privacy – stehen in Beziehung zueinander und können sich gegenseitig unterstützen, tragen aber auch das Potenzial von Zielkonflikten in sich. So kann sich beispielsweise ein hohes Maß an Datenschutz negativ auf das Sicherheitsziel der Verfügbarkeit auswirken. In dem Fall des Germanwing-Absturzes 2015 wurde ein „Security-Feature“, der vor Terorristen (Angriffen) schützende Verriegelungsmechanismus des Cockpits, zu einem „Safety Problem“, das das Leben der Passagiere bedroht. Diese Zusammenhänge müssen bei Design und Betrieb analysiert und berücksichtigt werden. Klassischerweise geschieht dies im Rahmen einer Risikoanalyse.

Jedes KI-System erfordert eine individuelle Analyse seiner Sicherheit. Weitere Forschung erscheint in diesem Umfeld erforderlich und der Industrie wird empfohlen, die entspre-

chenden Security-Level mitzuentwickeln. Die geltenden Regulierungen, Normen und Standards für IKT-Systeme sind zu berücksichtigen.

KI-Systeme sind aus Sicht der IT-Sicherheit spezielle IT-Systeme, auf die die Grundprinzipien der Informationssicherheit uneingeschränkt Anwendung finden.

In der IT-Sicherheitsforschung versteht man unter IT-System ein technisches System zur Speicherung und Verarbeitung von Informationen. Ein IT-System ist geschlossen, wenn seine Technologie von einer Quelle stammt, wenn es nicht kompatibel zu anderen IT-Produkten ist und wenn seine Ausdehnung räumlich begrenzt ist. Ein IT-System heißt offen, wenn es vernetzt, physisch verteilt und auf der Basis von Standards zum Austausch von Informationen bereit ist. Offene IT-Systeme sind meist nicht zentral administriert, ihre Teilsysteme sind heterogen. IT-Systeme sind Bestandteil soziotechnischer Systeme. Sie sind in gesellschaftliche, ökonomische und politische Strukturen eingebettet und werden für unterschiedlichste Zwecke benutzt, mit denen übergeordnete Absichten verfolgt werden. Bei der Betrachtung von IT-Systemen können normative, gesetzliche und organisatorische Regelungen und Vorschriften und Fragen der individuellen Nutzerakzeptanz und der gesamtgesellschaftlichen Akzeptanz eine Rolle spielen.

IT-Systeme verarbeiten und speichern Informationen, die als Daten dargestellt werden. Datenobjekte besitzen die Fähigkeit, Informationen zu speichern, und werden durch technische Prozesse, d. h. durch aktive Subjekte erzeugt, gelöscht, gelesen und verändert. Datenobjekte, die in ihnen enthaltenen Informationen und die Subjekte zu ihrer Verarbeitung sind die schützenswerten Güter innerhalb eines IT-Systems.

Die Disziplin der IT-Sicherheit umfasst alle Ziele, Verfahren und Maßnahmen, um informationstechnische Systeme so zu entwerfen, herzustellen, zu betreiben und zu erhalten, dass ein Maximum an Schutz gegenüber Bedienungsfehlern, technischem Versagen, katastrophenbedingten Ausfällen und absichtlichen Manipulationsversuchen gegeben ist.

4.4.1 Status quo

Um die Chancen von Künstlicher Intelligenz zum Wohle aller Beteiligten richtig nutzen zu können, sollte man die Risiken kennen und diesen mit entsprechend geeigneten Maßnahmen begegnen – eine Aufgabe der IT-Sicherheit.

KI-Lösungen oder -Systeme sind im Kern komplexe Systeme der Informations- und Kommunikationstechnologie (IKT- bzw. im Weiteren IT- Systeme genannt). Es ist zu erwarten, dass KI-Systeme zum Ziel von Cyberattacken werden oder es bereits sind. Der Digitalverband Bitkom kam in seiner letzten Umfrage 2019 [182] zu dem Ergebnis, dass drei von vier Unternehmen Opfer von Cyberangriffen wurden mit einer Schadenshöhe von mehr als 100 Milliarden pro Jahr.

Die bisher existierenden IT-Sicherheitsanforderungen an ein IT-System sind als Status quo damit auch beim Einsatz von KI zu berücksichtigen.

4.4.1.1 IT-Sicherheitsnormen und -standards

Für IT-Systeme existieren bereits vielfältige Normen, Standards sowie Gesetze und Regulierungen zum Thema Sicherheit, IT-Sicherheit (IT-Security), Safety, Privacy mit unterschiedlichen Historien. Hinzu kommen die Normen und Standards zur Risikoermittlung und -behandlung, die teilweise eigenständig, teilweise enthalten sind. Der Umfang und die Vielfalt stellen eine Herausforderung für Unternehmen und Behörden beim Einsatz von KI und deren IT-Sicherheit dar. Zudem wachsen Themenbereiche und Branchen durch die zunehmende Digitalisierung zusammen, die bisher eigene Normen und Standards für IT-Sicherheit nutzen.

Die beispielhafte Nennung von sicherheitsrelevanten Normen und Standards im Kapitel 6 erhebt keinerlei Anspruch auf Vollständigkeit, zumal KI-Systeme auch in industriellen Produktionsumgebungen (operational IT = OT) und für Aufgaben mit Safety-Anforderungen zum Einsatz kommen.

Bitkom und DIN haben zudem gemeinsam einen Kompass für einen ersten Einblick entwickelt [183]. Weiterhin exisitiert die Normungsroadmap DIN/DKE für IT-Sicherheit [184], die um KI-Themen erweitert werden könnte. Normen, die die Besonderheiten von KI-Systemen in Bezug auf IT-Sicherheit beinhalten, sind noch nicht verfügbar, aber teilweise in Diskussion.

Die weitere Recherche und Harmonisierung für KI ist Teil des Normungsbedarfs.

4.4.1.2 Gesetze und Regulierungen

- Da IT-Sicherheit/Cybersecurity durch die steigende Vernetzung und Digitalisierung für kritische Infrastrukturen und Unternehmen, aber auch für die Verbraucher elementar geworden ist, wurden verschiedene Regulierungen und Gesetze erlassen:

→ auf europäischer Ebene

- NIS-Richtlinie (Netz- und Informationssicherheit) [185];
- DSGVO, Datenschutz-Grundverordnung [95];
- JI-Richtlinie (Justiz und Inneres) insbesondere zur Verarbeitung personenbezogener Daten durch Polizei und Justiz [186];
- Datenschutzrichtlinie für elektronische Kommunikation [187];
- Cyber Security Act [188];
- Maschinenrichtlinie [94];
- Produktsicherheits-Richtlinie [189]
- sowie konzeptionelle Ansätze (enthalten ebenfalls IT-Sicherheits-/Cyber-Security-Themen):
  - Ethik-Leitlinien der HLEG-KI [5]
  - Weißbuch KI der EU-Kommission als Vorlage für eine Regulierung [15]

→ in Deutschland

- IT-Sicherheitsgesetz [190] – aktuell in der Überarbeitung für Version 2.0;
- Zweites Datenschutz-Anpassungs- und Umsetzungsgesetz EU – 2. DSAnpUG-EU [191] (neues angepasstes Bundesdatenschutzgesetz), Regelungen auf Länderebene;
- Telemediengesetz (TMG) [192] für Internetdienste;
- Telekommunikationsgesetz (TKG) [193];
- Gesetz über das Bundesamt für Sicherheit in der Informationstechnik (BSI-Gesetz, BSIG) [194];
- Produktsicherheitsgesetz (ProdSG) [195];
- Energiewirtschaftsgesetz (EnWG) [196];
- diverse branchenspezifische Regulierungen;
- die Deutsche Datenethikkommission beschäftigt sich in ihrem Gutachten [10] u. a. mit der IT-Sicherheit hinsichtlich KI-Systemen.

Aufgrund der Bedeutung für die IT-Sicherheit auch für KI-Systeme wird der EU Cybersecurity Act, das EU-Weißbuch zu KI, die EU-Maschinenrichtlinie und die EU-Datenschutz-Grundverordnung kurz vorgestellt.

EU Cybersecurity Act

Der EU Cybersecurity Act wurde am 17. April 2019 durch die EU beschlossen [188].

Ziel des Cybersecurity Acts ist es, die IT-Sicherheit EU-weit mit einheitlichen Regularien zu etablieren und für sogenannte informations- und kommunikationstechnische (IKT) Systeme, Dienste und Prozesse zu stärken.

IKT-Systeme sollen zukünftig gemäß definierter Sicherheitslevel und deren Vertrauenswürdigkeit in drei Stufen klassifiziert und zertifiziert werden. Die Einordnung erfolgt auf Basis einer Risikoabwägung im Hinblick darauf, wie wahrscheinlich ein Sicherheitsvorfall eintritt und wie er sich auswirkt.

Die Vertrauenswürdigkeitstufen:

Stufe „niedrig“: Die grundlegenden Risiken für Sicherheitsvorfälle und Cyberangriffe werden als gering angenommen. Auf dieser Stufe ist es auch möglich, dass ein Hersteller seine Konformität selbst und alleinverantwortlich bewertet.

Stufe „mittel“: Zertifizierte Produkte, Dienstleistungen und Prozesse sollen bekannten Cybersicherheitsrisiken standhalten können.

Stufe „hoch“: Cyberangriffe können auf dem Stand der Technik gegen Angreifer mit umfangreichen Fähigkeiten und Ressourcen abgewehrt werden. Auf dieser Stufe ist eine Zertifizierung nur durch behördliche Stellen zulässig.

Die 110 Erwägungsgründe des Cybersecurity Act enthalten noch weitergehende Kriterien für die Risikobewertung, Ziele und Basisanforderungen an die Cybersecurity und Mindestbestandteile für die Vertrauenswürdigkeitsstufen. Unter anderem werden in Erwägungsgrund 12 „security by design“, in Erwägungsgrund 13 „security by default“ und in Erwägungsgrund 41 „privacy by design“ sowie in Erwägungsgrund 49 „sorgfältige Risikomethoden und messbare Sicherheit“ gefordert. Die Zertifizierungen sind freiwillig. Die EU-Kommission wird regelmäßig prüfen, ob Cybersicherheitszertifizierungen als verbindlich vorgeschrieben werden sollten, z. B. für Unternehmen im Energie-, Banken- oder Gesundheitswesen.

Der Cybersecurity Act sieht weiterhin für die Umsetzung die ENISA vor und regelt deren Geschäftstätigkeit. Die ENISA hat erweiterte Aufgabenbereiche und Kompetenzen erhalten, darunter auch die Erstellung von Zertifizierungsschemata für Informationssicherheit und die Berücksichtigung bestehender Normen und Standards. In Zusammenarbeit mit nationa-

len Gesetzgebern und Organisationen, Sicherheitsexperten, Herstellern von IKT-Produkten und Anwendern sollen die Sicherheitskriterien über die kommenden Jahre ausgearbeitet werden.

Mögliche IT-Sicherheitsrisiken aus KI sind nicht speziell beschrieben oder berücksichtigt. Eine Prüfung, inwieweit eventuell Ergänzungen erforderlich wären, ist empfehlenswert.

EU-Datenschutz-Grundverordnung (DSGVO)
Die DSGVO [95] stärkt bzw. vereinheitlicht den Datenschutz in IKT-Systemen für Personen. Artikel 5, 24, 25 und 32 enthalten Verantwortlichkeiten, die Erstellung einer Datenschutzfolgeabschätzung (Risikobetrachtung) und Anforderungen an eine datenschutzfreundliche und sichere Technik sowie Organisation (u. a. Pseudonymisierung und Verschlüsselung).

Für automatisierte Entscheidungsfindung z. B. aus Machine-Learning (ML)-Modellen, die Personen betreffen, ist folgender Passus entscheidend: „Werden personenbezogene Daten […] erhoben, so teilt der Verantwortliche […] Folgendes mit: das Bestehen einer automatisierten Entscheidungsfindung […] und […] aussagekräftige Informationen über die involvierte Logik […].“

Zur Bestimmung der Risiken betroffener Personen haben sich die Datenschutz-Aufsichtsbehörden europaweit auf neun Kriterien geeinigt:
1. Bewerten oder Einstufen,
2. automatische Entscheidungsfindung,
3. systematische Überwachung,
4. vertrauliche oder höchst persönliche Daten,
5. Datenverarbeitung im großen Umfang,
6. Abgleichen oder Zusammenführen von Datensätzen,
7. Daten zu schutzbedürftigen Betroffenen,
8. innovative Nutzung oder Anwendung neuer technologischer oder organisatorischer Lösungen,
9. Betroffene werden an der Ausübung eines Rechts oder der Nutzung einer Dienstleistung bzw. Durchführung eines Vertrags gehindert.

Die genannten Risikokriterien und deren Bewertung sind bei Verwendung einer KI relevant, wenn personenbezogene Daten verwendet werden. Über die verwendete Logik müssen aussagekräftige Informationen vorliegen, d. h. Transparenz über die Entstehung der Entscheidung einer KI. In der „Hambacher Erklärung zur Künstlichen Intelligenz“ [43] nehmen die deutschen Datenschutzaufsichtsbehörden zu den Anforderungen der DSGVO in Bezug auf KI konkret Stellung.

EU-Maschinenrichtlinie
Mit der EU-Maschinenrichtlinie [94] werden einheitliche Anforderungen für Maschinen und unvollständige Maschinen für ein einheitliches Schutzniveau zur Unfallverhütung beim Inverkehrbringen derselben geregelt. In Deutschland wurde die Richtlinie in das Produktsicherheitsgesetz (ProdSG) und die darauf gestützte Maschinenverordnung (9. ProdSV) umgesetzt. Folgende Anforderungen müssen umgesetzt werden (Auszug):
- → Die Maschine muss mechanisch und elektrisch sicher gestaltet und die funktionale Sicherheit (z. B. sichere Steuerkreise) muss umgesetzt werden,
- → zum Zeitpunkt des Inverkehrbringens ist die Maschine sicher und eine sichere Bedienung ist gewährleistet,
- → Sicherheits- bzw. Schutzeinrichtungen der Maschine können nicht einfach umgangen werden,
- → Konformitätsbewertungsverfahren mit Risikobeurteilung (§ 158 ff) werden durchgeführt,
- → nach erfolgreicher Bewertung erfolgt die Konformitätserklärung und das **Anbringen des CE-Kennzeichens**,
- → Erstellen einer technischen Dokumentation und Betriebsanleitung, die Benutzer und Bediener der Maschine deutlich auf die gekennzeichneten, vorhandenen Restrisiken aufmerksam macht.

Da, wo KI-Komponenten in oder für „Maschinen“ verbaut werden, gelten die Anforderungen der Maschinenrichtlinie. Spezielle Erwägungen zu Risiken aus KI und zugehörigen Maßnahmen sind nicht enthalten. Dieser Sachverhalt könnte sich in den kommenden Jahren dahingehend ändern, dass neben den Safety- auch Security-Aspekte aufgenommen werden, die dann auch für KI gelten.

EU-Weißbuch KI
Das EU-Weißbuch KI [15] beschreibt auf 31 Seiten die Basis eines möglichen allgemeinen Rechtsrahmens für die Entwicklung und Umsetzung von KI-Anwendungen. Die Kommission greift dafür die Empfehlungen und die sieben Kernanforderungen der „Hochrangigen Expertengruppe für eine vertrauenswürdige Künstliche Intelligenz“ auf:
- → Vorrang menschlichen Handelns und menschlicher Aufsicht
- → Technische Robustheit und Sicherheit
- → Privatsphäre und Datenqualitätsmanagement
- → Transparenz
- → Vielfalt, Nichtdiskriminierung und Fairness
- → Gesellschaftliches und ökologisches Wohlergehen
- → Rechenschaftspflicht

Die Regulierung soll u. a. ein „Ökosystem des Vertrauens“ schaffen. Die IT-Sicherheit im Sinne von Safety, Security und Privacy findet sich in den Kernanforderungen 1 bis 4 wieder. Das Weißbuch stellt fest, dass eine wirksame Umsetzung einer Rechtsvorschrift aufgrund bestimmter Besonderheiten einer KI (z. B. Opazität, Komplexität, Unvorhersehbarkeit und autonomes/teilautonomes Verhalten) erschwert sein kann.

Ein verbesserter, risikoadaptierter[25] Rechtsrahmen und dessen Durchsetzung erscheint der Kommission wünschenswert und soll das Vertrauen in die Sicherheit einer KI und damit deren Vermarktungsmöglichkeiten erhöhen. Die Anwendung des Rechtsrahmens ist grundsätzlich „nur“ für KI-Systeme „mit hohem Risiko“ geplant. Für eine Klarstellung, wann ein hohes Risiko besteht, werden Kriterien vorgeschlagen. Ebenso werden mögliche Maßnahmen sowie eine mögliche verpflichtende Konformitätsfeststellung genannt, z. B. gemäß „Cybersecurity Act“. Für die übrigen KI-Anwendungen sollen die allgemeinen Vorschriften gelten und diese können sich einer „freiwilligen Kennzeichnung“ in Form eines Gütesiegels für Vertrauenswürdigkeit unterziehen. Vorhandene Strukturen sollen berücksichtigt werden, sowohl für Governance als auch Konformitätsbewertung. Das Weißbuch befindet sich aktuell in der Konsultation.

Deutschland: IT-Sicherheitsgesetz
Das Gesetz zur Erhöhung der Sicherheit informationstechnischer Systeme (IT-Sicherheitsgesetz, IT-SiG, 2015, Artikelgesetz u. a. über BSI-Gesetz, EnWG, TMG, TKG) [190] verfolgt als Kernziel die Verbesserung der Verfügbarkeit und Sicherheit von IT-Systemen, digitalen Infrastrukturen und Diensten sowie einen besseren Schutz der Bürgerinnen und Bürger im Internet. Für kritische Infrastrukturen Deutschlands, deren Ausfall oder Beeinträchtigung erhebliche Auswirkungen für Wirtschaft, Staat und Gesellschaft haben, wie z. B. Energie- und Wasserversorgung sowie Gesundheitswesen, enthält es Regelungen zu Mindestanforderungen für die IT-Sicherheit, Nachweispflichten und Meldepflichten. Zurzeit wird das IT-Sicherheitsgesetz 2.0 vorbereitet, mit dem u. a. die operativen Befugnisse des BSI erweitert und weitere Teile der Wirtschaft zur Einhaltung der Mindestanforderungen an die IT-Sicherheit verpflichtet werden sollen. Darüber hinaus ist geplant, Anforderungen an die Vertrauenswürdigkeit von Kernkomponenten der IT-Infrastruktur zu stellen, die KRITIS-Betreiber einsetzen.

Das IT-Sicherheitsgesetz ist auch für KI-Anwendungen zu berücksichtigen, enthält aber keine speziellen Anforderungen für den Einsatz von KI.

Der Industrieverband Bitkom bietet mit seiner Studie aus 2019 [197] ebenfalls einen Überblick.

Gutachten der deutschen Datenethik-Kommission
Auszug aus dem Papier der Datenethik-Kommission [10]: „Robuste und sichere Systemgestaltung umfasst sowohl die Sicherheit des Systems gegen Einflüsse von außen (z. B. durch Verschlüsselung, Anonymisierung etc.) als auch den Schutz der Menschen und der Umwelt vor negativen Einflüssen durch das System (insbesondere durch einen systematischen Risikomanagementansatz, z. B. auf der Grundlage einer Risikofolgenabschätzung). Sie muss zudem alle Phasen der Datenverarbeitung und alle technischen und organisatorischen Komponenten einbeziehen. Risiken können sich dabei nicht nur aus der technischen Gestaltung, sondern auch aus Fehlern ergeben, die menschliche Entscheidungen im Umgang mit algorithmischen Systemen mit sich bringen. Da algorithmische Systeme und ihre Einbettung in die sonstige Informationstechnik einer Organisation nicht statisch sind, wird zudem ein Managementsystem benötigt, das die Wirksamkeit der Maßnahmen angesichts veränderter Bedingungen, beispielsweise neu bekannt gewordener Risiken, überprüft und sicherstellt.“

Die Datenethik-Kommission für Künstliche Intelligenz sieht u. a. die Sicherheit und Robustheit einer KI als wesentliche Voraussetzung an und stellt fünf Stufen einer Kritikalität als Pyramide vor (siehe Kapitel 4.1 bis 4.3). In Normen und Standards ist allerdings in der Regel eine Matrix die gängige und bewährte Basis für eine Risikobewertung. Aus diesem Grund wurde die folgende Darstellung von Krafft und Zweig [136], die die Basis für die Kritikalitätspyramide war, vorgezogen.

Mithilfe einer Risikomatrix (siehe Abbildung 21), basierend auf den zwei Merkmalen (Schadenspotenzial durch Fehlurteile und Re-Evaluierungsmöglichkeit) lassen sich die Anwendungsszenarien von ADM-Systemen (Algorithm Decision Making) leicht verorten, sodass man schnell einen ersten Überblick über mögliche Risiken des Systems bekommt. Das Risiko setzt sich aus zwei Anteilen zusammen: dem Gesamtschaden aller Individuen plus einem möglichen, superlinearen gesellschaftlichen Gesamtschaden. Die einzelnen Regulierungsklassen rufen unterschiedliche Transparenz- und Nachvollziehbarkeitsforderungen auf, die in Abbildung 21 zugeordnet werden.

25 entspricht Englisch „risk based“

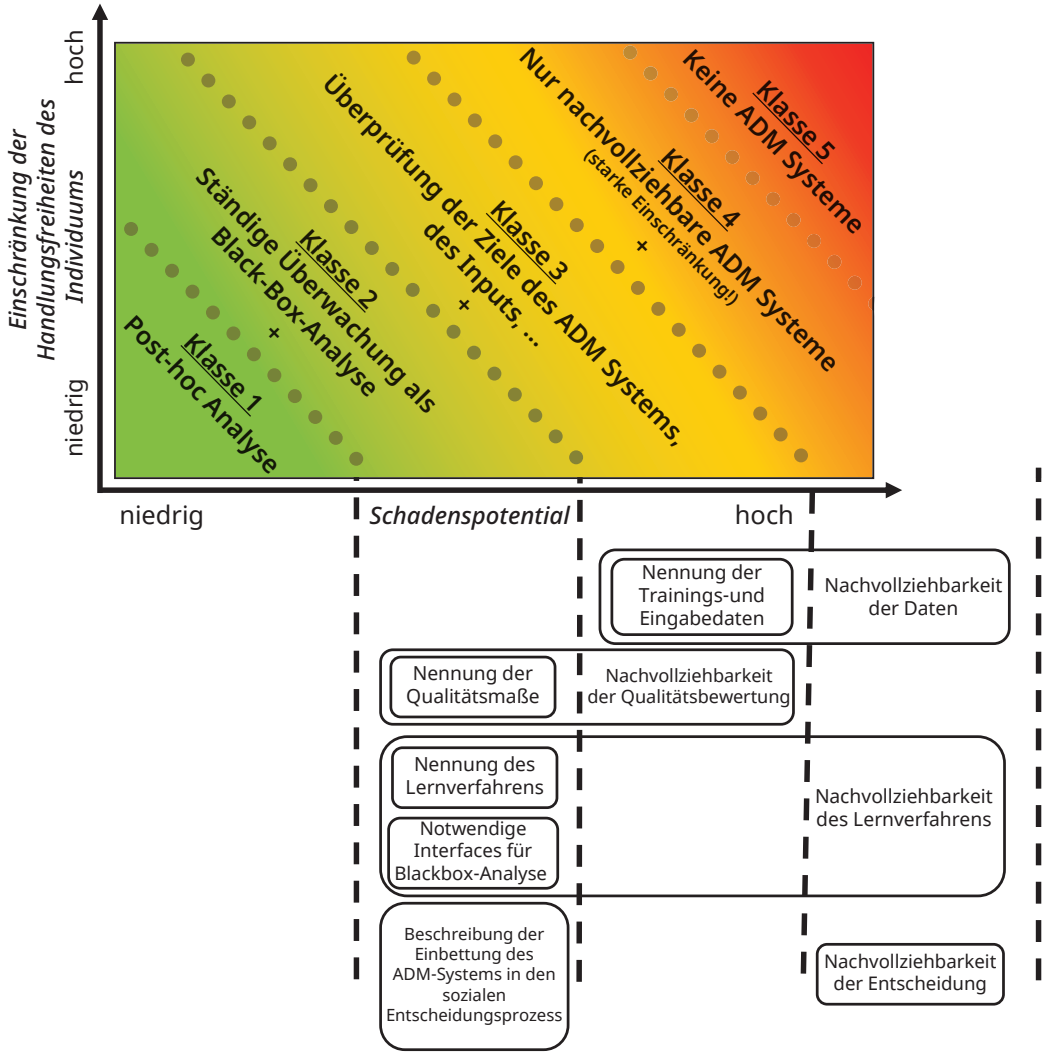


Abbildung 21: Risikomatrix mit Darstellung der Transparenz- und Nachvollziehbarkeitsanforderungen entsprechend der Regulierungsklassen [136]

4.4.1.3 Konformitätsbewertung und Zertifizierung für IT-Sicherheit

IT-Sicherheit kann durchaus als Teil von Qualität verstanden werden, hat allerdings eine umfangreiche eigene Historie mit unterschiedlichsten Wurzeln aus IT/OT Security, Safety und Privacy, die im Rahmen von KI berücksichtigt werden müssen.

IT-Sicherheit ist mit einem Nachweis (z. T. gesetzlich oder anderweitig gefordert) verbunden. Dafür stehen bereits umfangreiche anerkannte nationale und internationale Auditierungs- und Zertifizierungsverfahren zur Verfügung, insbesondere zu ISO/IEC 27001 [122], ISO/IEC 18045 [51], ISO/IEC 62443 [198]–[209] und BSI Grundschutz [184]–[186]. Nach der Maschinenrichtlinie [94] sind Konformitätsbewertungsverfahren (CE-Kennzeichen) für Produkte verfügbar.

Im Rahmen des EU CyberSecurity Act [188] sollen unter Berücksichtigung bestehender Verfahren weitere Zertifizierungsverfahren ausgearbeitet werden, die ggf. auch verpflichtend werden sollen.

Für diverse Branchen existieren zudem weitere Konformitätsbewertungen und Zertifizierungen wie beispielsweise in den Bereichen Mobilität, Gesundheit, Versorgung oder Arbeitsschutz. Weitere Betrachtungen zu Qualität, Konformität und Zertifizierung finden sich in Kapitel 4.3.

4.4.2 Anforderungen, Herausforderungen

Die wohl größte Herausforderung für den Einsatz von KI-Systemen durch die Wirtschaft liegt in der Schwierigkeit, Vertrauen in die (IT)-Sicherheit und in das KI-System herzustellen.

Vertrauen kann u. a. durch die Überprüfung und Zertifizierung der IT-Sicherheit eines KI-Systems geschaffen werden, deren Basis Normung sein kann.

IT-Sicherheit von KI-Systemen ist sowohl im sogenannten Office-Umfeld als auch im industriellen, operativen (z. B. Industrie 4.0; IoT) sowie Safety-Umfeld (z. B. automatisiertes Fahren oder Fliegen) relevant, wenn auch mit unterschiedlichen Herausforderungen und Anforderungen.

Für die IT-Security schaffen die bekannten IT-Schutzziele (Confidentiality, Integrity, Availability; CIA) eine Basis vor allem im Office-Umfeld.

Im industriellen Einsatz, z. B. in der Produktion, steht die Verfügbarkeit der Maschine, der Produktion, des Geräts oder Systems im Vordergrund. Ein Sicherheitsupdate oder ein Ausfall kann weitreichende Folgen haben, da so beispielsweise Fehlverhalten zu Gefährdungen für den Menschen oder zu erheblichen ökonomischen Einbußen führen können.

Aus Safety-Sicht ist wiederum der Schutz von Leib und Leben in jedem Fall das wichtigste Schutzziel, hier müssen für den Einsatz einer KI umfangreiche Regulierungen beachtet werden.

Allein diese unterschiedlichen Perspektiven für den Einsatz eines KI-Systems zu berücksichtigen, u. U. gleichzeitig, ist eine Herausforderung. Hinzu kommen die ergänzenden Anforderungen aus bestehenden Regulierungen und Normen.

Zusätzliche IT-Sicherheitsrisiken und damit verbunden mögliche zusätzliche Schutzmaßnahmen entstehen durch die technologischen Eigenschaften einer KI als lernendes und sich veränderndes IT-System. Dies ist einerseits bedingt durch den großen Einfluss der Daten auf das Training, den Test und den Wirkbetrieb einer KI sowie andererseits durch die Art der Datenverarbeitung in der KI (z. B. mit ML-Methoden unter Verwendung von neuronalen Netzen). Aus Sicht der IT-Sicherheit besteht Handlungsbedarf, um für diese zusätzlichen Risiken geeignete Schutzmaßnahmen zu entwickeln.

Der passende und erforderliche Grad der IT-Sicherheit des Gesamtsystems hängt von dem Zweck der KI und deren Kritikalität und Risikopotenzial bezüglich Eintrittswahrscheinlichkeit und Auswirkung eines Schadens ab. Hier besteht eine sehr große Spannweite von möglicherweise weniger problematischen Einsatzszenarien (beispielsweise als Marketingtool bis hin zu KI-gesteuerten Systemen in der Mobilität), die aufgrund ihrer Entscheidungen ein höheres Gefährdungspotenzial für Menschenleben bedeuten könnten. Daher ist immer die Frage zu stellen, welche Erwartungen und Anforderungen an die KI-Komponente sowie das Gesamtsystem, in welches die KI-Komponente integriert ist (z. B. in teilautonomen Fahrzeugen, in Videosystemen mit Mustererkennung, in IT-Systemen für automatisierte Abschlüsse von Versicherungen), hinsichtlich der ermittelten Gefährdungssituationen bestehen.

Im April 2019 veröffentlichte die europäische HLEG-KI ethische, rechtliche und technische „Key Requirements“ [22] für verlässliche KI-Systeme, u. a.

- → zur technischen Prüfung der Robustheit gegen Angriffe, der Sicherheit, Verlässlichkeit und Reproduzierbarkeit,
- → zum Schutz der Daten und ihrer Integrität und
- → zu Transparenz und Erklärbarkeit der Algorithmen bis hin zu Betrachtungen der Fairness.

Es besteht Untersuchungs- und Forschungsbedarf, welche zusätzlichen Risiken und neue Anforderungen aus IT-Sicherheitssicht an die Erstellung und den Betrieb eines KI-Systems sowie dessen Überprüfung und Zertifizierung zu erwarten sind.

Manipulierte oder mangelhafte Daten oder manipuliertes Training können IT-Sicherheit beeinflussen. Mitunter reicht schon eine gewisse Unschärfe der verwendeten Daten (z. B. minimale Pixelveränderung in Videodaten) aus, die zu Fehlinterpretationen führen kann. Der Umgang mit solch einer Unschärfe ist noch nicht erforscht. Davon kann jedoch die Erkennbarkeit, ob ein Datum (schadhaft) wahr oder falsch ist, beeinflusst sein.

Auf die Sicherheit eines KI-Systems hat auch das Umfeld, beispielsweise die Verfügbarkeit technischer Ressourcen, einen hohen Einfluss. Ein Angriffsszenario könnte die Manipulation der Ressourcen sein, um beispielsweise die Latenz bei Realtime KI und damit das Ergebnis zu beeinflussen. Es besteht ggf. eine Abhängigkeit der Ergebnisqualität des verwendeten Modells von den verfügbaren Ressourcen (z. B. Rechenleistung), der Hardware, der Infrastruktur, der Schnittstellen und der Umgebung.

Die IT-Sicherheitsanforderungen an KI-Systeme hängen auch davon ab, welcher Akteur/Marktteilnehmer mit der KI gerade arbeitet. Gemeint sind die möglicherweise unterschiedlichen Anforderungen an Hersteller, Zulieferer, Integratoren, Betreiber und Endnutzer, von denen die „nur“ mit einer KI-Soft-

ware, einem embedded System, einer KI-gesteuerten Anlage/Roboter, einem KI-System aus unterschiedlichsten Teilprodukten, einer KI-Cloudlösung, die möglicherweise mit einem KI-IoT-Gerät vernetzt ist, zu tun haben, bis hin zum Endnutzer eines KI-Systems. Das Einsatzumfeld kann dabei von unkritisch bis zu hohem Risiko reichen, wobei z. B. die KI-Software im Ursprung beim Hersteller diese Einsatzfelder noch nicht differenziert hat. Ein Beispiel wäre KI-gestützte Bilderkennung als Basis für unterschiedlichste Einsatzfelder.

4.4.2.1 Sichere Daten

Aktuell lässt sich davon ausgehen, dass in naher Zukunft Angreifende gefälschte Daten in KI-Systeme einbringen werden, sei es, um die Ergebnisse zu manipulieren, um Ressourcen aus legitimen Datenquellen umzuleiten oder um Industriesabotage zu begehen. Entgegnen ließe sich dem mit Mechanismen, welche gewährleisten, dass die auf diesen Daten operierenden Algorithmen diese identifizieren und zurückweisen können.

Das Problem der Erkennung gefälschter Daten ist nicht neu, und das maschinelle Lernen erbt dieses Problem im Wesentlichen von den Daten, auf denen es operiert. Zudem kann der Prozess der angemessenen Verifizierung der Herkunft der Datensätze für die Schulung und die Einspeisung in Algorithmen und der damit einhergehenden Frage des realistischen Einschätzens von Risiko und Haftung, die Entwicklung und Anwendung von KI-Technologien durchaus verzögern.

Eine klassische Anwendung des ML ist die Erstellung von Vorhersagen auf der Basis von Input-Daten. Allerdings basiert die Vorhersage eines KI-Systems auf der Qualität der Input-Daten. Falls diese verzerrt oder unvollständig sind, können Fehler in die KI gelangen, sodass die Vorhersage nicht verlässlich ist.

Beim maschinellen Lernen können schlechte Daten nur schwer erkannt oder entfernt werden. Ab einem gewissen Grad des Lernfortschritts ist es nahezu unmöglich herauszufinden, auf welchen Datenelementen welche systeminternen Entscheidungen basieren. „Vergessen“ oder „Verlernen“ ist gegenwärtig so gut wie unmöglich. Datenintegrität und Datenqualität sind entscheidende Qualitätsmerkmale für den Erfolg von ML-Systemen.

Betrachtet man die Qualitätsmerkmale von Daten und Metadaten nach der Definition des Fraunhofer Leitfadens NQDM 2019 [90] (vgl. auch Kapitel 4.1.2 und 4.3.2), können alle von ihnen bei absichtlicher oder unbeabsichtigter Einflussnahme bzw. mangelhafter Qualität Auswirkungen auf die Sicherheit von KI-Systemen haben. Je nach Nutzung der Daten im Training, beim Test, im Design oder im Betrieb sind deren Qualitätsmerkmale jedoch unterschiedlich zu betrachten. Verzerrte Trainings- und Testdaten können die Verlässlichkeit eines KI-Systems stark beeinträchtigen. Ein „Verlernen“ der verzerrten Daten ist gegenwärtig so gut wie unmöglich. Daten im Betrieb müssen gegebenenfalls bereits auf physikalischer Ebene vor Manipulation geschützt werden.

Speziell unter den Aspekten der Informationssicherheit (security) und Vertraulichkeit (privacy) sind insbesondere die Qualitätsmerkmale Transparenz und Vertrauenswürdigkeit bzw. Zugänglichkeit und Verfügbarkeit herauszugreifen. Letztere werden im Folgenden (siehe Kapitel 4.4.2.1.1) näher betrachtet.

Transparenz und Vertrauenswürdigkeit von Daten sind bei jeglicher Art von Datennutzung wichtig und sollten nachvollziehbar sein, da sie auch das Vertrauen in die Datenqualität und Datenintegrität erhöhen. Diese sind entscheidend für den Erfolg von ML-Systemen.

Datenintegrität bezeichnet die Konsistenz, die Richtigkeit, Vertrauenswürdigkeit und Rekonstruierbarkeit der Daten während der gesamten Lebensdauer der Daten in IT-Systemen. Sie umfasst Maßnahmen, damit geschützte Daten während der Verarbeitung oder Übertragung nicht durch unautorisierte Personen entfernt oder verändert werden können. In der IT-Sicherheitsforschung sind Datenintegrität, Datenschutz und Datensicherung unerlässliche Forderungen an verlässliche Informationssysteme.

Datenqualität ist ein essenzieller Bestandteil des Datenmanagements, weil die Qualität der Daten die Glaubwürdigkeit der Anwendungen bestimmt. Das gilt natürlich besonders für datengetriebene Technologien wie maschinelles Lernen oder BigData-Analytics-Applikationen. Deterministische Analytik und statistische Datenverarbeitung dokumentieren und fixieren Beziehungen zwischen Datenelementen. Erwartungen für die Analyse der Daten werden fest kodiert.

Machine Learning und insbesondere Deep Learning generieren und verfeinern in der Lernphase selbstständig Algorithmen. Um die für die genaue Entwicklung der Algorithmen erforderliche Variation zu gewährleisten, benötigt Deep Learning hinreichend große und im Regelfall viel größere Datenmengen als herkömmliche analytische Anwendungen.

Maschinelle Lernverfahren im Allgemeinen und Deep Learning-Modelle im Besonderen benötigen vertrauenswürdige Trainingsdatensätze. Hinreichend solide Prozesse zur Bereinigung der Daten sind unverzichtbar. Das benötigte Datenvolumen und die evolutionären Methoden beim maschinellen Lernen führen zu grundsätzlichen Fragen wie:

→ Woher kommen die Daten? Welche Systeme stellen Daten bereit?
→ Wie wird auf die Daten zugegriffen? Bleibt die Integrität der Daten erhalten?
→ Wie sind die Daten zu verstehen? Welche Beziehungen bestehen zwischen den Daten? Gibt es Abhängigkeiten und welcher Art sind sie?
→ Wie werden welche Daten in Analysevorgängen herangezogen? Wie werden Daten kombiniert? Wie können die Daten verbessert werden?

Die Qualität bei der Aufbereitung der Daten für ein KI-System wird als Kuratieren bezeichnet. Entwickler sollten die folgenden Eigenschaften dokumentieren, um Transparenz zu wahren über:

→ die Herkunft der Daten,
→ die Form der Veredlung (Definieren, Sammeln, Selektieren, Umwandeln, Verifizieren) und Anreicherung der Rohdaten zu Modell- oder Trainingsdaten,
→ der Lernstil (Supervised Learning, Unsupervised Learning, ggf. andere),
→ die verwendeten Lernmodelle,
→ der potenzielle Einsatz einer speziellen KI-Komponente,
→ menschliche Beteiligung (z. B. User-Feedback oder Labelling) an den Entscheidungsfindungen innerhalb einer Verarbeitung,
→ die Institutionen, die die Komponenten des KI-Systems herstellt und über die Auswahl, Konfiguration, Implementierung und den Betrieb der verwendeten KI-Technik,
→ das Kuratieren der Daten, das Training und die Auswahl der Modelle entschieden haben,
→ die Implementierung des KI-Algorithmus, insbesondere der regelbasierten Instruktionen und Entscheidungen,
→ den Einbau von Prüfankern und Prüfagenten.

4.4.2.1.1 Schutz von Daten, Methoden und Maßnahmen

Die Verarbeitung und Veredelung von Daten bei der Entwicklung und dem Betrieb von lernenden Systemen sollte den neun Kriterien des technischen Verständnisses der DSGVO Rechnung tragen. Es gibt grundsätzlich drei technische Verfahren, um Datenschutzverletzungen vorzubeugen:

Verschlüsselung von Daten ist eine Möglichkeit, sensible Informationen zu schützen. Im Regelfall wird dabei die Verarbeitbarkeit der Daten eingeschränkt oder sogar unmöglich. Homomorphe Verschlüsselungsverfahren können zwar Operationen auf verschlüsselten Daten ausführbar machen, dies ist aber oftmals zu kostenintensiv. Bei den meisten anderen Verschlüsselungsverfahren müssen Daten erst entschlüsselt werden, bevor Operationen durchgeführt werden können.

Aber Verschlüsselung verhindert Veröffentlichung von Daten auch dort, wo sie notwendig ist. Ein Verfahren zur Veröffentlichung ohne Datenschutzverletzung ist das **Anonymisieren** von Daten. Bei der Anonymisierung unterscheidet man drei Arten von Daten: Identifikatoren, Quasi-Identifikatoren und die sensiblen Werte.

Identifikatoren sind Angaben, mit denen eine Person direkt identifiziert werden kann, z. B. Namensangaben.

Als Quasi-Identifikatoren bezeichnet man Kombinationen aus Merkmalen, um eindeutige Zuordnungen durchzuführen. Für das Anonymisieren von Daten gilt es, gewisse Regeln und Richtlinien zu beachten, um zu gewährleisten, dass nicht trotz Anonymisierung Rückschlüsse auf Personen gezogen werden können. Allerdings können Angreifer mit genügend Hintergrundwissen die De-Anonymisierung von Daten vornehmen. Dieses Risiko kann wegen der Bandbreite möglichen relevanten Wissens kaum realistisch eingeschätzt werden, mit der Folge, dass Anonymisierung nicht unbedingt ausreicht, um Datenschutz zu gewährleisten.

Unter sensiblen Daten (gemäß DSGVO) versteht man personenbezogene Daten, die z. B. rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit enthalten, sowie die Verarbeitung von Gesundheitsdaten, genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer Person.

Das Konzept der **Differential Privacy** hingegen kann Datenschutz gewährleisten, indem es prinzipiell eine statistische Garantie darüber gibt, dass die Daten einzelner Personen keine Auswirkung auf das Ergebnis bestimmter Abfragen haben.

Als Grundsatz gilt, dass die Wahrung der Privatsphäre einer Person genau dann gewährleistet ist, wenn das Ergebnis einer Datenabfrage nicht von den Daten einer einzelnen Person abhängt. Dazu werden Funktionen benötigt, die Datenbankabfragen beantworten können und dabei sicherstellen, dass der Datenschutz nicht verletzt wird. Die Originaldaten werden während einer Abfrage mit „Rauschen“ versehen oder abgeändert. Die abgeänderten Daten sind nicht von den ursprünglichen Daten zu unterscheiden und statistische Zusammenhänge werden dabei nicht verfälscht. Die Auswirkungen auf Lernverfahren müssen allerdings noch genauer untersucht werden.

4.4.2.1.2 Sicherheit und Vertrauen in Authentizität, Integrität und Qualität von Daten, Methoden und Maßnahmen und Diskussion der Ansätze

KI-Systeme erhöhen die Komplexität der IT-Sicherheit gegenüber „normalen“ IT-Systemen. Dies liegt zum einen an den eingesetzten Softwaretools und deren eingeschränkter Nachvollziehbarkeit, z. B. für Machine Learning und neuronale Netze, sowie deren Wechselwirkung mit und Abhängigkeit von dem Umfeld. Andererseits erfordert die elementare Bedeutung der verwendeten Daten für das Training und im produktiven Einsatz über die DSGVO hinaus große Aufmerksamkeit bezüglich IT-Sicherheit.

Angesichts künftig zu erwartender Angriffsszenarien ist die Widerstandsfähigkeit gegenüber Angreifenden eine Schlüsselanforderung für cyber-physische Wertschöpfungsketten. Folgende vier beispielhafte Ansätze (siehe Tabelle 6) stellen mögliche Lösungsansätze dar, das Vertrauen in die Authentizität, Integrität und Qualität eines bestimmten ML-Datenlabels (maschinelles Lernen) herzustellen, und sollten beim Design und der Entwicklung künftiger ML-Anwendungen zur Erhöhung des Resilienzlevels in Betracht gezogen werden:

Tabelle 6: Vier beispielhafte Ansätze, um Vertrauen, Integrität und Qualität des ML-Datenlabels herzustellen

	Methode/ Ansatz	Beschreibung	Sicherheit gegen Manipulation
1	Reputations-systeme	Korrelation von Ereignissen und Feedback-Ergebnissen über die Identitätssubjekte, die die Datensätze erstellt haben	gering
2	Algorithmische Analyse	Analyse von Ausgabedatensätzen basierend auf maschinellem Lernen	gering bis mittel
3	End-to-End-Datenherkunft	Authentizitätsanalyse zur Überprüfung der konkreten Herkunft der Daten und der Integrität der Datenkette	mittel
4	Identifizierbare Datenherkunft mit Scoring-Mechanismus	Analyse der Authentizität und Integrität der Datenherkunft und Bewertung der beteiligten Entitäten auf der Grundlage ihrer Lebenszyklus-Zertifikate und historischer Ereignisse, sofern verfügbar	hoch

Andere hybride Modelle als Kombination aus diesen Verfahren können ebenfalls entwickelt werden.

Grundsätzlich sind jedoch alle Analysen resilienter, wenn sie mit einem auf Reputation/Datenherkunft basierenden Scoring-Mechanismus kombiniert werden. Aus diesem Grund erweist sich eine Diskussion darüber, wie bessere, globale Reputationssysteme entwickelt werden können, als unumgänglich (auch um verlässlich Daten für mehr Akteure am Markt sicher verfügbar zu machen), da durch die Verwendung von ML allein nicht angemessen überprüft werden kann, dass Eingabedaten oder darauf basierende Labels nicht durch ähnliche ML verfälscht wurden.

ANSATZ 1: REPUTATIONSSYSTEME

Zentralisierte Daten-Reputationssysteme sind in großem Maßstab beispielsweise auf monolithischen Plattformen abbildbar. Typischerweise verfügt ein Marktplatz über ein natives Reputationssystem, das unabhängig arbeitet und von eindeutigen persönlichen Identitäten abstrahiert ist. Das Fehlen von robusten Verifizierungs- und Bewertungsmechanismen ermöglicht es den Teilnehmern, diese Bewertungen zu manipulieren.

Die Integrität und Authentizität der Daten kann nicht ohne einen Zugang zu den Identitätsregistern einer zentralen Plattform überprüft werden, selbst wenn man die Integrität des Inhalts dieses idealerweise gut verwalteten Registers annimmt.

Dezentralisierte Daten-Reputationssysteme und pseudonymisierte, durch Token kuratierte Register, abgebildet auf einer Blockchain-Infrastruktur, wären in der Lage, eindeutige digitale Identitäten für alle Teilnehmer eines offenen Systems zu verifizieren und Reputationsdaten über alle Plattformen hinweg zu aggregieren, auf denen die betroffene Person zugestimmt hat, für Reputationszwecke korreliert zu werden.

Dieser sogenannte „Web of Trust"-Ansatz für die Veröffentlichung von Reputationsdaten befindet sich jedoch noch in einem frühen Stadium, und seine eigenen einzigartigen Angriffsvektoren müssen noch in der Praxis getestet werden. Bis solche Systeme ausgereift sind, können solche dezentralen Reputationsscores als eine Datenquelle unter vielen für ein probabilistisches, hybrides Scoring-Modell verwendet werden.

ANSATZ 2: ALGORITHMISCHE ANALYSE DER AUSGABE UND DEREN EINSCHRÄNKUNGEN

Ein weiterer Ansatz stellt die Analyse der Ausgabedaten eines IoT-Bausteins oder eines Algorithmus mit ML-Algorithmen dar. Folgende Techniken (siehe Tabelle 7) dienen zur Feststellung, ob ein gegebener Datensatz gefälscht oder echt ist:

Tabelle 7: Techniken zur Feststellung wahrheitsgetreuer Datensätze

Ausgabe-Vektor	Beschreibung	Beispiel Image Processing
Object Features	Analyse ausgewählter Merkmale oder Stellen eines Objekts, an denen Algorithmen, die das gefälschte Objekt erzeugen, typischerweise versagen	Sichtbare Artefakte an der Schnittstelle von Haar und Körper in einem Bild eines Menschen
Format Features	Analyse von Inhalten, die sich auf bestimmte Formatmerkmale beziehen	Gefälschte Bilder haben tendenziell glattere Texturen
Neural Monitoring, a.k.a. Reflexive Monitoring	Analyse von Neuronen und Schichten des Netzwerks, die bei der Identifizierung/Verarbeitung von echten und gefälschten Bildern aktiviert werden	Testen, wie andere fortschrittliche Algorithmen auf zuvor sortierte authentische und gefälschte Bilder reagieren

Statische Kriterien für alle drei dieser Analysevektoren können manuell bereitgestellt werden, aber da gegnerische Netzwerke auf historische Daten trainiert wurden, überwinden sie schnell jede Analyse. Dies führt dazu, dass alle drei Methoden zu drei separaten Fronten in einem „Wettrüsten" zwischen lernenden Algorithmen werden, bei dem keine der oben genannten Methoden eine endgültige, resiliente Lösung darstellen kann. Sie unterliegen vielmehr einer Zirkularität, welche mögliche Angriffsvektoren öffnet: da sich alle drei Methoden auf die Verwendung von maschinellem Lernen zur Identifizierung von Nebenprodukten einfacherer oder älterer maschineller Lernverfahren beziehen, und zwar in einem fortlaufenden Prozess, der nie vollständig ist.

ANSATZ 3: END-TO-END-DATENHERKUNFT (DATA PROVENANCE)

Die Nachvollziehbarkeit der Datenherkunft ist heute nur bedingt gegeben. Sofern die Daten nicht aus dem eigens kontrollierten Datensilo stammen, kann zugekauften Daten eine Fälschungssicherheit nicht gänzlich attestiert werden, da grundsätzlich Ursprünge und Datenspuren, die von außen nicht verifizierbar sind, noch leichter zu fälschen sind als die Daten selbst.

Stammen die Daten aus den eigenen, vertrauenswürdigen Quellen, lässt sich mit diesen ohne Frage ein effizientes KI-System aufbauen. Jedoch verfügen nur wenige Akteure am Markt über entsprechende eigene Datenmengen bzw. haben (sicheren) Zugriff auf diese.

Bei der End-to-End-Datenherkunft geht es im Grunde um die kryptografische Signierung aller Daten aus einer Datenquelle. Dies schafft die Fähigkeit, ein Datenpaket bis zu seiner Entstehung zurückzuverfolgen, bis zu genau dem Gerät, das als Erstes eine Messung vorgenommen oder ein Ereignis registriert hat. Nur durch die Identifizierbarkeit der Datenquelle kann die Herkunft des Datenflusses die Grundlage für überprüfbare Behauptungen bilden. Durch die Signierung der Datenpakete entsteht eine sogenannte Datenkette[26], welche es ermöglicht, die Vertrauenswürdigkeit, Zuverlässigkeit oder Risikometrik der Datenquelle zu beurteilen.

26 Eine Daten-„Kette" ist jede kryptografische Datenstruktur, die signierte Datenobjekte miteinander „verkettet" (mit unidirektionalen oder bidirektionalen „Verbindungen" zwischen Entitäten) und so eine Navigationsmethode für eine umfassende Datenflussprovenienz und -prüfung schafft. Die Datenflussprovenienz ermöglicht die Überprüfung der End-to-End-Integrität jedes Datenflussobjekts und seiner Transformationen.

ANSATZ 4: IDENTIFIZIERBARE DATENHERKUNFT UND GLOBALE BEWERTUNG

Nur ein Reputationssystem, das auf einem möglichst neutralen und vollständigen Prüfprotokoll basiert, kann die Vertrauenswürdigkeit von lernenden Systemen in einem angemessenen Maß bewerten.

Hierbei „bewertet" oder schätzt ein Scoring-Modell das Risiko der Verwendung von Datenketten ab, indem allen unbekannten Akteuren relative Werte der Vertrauenswürdigkeit oder Validierung zugewiesen werden. Hierbei werden Akteure und Agenten in einem System aufgrund ihrer historischen Bekanntheit oder Genauigkeit gescored.

Um eine vertrauenswürdige Umgebung für ML-Daten zu schaffen, sind im ersten Schritt von einer Datenquelle alle ausgegebenen Daten kryptografisch zu signieren (Ansatz 3). Nur durch die Identifizierbarkeit der Datenquelle kann die Herkunft des Datenflusses die Grundlage für überprüfbare Behauptungen und Bescheinigungen über den Datenfluss selbst sowie für Reputationsmechanismen bilden.

Der zweite Schritt beinhaltet die Verankerung der entstehenden Datenketten aus verifizierten Quellen, in einer dezentralisierten Identitäts-Meta-Plattform. Diese Plattform stellt eine Public-Key-Infrastruktur zur Verfügung, mit welcher öffentlich verankerte Datenidentitäten erzeugt werden können. Somit ließe sich die Datengenese und jedes Transformationsereignis elektronisch signieren, was die nachträgliche Verifizierung eines beliebigen Datenflusses und eine Abschätzung der Vertrauenswürdigkeit der Ausgabedaten eines maschinellen Lernalgorithmus ermöglicht. Diese Beurteilungen können direkt aus der Datenquelle und/oder indirekt anhand von öffentlichen/offenen Registern und Reputationssystemen erfolgen.

Ein Bewertungsalgorithmus kann von jedem verankerten Datenfluss eine Art „Lebenszyklus-Zeugnis" anfordern und so die allgemeine, aggregierte Vertrauenswürdigkeit, Datenflussprovenienz[27] und Genauigkeit eines Datenlabels für maschinelles Lernen widerspiegeln bzw. bewerten. Die öffentliche Verankerung der Datenreputationen ist eine Voraussetzung für eine angemessene Objektivität in einem ausgereiften Reputationssystem. Diese Funktion ließe sich gut über große Institutionen mit öffentlicher Aufgabe abbilden.

4.4.2.2 Robustheit und Sicherheit gegen Angriffe – Angriffsvektoren und Verteidigungsmechanismen

In Anlehnung an die ganzheitliche Betrachtungsweise von IT-Systemen hinsichtlich ihrer Absicherung sollte auch bei KI-Systemen neben Sicherheit und Schutz der Daten die gesamte Modell-Infrastruktur berücksichtigt werden. Dazu zählen im Kern drei Schutzmaßnahmen:

- → die Modell-Authentifizierung (z. B. durch eine rollenbezogene Beschränkung der Nutzung des Modells auf einem bestimmten Endgerät und für eine vorgegebene Dauer, diese ist insbesondere beim sogenannten verteilten Lernen (Federated Learning) zu beachten).
- → die sichere Modell-Haltung und -Verbreitung (z. B. durch eine Verschlüsselung des Modells und der Einhaltung sicherer Übertragungswege).
- → die Modell-Verifikation (z. B. mittels eines geeigneten Abgleichs zu Referenz-Modellen, um Auffälligkeiten bzgl. der erwarteten Arbeitsweise festzustellen).

Die hier umrissenen Maßnahmen decken bereits einen Großteil des Spektrums an Angriffen (u. a. den Diebstahl des Modells und Denial-of-Service-Attacken) ab, die schwerwiegende wirtschaftliche Folgen nach sich ziehen können.

4.4.2.2.1 Adversarial Machine Learning (AML)

Im Zuge der Betrachtung der Robustheit von KI-Systemen ist der Begriff AML von zentraler Bedeutung. Dieser bezeichnet ein aktuelles Forschungsgebiet[28], welches sich mit den Sicherheitsaspekten von ML beschäftigt. Im engeren Sinn werden darunter lediglich mögliche Angriffsszenarien diskutiert, in allgemeinerer Verwendung sind damit aber auch die Entwicklung entsprechend gehärteter (robuster) Modelle, detaillierte

27 Mit „Datenflussprovenienz" ist ein Mechanismus zur Verfolgung von Datenpunkten gemeint sowie der Umgang mit diesen Daten durch ein Verarbeitungssystem, das jede Transformation zu diesen Datenpunkten registriert. [Dazu gehören Flüsse mit mehreren Quellen, kollektive Sensorfusion und die Verarbeitung durch maschinell lernende Algorithmen. Eine umfassende Datenflussprovenienz beinhaltet nicht nur die Verfolgung der Aufbewahrung der Daten, sondern auch die Überprüfung der End-to-End-Integrität jedes Datenflusses, einschließlich aller Transformationen (Hinzufügungen, Löschungen, Modifikationen, Kombinationen und ML-Verarbeitung)].

28 Bereits in der überwiegend in englischer Sprache formulierten Forschungsliteratur etablierte Begrifflichkeiten werden hier ohne Übersetzung übernommen, da zu erwarten ist, dass sich diese auch im deutschen Sprachraum so durchsetzen werden.

Analysen von Verteidigungsmechanismen sowie die Bewertung der angriffsspezifischen Konsequenzen abgedeckt.

4.4.2.2.2 Angriffsvektoren und Verteidigungsmechanismen

Im Folgenden soll eine Übersicht sowohl über Angriffsvektoren (siehe Tabelle 8) als auch über Verteidigungsmechanismen dargelegt werden. Diese erhebt keinen Anspruch auf Vollständigkeit, bietet aber eine Einstiegs- und Orientierungshilfe, um unter Etablierung einheitlicher Begrifflichkeiten (Terminologie und Taxonomie) geeignete praktische Lösungsansätze zur Bewältigung sicherheitsrelevanter Herausforderungen in Bezug auf maschinelle Lernverfahren zu entwickeln. Die Ausführungen sind dabei bewusst so allgemein gehalten, dass sie unabhängig vom speziellen Lernparadigma sind. Lediglich an einzelnen Stellen werden dahingehend Konkretisierungen vorgenommen. Lernparadigmen sind beispielweise das überwachte, unüberwachte oder bestärkende Lernen. Speziell für bestärkendes Lernen soll, wann immer von einer Einflussnahme auf „Daten“ die Rede ist, auch die Betrachtung der Umwelt des Systems (in diesem Fall nach üblichem Sprachgebrauch: eines „Agenten“) mitgedacht werden.

Grundsätzlich können die Komponenten von ML-Systemen zum Ziel von Angriffen unter Benutzung verschiedener Techniken und dem Vorhandensein bestimmter Wissensstände werden. Die Ziele ergeben sich aus der typischen Struktur eines ML-Systems, nämlich der physikalischen Domäne und deren digitaler Repräsentation

→ der Eingangssensoren,
→ dem ML-Modell selbst sowie
→ der Ausgaben oder physikalischen Aktionen.

Der Wissensstand kann anhand aufsteigend inklusiver Mengen von Informationen, die dem Angreifer zur Verfügung stehen, beschrieben werden.

Im einfachsten Szenario besitzt dieser keine Kenntnisse über das ML-System, sondern kann lediglich Eingaben tätigen oder ggf. darüber hinaus Ausgaben abgreifen. Hierbei ist hinsichtlich der Stärke des Angriffs entscheidend, in welchem Umfang und in welcher Form (direkte Ausgabe, Wahrscheinlichkeiten, finale Klassifikation) Eingabe-Ausgabe-Paare abgegriffen werden können.

Weiterführend können dem Angreifer Informationen über die Modellfamilie (z. B. ein neuronales Netz, ein Ensemble von Entscheidungsbäumen oder ein hybrides System) sowie darüber hinaus noch deren konkrete Architektur bekannt sein.

Weiteres Detailwissen, beispielsweise in Form der Parameter oder gar des Lernalgorithmus (z. B. des Optimierungsverfahrens und der Verlustfunktion) einschließlich dessen Hyperparametern bilden schließlich die höchste Kenntnisstufe. Üblicherweise ist in diesem Zusammenhang von Blackbox, Greybox oder Whitebox Angriffen die Rede – je nachdem, wie weit die Kenntnis des Angreifers reicht.

Die Techniken sind hinsichtlich ihrer Verwendung während der Trainings- oder der Betriebsphase der Modellbildung bzw. -operation zu unterscheiden. Erstere zielen darauf ab, Einfluss auf die Daten, den Lernalgorithmus oder das Modell zu nehmen, während Letztere dies nicht tun, sondern neue Eingaben für das Modell generieren, die sich der angedachten Funktionsweise des Modells (Klassifikation, Regression, Verhalten eines Agenten) entziehen.

Die erste Frage, die es bei der Betrachtung von **Angriffen während des Trainings** (Poisoning Attacks) zu klären gilt, ist die nach dem Umfang, indem der Angreifer die Daten manipulieren, d. h. neue Daten hinzufügen, Daten löschen oder bestehende Daten verändern kann.

Für Letztere ist zudem entscheidend, ob der Angreifer lediglich Einfluss auf die Eingaben oder auch auf die dazugehörigen Ausgaben, d. h. auf vollständige Eingabe-Ausgabe-Paare, nehmen kann. Eine im Sinne des Angreifers gelernte Datenverteilung kann beispielsweise zu einer Verringerung der Genauigkeit der Klassifikation, der Installation einer Hintertür oder der gezielten Lenkung eines Agenten führen.

Bei **Angriffen während der Betriebsphase** bleiben die Trainingsdaten und das Modell selbst unberührt. Stattdessen entwerfen diese Angriffe neue Testdaten, die sich z. B. der vorgesehenen Klassifikation entziehen (Evasion Attacks) oder die genutzt werden können, um Informationen über die Trainingsdaten und/oder das Modell zu sammeln (Model Extraction Attacks).

Tabelle 8: Angriffsvektoren

Angriffsvektor	Charakteristik
Evasion Attacks	lösen üblicherweise ein beschränktes Optimierungsproblem, welches nach Eingabe-Beispielen sucht, die bei möglichst geringer Abweichung von regulären Trainings- oder Testdaten die Verlustfunktion des Modells maximieren. Dabei kommen meist ein- oder mehrschrittige (iterative) gradientenbasierte Verfahren zum Einsatz, seltener auch Methoden ohne Verwendung der Gradienten. Letztere benötigen im Falle von Klassifikations-Systemen in der Regel die entsprechenden Ausgabewahrscheinlichkeiten.
Model Extraction Attacks	die auf das Modell gerichtet sind, beabsichtigen oft seine Nachbildung zum Zwecke der eigennützigen Verwendbarkeit oder zum Erzeugen von Evasion Attacks mithilfe des Substitut-Modells. Für die Aggregation von Eingabe-Ausgabe-Paaren wird dabei eine hinreichend unbeschränkte Schnittstelle zum Modell benötigt. Werden im Zuge einer Model Extraction Attack Informationen über die Daten gesammelt, so können hierbei zum einen Statistiken über die Verteilung der Trainingsdaten gesammelt oder die Trainingsdaten selbst extrahiert werden (Model Inversion Attacks). Außerdem besteht die Möglichkeit, mittels geeigneter Anfragen festzustellen, ob bestimmte Datenpunkte zum Trainingsdatensatz gehören (Membership Inference Attack).
Beeinflussung des Modell-Outputs	bei kompletter oder auch nur teilweiser Bekanntheit der jeweiligen Entscheidungswege. Gemeint ist beispielsweise ein im Sinne des Angreifers positives Klassifikationsergebnis (z. B. bei der Bewertung der Kreditwürdigkeit) aufgrund der Kenntnis um die Verzweigungen in Entscheidungsbäumen, der Anfertigung und Auswertung sogenannter Saliency Maps (z. B. für neuronale Netze), das Ausnutzen der Belohnungsfunktion beim bestärkenden Lernen (Reward Hacking), aber auch bereits das Wissen um einen in den Trainingsdaten vorhandenen und demzufolge höchstwahrscheinlich vom Modell gelernten Bias. Das teilweise Wissen um das Zustandekommen der Entscheidung ist in dem Sinne für den Angreifer nutzbringend, als dass dadurch z. B. immerhin das Aufkommen bestimmter Kategorien beim Ergebnis einer Klassifikation ausgeschlossen werden kann.

Entsprechend der Unterscheidung der Angriffsvektoren ist für die **Verteidigungsmechanismen** eine differenzierte Betrachtung sinnvoll, je nachdem, ob sie gegen Angriffe während der Trainings- oder der Betriebsphase wirksam sind. Es sei an dieser Stelle angemerkt, dass die Absicherungsmaßnahmen oftmals die Leistungsfähigkeit, z. B. die Genauigkeit oder die Inferenz-Zeit des ML-Systems, beschränken und deswegen eine Abwägung zwischen beiden vorgenommen werden muss.

Die Eindämmung von Poisoning Attacks kann, neben allgemeinen Bestimmungen betrefflich der Datenhaltung und Datenbereitstellung, mittels einer Überwachung der Eingaben erfolgen. So sind manipulierte Daten mittels geeigneter statistischer Methoden unter Umständen schon vor Beginn des Trainings zu identifizieren und der Trainingsdatensatz um diese zu bereinigen. Darüber hinaus existieren Ansätze, die auch unter Beobachtung des Modells selbst, beispielsweise seiner gelernten Parameter, Unterschiede in der Folge von legitimen oder manipulierten Datensätzen feststellen können. Um Angriffe während der Betriebsphase zu mitigieren, wurde eine Vielzahl an potenziellen Vorgehensweisen entwickelt. Die folgende Liste enthält eine beispielhafte Zusammenstellung mit dazugehörigen Erläuterungen:

- → Der Trainingsdatensatz wird um Eingabe-Ausgabe-Paare ergänzt, die mittels eines spezifischen Angriffs erstellt wurden (Adversarial Training).
- → Dem Angreifer wird der Zugang zu verwertbaren Informationen über die Gradienten des Modells verwehrt (Gradient Masking).
- → Ausgehend von einem bestimmten Modell wird ein neues, weniger komplexes Modell trainiert, dessen Entscheidungsgrenzen glatter verlaufen (Defensive Distillation).
- → Die Ausgaben eines Ensembles von Modellen wird kombiniert (Ensemble Methods).
- → Auf dem Trainingsdatensatz oder innerhalb des Modells wird Zufall eingebracht (z. B. in Form von Gaußschem Rauschen), um den Informationsgewinn bei fester Zahl abgegriffener Eingabe-Ausgabe-Paare zu verringern.
- → Um Manipulationen der Eingaben unwirksam zu machen, werden diese auf verschiedene Art und Weise transformiert, z. B. durch Kompressions- oder Glättungsoperationen (Feature Squeezing).
- → Eingaben werden mithilfe eines vorgeschalteten Modells (Reformer) in die Nähe der nächstgelegenen Datenpunkte im Trainingsdatensatz geschoben.
- → Das ML-System wird nur mit geeignet verschlüsselten Datensätzen trainiert und betrieben (Homomorphic Encryption).

An dieser Stelle ist jedoch zu betonen, dass viele der genannten Herangehensweisen, beispielsweise das Gradient Masking, mittels eines Substitut-Modells als Ergebnis einer Extraction Attack einfach umgangen werden können.

Abschließend sei gesagt, dass Angreifer und Verteidiger in einem dynamischen Wechselspiel stehen, also eine Verzahnung der betrachteten Angriffsvektoren und Verteidigungsmechanismen vorliegt. Im Fall von ML-Systemen kann dieses beispielsweise anhand sogenannter Sicherheits-Evaluations-Kurven (Genauigkeit des Modells vs. Stärke des Angriffs) quantifiziert werden. Nach momentanem Forschungskonsens kann zu jedem erdenklichen Angriffsvektor zwar ein entsprechender Verteidigungsmechanismus konstruiert werden, zu diesem wiederum lässt sich jedoch stets ein adaptiver Angriff (im simpelsten Szenario bereits der gleiche Angriffsvektor, bloß mit geänderten Parametern) konstruieren, der die Absicherung wirkungslos macht. Für alle weiteren Forschungsbemühungen bedeutet dies zunächst, dass sämtliche neuen Verteidigungsmaßnahmen umfassend und sorgfältig gegen adaptive Angriffe ausgewertet werden müssen. Obwohl solche bereits in Ansätzen existieren, beruhen alle bisher getroffenen Garantien auf unzureichenden Metriken. So existiert insbesondere im Bereich der Bildverarbeitung bisher keine Metrik, die deckungsgleich mit der visuellen Wahrnehmung eines Menschen legitime von manipulierten Eingaben zuverlässig unterscheiden kann.

Aus dieser Herangehensweise können kurzfristig Fortschritte bezüglich des Verständnisses der Wirkungsweise von Angriffsmodellen und den Anfälligkeiten von ML-Systemen gewonnen werden. Langfristig jedoch ist diese aus Sicht der IT-Sicherheit nicht zufriedenstellend, und der Fokus muss zwangsläufig auf der Entwicklung formalerer Robustheits-Garantien liegen. Es besteht Forschungs- und Standardisierungsbedarf, um zukünftig verlässlicher zu sein und mehr IT-Sicherheit bieten zu können.

4.4.2.3 Risikobewertung IT-Sicherheit

KI kann als singuläres Produkt (Software für Spracherkennung) verstanden werden oder als Einbettung in ein System mit verschiedenen Bestandteilen (z. B. Teil einer Online-Plattform einer Versicherung oder Teil der Steuerung eines Industrieroboters). Das KI-System wiederum ist Teil einer Umwelt, hat zu diesen Schnittstellen (z. B. Wissensdatenbank oder Videosensoren als Datenquelle), die auf die KI wirken. Andererseits hat die KI Auswirkungen (Entscheidung über Versicherungsleistung oder Funktionsweise des Industrieroboters) auf ihre Umwelt. Die Auswirkungen können neben den vielfältigen positiven Ergebnissen auch Risiken unterschiedlichen Schweregrades bergen, u. a. auf die IT Sicherheit.

Die Risikobewertung IT-Sicherheit für KI betrifft Safety-, Security- und Privacy-Aspekte (siehe Kapitel 4.4) und es bieten sich verschiedene Vorgehensweisen an.

Zum Beispiel kann das Meta-Modell der DIN SPEC 92001-1 [87] (siehe Abbildung 14) hilfreich sein. Dieses empfiehlt, Sicherheitsanforderungen (Safety, Security und Privacy) zu den drei Säulen „Functionality & Performance“, „Robustness“ und „Comprehensibility“ zu ermitteln sowie die Einstufung des jeweiligen Risikos .

Für die nachhaltige IT-Sicherheit eines KI-Systems muss der gesamte „Life Cycle“ Berücksichtigung finden. Jedes Stadium beinhaltet mögliche Risiken, die analysiert und bewertet werden sollten. Dies beginnt mit der Konzeption und der Entwicklung und erstreckt sich über Test, Training, Verteilung und Betrieb bis zur Stilllegung. Dabei sind sowohl das KI-Modell z. B. mit Maschine Learning als auch die Daten für Test, Training und Betrieb, das Produkt, das IT-Gesamtsystem und die Interaktionen mit der Umgebung in der Risikobetrachtung zu berücksichtigen. Ein Management der IT-Sicherheit, ein KI-System in Bezug auf Safety, Security und Privacy wie beispielsweise mit ISO/IEC 27001 [122] und ISO/IEC 27005 [210] kann große Unterstützung bieten. Inwieweit die Norm für KI-Systeme erweitert werden kann oder sollte, müsste geprüft werden.

BEISPIEL: Eine intelligente, digitale Videokamera mit eingebettetem trainierbarem KI-basiertem Analysemodul sollte IT-Sicherheit für Safety, Security und Privacy bereits im Design und der Entwicklung berücksichtigen, um als intelligentes Endgerät (IoT) in verschiedenen Einsatzszenarios, z. B. im selbstfahrenden Auto oder als Überwachungskamera im sensiblen Zutrittsbereich – Bereiche mit höherem Risiko –, eingesetzt werden zu können.

Der Schutz der verwendeten Daten, deren Qualität und Vertrauenswürdigkeit sind von großer Bedeutung, da sie die Funktion und Ergebnisse im Test, Training und Betrieb intensiv beeinflussen. Die Aufgabe der IT- Sicherheit besteht darin, dass weder die Daten noch das Modell, das System und das Umfeld während der Test-, Trainings- oder Betriebsphase Ziel eines erfolgreichen Angriffs werden. Um vorbeugende Maßnahmen ergreifen zu können, ist eine Risikobewertung zwingende Voraussetzung.

Vielfalt der Gesetze, Normen und Standards zur Risikobewertung für IT-Sicherheit

Für die Risikobewertung und Klassifizierung stehen aktuell verschiedenste Ansätze, Empfehlungen, Modelle, Vorgehensweisen, Standards und Gesetze sowohl in Deutschland als auch auf EU- oder internationaler Ebene zur Verfügung. Auch hier empfiehlt sich eine Prüfung und Bewertung für KI-Systeme sowie die Entwicklung einer praxisorientierten gemeinsamen Basis. Im Folgenden werden nur einige beispielhaft genannt, nicht alle sind spezifisch nur für IT- Sicherheit und/oder KI ausgeprägt:

→ **Deutsche Datenethik-Kommission** [10] empfiehlt fünf Kritikalitätsstufen von Anwendungen ohne oder mit geringem bis zu unvertretbarem Schädigungspotenzial und sieben Bewertungskriterien:
- Die Würde des Menschen
- Selbstbestimmung
- Privatheit
- Sicherheit
- Demokratie
- Gerechtigkeit und Solidarität
- Nachhaltigkeit

→ **EU-Weißbuch KI** [15] beschreibt zwei Risikostufen, d. h. KI mit „hohem Risiko“ und die übrigen KI-Anwendungen. Es sind Hinweise auf KI-Anwendungen mit hohem Risiko enthalten, z. B. biometrische Fernindikation im öffentlichen Raum. Folgende Kriterien werden vorgeschlagen:
- Vorrang menschlichen Handelns und menschlicher Aufsicht
- Technische Robustheit und Sicherheit
- Privatsphäre und Datenqualitätsmanagement
- Transparenz
- Vielfalt, Nichtdiskriminierung und Fairness
- Gesellschaftliches und ökologisches Wohlergehen
- Rechenschaftspflicht

→ **EU Cyber Security Act** [188] (siehe Kapitel 4.4.1) nennt die drei Stufen der Vertrauenswürdigkeit Basic, Substantial, High.

→ **Maschinenrichtlinie** [94] (siehe Kapitel 4.4.1)

→ **ISO/IEC 27001** [122] **bzw. ISO/IEC 27005** [210] **Risikomanagement IT-Security** nimmt Bezug zu ISO 31000 [93], ergänzt diese allerdings um konkretere Punkte (Auszug):
- „the strategic value of the business information process;
- the criticality of the information assets involved;
- operational and business importance of availability, confidentiality and integrity;
- stakeholders’ expectations and perceptions, and negative consequences for goodwill and reputation; Additionally, risk evaluation criteria can be used to specify priorities for risk treatment.“

→ **IEC 61508** [79]–[86] **und IEC61511** [211] bezeichnet vier Sicherheitsstufen oder Sicherheits-Integritätslevel (SIL) aus der Funktionalen Sicherheit, die der Beurteilung von E/E/PE-Systemen in Bezug auf die Zuverlässigkeit von Sicherheitsfunktionen dienen. Bewertungskriterien sind hier die Gefährdung von Leib und Leben über das Schadensausmaß und die Eintrittswahrscheinlichkeit.

→ **ISO/IEC 15408** [48]–[50] kennt
- sieben Stufen der Vertrauenswürdigkeit (EAL, siehe Kapitel 4.1.2.2.2 „Common Criteria“) in eine Sicherheitsleistung eines IT-Systems/Produktes über
- elf definierte Funktionsklassen (z. B. die Sicherheitsprotokollierung, Kommunikation, kryptografische Unterstützung, Identifikation und Authentisierung, den Schutz der Benutzerdaten und Sicherheitsfunktionen und den Evaluierungsgegenstand (EVG)-Zugriff) und
- sieben organisatorische Klassen für Auslieferung und Betrieb, Entwicklung, Qualität der Handbücher, Funktionstests und die Schwachstellenbewertung.

→ **DIN SPEC 92001** [87] unterteilt die Bewertungskriterien in drei Säulen (Functionality & Performance, Robustness, Comprehensibility) und zwei Risikostufen („low, high“).

→ **ISO/IEC 23894 AI Risk Management** (in Bearbeitung) nennt mögliche Bewertungskriterien (Auszug):
- Security
- Privacy
- Robustness
- Availability
- Integrity
- Maintainability
- Availability and quality of data
- AI expertise

→ **ISO 31000** [93] Risikomanagement Leitlinien

→ Die Veröffentlichung von **VDE und Bertelsmann Stiftung** sowie weiteren Autoren „From Principles to Practice; An interdisciplinary framework to operationalise AI ethics“ [123] stellt eine Risiko Matrix mit fünf Klassen von KI-Application-Bereichen mit Risikopotenzial vor. Die Bereiche erstrecken sich von „no ethics rating required“ in Klasse 0 bis zu „prohibition of AI systems“ in Klasse 4.

→ **Roadmap SafeTRANS „Safety, Security, and Certifiability of Future Man-Machine Systems“**, als Beispiel zur Verzahnung von Security und Safety (siehe Kapitel 11.3)

→ **DGSVO** (siehe Kapitel 4.4.1) mit der Herausforderung einer Datenschutzfolgeabschätzung (Risiko) bei „hohem“ Risiko nennt Risikobereiche, z. B. Gesundheitsdaten.
→ **ISO/IEC 29134** [212] Datenschutzfolgeabschätzung
→ **verschiedene Schutzstufenkonzepte für personenbezogene Daten**, z. B.
  - des LfD Niedersachsen oder
  - des Unabhängigen Datenschutzzentrums Saarland oder
  - das Standard-Datenschutzmodell (SDM) der Konferenz der unabhängigen Datenschutzbehörden des Bundes und der Länder (DSK)

Für KI-IT-Sicherheit werden sowohl Sicherheitskriterien und Risikobewertung für das sichere Produkt oder System benötigt als auch für einen sicheren Life Cycle. Die Kriterien und Risiken sind zudem je nach Anwendung (Use Case) und Akteur wie Hersteller des KI-Produkts, Test, Training, Konfiguration, Installation, Integration, Betrieb und Nutzung des KI-Systems durchaus unterschiedlich. Dabei erfordern die verwendeten lernenden Systemen und Daten zusätzliche Beachtung und möglicherweise zusätzliche Kriterien.

Die vorhandenen Gesetze, Normen und Standards bieten unterschiedliche Kriterien und Bewertungsmaßstäbe. Für die praktische Umsetzung im Sinne der Wirtschaft ist dies eine Herausforderung und es wäre eine Basis mit gemeinsamen Kriterien und Bewertungsmaßstäben für die IT-Sicherheit und die Risikobewertung hilfreich. Diese könnten dann je nach Branche vertieft und ergänzt werden.

4.4.3 Normungs- und Standardisierungsbedarfe

Als konkrete Handlungsempfehlungen an Normung und Standardisierung, Forschung und öffentliche Hand haben sich folgende Themen ergeben:

4.4.3.1 Basis für Normung und Standardisierung

BEDARF 1:
Recherche/Prüfung/Bewertung bestehender Normen, Konformitäts- und Zertifizierungsverfahren sowie vorhandener Gesetze
Für IT-Systeme existieren bereits vielfältige Normen, Standards und Regulierungen, die für Systeme mit KI berücksichtigt werden können bzw. müssen. KI-Systeme kommen zunehmend in industriellen Produktionsumgebungen (operational IT = OT) und für Aufgaben mit Safety-Anforderungen zum Einsatz. Der erste Schritt sollte eine angemessene Recherche, Prüfung und Bewertung vorhandener Normen, Konformitäts- und Zertifizierungsverfahren sowie Regulierungen sein für die IT-Sicherheit und Risikobewertung, um diese um KI-Spezifika und ggf. um Dokumentationsanforderungen zu erweitern. Vorschläge für Harmonisierungen und Konsolidierung sind ebenfalls empfehlenswert.

BEDARF 2:
Empfehlungen für Akteure/Marktteilnehmer
Hersteller, Inverkehrbringer oder Nutzer von KI-Systemen könnten mit Normen und konkreten Handlungsempfehlungen dahingehend unterstützt werden, dass auch technisch weniger Versierte geeignete IT-Sicherheit umsetzen können. Bei KI-Systemen sind die Zuordnung des Risikos, der Kritikalität und Vertrauenswürdigkeit u. U. schwierig. Beispielsweise
→ wird eine universellere KI entwickelt, vortrainiert und vom Hersteller in den Verkehr gebracht (Mustererkennung per Video).
→ Im Weiteren wird diese KI durch einen Integrator für speziellere Anwendungen trainiert/angepasst/customized/eingebaut und als eigenes Produkt vermarktet (z. B. Erkennung von Straßenschildern).
→ Dieses Produkt wird von einem Unternehmen erworben, eventuell eingebaut, weiter optimiert/trainiert (z. B. in einem Fahrzeug) und als Gesamtsystem Endnutzern (Fahrern) zur Verfügung gestellt.
→ Dabei werden Daten aus dem Fahrzeug an einen Cloud-Dienstleister übertragen, der die KI-Software in seinem Rechenzentrum hostet.

Hieraus ergeben sich komplexe Fragestellungen zur IT-Sicherheit, zu Risiken und damit verbundenen Haftungsfragen. Im Rahmen der Normung, beispielsweise durch praktikable Handlungsanleitungen, könnte hier Unterstützung geleistet werden.

BEDARF 3:
Erarbeitung von Ergänzungen/Anpassungen im Risikomanagement
Für die Risikobewertung von IT-Sicherheit hinsichtlich KI-Systemen wird es voraussichtlich erforderlich sein, weitere Normungsinhalte zu erarbeiten. Innerhalb der IT-Security-Management-Normenfamilie ISO/IEC 27000 könnte für KI-Systeme z. B. eine „eigene“ KI-spezifische Norm ISO/IEC 2700x oder eine Ergänzung in der vorhandenen

ISO/IEC 27005 (Risikobetrachtung) [210] erarbeitet werden. Das passende Vorgehen ist abzuwägen.

4.4.3.2 Allgemeiner Rahmen für IT-Sicherheit

BEDARF 4:
Kritikalitätsstufen und IT-Sicherheit verbinden
Die von der Daten-Ethikkommission vorgeschlagenen Kritikalitätsstufen sollten hinsichtlich der Verwendung für IT-Sicherheit geprüft werden und um möglichst präzise Vorgaben im Rahmen einer Normung aufzuzeigen.

BEDARF 5:
IT-Sicherheitskriterien für Trainingsmethoden definieren
Für das Lernen von KI-Systemen fehlt es im Sinne der IT-Sicherheit derzeit an eindeutigen Kriterien. Diese Lücke an IT-Sicherheitskriterien für Trainingsmethoden kann zu einem unbeabsichtigten Eingriff Externer, dem Einsatz schadhafter Daten oder anderer Manipulation führen.

BEDARF 6:
Explainable AI schaffen
Es ist notwendig, die relevanten Aspekte für Transparenz zu definieren. Dies schließt die beiden Begriffe „Nachvollziehbarkeit“ und „Nachweisbarkeit“ mit ein. Hier ist das Ziel klar: ein erklärbares KI-System.

BEDARF 7:
Controls für IT-Sicherheit definieren
Für die Normung und Prüfung von KI-Systemen ist es erforderlich, passende Maßnahmen (Controls) für IT-Sicherheit (enterprise IT security, OT security, safety IT security; privacy) zu definieren. Diese sind für die Umsetzung hilfreich und sind prüf- und zertifizierbar.

BEDARF 8:
KI-Security-by-Design und KI-Security-by-Default
Wirkungsvolle IT-Sicherheit und Datenschutz in einem KI-System müssen bereits im ersten Schritt der Entwicklung („Design“) ganzheitlich berücksichtigt und gestaltet werden sowie der Funktion zugrunde liegen („Default“). Deshalb sind für die Normung passende Engineering Requirements unbedingt erforderlich („Security-by-Design and -Default“) und könnten ggf. als Ergänzung in vorhandene Normung für sichere Softwareentwicklung aufgenommen werden.

4.4.3.3 Daten

BEDARF 9:
Verifikation der Herkunft und Schutz der Daten
Der Schutz der Daten vor Manipulation in einem KI-System ist von hoher Relevanz bedingt durch die spezifischen Charakteristika von KI-Systemen. Es müssen Mechanismen implementiert werden, welche gewährleisten, dass eingebrachte gefälschte Daten (durch Angreifende) in die operierenden Algorithmen von diesen identifiziert und zurückgewiesen werden können – hier besteht Forschungsbedarf. Die eindeutige Nachvollziehbarkeit und/oder Verifikation der Datenherkunft und -verwendung sollte gegeben sein und durch Methoden unterstützt werden. Infrage kommen könnten hier Technologien wie etwa Blockchain oder andere kryptografische Verfahren.

Darauf aufbauend gilt es zu klären, wie der Umgang mit (Trainings-)Daten sein soll, wenn diese außerhalb des späteren Anwendungsfeldes gesammelt bzw. genutzt werden. Die Risikobetrachtung sollte unabsichtliche und absichtliche Beeinflussungsmöglichkeiten und die Gewichtung der (Trainings-)Daten (Online-Daten und Offline-Daten) berücksichtigen. In diesem Zusammenhang sollten das Design und der Import der Eingangsdaten, die Auswahl und Herkunft der Test- und Trainings-Daten, die sicheren und richtigen Daten im Betrieb sowie die Ausgabe-Daten ein Handlungsfeld sein.

BEDARF 10:
IT-Sicherheit der Trainings-Daten
Daten sowie die Trainings-/Betriebsmodelle sollten verifizierbar sein. Für das Training selbst braucht es eine festgelegte Semantik bzw. einen semantischen Kontext. Es gibt beispielsweise unterschiedliche Trainingsarten wie „Einzeltraining“ oder mehrere KI-Systeme in einem „Gruppentraining“. Das Trainingskonzept sollte in die IT-Sicherheit einbezogen, auf diesbezügliche Risiken untersucht und mit möglichen Schutzmaßnahmen versehen werden.

4.4.3.4 Lernende Systeme

BEDARF 11:
IT-Sicherheitskriterien für lernende Systeme definieren
Ein Kernelement einer Künstlichen Intelligenz sind die lernenden Systeme (ML, Deep Learning) und deren IT-Sicherheit. Hier werden neue IT-Sicherheitsuntersuchungen und -vorgaben erforderlich. Jegliche Kenntnisse über Einflussgrößen und Risiken müssen bei Definitionen, Festlegungen

und Gewichtungen, die allgemein im Zusammenhang stehen, für die IT-Sicherheit Beachtung finden. Security in hybriden KI-Systemen (wissensbasierte und datenbasierte Vorgehensweise) ist ein weiteres Handlungsfeld.
Für die neuen Anforderungen an Kriterien für lernende Systeme und deren Komponenten sollte ermittelt und ggf. erforscht werden, welche Security-Controlls, Prüfverfahren, Auditierung und Zertifizierungen erforderlich sind.

BEDARF 12:
Verifizierbare Identität für KI-Algorithmen
Zur Stärkung des Vertrauens in KI-Systeme müssen KI-Algorithmen mit einer verifizierbaren Identität versehen und ihre Funktion und Arbeitsweise in Dokumentationen festgehalten werden. Ergebnisse sollten möglichst nicht nur als Wahrscheinlichkeitswert einer Ergebnisklasse als Grundlage einer Entscheidung, sondern als Konfidenzintervall ausgewiesen werden.

4.4.3.5 Forschungsthemen

BEDARF 13:
IT-Sicherheitsmetriken für lernende Systeme und Adversarial Machine Learning (AML)
Auf dem Gebiet des AML, speziell mit Hinblick auf Anwendungen in der Praxis, ist weitere Forschung und Entwicklung erforderlich, da die Analyse der IT-Sicherheit eines jedes KI-Systems eine komplexe, individuelle Analyse seiner Sicherheit erfordert.

Unter anderem sollte auf eine Entwicklung aussagekräftiger Metriken zur Bewertung der Robustheit abgezielt werden, welche perspektivisch als Grundlage für Normen und Standards zugrunde gelegt werden können, um eine Vergleichbarkeit von Robustheit zu ermöglichen.

Die Erforschung und Entwicklung einer allgemeinen Taxonomie für AML wird empfohlen, die sowohl Angriffs- als auch Verteidigungsverfahren auflistet, die bei der Entwicklung von Modellen berücksichtigt werden sollten. In einem weiteren Schritt könnte eine „Toolbox“ zur Verfügung gestellt werden, welche Angriffsvektoren enthält und die auf bestehende trainierte Systeme angewendet werden kann. Hier wäre insbesondere an das Zusammenführen und Vereinheitlichen von schon bestehenden Tools, Verteidigungsmechanismen und Robustheits-Konzepten zu denken. Auch standardisierte automatisierte Verfahren wären denkbar und hilfreich.

BEDARF 14:
Auswirkungen von Verfügbarkeit von Ressourcen
Die Verfügbarkeit von Ressourcen, wie z. B. Prozessorleistung, hat auf die IT-Sicherheit eines KI-Systems Auswirkungen und ist bei der Betrachtung von verschiedenen Phasen relevant. Zunächst ist dies die Entwurfszeit während der Entwicklung, im Weiteren die Test- und Trainingszeit und dann die Laufzeit, also während des Betriebs. In allen Phasen können mangelnde Ressourcen zu fehlerhaften/falschen Ergebnissen führen. Dies könnte auch als Angriffsvektor im Sinne der IT-Sicherheit infrage kommen. Hier sollte mithilfe der Forschung eine Simulation möglicher Angriffsszenarien auf die Ressourcen und Arbeitsweisen der IT-Sicherheit eines KI-Systems untersucht und eine Sammlung an Angriffsvektoren erstellt werden. Diese Simulationen bedürfen wiederum einer Vielzahl von (Trainings-)Daten (z. B. Gesichtserkennung, Rotation von Gesichtern, Hintergründe von Gesichtern, Hautfarbe etc.).

4.5 Industrielle Automation

Künstliche Intelligenz (KI) stellt eine wichtige und wesentliche Schlüsseltechnologie dar, um den Erhalt der wirtschaftlichen Leistungsfähigkeit Deutschlands zu sichern. Insbesondere weist KI ein besonders hohes Potenzial auf, um Abläufe und Prozesse in der produzierenden Industrie – d. h. Industrie 4.0 [213] – nachhaltig zu gestalten und die Wertschöpfung durch Dynamisierung und Flexibilisierung zu steigern sowie Geschäftsmodelle in der produzierenden Industrie zu verändern. Dabei können sowohl traditionelle, aber auch neu gestaltete Produktionsabläufe und Sekundärprozesse wie beispielsweise Logistikprozesse durch KI verbessert, optimiert und flexibilisiert werden.

Im Englischen und vermehrt auch im deutschsprachigen Raum wird auch von Industrial Artificial Intelligence oder Industrial AI gesprochen und diese umfasst alle Anwendungsfelder von Künstlicher Intelligenz in der industriellen Anwendung [214]. Im Rahmen der Arbeiten des Zukunftsprojekts Industrie 4.0, welches durch die Deutsche Bundesregierung im Jahr 2015 initiiert wurde, wurde das Themenfeld strukturiert aufgearbeitet. Ausgehend von existierenden Wertschöpfungsprozessen der produzierenden Industrie [215] wurden entsprechende zukünftige Anwendungsszenarien definiert [216]. Dabei umfassen die Anwendungsszenarien einen breiten Anwendungsbereich wie beispielsweise die auftragsgesteuerte Produktion auf Basis dynamischer Wertschöpfungs- und Liefernetzwerke, wandlungsfähige Fabriken, welche eine flexible Adaption von Fertigungsressourcen einer Fabrik ermöglichen, smarte Produktentwicklung, u. v. m. Diese Anwendungsszenarien stellen dabei die Grundlage für weiterführende Verfeinerungen und Analysen zur Ableitung etwaiger Forschungs- und Normungsbedarfe.

Eine wichtige Rolle in der digitalen Transformation wird der digitalen Abbildung der physischen Realität zugeschrieben, dem sogenannten digitalen Zwilling. Um die Interoperabilität innerhalb eines digitalen Ökosystems sicherzustellen, erarbeitet die Plattform Industrie 4.0 gemeinsam mit allen beteiligten Institutionen die Spezifikation der sogenannten Verwaltungsschale als digitales Abbild jedes relevanten Gegenstands (Asset) in der vernetzten Produktion. Eine Verwaltungsschale speichert alle wesentlichen Eigenschaften eines Assets wie beispielsweise physische Eigenschaften (Gewicht, Größe), Prozesswerte, Konfigurationsparameter, Zustände und Fähigkeiten. Dabei ist die Verwaltungsschale nicht nur Informationsspeicher, sondern auch Kommunikationsschnittstelle, über die ein Asset in die vernetzt organisierte Industrie-4.0-Produktion eingebunden wird. Hierdurch ist es möglich, auf alle Informationen in einem Asset zuzugreifen und dieses zu kontrollieren. Dies stellt eine wichtige Grundlage für die Anwendung von Künstlicher Intelligenz für Industrie 4.0 dar, da hierdurch auf Daten- und Metadaten relevanter Assets einheitlich zugegriffen werden kann und diese in einem strukturierten Datenformat zur Verfügung stehen.

Ferner werden im Kontext von Industrie 4.0 weitere Anwendungen der Künstlichen Intelligenz betrachtet. Neben der autonomen Intralogistik (siehe hierzu auch Kapitel 4.6 Mobilität und Logistik) werden beispielsweise auch die industrielle Bildverarbeitung und Bilderkennung sowie die Verbesserung der Interaktion und Integration von Mensch und Maschine berücksichtigt. Zum einen durch den Einsatz neuer Interaktionsmechanismen wie Sprache und Geste, durch neue Darstellungsmöglichkeiten wie Augmented Reality und die Stärkung der Zusammenarbeit wie beispielsweise durch kollaborative Robotik. Hierbei finden KI-Technologien durchweg intensive Anwendung. Normungsaspekte, welche für die industrielle Automation spezifisch sind, werden aktuell untersucht. Im Rahmen der Arbeiten der Arbeitsgruppe 2 der Plattform Industrie 4.0 wurde der Einfluss von KI in ausgewählten Anwendungsszenarien bereits detailliert betrachtet [23], [24]. Bekannte Anwendungsbeispiele für KI in Industrie 4.0 sind u. a. die vorhersagende Wartung, wobei auf Basis symbolischer Modelle und gesammelter Betriebsdaten die Lebensdauer und der notwendige Wartungszeitpunkt von Komponenten vorhergesagt wird. Zudem werden die weiterführende Beobachtung der Produktionsprozesse, die Vorhersage von Prozess- und Produktqualität und (aktuell noch semi-automatisch) die Parametrierung und Konfiguration der technischen Systeme zur Prozess- und Qualitätsoptimierung prognostiziert. Abbildung 22 stellt eine Auswahl dieser Anwendungsfälle von Industrie 4.0 exemplarisch möglichen KI-Technologien gegenüber.

Die Normung und Standardisierung von KI hat in Deutschland einen wesentlichen Stellenwert – nicht zuletzt aufgrund der nationalen KI-Strategie der Bundesregierung.

Aus diesem Grund wurde das Thema KI bereits im Rahmen der Version 4 der DIN/DKE Normungsroadmap Industrie 4.0 [217] explizit und dediziert adressiert. Bei der Normung zu KI in industriellen Anwendungen ist zwischen horizontalen und vertikalen Aspekten zu unterscheiden. Horizontale Normen und Standards weisen eine anwendungsbereichsübergreifende Gültigkeit auf, z. B. allgemeingültige technische Regeln zur Qualitätsmessung von (technischen oder informationellen) Systemen. Im Gegensatz dazu existieren Normen und Standards in unterschiedlichen Anwendungsbereichen (vertikale

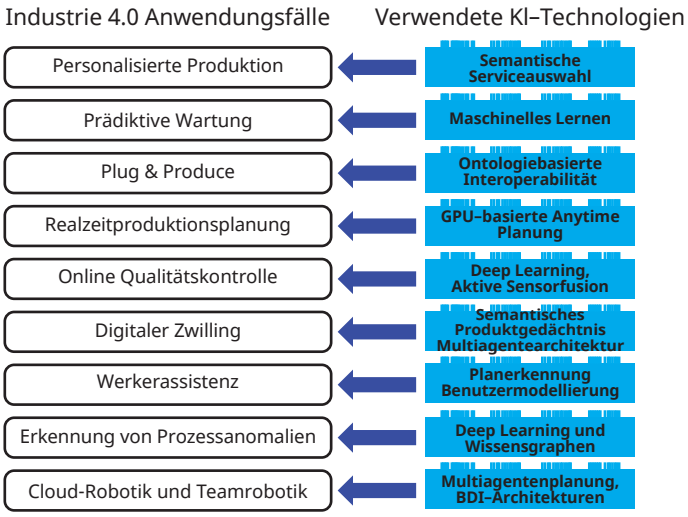


Abbildung 22: Exemplarischer Einsatz von Industrieller KI in ausgewählten Industrie-4.0-Anwendungsfällen

Normen) wie z. B. Industrie 4.0. In diesen Anwendungsbereichen werden spezifische technische Regeln erarbeitet, welche die konkreten Anwendungen und spezifischen Anforderungen des Anwendungsbereichs widerspiegeln.

4.5.1 Status quo

In der deutschen Industrie haben bereits seit mehreren Jahren das Thema KI und damit verwandte Themengebiete einen hohen Stellenwert. Verbände wie VDMA, ZVEI und Bitkom sowie der VDI e. V. und VDE e. V. befassen sich in diversen Arbeitsgruppen mit unterschiedlichen Aspekten und verschiedenen Anwendungen von KI. Hierbei werden eine Vielzahl unterschiedlicher Beschreibungen der Anwendung von KI in Form von Anwendungsszenarien oder Anwendungsbeispielen betrachtet. Die Austauschbarkeit und Vergleichbarkeit ist national und international aus diversen Gründen (z. B. fehlende einheitliche Beschreibungsmethodik, heterogene Betrachtungswinkel und sehr unterschiedliche Abstraktionsniveaus) bis dato nicht gegeben.

Der Einsatz von KI in industriellen Anwendungen kann, je nach Anwendungszweck und Funktion der KI, Einfluss auf die Erfüllung von in Normen und anderen technischen Regeln beschriebenen Anforderungen haben. Wird beispielsweise KI-Technologie eingesetzt, um z. B. das Verhalten automatisierter Funktionen anzupassen, muss der Einfluss der KI auf das automatisierte System im Zuge eines Konformitätsbewertungsverfahrens berücksichtigt werden. Dies gilt insbesondere für industrielle Anwendungen mit Anforderungen hinsichtlich funktionaler und industrieller Sicherheit. Demzufolge ist es notwendig, stets die Erfüllung normativer Rahmenbedingungen unter Berücksichtigung der Funktion und des Einflusses von KI zu prüfen und sicherzustellen. Eine objektive Bewertung des Einflussbereichs der KI ist in diesem Zusammenhang zwingend notwendig.

Im Rahmen des DKE/AK 801.0.8 wurde eine VDE Anwendungsregel VDE-AR-E 2842-611 „Spezifikation und Entwurf autonomer/kognitiver Systeme“ [218] entwickelt, in die Begriffe und Konzepte für den Umgang mit autonomen/kognitiven Systemen eingeflossen sind. Es wird ein Referenzmodell für System- und Applikationsarchitekturen erarbeitet, das den gesamten Lebenszyklus betrachtet. Einige Ansätze aus dem Bereich der funktionalen Sicherheit werden auf dieses Referenzsystem übertragen. Das IEC SEG 10 befasst sich mit Ethikaspekten in autonomen Anwendungen und KI als wichtigen Ansatz zur Technikakzeptanz. Hierbei werden insbesondere gesellschaftlich relevante Aspekte betrachtet und Empfehlungen für das IEC Standardization Management Board (SMB) erarbeitet.

Der VDI e. V. deckt ein breites Spektrum der Ingenieurwissenschaften ab. KI stellt in diesem Zusammenhang ein Querschnittsthema dar, weshalb sich zahlreiche Fachgesellschaften und Fachbereiche mit diesem Thema aus unterschiedlichen Blickwinkeln auseinandersetzen. Im Rahmen der Arbeiten des VDI/VDE-GMA-Fachbeirats 8 Optische Technologien wurde ein Statusreport Maschinelles Lernen [219] erarbeitet, welcher die Anforderungen der Anwender an die wissenschaftliche Forschung und den Wissenstransfer fokussiert. Die Richtlinienreihe VDI/VDE/VDMA 2632 Industrielle Bildverarbeitung [220]–[223] beschreibt die Erstellung von Lasten- und Pflichtenheften und die Abnahme von klassifizierenden Bildverarbeitungssystemen. Diese befindet sich aktuell in der Überarbeitung, da sich Bildverarbeitungssysteme mit Künstlicher Intelligenz grundlegend anders verhalten als konventionelle Verfahren.

Die Taskforce „Usage of new technologies“ des IEC/TC 65/JWG 23 (gemeinsam mit ISO/TC 184 SC1) führt eine Evaluation neuer Technologien und deren Normungsrelevanz im Bereich Smart Manufacturing durch. Hierbei wird als eine Zukunftstechnologie die KI in industriellen Anwendungen betrachtet. Die nationale Spiegelung dieses vertikalen KI-Bereichs erfolgt im Arbeitskreis DKE/AK 931.0.14.

Auf europäischer Ebene wurde bei CEN/CENELEC im April 2019 die Focus Group on Artificial Intelligence gegründet. Sie berät CEN und CENELEC bei der Entwicklung und Verbreitung von KI in Europa und konzentriert sich bei ihrer Arbeit auf besondere europäische Bedürfnisse, während allgemein global relevante Fragen nach Möglichkeit auf globaler Ebene gelöst werden. Unter anderem soll die CEN-CENELEC-Fokusgruppe die Leitlinien der High Level Expert Group on Artificial Intelligence, eingesetzt durch die Europäische Kommission [224] zum Thema Artificial Intelligence for Europe, berücksichtigen. Die CEN-CENELEC-Fokusgruppe erarbeitet eine gemeinsame Vision hinsichtlich der europäischen KI-Normung. Im Rahmen des technischen Komitees CLC/TC 65X bei CENELEC werden Aspekte der Nutzung von KI in der industriellen Automation auf europäischer Ebene betrachtet.

Das derzeitig hohe Interesse an KI führt zu einer Vielzahl unterschiedlicher Aktivitäten bei verschiedenen Verbänden, Institutionen, Konsortien und Vereinen hinsichtlich Anwendung und Normung von KI im industriellen Bereich. Um parallele Mehrarbeit im Kontext der Normung von KI für industrielle Anwendungen zu vermeiden, den Austausch zwischen diesen verschiedenen Aktivitäten zu fördern und letztendlich ein möglichst harmonisiertes nationales Meinungsbild zu erarbeiten, wurde der Expertenrat für Künstliche Intelligenz in industriellen Anwendungen durch das Standardization Council Industrie 4.0 ins Leben gerufen. Zielstellung ist die nationale Koordination und Harmonisierung der Standardisierungsaktivitäten zur Entwicklung eines konsolidierten Bildes der Anforderungen und Normungsbedarfe im Kontext von KI in Industrie 4.0 der deutschen Wirtschaft und die Koordination geeigneter Aktivitäten in der technischen Regelsetzung. Der Expertenrat für Künstliche Intelligenz in industriellen Anwendungen stellt den Dreh- und Angelpunkt für Diskussionen und die Koordination der technischen Regelsetzung im Bereich der Künstlichen Intelligenz für industrielle Anwendungen dar. Die Aufgaben umfassen dabei die Sammlung von Anwendungsfällen und der darauf aufbauenden Ableitung von Normungs- und Standardisierungsbedarfen, die Erarbeitung und Spezifikation von Handlungsempfehlungen, deren Einbringung in unterschiedliche, aktuell und zukünftig in Erarbeitung befindliche, nationale und internationale Normungsroadmaps und die Koordination nationaler und internationaler Standardisierungsaktivitäten.

4.5.2 Anforderungen, Herausforderungen

Die Handlungsfelder und die Struktur dieses Kapitels orientieren sich an der thematischen Organisation der Plattform Industrie 4.0 und analysieren dessen inhaltliche Ergebnisse aus normativer Sicht: Grundlegende Anforderungen (siehe Kapitel 4.5.2.1), Anwendungsszenarien und Use Cases (siehe Kapitel 4.5.2.2), sichere vertrauenswürdige KI-Systeme (siehe Kapitel 4.5.2.3), Datenmodellierung und Semantik (siehe Kapitel 4.5.2.4) sowie Mensch und KI (siehe Kapitel 4.5.2.5).

4.5.2.1 Grundlegende Anforderungen und Herausforderungen aus Sicht der Industrie 4.0

Heute ist noch keine eindeutige Aussage möglich, wann KI-spezifische Normen/Regulationen bei einem System bzw. einer Komponente greifen. Eine Möglichkeit der Klassifikation besteht in der Definition von Autonomieklassen. Bestehenden Definitionen von Autonomieklassen mangelt es derzeit noch an objektiven Kriterien der Zuordnung und damit an der Möglichkeit, zu bewerten, wann KI-spezifische technische Regeln bei neuen Systemen bzw. Komponenten Anwendung finden. Dieser Bedarf wurde bereits im Rahmen der Aktivitäten der Projektgruppe KI der Plattform Industrie 4.0 identifiziert und im Rahmen der Arbeiten des SCI4.0-Expertenrats für Künstliche Intelligenz in industriellen Anwendungen weiter konkretisiert.

In der Entwicklung und im Betrieb moderner industrieller Anwendungen, in welchen neue Technologien insbesondere aus dem Umfeld der KI zum Einsatz kommen, nehmen neben klassischen technologischen Aspekten auch Aspekte aus weiteren Disziplinen wie z. B. Recht, Ethik oder Wirtschaft zunehmend größeren Platz ein. Dabei sind Begriffe in unterschiedlichen Disziplinen teils mit unterschiedlicher Bedeutung versehen, deren spezifische Interpretationen interdisziplinär unbekannt oder zumindest schwer zu erfassen sind. Dies gilt insbesondere auch in der Normung und Standardisierung von Künstlicher Intelligenz im Kontext von Industrie 4.0. Daher benötigt die technische Regelsetzung für die industrielle Automatisierung eine geeignete und akzeptierte gemeinsame Sprache und Begriffsverzeichnisse (Glossare, Ontologien). Hierdurch können alle interessierten Kreise möglichst umfassend mit den Begriffen bekannt gemacht werden, damit einerseits ein gemeinsames interdisziplinäres Verständnis entsteht und andererseits die Anwendersicher-

heit bei der Kooperation zwischen Mensch und Maschine, welche unter Umständen von einer eingebetteten KI gesteuert wird, gewährleistet ist.

4.5.2.2 Anforderungen und Herausforderungen hinsichtlich der Aufbereitung und Konkretisierung des Anwendungsfelds KI in industriellen Anwendungen

Um den Anwendungsbereich von Künstlicher Intelligenz strukturiert aufzubereiten, den erfolgreichen wirtschaftlichen und technischen Einsatz von KI in industriellen Anwendungen zu konkretisieren und anwendungsnah Normungsbedarfe ableiten zu können, wurde bereits in der Normungsroadmap Industrie 4.0 eine strukturierte Aufbereitung mittels Use Cases und Szenarien empfohlen. Dieser Empfehlung folgend wurden bereits existierende KI-Use-Case-Sammlungen mit Fokus auf die produzierende Industrie untersucht und eine Strukturierung bzw. Klassifikation dieser Use-Case-Sammlungen vorgenommen. Dabei hat sich insbesondere gezeigt, dass einzelne Use Cases nicht deutlich genug herausarbeiten konnten, welche Neuerungen sich aus dem Einsatz der KI im Vergleich zur Nutzung klassischer Modelle und Methoden ergeben bzw. warum das einem Use Case zugrunde liegende Problem nicht auch ohne KI gelöst werden kann. Dies ist u. a. auf ein unscharfes Verständnis des Begriffs KI zurückzuführen, was durch ein Scoping des Begriffs „KI“ und durch eine klare Klassifikation von Use Cases (insbesondere für die produzierende Industrie) adressiert werden kann, um damit zu verdeutlichen, wann es sich um einen KI Use Case handelt.

Auch wurde im Rahmen der Analyse eine Kategorisierung vorgenommen, um Anwendungsfälle von Anwendungsbeispielen zu unterschieden. Der Fokus gesammelter Beispiele liegt dabei auf der Gestaltung und Optimierung interner Produktionswertschöpfungsprozesse mithilfe von Künstlicher Intelligenz. Der Aspekt neuer Geschäftsmöglichkeiten durch KI ist dabei nicht adäquat angesprochen worden, weshalb durch geeignete Beispiele illustriert werden sollte, wie durch KI neue Geschäftsmöglichkeiten in der industriellen Produktion erschlossen werden können.

Im Umfeld von KI wird häufig von sogenannten KI-Firmen gesprochen, welche als Anbieter von Technologien, Algorithmen und Methoden aus dem Bereich KI branchenübergreifende Lösungen anbieten. Aktuell beauftragen etablierte Firmen aus der produzierenden Industrie KI-Firmen, übernehmen dabei jedoch auch weiterhin das geschäftliche Risiko und/oder erweitern ihr Portfolio durch KI, werden aber nicht durch KI-Firmen verdrängt. Bis dato sind allgemein nur sehr wenige konkrete Beispiele bekannt, in denen eine KI-Firma die Geschäftsverantwortung und das damit einhergehende geschäftliche Risiko in industriellen Anwendungen übernommen hat. Dieser Aspekt sollte durch geeignete Beispiele besser illustriert und damit näher untersucht werden.

Wie eingangs bereits angeführt dient die strukturierte Auseinandersetzung mit KI Use Cases der produzierenden Industrie u. a. dazu, etwaige Normungs- und Standardisierungsbedarfe zu identifizieren. Jedoch sind die meisten Use Cases sehr knapp und generisch beschrieben und haben deshalb keine ausreichende Detailtiefe, um etwaige Anforderungen an eine Normung oder Standardisierung ableiten zu können. Daher sollten Use Cases in einer hinreichenden Detailtiefe beschrieben werden, um daraus Anforderungen an eine Standardisierung ableiten zu können.

Die bisher im Rahmen der Arbeiten des SCI4.0-Expertenrats für Künstliche Intelligenz in industriellen Anwendungen durchgeführte Sammlung von KI Use Cases ermöglichte bereits die Ableitung der zuvor beschriebenen Handlungsempfehlungen. Um konkrete Handlungsempfehlungen bezüglich relevanter Normen und Standards systematisch ableiten zu können, sollten die bereits durchgeführten Arbeiten nun in einem systematischen Konsolidierungsprozess (im Hinblick auf Anzahl, Abdeckungsgrad und Qualität der Use Cases) so weitergeführt werden, dass eine repräsentative Use-Case-Sammlung entsteht. Ferner sollten die existierenden Use Cases im Hinblick auf eine funktionale, technische Perspektive detailliert werden, um relevante Normungsbeziehungen zu identifizieren und um eine zielgerichtete Kooperation mit thematischen Fachgremien zu etablieren, wodurch letztendlich konkrete Handlungsempfehlungen in der technischen Regelsetzung abgeleitet werden können.

Betrachtet man die Use Cases, welche bereits im Rahmen des IEC TC65 WG23 erarbeitet wurden und werden, decken diese weitestgehend die in der produzierenden Industrie vorherrschenden Wertschöpfungsprozesse production planning/engineering, production execution und product service ab, jedoch werden dabei die ebenfalls grundlegenden Wertschöpfungsprozesse product design und product configuration/sales nicht abgedeckt. Um eine vollständige Überdeckung der Wertschöpfungsprozesse in Industrie 4.0 zu erreichen, bedarf es einer Vervollständigung der bereits im Rahmen von IEC TC65 WG23 erarbeiteten Use Cases im Hinblick auf das Thema KI.

4.5.2.3 Anforderungen und Herausforderungen hinsichtlich sicherer, vertrauenswürdiger KI-Systeme

Im industriellen Kontext ist der Nachweis notwendiger Eigenschaften (z. B. Robustheit, Erklärbarkeit etc.) von essenzieller Bedeutung für IT-Security und -Safety. In der KI werden häufig sogenannte Blackbox-Machine-Learning-Methoden wie z. B. neuronale Netze eingesetzt, welche sich u. a. sehr anfällig für kleine Änderungen in den Eingabedaten zeigen. Dies machen sich u. a. sogenannte Adverserial Attacks zunutze. Etablierte Verfahren zur Verifikation wie z. B. Code Reviews sind durch den Einsatz von Blackbox-Technologien nun jedoch nicht mehr möglich. Einen möglichen Lösungsansatz, um Kenntnisse über die internen Zusammenhänge zu erlangen, stellt der Einsatz von formalen (mathematischen) Methoden dar. Zwar gibt es in verschiedenen Gremien Vorhaben, welche entweder auf den Einsatz von KI ohne Bezug zur industriellen Anwendung fokussieren (Vorhaben 24029-2 des ISO/IEC JTC 1 SC 42) oder, wie im Falle des IEC TC 65/SC 65A die in der Entstehung befindliche IEC/TS 61508-3-2 zwar die industrielle Anwendung berücksichtigen, jedoch nicht den Einsatz von KI betrachten. Demzufolge ergibt sich ein Handlungsbedarf hinsichtlich der Eignungsprüfung formaler Methoden für den Nachweis spezifischer Eigenschaften für die Nutzung von KI in der industriellen Anwendung. Konkret wurde ein entsprechender Handlungsbedarf beispielsweise bei der Richtlinienreihe VDI/VDE/VDMA 2632 [220]–[223] zur industriellen Bildverarbeitung identifiziert. Lasten- und Pflichtenhefte müssen anders erstellt werden, wenn Systeme mit künstlichen neuronalen Netzen zum Einsatz kommen. Gleiches gilt für die Abnahme und die Prüfung der Klassifikationsleistung eines Bildverarbeitungssystems.

Durch die zunehmende Dynamisierung vom Wertschöpfungsnetzen und die zugehörige Kooperation von Systemen im laufenden Betrieb steigt die Anzahl potenzieller Konfigurationsvarianten massiv an und es ist nicht mehr möglich, jede einzelne Konfiguration a priori zu betrachten. Digitale Zwillinge bzw. Verwaltungsschalen, welche die Rekonfiguration zur Laufzeit aus funktionaler Sicht ermöglichen, müssen deswegen um Safety-Eigenschaften erweitert werden, sodass die Risikobetrachtungen einer Konfiguration zur Laufzeit durchgeführt werden kann. Ganz allgemein geht es darum, Worst-Case-Annahmen über die Systemumgebung zu minimieren, um nicht unnötig die Performanz zu beeinträchtigen. Mögliche Lösungsansätze dazu sind u. a. Conditional Safety Certificates (ConSerts) [225] und Digital Dependability Identities [226]. Eine Grundidee dabei ist, Worst-Case-Annahmen, die in allen Situationen gelten, durch situationsabhängige Annahmen zu ersetzten, die zur Laufzeit überprüft werden können. Solche Ansätze müssen in der industriellen Praxis erprobt werden. Neben den Unsicherheiten in der Systemumgebung stellen die Unsicherheiten im Systemverhalten eine große Herausforderung für die Verlässlichkeit da. Die Anwendung von Methoden wie dem maschinellen Lernen führen zu unvorhersehbarem Systemverhalten. Einfache Überwachungsmechanismen, die das Systemverhalten bezüglich Verlässlichkeit einschränken, sind oft nicht anwendbar, weil sie nicht situationsspezifisch sind und in vielen Situationen die Leistungsfähigkeit einschränken. Deswegen müssen Überwachungsmechanismen erforscht werden, die Risiken der aktuellen Situation erkennen und kontrollieren können. Durch KI-Methoden können sicherheitskritische Aufgaben automatisiert werden, die zuvor nur durch Menschen übernommen werden konnten. Aufgrund der Komplexität dieser Aufgaben kann man nicht davon ausgehen, dass die Fehlerrate genauso gering ist wie bei sehr einfachen Sicherheitsfunktionen wie einem Not-Aus-Schalter. Die Anzahl der Unfälle könnte jedoch trotzdem signifikant reduziert werden, wenn eine Aufgabe durch eine KI signifikant sicherer erledigt werden könnte als durch den Menschen. Das wirft die Forschungsfrage auf, ob existierende Risikoakzeptanzkriterien für KI-basierte Sicherheitsfunktionen geeignet sind oder ob neue Konzepte eingeführt werden sollten, um die Unfallzahlen zu minimieren.

Die Anwendung von KI für (industrielle) Systeme mit Sicherheitsfunktionen stellt aktuell eine große Herausforderung dar, da entsprechende Normen und Richtlinien den Einsatz von KI nicht ausreichend berücksichtigen. So wird beispielsweise häufig missverstanden, dass ab SIL2 der Einsatz von KI verboten sei. Darüber hinaus steht in IEC 61508, dass Sicherheit erreicht ist, sobald die Sicherheitsfunktion umgesetzt worden ist. Die Kernfragestellung ist dabei, was für eine KI richtig oder falsch ist. Klare Regelungen, was eine KI darf und was nicht, sowie welche Nachweise von einer KI erbracht werden müssen, fehlen. Hierdurch ergibt sich eine deutliche Unklarheit für diverse Stakeholder wie z. B. Anwender, Lösungsanbieter, Zertifizierer etc. Aus diesem Grund ist eine Überarbeitung relevanter Normen und Richtlinien im Bereich Safety und Security, insbesondere z. B. IEC 61508 [79]–[86], für den Einsatz von KI zwingend notwendig. Demzufolge ergibt sich ein Handlungsbedarf hinsichtlich der Eignungsprüfung formaler Methoden für den Nachweis spezifischer Eigenschaften für die Nutzung von KI in der industriellen Anwendung. Konkret wurde ein entsprechender Handlungsbedarf beispielsweise bei der Richtlinienreihe VDI/VDE/VDMA 2632 [220]–[223]

zur industriellen Bildverarbeitung identifiziert. Lasten- und Pflichtenhefte müssen anders erstellt werden, wenn Systeme mit künstlichen neuronalen Netzen zum Einsatz kommen. Gleiches gilt für die Abnahme und die Prüfung der Klassifikationsleistung eines Bildverarbeitungssystems.

Der Einfluss von KI auf rechtliche und regulatorische Rahmenbedingungen wird momentan auf politischer Ebene in Deutschland und der EU diskutiert (siehe dazu „Weißbuch zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen“ [15] und den „Bericht über die Auswirkungen Künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung“ der Europäischen Kommission [227]). Dabei werden insbesondere Sicherheits- und Haftungsthemen betrachtet und spezifische Gesichtspunkte diskutiert wie z. B. spezifische regulatorische Anforderungen an KI-Anwendungen mit hohem Risiko, Änderungen in der Funktion eines Produkts nach der Inverkehrbringung durch KI-Systeme, die im Betrieb lernen, spezifische Anforderungen an die menschliche Aufsicht über den gesamten Lebenszyklus der KI-Produkte und -Systeme und die Transparenz bezüglich der Entwicklung und des Verhaltens von KI-Systemen. Hieraus können sich Wechselwirkungen mit KI-Normungsaktivitäten ergeben, die zu betrachten sind.

4.5.2.4 Anforderungen und Herausforderungen im Hinblick auf Datenmodellierung und Semantik für KI-Systeme in industriellen Anwendungen

Heute werden KI Use Cases vorrangig syntaktisch beschrieben, d. h. die Ausdrücke haben nur eine freigewählte (ontologische) Bedeutung, was die Beschreibung der Dynamik von Anwendungsfällen erschwert. Es werden aktuell semantische Modelle auf Basis etablierter oder im Verlauf von Projekten neu entwickelter Vokabulare als vorwiegend statische Architekturen von Instanzen definiert. Häufig werden mögliche Folgen von Interaktionen zwischen Modellinstanzen (sogenannte Narrationen) nicht ausreichend beschrieben. Aktuell können Interaktionen zwischen Modellen im Sinne einer zielgerichteten (Re-)Kombination nur in hochgradig individuellen und damit wenig übertragbaren Ansätzen formalisiert werden. Um diese Herausforderung zu adressieren, sollte eine narrative Darstellung basierend auf einem deklarativen Semantikstil verwendet werden, um eine durchgängige Beschreibung für die (Re-)Kombination bzw. Kompatibilität von Teilmodellen zu ermöglichen.

Gegenwärtig werden mögliche Beziehungen zwischen Komponenten bestenfalls fallweise annotiert, wie z. B., welche Anteile der Schnittstellen zur Unterstützung von Interaktionen dienen können. Etwaige Muster, nach denen solche Interaktionsmöglichkeiten gestaltet werden können, sind ebenso höchst individuell und erlauben wenig Sicherstellung des Zusammenpassens auf Modellebene. Auch kann vorab nicht überprüft werden, ob Modelle aus verschiedenen Quellen im Rahmen von etwa Qualitätszusicherungen zusammenwirken können. Aus diesem Grund wird die Anwendung von Narrativen im Entwicklungsprozess empfohlen, um in Modellen damit auszuzeichnen, welche Elemente in welchem Maß veränderlich sein können.

4.5.2.5 Anforderungen und Herausforderungen für die Kooperation zwischen Menschen und KI in industriellen Anwendungen

Wie in Kapitel 4.5.2.1 beschrieben, besteht ein Bedarf an einem gemeinsamen, interdisziplinären Verständnis aller, teils sehr heterogenen Aspekte der Anwendung von KI in der produzierenden Industrie . Aus diesen Anforderungen nach einer einheitlichen semantischen Betrachtung von Industrieanlagen samt Daten, Vorgängen, Interoperationen zwischen Mensch und Maschine und dem sprachlichen Ausdruck zur ontologischen Charakterisierung ergibt sich ein neuer Bedarf nach einem Vokabular mit Anwendungsregeln (Leitfäden), womit formale und berechenbare Ausdrücke bzw. eine Sprache sowohl von der Maschine als auch vom Menschen jeweils auf ihre eigene Art verstanden werden können. In diesem Zusammenhang werden häufig die Prinzipien der Common Logic genannt, welche für Fachexperten teils nur schwer verständlich sind und daher die Anwendung der Formalismen nicht bewerten können. Hieraus ergibt sich der Bedarf, die Prinzipien der Common Logic und deren Rolle in der Normung von KI für Industrie 4.0 anwendungsorientiert zu beschreiben.

Zu beantworten ist auch die Frage, welchen Einfluss der Einsatz von KI auf die Arbeit der Ingenieure in den unterschiedlichen Fachgebieten hat. Da KI für eine Vielzahl von Aufgaben eingesetzt wird oder eingesetzt werden kann, wird diese Frage in immer mehr Fachgebieten relevant. Entsprechend werden sich immer mehr Arbeitsgruppen mit dem Einfluss von KI auf die Arbeit von Ingenieuren und Ingenieurinnen in deren jeweiligem Fachgebiet auseinandersetzen. Dieser Herausforderung werden sich diverse Gremien – insbesondere auch Verbände und Vereine – im Kontext der

Anwendung von KI im industriellen Umfeld stellen. Der VDI hat bereits angekündigt, sich auf diesem Gebiet besonders zu engagieren.

4.5.3 Normungs- und Standardisierungsbedarfe

Normung, Standardisierung und Technische Regelsetzung

BEDARF 1:
Kriterien zur Klassifikation von Systemen bzw. Komponenten im Rahmen der KI
Es wird vorgeschlagen, Kriterien zur Abgrenzung (u. a. zu bestehenden Automationssystemen) zu definieren. Bestehende Einteilungen nach Autonomieklassen könnten hierzu erweitert werden.

BEDARF 2:
Kriterien für die Einordnung von Use Cases unter Berücksichtigung der Rolle von KI
Es werden klare Kriterien benötigt, wann es sich um einen KI Use Case handelt und wann nicht. Es ist eine klare Argumentation notwendig, warum gerade aufgrund von KI gewisse Handlungsbedarfe identifiziert wurden.

BEDARF 3:
Anpassung bestehender Normen, Standards und Richtlinien
Es gilt zu evaluieren, ob und wieweit es einer Anpassung existierender Normen, Standards oder technischer Regeln im Hinblick auf die Anwendung von KI bedarf. Als Beispiel dieser Aktivitäten kann die Klarstellung der IEC 61508 [79]–[86], die Überarbeitung der Maschinenrichtlinie 2006/42/EG [94] sowie der Schlussentwurf zum ISO/CD TR 22100-5 betrachtet werden.

BEDARF 4:
Normung eines Datenstandards für eine ökonomische Interoperabilität von deklarativen Modellen
Es wird vorgeschlagen, in Modellen die lebenszyklus- und interaktionsrelevanten Elemente und ihre Zusammenhänge (Patterns) explizit auszuzeichnen, um eine einheitliche automatisierte Verarbeitung zu unterstützen. Dies umfasst neben Schnittstellenbeschreibungen, ähnlich wie mittels OWL-S, auch die damit verbundenen Strukturen. Für diese Auszeichnungen und ihre Handhabung werden Standards benötigt.

BEDARF 5:
Normung einer formalen I4.0-Methodik, die die Prinzipien der Deklaration und Narration kombiniert unterstützt
Es soll ein(e) standardisierte(s) Vorgehensweise/Verfahren definiert werden, wie die unter Handlungsempfehlung 4 benannten Auszeichnungen in einzelnen Modellen gezielt vor einem anstehenden Anwendungskontext vorgenommen, ausgewertet und bewertet werden können. Damit sollen bereits a-priori-Untersuchungen späterer möglicher Interaktionen auf Plausibilität hin durchführbar werden.

BEDARF 6:
Normung eines Gestaltungsprozesses für Modelle, die mit semantischen Formaten wie Deklaration und Narration beschrieben werden sollen
Es wird vorgeschlagen, einen Entwicklungsprozess zu definieren, der die Gestaltung von Modellen zum Zweck einer späteren dynamischen Verschaltung unterstützt (d. h. es ist zum Zeitpunkt der Modellerstellung zwar bekannt, dass das Modell mit anderen Modellen in Interaktion treten soll, aber noch nicht, wie).

BEDARF 7:
Regulierung und Haftung
Der Einfluss von KI-spezifischen Anpassungen der Regulierung und des Haftungsrechts auf KI-Standardisierungsaktivitäten ist zu betrachten.

BEDARF 8:
Überprüfung der rechtlichen und regulatorischen Rahmenbedingungen für sicherheitskritische Aufgaben
Eine geeignete Anpassung rechtlicher und regulatorischer Rahmenbedingungen stärkt den Einsatz von KI-Technologien und ermöglicht u. a. auch mittelständischen Unternehmen mit kalkulierbarem wirtschaftlichem Risiko den Einsatz dieser Technologie.

Forschung

BEDARF 9:
Risikobeurteilung durch KI/neue Methoden für die Risikobeurteilung
Es gilt zu untersuchen und zu bewerten, in wieweit aktuelle Methoden der Risikobeurteilung ausreichend sind bzw. welche Anforderungen durch aktuelle Normen, Standards und technische Regeln noch nicht ausreichend adressiert werden. Als Beispiel dieser Aktivität können die Arbeiten an ISO/CD TR 22100-5 betrachtet werden, in der die Risikobewer-

tungsmethodik entsprechend ISO 12100 [137] (und damit der Maschinenrichtlinie 2006/42/EG [94]) hinsichtlich möglicher KI-Einflüsse auf Safety betrachtet wird.

BEDARF 10:
Erfassung von Begriffen aus unterschiedlichen Disziplinen (Glossar)
Es wird vorgeschlagen, eine gemeinsame Sprache (d. h. Semantik) in Form eines Glossars zu erarbeiten mit Regeln, Gesetzen und Axiomen, welche sowohl disziplinspezifisch (z. B. Recht, Technik, Wirtschaft) als auch Teile davon branchenübergreifend eindeutig definiert.

Anwendung

BEDARF 11:
Erweiterung der Sammlung von Use Cases hinsichtlich neuer Geschäftsmöglichkeiten durch KI
Bei zukünftig zusammengestellten Beispielen sollten neben einer Gestaltung und Optimierung der internen Produktions-Wertschöpfungsprozesse auch Beispiele Berücksichtigung finden, in denen auf Basis von KI neue Geschäftsmöglichkeiten erschlossen werden.

BEDARF 12:
Identifikation von Business Scenarios für die Rolle von KI-Firmen in der industriellen Automation
Es wird vorgeschlagen, gezielt Business Scenarios zu identifizieren und entsprechend aufzubereiten, in denen eine KI-Firma ein geschäftliches Risiko im Hinblick auf das Wertversprechen von KI übernimmt.

BEDARF 13:
Standardisierte Aufbereitung von Use Cases
Es wird vorgeschlagen, Use Cases gemäß des IEC/TC 65/WG 23-Templates aufzubereiten und durch Aufbereitung der Use Cases gemäß den Usage Viewpoints entsprechend IIRA-Template eine notwendige Detailtiefe im Umfang von ca. 20 Seiten je Use Case zu erreichen.

BEDARF 14:
Betrachtung spezifischer Use Cases und die Rolle von KI für das Produktdesign und die Produktkonfiguration
Es wird vorgeschlagen, weitere Use Cases, beispielsweise für die Wertschöpfungsprozesse product design und product configuration/sales zu entwickeln, welche derzeit im Rahmen des IEC/TC 65/WG 23 nicht adressiert werden.

BEDARF 15:
Prüfung der Überdeckung von gesammelten Use Cases mit dem Betrachtungsscope des IEC/TC 65/WG 23
Es wird vorgeschlagen, vor der Ausarbeitung weiterer Use Cases (beispielsweise auskunftsfähige Maschine, adaptive Logistik) zunächst zu prüfen, inwieweit diese bereits durch bestehende Use-Case-Beschreibungen (beispielsweise IEC/TC 65/WG 23) adressiert werden.

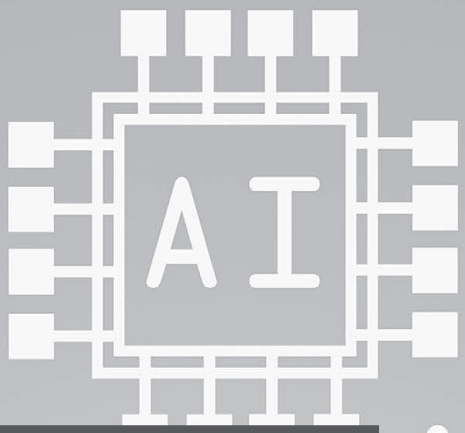
BEDARF 16:
Fortschreibung der Sammlung von Use Cases in einem nationalen Koordinierungsgremium
Es wird empfohlen, die im Rahmen der Arbeiten des SCI4.0 Expertenrats KI in industriellen Anwendungen erstellte Use-Case-Sammlung weiter fortzuschreiben.

BEDARF 17:
Detaillierung bestehender Use Cases
Weitere Detaillierung der Use Cases gemäß Usage View in Form von Functional Views.

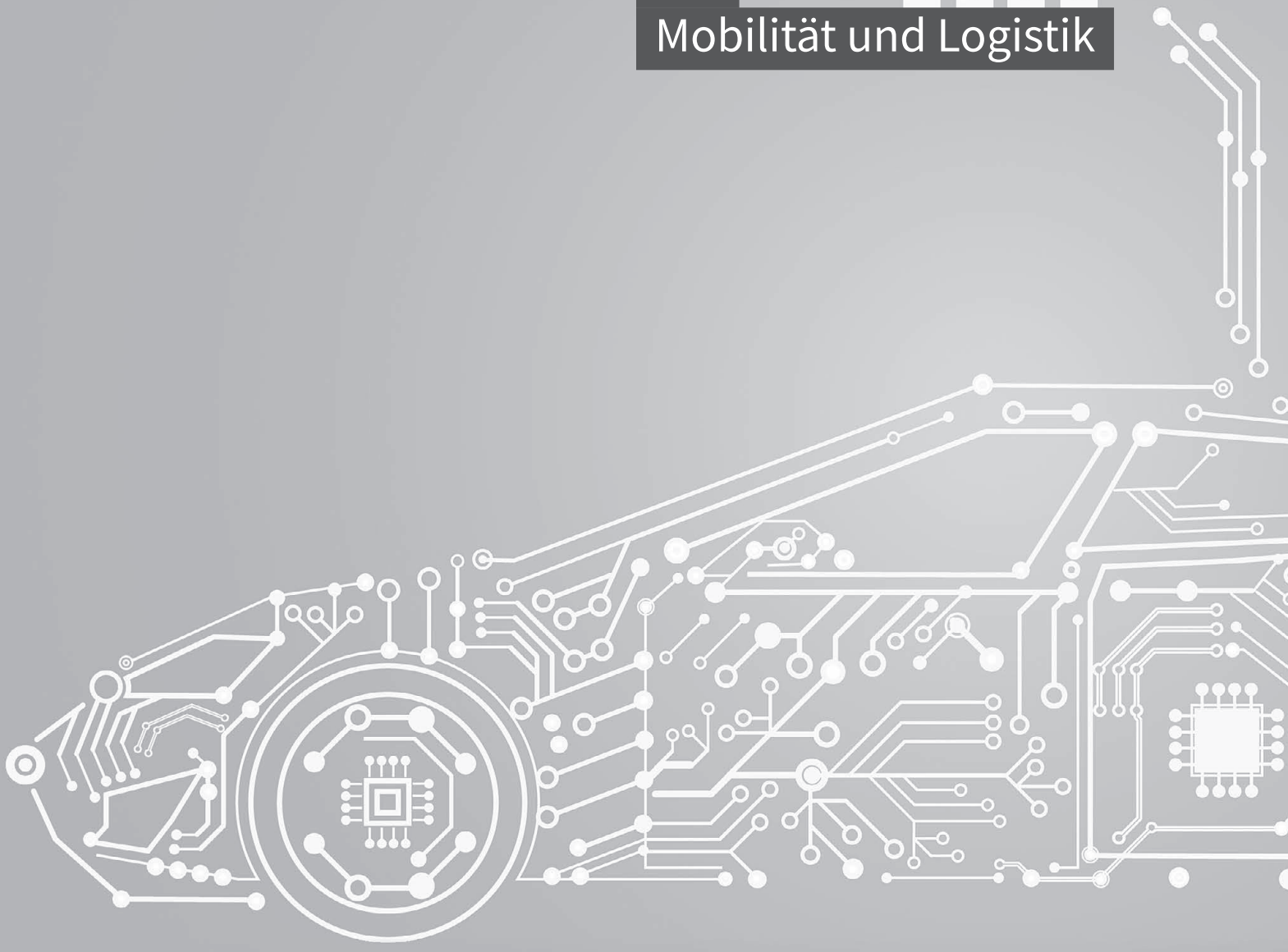
BEDARF 18:
Formale Methoden
Prüfung der Eignung von mathematischen Verfahren für den Nachweis notwendiger Eigenschaften (z. B. Robustheit, Erklärbarkeit etc.) von Blackbox-Machine-Learning-Modellen.

BEDARF 19:
Anwendung der Plastizität und Elastizität von Modellen hinsichtlich einer Konflikterkennung/Kritikalität
Es wird vorgeschlagen, im Entwicklungsprozess von Modellen eine explizite Auszeichnung anzuwenden, welche Elemente in welchem Maß veränderlich sein können, um mit anderen Modellen interagieren zu können. Diese Auszeichnungen sollen verwendet werden, um zur Laufzeit eines Systems Vorhersagen über die zu erwartende Passgenauigkeit zweier Modelle treffen zu können.

BEDARF 20:
Evaluation der Prinzipien der Common Logic und deren Rolle in der Normung von KI in I4.0
Beschreibung und Skizzierung der Anwendung der Prinzipien einer Common Logic/Semantics und deren Rolle in der Normung. Stärkung der Vernetzung mit allen Beteiligten, Vergleichbarkeit mit alternativen Ansätzen herstellen und Community-Aufbau, um diese Aktivitäten der breiten Masse und allg. Spektrum von I4.0 zuzuführen.



4.6 Mobilität und Logistik



Dieses Kapitel erläutert das massive Innovationspotenzial und einige damit einhergehende rasante Veränderungen, die der Einsatz von KI für die Domäne „Mobilität und Logistik“ mit sich bringt. Das Kapitel ist strukturiert entlang dreier wesentlicher Gesichtspunkte:

Rechtsrahmen: Relevante Aspekte des bestehenden Rechtsrahmens für Mobilität und Logistik werden kurz skizziert. Dies dient als Grundlage, um die Potenziale von Normung und Standardisierung aufzuzeigen. Ein entsprechender Schulterschluss aller Prozessbeteiligten (Industrie, Prüforganisationen, Gesetzgeber und Genehmigungsbehörde) wird empfohlen, um KI-spezifische Anwendungsfälle im Lichte des vorherrschenden gesetzlichen Rahmens zu durchleuchten.

Erklärbarkeit und Validierung: Wie in anderen Domänen spielt auch in der Mobilität und Logistik das Thema Erklärbarkeit und Validierung von KI-Systemen eine große Rolle. Eine Vielzahl von Normungs- und Forschungsgremien beschäftigt sich derzeit mit der Frage, in welcher Tiefe die verwendeten KI-Algorithmen dokumentiert und nachvollziehbar dargelegt werden müssen, sodass eine klare funktionale Beziehung zwischen Input und Output erkennbar ist. Daran geknüpft ist die Frage des Nachweises der korrekten Funktionalität (Validierung). Aufgrund der Komplexität der Fragestellungen, des gestiegenen Bedarfs der Anwendungen und des Anspruchs der Normung, den aktuellen Stand der Wissenschaft und Technik widerzuspiegeln, besteht die Aufgabe in der Definition weiterer Forschungsfelder, um Normungsthemen vorzubereiten.

Interoperabilität: IT-Systeme interoperabel zu gestalten ist gängige Praxis, nicht nur in der Domäne „Mobilität und Logistik“. Dennoch verdient das Thema Interoperabilität besondere Aufmerksamkeit bei der Diskussion von KI-Systemen für Mobilität und Logistik. In dieser Domäne gibt es heterogene und multimodale Dienste und Anwendungen basierend auf Systemen unterschiedlicher Betreiber und Anbieter, welche in ihrer Funktionalität durch KI massiv wachsen werden. Deshalb werden normierte Datenmodelle und Schnittstellen hier in besonderem Maße zu Innovation und Effizienz beitragen.

4.6.1 Status quo

Straßenverkehrsrecht: In der Bundesrepublik wurde das Straßenverkehrsgesetz (StVG) [228] im Jahr 2017 für das automatisierte Fahren (welches noch die Anwesenheit eines Fahrers für die Übernahme voraussetzt), also für die Hoch- und Vollautomatisierung, erweitert. Neben den an den Fahrer gerichteten Pflichten und der Definition der Hoch- und Vollautomatisierung enthält das neue StVG Verweise auf das Zertifizierungsrecht, denn nur Fahrzeuge mit zertifiziertem hoch- bzw. vollautomatisiertem Fahrsystem fallen unter das novellierte Straßenverkehrsgesetz. Insbesondere wird im Gesetz gefordert, dass das hoch- bzw. vollautomatisierte Fahrsystem die Verkehrsvorschriften einhält. Diese Anforderung findet sich seit Neuestem auch in der technischen Regulierung der Automated Lane Keeping Systems (ALKS) und ist im Umfang der Typprüfung enthalten. Würden KI-Mechanismen nun das Fahrverhalten der hoch- bzw. vollautomatisierten Fahrfunktionen im Sinne des Straßenverkehrsgesetzes verändern, so würde ggf. die bereits erteilte Typgenehmigung das neue Fahrverhalten nicht länger abdecken. Wollte man diesen Mechanismus zukünftig ändern, hätte dies weitreichende (rechtliche) Konsequenzen.

Auf internationaler Normungsebene arbeitet das Technische Komitee ISO/TC 22 „Road vehicles“ bereits an ISO/TR 4804, **Road vehicles – Safety and security for automated driving systems – Design, verification and validation methods**. KI wird jedoch auch in weiteren Standards des ISO/TC 22 adressiert. Der Umfang ist in Diskussion.

Zusätzlich zu den Aktivitäten in der Normung und Standardisierung befindet sich der Konsortialstandard UL4600 „Standard for Safety for the Evaluation of Autonomous Products“ [157] derzeit in der Erarbeitung. Verantwortlich hierfür ist die gemeinnützige Organisation Unterwriters Laboratories (UL).

Ferner haben Arbeiten im Bereich Automated Driving in einer Arbeitsgruppe (IEEE P2846) begonnen.

Vorrangig für das Thema Logistik fundamentiert sich eine Kooperation zwischen den Normungsfeldern. Unter Beteiligung von DIN und DKE ist auf der europäischen Normungsebene eine Joint Working Group (Kooperation zwischen CEN und CENELEC) beantragt. Diese Initiative erhält eine bedeutende Relevanz, da sich das Verhältnis der Transportwege (aktuell ca. 70 Prozent aller Güter weltweit noch über die Straße) ändert und sich Kooperationen dynamischer und vielschichtiger entwickeln.

Daran anknüpfend ist China bereits bei der Europäischen Normungsorganisation CEN als Kooperationspartner involviert. So beabsichtigen beide Partner das Erarbeiten gemeinsamer Standards für den europäischen Warentransfer.

Speziell für den Lebensmittelbereich finden internationale Arbeiten für die künftige Norm ISO 23412, Indirekte, temperaturgeführte Kühllieferdienste – Landtransport von Paketen mit Zwischenübergabe – unter Federführung von Japan statt, in der der Transport von Tiefkühlware (hauptsächlich Fisch und Rindfleisch) betrachtet wird. Der beabsichtigte Standard lässt sich als weiterer Grundstein für die automatische Warendistribution weltweit verstehen.

4.6.2 Anforderungen, Herausforderungen

Die Herausforderungen für KI-Systeme in Mobilität und Logistik sind vielschichtig und tiefgreifend. Hierunter fallen z. B. das Automatisierte Fahren, Steuerung internationaler Warenströme, Optimierung von Lagerlogistik, Fahrplandisponierung von Schienenfahrzeugen.

4.6.2.1 Rechtsrahmen

Die Homologation/technische Regulierung im Bereich der Kraftfahrzeuge stellt sich wie folgt dar: Kraftfahrzeuge, die ohne Einschränkungen am öffentlichen Straßenverkehr teilnehmen dürfen, müssen in Europa die Anforderungen einer Typgenehmigung erfüllen. Diese ist geregelt in der Rahmenrichtlinie 2007/46/EG (ab dem 1.9.2020 in der EU VO 2018/858 [229]) und im Übereinkommen der UNECE vom 20. März 1958 [230] über die Annahme einheitlicher Bedingungen für die Genehmigung der Ausrüstungsgegenstände und Teile von Kraftfahrzeugen und über die gegenseitige Anerkennung der Genehmigungen. Dieses Abkommen, das über 150 technische Regelungen umfasst, war ein Meilenstein auf dem Weg zu einheitlichen technischen Zulassungsvorschriften. Diese betreffen neben Systemen und Bauteilen für die aktive und passive Sicherheit auch umweltrelevante Vorschriften zum Schutz aller Verkehrsteilnehmer. Im Rahmen der Typgenehmigung wird die Erfüllung aller erforderlichen Vorschriften gemäß der EU-Typgenehmigungsverordnung nachgewiesen und von einer unabhängigen und von einer nationalen Typgenehmigungsbehörde benannten Prüfinstitution (TÜV, DEKRA etc.) geprüft und gutachterlich bestätigt. Parameter, die nicht typprüfrelevant sind bzw. diese nicht beeinträchtigen, sind davon ausgenommen. Eine Genehmigungsbehörde stellt auf Basis dieser Bestätigung die Typgenehmigung aus, ohne die die Fahrzeuge nicht in den Verkehr gebracht werden dürfen. Der Hersteller bestätigt dies mit der sogenannten Übereinstimmungsbescheinigung (CoC – Certificate of Conformity). Typgeprüfte Fahrzeuge mit gültiger Übereinstimmungsbescheinigung erhalten in Deutschland bzw. in weiteren EU-Mitgliedstaaten eine Zulassung für die Teilnahme am Straßenverkehr und damit ein Kennzeichen. Die periodische technische Überwachung nach der EU-Richtlinie 2014/45/EU [231] stellt sicher, dass die Fahrzeuge auch nach vielen Jahren des Gebrauchs allen geltenden Anforderungen an Sicherheit und Umwelt entsprechen. Sie ist somit ein wichtiges Mittel, um Vertrauen beim Verbraucher zu schaffen.

Im Zuge der technischen Weiterentwicklung und der Digitalisierung können Produkte oder deren Funktion jedoch schon heute während ihrer Nutzungsdauer verändert werden. Die aktuell gültigen Prozesse bei Zulassung, Typgenehmigung und periodischer Überwachung müssen also künftig sicherstellen, dass auch diese Veränderungen die Sicherheit im Straßenverkehr nicht beeinträchtigen und der Schutz der Umwelt weiterhin garantiert ist. Es müssen neue und effiziente Mechanismen entwickelt werden, mit denen Änderungen am Produkt während der Nutzungsphase erkannt, geprüft und freigegeben werden können. Die Prozesse der Zertifizierung werden sich in Zukunft sehr viel produktspezifischer gestalten. Eine Herausforderung ist dabei die Anpassung der Typgenehmigung und der periodischen technischen Überwachung an automatisierte und vernetzte Fahrzeuge.

Zur Zulassung von teilautomatisierten Fahrzeugen wurde die Regulierung für Lenkanlagen (UN-R79 [232]) seit dem Jahr 2016 umfassend überarbeitet und erweitert. In einem ersten Schritt wurden Anforderungen für Systeme zum assistierten Spurhalten und Spurwechseln, für korrigierende Lenkeingriffe bei drohendem Verlassen der Fahrspur oder zum Ausweichen vor Objekten sowie für das automatisierte oder ferngesteuerte Parken von Fahrzeugen ergänzt. Die beschriebenen Systeme gehören maximal der Teilautomatisierung an (Automatisierungsstufe zwei).[29] In der Zwischenzeit wurde entschieden, die Anforderungen für hoch- oder vollautomatisierte Spurführungssysteme (Automatisierungsstufen drei und vier) nicht als Ergänzung der UN-R79, sondern als eigene Regulierung für ALKS zur Entscheidung zu bringen.

Die ALKS-Vorschrift formuliert wichtige Randbedingungen:

- → Beschränkung auf Autobahnen
- → Beschränkung auf Stau und niedrige Geschwindigkeiten von 0 bis 60 km/h
- → Beschränkung auf spurführende Systeme, kein automatischer Spurwechsel möglich

29 https://www.sae.org/standards/content/j3016_201806/

→ Beschränkung auf Fahrzeuge der Zulassungskategorie M1 (Pkw), Anwendung im NFZ-Bereich ist angedacht.

Um diese weitreichenden Beschränkungen aufzulösen, wird seit Oktober 2019 in der UNECE-Arbeitsgruppe „Functional Requirements for Automated Vehicles“ (FRAV) an generischen Anforderungen für die Zulassung automatisierter Fahrfunktionen gearbeitet. Parallel dazu befasst sich die Arbeitsgruppe „Validation Methods for Automated Driving“ (VMAD) seit März 2018 mit einheitlichen Methoden, um die Erfüllung dieser Anforderungen nachzuweisen. Dieser Nachweis soll dabei – laut aktuellem Diskussionsstand – aus einem Audit und einem Simulationsteil sowie zusätzlich aus physischen Tests auf Prüfgeländen und einer Beurteilungsfahrt auf öffentlichen Straßen bestehen.

Für die Beschreibung möglicher Validierungsszenarien wurde im deutschen Förderprojekt PEGASUS mit openSCENARIO und openDRIVE ein einheitliches Format entwickelt, welches inzwischen über ASAM e. V. als frei zugänglicher Standard verfügbar ist und das in die Arbeiten der UN eingeflossen ist. Dieser Textabschnitt wurde im Kern von der Arbeitsgruppe „AG 6 – Standardisierung, Normung, Zertifizierung und Typgenehmigung“ der Nationalen Plattform Zukunft der Mobilität, NPM AG 6 (siehe Kapitel 3.2), konsolidiert und ist zwischenzeitlich bereits veröffentlicht.

Hinsichtlich der europarechtlich regulierten Verfahren Typgenehmigung und periodisch technische Überwachung (Hauptuntersuchung) von KI-Systemen in Kraftfahrzeugen ist der Teilaspekt des maschinellen Lernens (ML) insbesondere relevant. Es bietet sich an, KI-Systeme danach zu unterscheiden, ob dieses Lernen in der Entwicklung oder im Betrieb der Systeme erfolgt. In Anlehnung an das laufende Projekt ISO/IEC CD 22989 sollte im Fall, dass maschinelles Lernen ausschließlich in der Entwicklung stattfindet und sich das System nach Inverkehrbringen nicht mehr ändert, ausgehend vom Begriff „trained model“ von „angelernten Systemen“ oder „Offline Learning“ gesprochen werden. Analog sollte von „kontinuierlich lernenden Systemen“ oder „Online Learning“ gesprochen werden, wenn maschinelles Lernen im Betrieb erfolgt, wenn sich also das System nach dem Inverkehrbringen verändert.

Besonderes Augenmerk liegt hierbei auf den Trainingsdaten von KI-Systemen. Es bedarf Regelungen bzw. Normen über das Zur-Verfügung-Stellen und Verwenden von Datensammlungen (freiwilliger Datenaustausch oder Datapools), die zu Trainingszwecken von KI-Systemen eingesetzt werden können.

Produkthaftungs- und produktsicherheitsrechtliche Verantwortung sowie die Verantwortung für die Einhaltung von Typgenehmigungsvorschriften und Marktzugangsregeln (zusammenfassend: „Product Compliance“) liegen beim Hersteller des Produkts, in dem das KI-System zur Anwendung kommt bzw. „eingebaut“ ist. Bei einem Produkt, das Normen oder anderen technischen Spezifikationen entspricht, wird gemäß § 9(2) ProdSG [195] vermutet, dass es den Anforderungen an die Produktsicherheit genügt, soweit diese von den betreffenden Normen oder anderen technischen Spezifikationen abgedeckt sind. Dieser Absatz gilt analog für Straßen- und Schienenfahrzeuge.

4.6.2.2 Erklärbarkeit und Validierung

Die Möglichkeit zur Validierung des Verhaltens von KI-Systemen ist Voraussetzung für deren Integration in sicherheitsrelevante Produkte. Bei der Validierung ist entsprechend dem Stand der Technik vorzugehen. Der Stand der Technik zur Validierung von KI-Systemen entwickelt sich noch dynamisch, sodass keine entsprechenden Normen existieren. Damit ist der Stand der Technik nur implizit definiert. Dies erschwert zu entscheiden, wann die Validierung hinreichend ist.

Ferner ermöglicht die Weiterentwicklung des Mobilfunkstandards die Übertragung von großen Datenmengen vom Fahrzeug. Diese Entwicklung ermöglicht ein neues und zeitgerechtes Werkzeug zur kontinuierlichen Validierung von mobilen automatisierten Systemen im Rahmen einer Überwachung im Feld. Es gilt auszuloten, wie diese neuen Möglichkeiten zielführend eingesetzt werden können. Insbesondere könnte die kontinuierliche Validierung nach Markteinführung helfen, Abweichungen vom Soll-Verhalten zu erkennen und Optimierungspotenziale zu identifizieren.

Erklärbarkeit unterstützt die Validierung, Transparenz und Nachvollziehbarkeit von Entscheidungen, welche vom KI-System getroffen werden, sowie des Entscheidungsprozesses, den das KI-System durchläuft. Um Entscheidungen von KI-Systemen nachvollziehen zu können, müssen sowohl Aspekte der Entwicklung (z. B. verwendete Daten und Trainingsmethoden) wie auch der Ausführung (z. B. ausschlaggebende Merkmale) des KI-Systems betrachtet werden. Maßstäbe und Metriken für die Nachvollziehbarkeit und Transparenz von KI-Systemen sollten entwickelt werden.

Bislang fehlen klare Definitionen für notwendige Schnittstellen eines KI-Systems zum Menschen bzw. der Gesellschaft und zu (einem) anderen KI-System(en). Die zunehmende Komplexität und Breite von automatisierten Funktionen erschwert für einen menschlichen Beobachter die frühzeitige Unterscheidung zwischen bestimmungsgemäßer Funktion einerseits und Abweichungen, die ein Eingreifen erfordern, andererseits. Da die Entscheidung eines automatisierten Systems etwa auf heterogenen Sensordaten, internen Modellrechnungen und vernetztem Datenaustausch fußen kann, ist diese für den Menschen nicht per se unmittelbar nachvollziehbar. Darüber hinaus beeinflusst der Bezug eines Menschen zum System seine Urteilsfähigkeit und seine betrachteten Eingriffsmöglichkeiten. Dieser reicht vom Betreiber des Systems (etwa dem Besitzer eines automatisierten Pkw) über den beiläufigen Nutzer (etwa den Passagier einer automatisierten U-Bahn) bis hin zum unfreiwillig Betroffenen (etwa einem Fußgänger gegenüber einem automatisierten Fahrzeug). Somit interagiert ein System potenziell mit einer heterogenen Menge an betroffenen Menschen mit unterschiedlichen Erwartungen an das System. Normung von Mensch-Maschine-Schnittstellen kann Menschen helfen, unabhängig von Erfahrungen mit dem konkreten System dessen inneren Zustand zu verstehen, und abzuschätzen, ob eigener Handlungsbedarf besteht und welche Handlungen oder Eingriffe möglich, notwendig und/oder angemessen sind.

Hierbei schließt sich auch der Kreis zur Transparenz und zur Nachvollziehbarkeit. Wenn ein KI-System die Entscheidung dem Menschen überlässt, muss dieser wissen, wie, wann und zu welchen Entscheidungen die Übergabe erfolgt.

4.6.2.3 Interoperabilität

Eine Vielzahl an Geschäftsprozessen in der Mobilität und der Logistik (beispielsweise intermodale Transporte, Third-Party-Logistics-Dienstleistungen (3PL-Dienstleistungen), öffentlicher Personenverkehr oder Verkehrsflusssteuerung) sind darauf angewiesen, dass die jeweiligen Akteure organisationsübergreifend eng kooperieren. Wo die dafür notwendigen Systeme und Prozesse nicht direkt miteinander verknüpft sind (oder sein sollen), wird Interoperabilität zum Schlüsselfaktor erfolgreicher Geschäftsbeziehungen. Das Ziel der Interoperabilität besteht darin, die Zusammenarbeit zwischen Akteuren möglichst effizient und effektiv zu gestalten, um etwaige Reibungsverluste bei Interaktionen (beispielsweise zeitliche Verzögerungen, Nachfragen, Missverständnisse, Formatumwandlungen) zu minimieren. Je interoperabler heterogene Systeme sind, desto geringer sind Interaktionsaufwand sowie Fehlerquote und desto größer sind die Flexibilität sowie Belastbarkeit des Gesamtsystems – Eigenschaften, die in Zeiten fortschreitender Digitalisierung, der entstehenden Industrie 4.0, anstehender Strukturveränderungen in der Logistik, Big Data und Künstlicher Intelligenz immer wichtiger werden. Dies gilt umso mehr, da Kollaboration – wofür Interoperabilität ein maßgeblicher Erfolgsfaktor ist – im aktuell hochdynamischen Umfeld zunehmend erfolgskritisch wird. Zukünftige Normungsaktivitäten, insbesondere im Kontext der Künstlichen Intelligenz, sollen daher Interoperabilität als Prinzip berücksichtigen und entsprechende Standards eindeutig definieren, um Interpretationsspielräume zu minimieren und so die Kompatibilität von Geschäftsprozessen zu steigern.

Schnittstellen haben eine besondere Bedeutung für die Interoperabilität von KI-Systemen. Dabei geht es um die intermodalen Transportketten für Personen und Güter sowie die Planung solcher Transportketten, also z. B. Güter aus Lagern auf öffentliche Straßen, Empfehlungen und Umsetzungen für individuelle Fortbewegung von Personen. Diese Beispiele werden im Anwendungsszenario „Intelligent vernetzt unterwegs“ der Plattform Lernende Systeme genauer dargestellt [233]. Sobald KI-Systeme miteinander kooperieren, bedarf es klarer Regeln, wie diese Kooperation aussieht, da jedes KI-System auch selbstständig ist. So ist z. B. davon auszugehen, dass über die Schnittstellen zwischen den KI-Systemen hinausgedacht und z. B. die Zielfunktionen beteiligter Systeme genannt werden müssen, um ungewollte Wechselwirkungen auszuschließen. Eine weitere große Herausforderung wird bei den Daten gesehen. Dabei ist u. a. noch offen, wie deren Qualität sein, das Sammeln erfolgen, mit welchen Daten ein KI-System anfangs lernen (Trainingsdaten), wie die technisch und rechtlich einfache Datenverteilung über Kooperationsketten für die Nutzbarkeit durch alle aussehen soll.

Der Datenaustausch zwischen einzelnen Systemen bildet die Grundlage automatisierter Prozesse, sodass Interoperabilität ein wichtiger Erfolgsfaktor für Industrie 4.0 und Digitalisierung ist. Im Kontext der Künstlichen Intelligenz bezieht sich die Interoperabilität primär auf Software, Softwareschnittstellen und Daten sowie deren Kompatibilität zueinander. Aufgrund der Entwicklungsgeschwindigkeit in diesen Bereichen sowie der Vielzahl unterschiedlicher Entwicklungsstrategien ergeben sich aus (neuen) Softwarelösungen und Schnittstellen oft auch neue Herausforderungen für die praktische Umsetzung bzw. Sicherstellung von Interoperabilität. Dies bedeutet für die Standardisierung, dass Normen immer bestehende

Best-Practice-Ansätze und/oder weitverbreitete Lösungen aufgreifen müssen. Zudem sollten systemunabhängige Gestaltungsprinzipien und „Spielregeln“ für Interoperabilität gefunden werden, die – so weit wie möglich – zeitlos sind.

Die besondere Herausforderung beim Einsatz von Künstlicher Intelligenz besteht vor allem darin, **jene Daten kenntlich zu machen, die durch KI-unterstützte Anwendungen entstanden sind**. Kooperationspartner müssen bei übermittelten Daten unterscheiden können, ob diese sich direkt aus Realdaten beispielsweise eines ERP-Systems oder aus einer Berechnung eines KI-Systems ergeben. Ebenso muss eine Beurteilung möglich sein, in welchem Kontext die Daten entstanden und wie zuverlässig diese sind (analog zum Prinzip „Quality of Service“). Dies gilt insbesondere dann, wenn aktuell und zukünftig eingesetzte Systeme über einen hohen Grad an Autonomie verfügen, also weitgehend selbstständig agieren: Je geringer der Mensch in Steuerung und Überwachung dieser Systeme eingebunden ist, desto höher sind die Anforderungen an die Interoperabilität und insbesondere an die Quality of Service der KI-unterstützt generierten Daten.

Bei allen Standardisierungsvorhaben ist zu beachten, dass Interoperabilität grundsätzlich eine **ganzheitliche Herausforderung** ist. Sie sollte von Anfang an – vom Design über Testverfahren bis zur Implementierung und den täglichen Betrieb – mitbedacht und berücksichtigt werden. Je mehr dies gelingt, desto weniger Aufwand entsteht für die nachträgliche Verbesserung der Interoperabilität verschiedener Systeme, was die Realisierung von ökonomischen, ökologischen, sozialen und sicherheitsbezogenen Potenzialen erleichtert.

Bei allen Tätigkeiten zur Sicherstellung der Interoperabilität sowie zu ihrer Realisierung in der Praxis (z. B. Datenerhebung, -speicherung und -austausch) sind die jeweils geltenden Rahmenbedingungen des **Datenschutzes** einzuhalten.

4.6.3 Normungs- und Standardisierungsbedarfe

4.6.3.1 Rechtsrahmen

BEDARF 1:
Schulterschluss der Prozessbeteiligten implementieren
Es ist zielführend, Industrie, Prüforganisationen, Gesetzgeber und Genehmigungsbehörde gleichermaßen einzubinden, um KI-Anwendungen aus dem Bereich „Mobilität und Logistik“ unter den Gesichtspunkten der vorherrschenden gesetzlichen Anforderungen (fahrzeugtechnische Vorschriften, rechtliche Anforderungen – z. B. aus der StVO [234], Waren- und Gütertransport – sowie tangierende Regelwerke wie z. B. die DSGVO [95] im Bereich Datenaustausch, kontinuierliche Hauptuntersuchung) zu erörtern. Das Ziel sollte eine Interpretation des Rechtsrahmens zukünftiger KI-Anwendungen sein sowie – falls zutreffend – eine Strategie für die Anpassung des Rechtsrahmens, um neuartige KI-Anwendungen in der Gesellschaft zu ermöglichen.

BEDARF 2:
Klärungsbedarf für „safe and compliant“ zum Bereitstellen auf dem Markt
Alle KI-Systeme, die im Bereich „Mobilität und Logistik“ Anwendung finden, müssen vor dem Bereitstellen auf dem Markt sowie während der Nutzungsdauer (Nutzungsdauer in diesem Zusammenhang meint den Lebenszyklus des Fahrzeugs bzw. den Lebenszyklus der Funktion/Dienstleistung) „safe and compliant“ sein. Hier existiert die Unterscheidung zwischen Produkten und Dienstleistungen. Produkte müssen den einschlägigen Gesetzen und Vorschriften folgen. Ob diese Gesetze und Regeln für eine reine KI-System-Dienstleistung („Software-Service“) ausschlaggebend sind, bedarf weiterer Klärung.

BEDARF 3:
Widerspruch zwischen „statischem“ Anknüpfungspunkt und „dynamischem“ Lernen auflösen
Der Einsatz von zur Ausführungszeit selbstlernenden KI-Systemen würde unter geltendem Recht erhebliche Herausforderungen erzeugen, weil sich die (Funktions-)Eigenschaften durch das Selbstlernen schwer rückverfolgbar und schwer vorhersehbar verändern könnten. Die produkthaftungsrechtliche Behandlung derartiger Konstellationen wird derzeit noch diskutiert. Auch zertifizierungs- bzw. regulierungsrechtlich wären derartige Produkte nur schwer zu erfassen, da das Recht bislang stets einen festen Anknüpfungspunkt für Produkteigenschaften braucht/erfordert.

Wichtig dabei ist, vor Aufspielen von Software-Updates auf bereits im Verkehr befindlichen Systemen deren aktuellen Zustand hinsichtlich möglicher Veränderungen zu analysieren, um unbeabsichtigte sicherheitsrelevante Wechselwirkungen zu minimieren. Zudem können beispielsweise Veränderungen von Kraftfahrzeugen auch genehmigungsrelevant werden, die eine vorherige Prüfung durch die Prozessbeteiligten erfordern könnten. Es ist notwendig, dass Prozessbeteiligte die systembezogene Gesetzes-/Vorschriftenlage für die Fälle erörtern,

in denen der Einsatz von KI-Systemen Veränderungen zur Laufzeit generiert.

4.6.3.2 Erklärbarkeit und Validierung

BEDARF 4:
Forschung fördern
Bei der Prüfung von KI-Systemen im Bereich „Mobilität und Logistik" bedarf es klarer Definitionen der Prüfkriterien, des Prüfprozesses, der Prüfidentität und der genauen Prüfinhalte, aus denen sich die Prüfung aufbaut.

Zur Vorbereitung entsprechender Standardisierung und Normung empfehlen sich diesbezüglich folgende Forschungsaufgaben und deren Förderung:

→ Erforschung der Gefahr, dass zu prüfende Systeme speziell auf Prüfungen optimiert werden, z. B. dass KI-Systeme auf singuläre Situationen trainiert werden und sich auf Prüfungsinhalte überanpassen („auswendig lernen", „over-fitting")
→ Entwicklung von Prüfungen inklusive dynamischer Prüfverfahren, die der o.g. Gefahr der Optimierung entgegenwirken
→ Charakterisierung von KI-Systemen, die sich durch Lernen im Einsatz selbst verändern und/oder in sich verändernden Umgebungen eingesetzt werden; entsprechende Auswirkungen auf kontinuierliche Prüfverfahren

BEDARF 5:
Forschung begleiten
Forschungsprojekte für eindeutige, allgemeingültige und objektive Bewertungskriterien und -methoden unter Berücksichtigung einer kontinuierlichen Validierung der Sicherheit und Leistungsfähigkeit von automatisiertem und vernetztem Fahren mit zunehmendem KI-Einsatz müssen aktiv unterstützt und begleitet werden. Hierbei sollten geeignete Algorithmen für die Bewertung der Fahraufgabe entwickelt und deren Schnittstellen definiert werden. Dabei können insbesondere KI-Methoden zum Einsatz kommen. Die Ergebnisse – z. B. aus dem Projekt PEGASUS, gesetzgebenden Arbeitsgruppen (z. B. IWG FRAV und VMAD der UNECE) – sollen in Normung und Standardisierung eingebracht werden und sich am menschlichen Fahrverhalten orientieren.

BEDARF 6:
Transparente Ausgestaltung von KI-Systemen
Um KI-Systeme transparent zu gestalten, braucht es Vorgaben für die Ausführungszeit. Relevant sind dabei Punkte, die die Sinnhaftigkeit von Interaktionen mit anderen Systemen sowie die Kompetenz des KI-Systems für die aktuelle Situation bewerten.

Zur Vorbereitung von Normung und Standardisierung in diesem Bereich werden die folgenden Forschungsaufgaben empfohlen:

→ Methoden zur Identifizierung und Beschreibung des eigenen Kompetenzbereichs des KI-Systems (Bsp. adversarial examples, Kontext und Grenzen), insbesondere bei sicherheitsgerichteten Funktionen bzw. dem Übergang in einen sicheren Zustand
→ Umfassende Analyse der Mensch-KI-Interaktion (z. B. KI schlägt verschiedene Optionen vor, Mensch wählt eine aus oder Spannungsfeld „Safety vs. Security") in einer bestimmten Aktion, also die Nachvollziehbarkeit der Handlung des KI-Systems.
→ Erforschung, wie neuronale Netze für sicherheitsgerichtete Funktionen nutzbar sein können. Dies betrifft deren Entwicklungs- und Freigabeprozess sowie Nachweismethoden für Eigenschaften und Erklärbarkeit. Ferner stellt sich die Frage, welche Architekturmuster zur Integration neuronaler Netze in sicherheitsgerichtete Funktionen zielführend sind.

Für den Automobilbereich adressiert das Projekt „KI-Absicherung" aus der VDA-Leitinitiative Autonomes und Vernetztes Fahren Fragen zu Absicherung und Freigabe von KI-Systemen für einen konkreten Anwendungsfall. Für andere Anwendungsbereiche werden ähnliche Initiativen empfohlen.

4.6.3.3 Interoperabilität

BEDARF 7:
Datenreferenzmodell für Interoperabilität schaffen
Für die Interoperabilität sind Daten und deren korrekte Verwendung ein entscheidendes Erfolgskriterium. In Logistik und Mobilität existieren bereits vielfältige Best Practices zu Daten, Datenarten, Datenmodellen und Datenbanken (beispielsweise Stamm-, Änderungs- und Bewegungsdaten, deren Beziehungen zueinander sowie Möglichkeiten der Einbindung in Softwarelösungen). Diese Best Practices verändern sich aktuell durch neue Technologien, Anforderungen, Möglichkeiten und Lösungen. Standardisierungsgremien und Anwender sollten die tägliche Praxis beobachten und, bei erkennbarer Verfestigung einer neuen Best Practice beispielsweise der Datenarten, diese unter Berücksichtigung einer gewissen kurz- und mittelfristigen Flexibilität einheitlich

definieren, um ein Datenreferenzmodell für Interoperabilität in Mobilität und Logistik vorzuschlagen: In einem solchen Modell sollten grundlegende, für Interoperabilität relevante Datenarten, ihre Strukturen und Beziehungen zueinander dargelegt werden, wobei gewisse Freiheitsgrade (beispielsweise durch „sollte"- oder „kann"-Anforderungen) Berücksichtigung finden. Hierbei sollten bestehende Arbeiten, etwa zu Metadaten (z. B. ISO/IEC 11179 Metadata Registry [235]–[242]), aufgegriffen werden. Ein Datenreferenzmodell würde es ermöglichen, Anwendungsfälle sowie Schnittstellen mit größerer Geschwindigkeit und Kompatibilität zu entwickeln und zudem eine einheitliche (Kommunikations-) Grundlage darzustellen, auf der die Vielzahl der in Mobilität und Logistik involvierten Akteure den Datenaspekt der Interoperabilität standardisieren können. Dieses Datenreferenzmodell kann in Kombination mit einem Funktionsreferenzmodell (s. u.) eine verlässliche Grundlage der Interoperabilität sein.

BEDARF 8:
Funktionsreferenzmodell für Interoperabilität schaffen
Durch die notwendigerweise ganzheitliche Betrachtung von Interoperabilität bei gleichzeitiger Vielfalt an Standardisierungsorganisationen und -initiativen, gekoppelt mit der großen gesellschaftlichen und wirtschaftlichen Bedeutung von Mobilität und Logistik (nicht zuletzt als kritische Infrastruktur – KRITIS) sollte zeitnah ein Funktionsreferenzmodell Interoperabilität konzipiert und in Standards verankert werden, um einheitliches Verständnis dafür zu schaffen, was Interoperabilität im Kontext von KI-Anwendungen auszeichnet und wie sie realisiert sowie sichergestellt werden kann. Weiterhin sollten die für die Herstellung von Interoperabilität erforderlichen Funktionen wie etwa Datenerfassung, -verarbeitung, -auswertung, -transfer usw. definiert und deren grundlegende systemische Anforderungen dargelegt werden. Zudem sollten Vorschläge erarbeitet werden, wie die anforderungsgerechte Ausführung diese Interoperabilitätsfunktionen von der Konzeptionsphase eines Systems über den gesamten Lebenszyklus hinweg realisiert und sichergestellt werden kann. Dieses Funktionsreferenzmodell sollte das Datenreferenzmodell Interoperabilität (s. o.) integrieren und zudem bestehende Arbeiten (z. B. ISO/IEC 19763 zum Metamodel framework for interoperability (MFI) [243]–[252]) sowie bewährte Methoden und Werkzeuge (z. B. Systems Modeling Language (SysML) oder Unified Modeling Language (UML) für die Modellierung) aufgreifen. Dabei ist die fortwährende Eignung dieser Modelle unter der Maßgabe der weiteren Entwicklung und Integration von KI-Lösungen in der täglichen Praxis zu evaluieren. Die Erkenntnisse und Ergebnisse dieses Prozesses sollten auch in nicht KI-bezogene Arbeiten zur Interoperabilität in nationalen sowie internationalen Standardisierungsorganisationen eingespielt werden.

BEDARF 9:
Verfahren für Datenaustausch festlegen
Die Schnittstellen stellen den tragenden Punkt für die Interoperabilität dar. Vor dem Hintergrund zunehmender Datenmengen und Datenkomplexität sowie der fortschreitenden Verwendung von Daten durch KI-unterstützte Werkzeuge sollten daher ebenso Verfahren des Datenaustauschs standardisiert werden, insbesondere im Hinblick auf Syntax, Semantik, Formate, Konsistenz, Kohärenz, Vollständigkeit (z. B. Quality-of-Service-Angaben zu KI-generierten Daten oder Qualitätsniveau) und Art der Datenübermittlung, damit Akteure ihre Systeme und Schnittstellen entsprechend optimieren können. Weiterhin sollte der Austausch von Ergänzungsdaten standardisiert werden, um beispielsweise Datenmodelle, Inferenzmaschinen oder Angaben zur Autonomie der eingebundenen Systeme zwischen Akteuren bei Bedarf transferieren zu können und es so interessierten Anwendern zu erlauben, Datenqualität usw. selbstständig nachprüfen zu können. Ergänzend dazu sollten Gütesiegel sowie Methoden zu deren Vergabe (inkl. den dafür erforderlichen Qualitätskriterien, Prüfmechanismen usw.) definiert werden, um eine Möglichkeit zu schaffen, Interoperabilität auf einer belastbaren Vertrauensbasis aufzubauen. Dies würde den Bedarf der selbstständigen Prüfungen reduzieren und sich somit positiv auf Akteursbeziehungen und die Nachhaltigkeit auswirken.

BEDARF 10:
Art und Qualität von Daten definieren
Aufgrund der zu erwartenden Zunahme der Vernetzung zwischen Akteuren sowie des verstärkten Einsatzes (teil-) autonom agierender Systeme sollte daher zudem die für die Sicherstellung der Interoperabilität mindestens erforderliche Art und Qualität von Daten standardisiert werden. Dazu gehören klare Richtlinien, wie und in welchem Umfang gegebene Daten(arten) mit z. B. weiteren Informationen zum Datenkontext anzureichern sind sowie welche Grenzen für die autonome Weiterverarbeitung oder Nutzung dieser Daten gelten bzw. ab welchem Qualitätsniveau Daten für die autonome Verarbeitung geeignet sind. Dazu gehört auch die Definition von Qualitätskriterien (Quality of Service), die von Daten im Sinne der Interoperabilität und Sicherheit einzuhalten sind. Entsprechende Vorgaben können Eingang in das Datenreferenzmodell finden.

4.7 KI in der Medizin

In der Medizin entstehen durch KI weitere Optionen für Prävention, Diagnostik und Therapie: von smarten Apps für die – derzeit – assistierende Früherkennung von Krankheiten bis hin zu noch differenzierteren, personalisierten onkologischen Therapien. Um derartige Chancen zu nutzen, sind geeignete sichere Rahmenbedingungen zu schaffen. Auch sind noch Herausforderungen zu meistern, die KI mit sich bringt im Spannungsfeld von Ethik, rechtlichem Kontext, Ökonomie, technischen Aspekten, auch Akzeptanz sowie Empathie. Ist es ethisch, in sensiblen Fragen zu Modalitäten des Überlebens sowie zu Leben und Tod „auf eine Maschine“ zu hören? Welchen Regelungen soll es geben, damit die Technik stets dem Menschen dient – und nicht etwa andersherum?

4.7.1 Status quo

Für den Bereich der Medizin sind die normungsbezogenen Vorarbeiten und Ergebnisse bisher übersichtlich; wohingegen die Technologien für Medizin und Gesundheit auf dem Markt bereits äußerst vielschichtig sind. So bieten Gesundheits-Apps Beratung bei medizinischen Fragen von Privatpersonen und professionellen Anwendern weltweit an. Auch existieren sogenannte Chatbots in der Medizin für die Analyse von Erkrankungen.

Die derzeitigen Systeme fallen in den Bereich der sogenannten „schwachen KI“ und werden in den Bereichen „wissensbasierter Systeme“, „Musteranalyse und Mustererkennung“ sowie „Robotik“ entwickelt. (Eingruppierung nach: [12], S. 4 ff.).

Mit verschiedenen Klassifizierungssystemen bei bildgebenden Verfahren lassen sich Diagnosen – z. B. in den Bereichen der Labordiagnostik, Parasitologie, Radiologie, Pathologie, Zytologie, Dermatologie, Ophthalmologie – schneller und präziser oder Wege in z. B. minimalinvasiver Chirurgie sicherer gestalten. Bei einigen operativen Verfahren haben Chirurgen bereits heute die Möglichkeit, Robotik-Systeme auch kooperierend via Telemedizin indikationsbezogen einzusetzen.

In Krankenhäusern, medizinischen Versorgungszentren (MVZ), Praxen und Instituten finden KI-Systeme nicht nur in Diagnostik oder Therapie Anwendung. KI-basierte Systeme können auch in anderen Bereichen unterstützen. Dazu gehören Anwendungsbereiche wie Exo-Skelette und Prothesen, sensorbasiertes Monitoring und Therapiemonitoring sowie Lösungen, die die Prozesse in Medizin und/oder Administration verbessern und so helfen, die Patientenversorgung effizienter und in einigen Bereichen überhaupt erst möglich zu machen (z. B. „intelligente“ Prothesensysteme).

Die Weltgesundheitsorganisation (WHO) und die Internationale Fernmeldeunion (ITU) kooperieren in einer Fokusgruppe namens AI4Health und haben bereits ein White Paper [253] herausgegeben.

Voraussetzungen für die Anwendung von KI in der Medizin können sein:

1. Klassifikationen, Standards und Terminologien für Daten des Gesundheitswesens vereinheitlichen;
2. Interoperabilität von Daten gewährleisten und Informationen und Möglichkeiten der Datengewinnung prüfen (z. B. mobile Geräte von Patienten, technische Geräte in Krankenhäusern, MVZ, Arztpraxen, Institutionen des Gesundheitswesens);
3. Zusammenarbeit der verschiedenen Akteure zur verbindlichen Etablierung von Standards fördern und klären bzgl. Datenzugriff, Datenherkunft, Interessenlagen/Ansprüche an Daten und Schnittstellen;
4. offene Fragen zu Infrastruktur-Aspekten klären hinsichtlich „eigenständiger“ Medizin-Infrastruktur oder „allgemeiner“ Infrastruktur mit einem spezifischen Modul für Medizindaten klären, wobei auch der staatenrechtliche und föderalistische Kontext (Bundesländerbene, national, europäisch, international) klärungsbedürftig erscheint.

4.7.2 Anforderungen, Herausforderungen

KI-Systeme in der Medizin werden durch drei Faktoren maßgeblich beeinflusst:

1. die Verfügbarkeit und Qualität von Gesundheitsdaten;
2. den Rechtsrahmen und
3. die Medizinethik.

Alle drei bedingen die Vertrauenswürdigkeit eines KI-Systems, wenn Klarheit über das Gewährleisten von Privatsphäre und Transparenz existieren.

So empfiehlt das Whitepaper „Sichere KI-Systeme für die Medizin“ von Expertinnen und Experten der Plattform Lernende Systeme für einen sicheren Einsatz der KI-Systeme in der Medizin und zum Wohle von Patienten u. a. eine Zertifizierung von KI-Systemen – etwa für die Sicherstellung unverfälschter Trainingsdaten [254]. Wichtig erscheint hier, gemeinsame

Leitlinien und Prüfvorschriften für die Zulassung und Zertifizierung für KI-Datenbanken sowie für deren Betreiber zu entwickeln. Darüber hinaus sollten Hersteller gesetzlich zur Mängelbehebung verpflichtet und neutrale Einrichtungen mit dem Betrieb des KI-Assistenzsystems beauftragt werden; alles in allem ein hochkomplexes, herausforderndes Metier. Ein unabhängiges Prüfkomitee (z. B. Benannte Stellen) kann zudem in regelmäßigen Abständen die Funktionsweise der zertifizierten und eingesetzten KI-Systeme überprüfen. Auch Rückrufprozesse könnten etabliert werden.

In der Medizin handeln Ärzte sowie Angehörige weiterer Heil- und Gesundheitsfachberufe nach ethischen Prinzipien. Bereits seit längerer Zeit finden darauf aufbauend Diskussionen statt, inwieweit es zulässig sein kann, dass Geräte Entscheidungen maßgeblich beeinflussen bzw. sogar übernehmen. Letztendlich stellen Ärzte Diagnosen und legen Therapien fest. Ferner kommt zum Tragen, dass eine Debatte stattfinden muss, wann ein „menschliches Handeln“ und wann lediglich eine „menschliche Aufsicht“ (schwache KI) notwendig ist oder ein autonomes Handeln erfolgt (starke KI). Hier fehlen schlicht Grundsätze und Regelungen für die Mensch-Maschine-Interaktion im medizinischen Bereich.

Die ethischen Aspekte sind umso komplexer, je globaler ein KI-System agieren soll. Weltweit existieren unterschiedliche Herangehensweisen aufgrund unterschiedlicher kultureller und historischer Hintergründe sowie medizinischer Versorgungsstrukturen. Für Europa sind an dieser Stelle besonders die „Guidelines for Trustworthy AI“ der Europäischen Kommission und für Deutschland das Gutachten der Datenethikkommission zu erwähnen. Weltweit hat die World Medical Association(WMA) mit der Deklaration von Helsinki (1964, zuletzt aktualisiert 2008 [255]) für alle Kulturkreise einen Nenner definiert.

Wie in anderen Bereichen stellt sich auch für die Medizin die Frage nach der Haftung. Dies ist z. B. beim Stellen einer Fehldiagnose der Fall oder beim Verursachen von Personen- oder ökonomischen Schäden. Daraus ergeben sich außerdem Unklarheiten über die Beweislastumkehr. Im potenziellen – nicht zu wünschenden Schadensfall – ist klärungsbedürftig, ob Kunde, Hersteller, Betreiber/Anwender oder eine ganz andere involvierte Stelle beweispflichtig ist. Hierzu bedarf es einer Risikoeinschätzung, z. B. in Szenarien (siehe Tabelle 9) [87]):

Tabelle 9: Anforderungen und Herausforderungen der KI-Systeme in der Medizin

AI Module class Requirement class	**High risk**	**Low risk**
Mandatory	No deviation from requirements allowed	No deviation from requirements allowed
Highly recommended	Deviation from requirements with justification only	Deviation from requirements with justification only
Recommended	Deviation from requirements with justification only	Deviation from requirements without justification allowed

Zum einen ist die Zulassung von Unklarheit im Rechtsrahmen betroffen. In Deutschland gibt es ein hoch reguliertes Medizinproduktegesetz (MPG) [256], in Europa die Medizinprodukteverordnung (englisch: Medical Device Regulation) [141] mit Detailregelungen. Jedes Medizinprodukt darf bisher nur auf dem Markt bereitgestellt werden, wenn es eine Zulassung hat, die wiederum nur einen bestimmten Zustand in Betracht zieht. Für die Zulassung selbst müssen Medizinprodukte alle relevanten rechtlichen Anforderungen erfüllen und ein Konformitätsbewertungsverfahren durchlaufen haben, ggf. unter Einbeziehung einer Benannten Stelle. Das Konformitätsbewertungsverfahren bezieht sich auf einen bestimmten technischen Zustand des Produkts mit entsprechenden Funktionen. Mit dem stetigen Weiterlernen von KI-Systemen und damit der Veränderung des Produkts selbst ist der Zustand zum Zeitpunkt der Zulassung (und ggf. Zertifizierung durch eine Benannte Stelle) mitunter bereits verlassen.

Zum anderen besteht Unklarheit beim Rechtsrahmen für den Einsatz von KI in der Medizin, beispielsweise hinsichtlich zivilrechtlicher Haftung für potenzielle, nicht zu wünschende Behandlungsfehler. Untrennbar damit verbunden ist die Klärungserfordernis einer Zulässigkeit des Einsatzes von KI-Systemen zur Entscheidungsfindung und auch der Verzicht auf eine Entscheidungsunterstützung durch KI-Systeme. Vor diesem Hintergrund bedarf es eines breiten gesellschaftlichen Konsenses über Zulassung sowie Einsatz kontinuierlich lernender Systeme.

Zum dritten ist der Marktzugang von Unklarheit im Rechtsrahmen gekennzeichnet. Medizinprodukte sind teils sehr

komplex und streng reguliert. In Deutschland gilt derzeit noch das MPG, welches jedoch zeitnah durch die Medizinprodukteverordnung abgelöst wird.

Jedes Medizinprodukt darf bekanntlich nur mit einer – nota bene gültigen – CE-Kennzeichnung auf dem Markt bereitgestellt werden.

Für die CE-Kennzeichnung müssen Medizinprodukte alle relevanten rechtlichen Anforderungen erfüllen und ein Konformitätsbewertungsverfahren, ggf. unter Einbeziehung einer Benannten Stelle, durchlaufen haben. Das Konformitätsbewertungsverfahren bezieht sich auf einen bestimmten technischen Zustand des Produkts mit entsprechenden Funktionen. Mit dem stetigen Weiterlernen von KI-Systemen und damit der Veränderung des Produkts selbst wird der Zustand zum Zeitpunkt der CE-Kennzeichnung (und ggf. Zertifizierung durch eine Benannte Stelle) mitunter verlassen, wodurch die Voraussetzungen für den Marktzugang de iure nicht mehr erfüllt werden.

Entscheidungen von KI-Systemen erfolgen basierend auf der Interpretation von vorhandenen Daten mit notwendigerweise hoher Qualität; dies scheint derzeit jedoch häufig noch nicht adäquat gegeben. KI-basierte Anwendungen sind naturgemäß der DSGVO [95] unterworfen; dies wirft Fragen für die Entwicklung und Verwendung von KI-Systemen auf. Klärungsbedürftig sind auch im aktuellen Kontext Dateneigentum zur Datenintegrität und zur Einwilligung zur Datenverwendung. Personen, die der Verwendung ihrer Daten zustimmen, tun dies in der Regel für einen bestimmten Zweck. Durch die Verarbeitung von Daten in einem KI-System kann sich dieser Zweck systemimmanent ändern. Daraus erwächst weiterer Klärungsbedarf zu rechtskonformem, ethisch abgesicherten Prozedere, da ein KI-System sich mit jedem Lernschritt ändern kann und auch bei „nicht freigegebenen“ Daten getätigte Interpretation nicht „vergessen“ kann, sofern dies nicht in der KI-Programmierung vorgesehen wurde. Potenziell entsteht durch die Verwendung dieser Daten, z. B. als Bestandteil eines gelernten Modells, ein gesellschaftlicher Nutzen, beispielsweise im Sinne einer Verbesserung von Diagnose- und Therapiemöglichkeiten über das Individuum hinaus, mit weitergehenden Entlastungen direkt und indirekt Beteiligter im Gesundheitswesen.

Daten müssen repräsentativ, konsistent und genau sein, darüber hinaus bedarf es technischer Verfügbarkeit (Datenformate, Maschinenlesbarkeit, Sicherheit sowie Zugriffsmöglichkeiten).

4.7.3 Normungs- und Standardisierungsbedarfe

BEDARF 1:
Fehlerklassifikationen, Fehleinordnungen und Lernen aus Fehlern definieren
Resultierend aus Betrachtungen der Medizinethik sind Erkenntnisse darüber zu erwarten, welche Stufe im KI-System ausschlaggebend ist für das Lernen. Maßgeblich hierfür sind Ausgangsdaten (Input – mit Unterscheidung in Trainingsdaten und Volldaten, siehe DIN SPEC 13266), die eigentliche Lernstrategie (Verarbeitung) im engeren Sinne, das Ergebnis (Outcome) und die Verwendung (Impact). Obligat bedarf es einer Konkretisierung zum präventiven Umgang mit Fehlklassifikationen/Fehleinordnungen durch ein KI-System, da ein KI-System aufgrund seiner Daten und seiner bisherigen Lernergebnisse Vorurteile/ethisch kontraproduktive Prozesse entwickeln und somit fehlgeleitet und/oder diskriminierend agieren kann. Relevant für sozial-ethisch konformes Lernen ist des Weiteren der adäquate Umgang mit Fehlversuchen bzw., ob Lernen durch Fehlversuche überhaupt zulässig sein darf.

BEDARF 2:
Medizinethische Werte definieren
KI-Systeme in der Medizin müssen ethische Werte im sozialkulturellen Kontext einhalten, wobei die Festlegungen dafür zu definieren sind – wenn dies nicht schon gegeben ist. Eine Orientierung ist an den Grundlagen und ethischen Empfehlungen der World Medical Association (WMA) möglich, die für alle Kulturkreise einen Nenner definiert. Diese sind in der jeweils aktuellen Fassung der Deklaration von Helsinki [255] enthalten und erfahren stetige Anpassung. Eine Beachtung dieser Deklaration in der jeweils dann gültigen Fassung muss dem KI-System möglich sein. Hierbei zeigt sich u. a. die Relevanz der Verbindung von Big Data und ethischen Aspekten.

BEDARF 3:
Prüfprozess zum Evaluieren vorhandener Prinzipien schaffen
Eine Bewertungsgrundlage zur Adäquanz gegebener Prinzipien erscheint dringend erforderlich. Es existieren bekanntlich trotz einer Vereinheitlichung durch die World Medical Association (WMA) noch verschiedene nationalstaatliche Prinzipien, Forschungsergebnisse in differenten Populationen, unterschiedliche Erkenntnisse etc. Für ein KI-System gilt es, einen Prüfprozess zu entwickeln, der die Eignung von Prinzipien, Forschungsergebnisse, Erkenntnissen, weitere Variablen etc. bewertet und dabei die Dynamik der KI-Systeme (insbesondere selbstlernender Systeme) integriert. Dies gilt nicht nur

für das KI-System selbst, sondern auch für die konnotierte Datenqualität.

BEDARF 4:
Rechtliche Definitionen und Vorgaben für selbstständig lernende und sich selbstständig weiterentwickelnde/verändernde KI-Systeme eindeutig festlegen
Es muss klargestellt werden, ob und ggf. wie der Nachweis (Dokumentation) über die Sicherheit und Leistungsfähigkeit solcher KI-Systeme über den gesamten Lebenszyklus hinweg durch den Hersteller erfolgen soll. Nach derzeitiger Auffassung der Interessengemeinschaft Notified Bodies (IG-NB) ist eine Zertifizierung selbstständig lernender und sich selbstständig weiterentwickelnder/verändernder KI-Systeme nach gültigen rechtlichen Rahmenbedingungen nicht möglich. Hier ist der Gesetzgeber gefordert.

Im Kontext mit ärztlichen Entscheidungen und Eingriffen am Menschen entwickeln solche „continuous learning systems" nicht zuletzt eine bisher nicht eindeutig zugeordnete, jedenfalls andere ethische Dimension. Es sind rechtliche Verantwortungs-Definitionen notwendig – ebenso Festlegungen zur Haftung für ein KI-System, welches sich selbstständig weiterentwickelt/verändert – fokussierend auf Hersteller oder Anwender oder „third parties".

BEDARF 5:
Proprietät, Allokation und Widerruf von Daten
Die Legislative ist ebenfalls gefordert zur Regelung der Proprietät von (Gesundheits-)Daten und deren Allokation sowie vor allem des legitimierten Verfahrens beim Widerruf bzw. Entzug einer Datenüberlassung integrierend in ein bereits stattgefundenes, datenbasiertes Lernen des KI-Systems.

Zu beachten ist die Tatsache, dass es aus medizinischer Sicht wertvoll und wertschöpfend sein kann, ein KI-System in weiteren Anwendungen als ursprünglich gedacht einzusetzen, zu nutzen oder innovative Lösungen aus einem anderen Rechtsraum einzusetzen (z. B. Verwendung in Asien, obwohl Daten aus europäischen Einrichtung stammen). Zur Nutzbarmachung von Daten sind gesonderte Anforderungen an deren Anonymisierung oder Pseudonymisierung denkbar. Zweites stellt eine weitere Option bei Änderung des ursprünglichen Verwendungszwecks dar.

Für Einrichtungen, die Daten zu Zwecken der Forschung bereitstellen wollen, ergeben sich prozessuale, rechtliche und technische Fragen in der Datenausleitung, der Qualitätssicherung, der Datenbereitstellung, die heute im besten Fall von großen Einrichtungen identifiziert, aber nur teilweise beantwortet werden können.

Darüber sind bei gegebenem föderalem Prinzip die Landesdatenschutzgesetze und die konfessionellen Datenschutzgesetze eine besondere Herausforderung im Interesse „nur" einer nationalen Rechtssicherheit; es gilt jedenfalls, insbesondere die europäische Perspektive (z. B. Health Data Space, Gaia-X etc.) zu beleuchten.

BEDARF 6:
Definition von Daten und Verwendung festlegen
Die Daten an sich benötigen ebenfalls weitere Spezifikationen, welche den verantwortlichen Stellen gegenüber transparent zu machen sind. So gilt es, durch die Hersteller datenbasierter Anwendungen Ein- und Ausschlusskriterien zu definieren, eine Beschreibung von Trainings-, Validierungs- und Testdaten anzugeben und darzulegen, wie lösungsorientiert und rechtskonform mit statistischen „Ausreißer-Daten" umgegangen wird. Gleiches gilt für weitere Ergänzungen, beispielsweise im Zusammenhang mit Einschränkungen der Arbeit mit analytischen Daten – wobei es zu klären gilt, ob Einschränkungen notwendig sind oder zu vermeiden. Relevant dafür ist eine Datensatzbetrachtung zur Auswirkung von Zufallsvariablen mit dem Ziel, daraus Festlegungen auf Einzeldaten, statistische Auswertungen etc. zu formulieren.

BEDARF 7:
Einschränkungen bei Big Data festlegen
Die Einschränkung bei Big Data ist zu klären. Die Daten bei Big Data sind zunächst die gleichen wie bei traditionellen Daten, nur dass sie in gigantischer Anzahl anfallen. Dies hat dann aber Auswirkungen auf die Einschätzung von Transparenz und Richtigkeit. Während die Richtigkeit aber aufgrund der größeren statistischen Stichprobe steigt, sinkt die Transparenz aufgrund der Datenvielfalt und mangelnder Reproduzierbarkeit der herangezogenen Datenbasis.

BEDARF 8:
Datenschutz und Datenqualität im Gleichgewicht betrachten
Aktuell empfiehlt es sich, wegen des Rechtsrahmens verstärkt den Datenschutz einzubeziehen. Darunter kann die Qualität leiden. Somit braucht es definierte Vorgaben für ein Gleichgewicht zwischen Datenschutz und Datenqualität. Neben berechtigten datenschutzrechtlichen Aspekten ist an dieser Stelle jedoch ebenfalls das ethische Gebot zur Datennutzung hervorzuheben, sofern dieses dem Allgemeinwohl dient.

Daten sollen fair sein und dürfen nicht diskriminieren. Dafür fehlen noch Vorgaben, wie sich dies bei der Datenerhebung einhalten lässt. Daten selber können keine Eigenschaft wie „fair“ annehmen, sondern lediglich der Datensatz aufgrund seiner Erhebung (z. B. Auswahl der Population bei statistischer Erhebung, Bewertung des/der auswertenden Person etc.). Definierte Vorgaben senken die Gefahr, dass ein KI-System ausgehend von seinen Daten diskriminierend bewertet oder handelt.

BEDARF 9:
Generieren und Konsentieren von Grundsätzen für die Mensch-Maschine-Mensch-Interaktion im medizinischen Bereich
Eine Orientierungsoption bieten die Kriterien für die Gestaltung der Mensch-Maschine-Interaktion [257] insbesondere fokussierend/die Kriterien zu den Themen Schutz des Einzelnen und Vertrauenswürdigkeit. Dabei stehen u. a. Datensicherheit, Datenschutz und Diskriminierungsfreiheit sowie Qualität verfügbarer Daten, strukturierte Transparenz, Erklärbarkeit und Widerspruchsfreiheit der KI-Systeme im Mittelpunkt.

BEDARF 10:
Innovationen für Einsatz von KI-Systemen fördern
KI-Systeme und -Anwendungen bieten nicht nur enorme Chancen für eine bessere oder konsolidierbare Gesundheitsversorgung, sondern schlicht nachhaltige Perspektiven für den Wirtschaftsstandort Deutschland. Ein Kulturwandel im Sinne einer maximalen Förderung von Innovationen für den Einsatz von „Künstlicher Intelligenz – Made in Germany“ – auch im Gesundheitswesen – ist conditio sine qua non im Interesse einer international führenden Rolle; dabei sind Kooperationen von Wissenschaft und Wirtschaft zu fördern, die sich zu strukturierter Transparenz und konsentierter Offenheit bei der Entwicklung von KI-Systemen und -Anwendungen verpflichten und sowohl Start-ups, den Mittelstand als auch Großunternehmen vernetzen.

BEDARF 11:
KI-Normen und KI-Excellence-Cluster zur Verarbeitung von medizinischen Daten
Deutschland hat eines der besten Gesundheitssysteme der Welt. Gerade das hohe Ausstattungsniveau an technischen Geräten bildgebender Verfahren zur Diagnose und Therapie spielt hierfür eine wichtige Rolle. Trotz großer Investitionen in moderne Röntgengeräte, Ultraschallgeräte, Magnetresonanztomografen und andere nuklearmedizinische Bildverfahren findet nur ein sehr geringer Anteil der täglich generierten Bilder Verwendung zur KI-Forschung und zum Training von medizinischen KI-Klassifikatoren, verschärft durch das Fehlen von Normen zur Verarbeitung von medizinischen Bildern. Weiter erlauben die aktuellen deutschen Regelungen zum Daten- und Patientenschutz keine Nutzung von Bildern, Texten oder Sprachdokumentationen und folglich keine wettbewerbsfähige Forschung. Durch all diese Faktoren schwindet das Potenzial, KI-basierte Assistenten ganzheitlich für das Gesundheitssystem, z. B. zur Aufklärung hinsichtlich Prävention sowie zur Unterstützung der Ärzte bei Diagnostik und Therapie zu entwickeln.

Daher braucht es eine gesetzliche Initiative, welche die Nutzung medizinischer Daten zu regionaltypischen, populationsbezogenen Forschungs- und Entwicklungszwecken im Gesundheitswesen ermöglicht und bewusst fördert. Weiter sind definierte KI-Normen und KI-Standards nötig (siehe Kapitel 4.7). Zusätzlich könnte ein KI-Excellence-Cluster für medizinische Bildgebungsverfahren helfen, diese Normen und Standards zu entwickeln. Erste Projekte sollten in 2021 starten, um den Rückstand gegenüber anderen Ländern zu verringern.

BEDARF 12:
Daten der Forschung zur Verfügung stellen
Darüber hinaus bedarf es eines verbesserten Zugangs zu hochqualitativen Daten sowohl für Universitäten, Hochschulen und andere wissenschaftliche und forschende Einrichtungen sowie für Unternehmen, die durch die Entwicklung von Innovationen zur zukunftsorientierten Ausgestaltung des Gesundheitswesens beitragen – gleichermaßen für Diagnostik, Therapie, pharmazeutische Industrie. In diesem Kontext sind entsprechende Aktivitäten auf nationaler und europäischer Ebene (bisher Medizininformatik-Initiative, europäischer Gesundheitsdaten-Raum) zu definieren, die besonders die ethischen Rahmenbedingungen für KI inkludieren und an denen sich künftige Maßnahmen orientieren sollten.

BEDARF 13:
Automatisierte Text- und Spracherkennungsverfahren bei Zulassung
Die Regulierung für die Zulassung von Medikamenten, Medizinprodukten und diversen Maßnahmen ist ein komplexes Konstrukt aus Entwicklung, Prüfung, Konformitätsbewertung und Zertifizierung. Im Zusammenhang damit steht ein großer Dokumentationsaufwand, der die Zulassung trotz dringendem Bedarf bremst – besonders in Krisenzeiten. Hier kann ein KI-System den Prozess deutlich beschleunigen, das vollautomatisiert Texte oder Sprache erkennt.

Epilog/Zusammenfassung und Ausblick

Die skizzierten Themen, Handlungsempfehlungen und Anregungen sind nicht auf die zentralen medizinischen Bereiche Diagnostik und Therapie begrenzt, sondern beinhalten u. a. die Bereiche Prävention, Gesundheitsförderung, „Care“ (englischer Begriff, umfassender als deutsche Begrifflichkeit „Pflege“ umfasst), Screening und multiprofessionelle Klienten-/Patientenorientierung.

Wichtig erscheint die von (ärztlicher) Ethik getragene Balance zwischen „high tech und high touch“.

Für die Nutzung des von KI u. a. in der Medizin ausgehenden großen Potenzials ist ein gesamt-gesellschaftlicher Konsenses unabdingbar – mit eben einer Normungsroadmap für einen möglichst sicheren, gemeinsamen Weg.

5 Anforderungen an die Erarbeitung und Nutzung von Normen und Standards

5.1 Überprüfung und Entwicklung von Normen und Standards im Bereich KI

5.1.1 Überprüfung bestehender Normen und Standards

Die Einsatzfelder für KI sind äußerst vielfältig. In nahezu allen Wirtschaftsbereichen und auch sonstigen Anwendungsfeldern finden KI-Technologien sowohl in Form von Komponenten in Endprodukten und Dienstleistungen als auch in den produktiven Kern- und Unterstützungsprozessen innerhalb der Unternehmen Anwendung. Nach Einschätzung der Bundesregierung wird Künstliche Intelligenz damit früher oder später für alle wirtschaftlichen und gesellschaftlichen Bereiche von großer Bedeutung sein. Mit Normen und Standards verhält es sich ähnlich. Auch diese existieren für nahezu alle Wirtschaftsbereiche und Anwendungsfelder. Aktuell umfasst das deutsche Normenwerk mehr als 30.000 Normen (DIN, DIN EN, DIN EN ISO/IEC). Kombiniert man beide Thesen, so bedeutet das, dass ein Großteil der über 30.000 existierenden Normen überprüft und um KI-Aspekte ergänzt werden muss.

Die KI-Strategie der Bundesregierung [12] adressiert diesen Aspekt in Handlungsfeld 12 und empfiehlt die Maßnahme zur „Überprüfung von bestehenden Normen und Standards auf KI-Tauglichkeit“. Zwar ist der Begriff „KI-Tauglichkeit“ nicht definiert, gemeint ist hier aber:

Normen und Standards, die eine Relevanz bei der Anwendung von KI haben, sind zu identifizieren und schließlich um KI-Spezifika zu erweitern. Durch die Erweiterung der Anwendungsbereiche der Normen und Standards können KI-Lösungen unter Zuhilfenahme dieser sicher und zuverlässig zum Einsatz gebracht werden. Zur Umsetzung der Maßnahme bedarf es einer Methodik, nach der bestehende Normen und Standards hinsichtlich ihrer KI-Relevanz identifiziert werden. Denkbar ist ggf. der Einsatz von IT-Tools zur Unterstützung dieser Recherche. Parallel dazu ist eine Systematik zu entwickeln, um etwaige Handlungsbedarfe zur Optimierung bestehender Normen und Standards zu ermitteln. Schließlich sind auf Basis dieser Vorarbeiten Maßnahmen zu konzipieren, die auf eine umfassende Einbeziehung von KI-Aspekten zielen. Eine der größten Herausforderungen in diesem Kontext dürfte die fehlende KI-Expertise sein. Die oft vertikal ausgerichteten Gremienstrukturen, in denen Normen insbesondere für traditionelle Branchen entwickelt werden, bedingen ein tiefes Domänenwissen. Dieses gilt es um das KI-Technologiewissen zu erweitern. Zu beachten ist hierbei, dass die in Rede stehenden relevanten Normen ganz überwiegend europäischen oder internationalen Ursprungs sind. Prinzipiell schwieriger dürfte sich die Überprüfung bestehender Standards gestalten, da nur wenige von den etablierten Normungsorganisationen und die ganz überwiegende Anzahl der relevanten Standards von diversen Konsortien entwickelt wurden.

5.1.2 Agile Entwicklung von Normen und Standards für KI

Eine bislang große Herausforderung in Bezug auf die Entwicklung von Normen und Standards für KI-Systeme stellt die große Dynamik der KI-Technologieentwicklung dar. Viele Branchen verwenden abhängig vom Einsatzfeld der KI-Lösung unterschiedliche und auf den Anwendungsfall bezogene KI-Technologien. Hybride KI-Lösungen beruhen oftmals sogar auf einer Kombination von KI-Methoden. Die Anwendungsspezifika werden dabei in den meisten Fällen von modernsten Ansätzen aus KI-Teildisziplinen erfüllt, die individuell angepasst und verfeinert werden. Folglich ist die Dynamik an der Schnittstelle zwischen KI-Forschung und industrieller Entwicklung und Anwendung besonders hoch. Auf diese Weise wird die angewandte KI ständig weiterentwickelt und industriell evaluiert. Die Standardisierung muss diesem Spannungsbogen zwischen angewandter Forschung und industriereifer Entwicklung Rechnung tragen und pragmatische, bidirektionale Ansätze bei der Analyse der Standardisierungsbedarfe sowie der Entwicklung marktreifer Standards verfolgen. Hierfür braucht es einen iterativen Prozess, der bei der Gestaltung von Normen und Standards wechselseitig Impulse aus Forschung, Industrie, Gesellschaft und Regulierung einbezieht und ein kontinuierliches und gegenseitiges Lernen zwischen den Akteuren unterstützt. Im Zentrum dieses Ansatzes steht die Erprobung und sukzessive Verfeinerung der entwickelten Standards entlang von Use Cases. So können anwendungsspezifische Bedarfe frühzeitig erkannt und marktfähige Standards für KI realisiert werden. In der Folge wird damit die Akzeptanz dieser Standards in Wirtschaft, Wissenschaft und Gesellschaft sichergestellt.

5.2 SMART Standards – Neugestaltung von Normen für KI-Anwendungsprozesse

Das vorliegende Kapitel stellt die Motivation für SMART Standards in Bezug auf KI dar sowie den aktuellen Stand der Entwicklungen eines zukünftigen Modells und mögliche technologische Ansätze zur Realisierung von SMART Standards.

Im Anhang 11.4 findet sich darüber hinaus eine detailliertere Darstellung der technologischen Ansätze im Sinne einer weiterführenden Lektüre.

5.2.1 Motivation

Definition SMART Standards: Norm (Standard), deren Inhalte für Maschinen, Software oder sonstige automatisierte Systeme anwendbar (applicable) und lesbar (readable) sind und darüber hinaus anwendungs-/nutzerspezifisch digital bereitgestellt werden können (transferable).

SMART Standards – das ist ein Thema, das seit drei Jahren nicht nur national in DIN/DKE, sondern auch bei CEN/CENELEC (CCMC) und ISO/IEC an Bedeutung gewinnt, siehe Abbildung 23.

Über aufeinander abgestimmte Pilotprojekte und wissenschaftliche Untersuchungen nähern sich in ersten Schritten die Promotoren dieser innovativen Aufgabe und deren möglichen Lösungen. Eine Herausforderung wird die Konsolidierung eines gemeinsamen Zielbildes von Normenerstellern und Normennutzern sein: Wie gestaltet sich ein zukünftiger Entwicklungsprozess von SMART Standards und welches Informationsmodell ist hierfür erforderlich? Welchen Input können SMART Standards für nachgelagerte KI-basierte Anwendungsprozesse liefern? Zu diesen Fragestellungen fasst der vorliegende Beitrag Bestehendes zusammen und liefert Ansätze und Impulse für weiterführende Untersuchungen, die im Text und Anhang mit Anforderungen mit KI-Relevanz gekennzeichnet sind.

5.2.2 Status quo

Die direkte Weiterverwertung von Normen und deren Inhalte in nachgelagerten Prozessen gewinnt zunehmend an Aufmerksamkeit. Von Normbestandteilen (Wertetabellen, Teilebeschreibungen, 3D-Modellen, Software, Anforderungsdefinitionen, Prüfverfahren), die von Maschinen direkt übernommen und ausgeführt werden können, versprechen sich Unternehmen zukünftig Effizienzgewinne [258]. Eine umfassende elektronische Bereitstellung von nationalen/europäischen/internationalen Normen erfolgt heute noch überwiegend über PDF-gestützte Normen-Management-Verfahren, die mittels Metadaten organisiert werden. Technologisch und anwenderseitig betrachtet befinden wir uns – für eine mittlerweile jedoch sehr breit angelegte Informationsbereitstellung in Bezug auf die Zahl der nachgewiesenen Regelwerke und der Indexierungstiefe – auf einem ausgereiften und zuverlässigen Stand (siehe Abbildung 24).

Der heutige seit Jahrzehnten etablierte Workflow funktioniert erfolgreich und ausgewogen aufgrund stiller oder vereinbarter Übereinkünfte der handelnden Prozesspartner. Die zugrunde liegenden Prinzipien sind sorgfältig und normungs- und rechtskonform aufeinander abgestimmt und garantieren ein zuverlässiges Management der Normungsergebnisse in kundenorientierten Systemen. Heutige Änderungen der Ausprägungen der Parameter „Prinzip“ und „Kennwert“ erfolgen gewissenhaft und im Konsens aller Beteiligten unter Berücksichtigung jeweils geltender Regeln, siehe Abbildung 25.

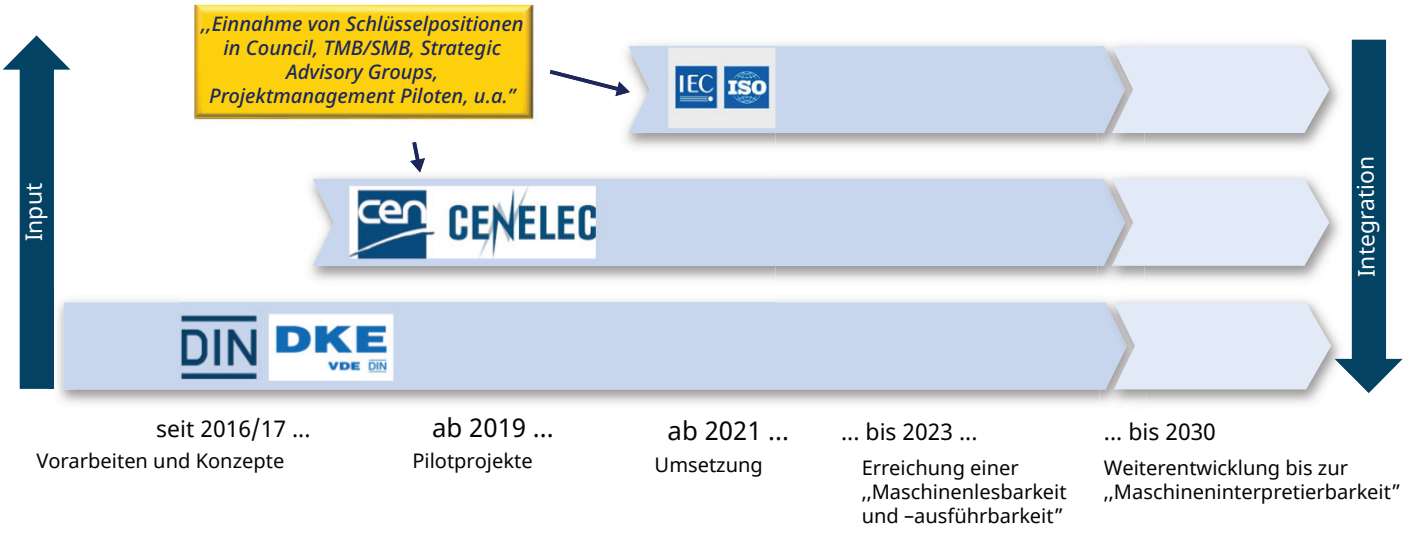


Abbildung 23: Beteiligte Normungsorganisationen bei SMART Standards

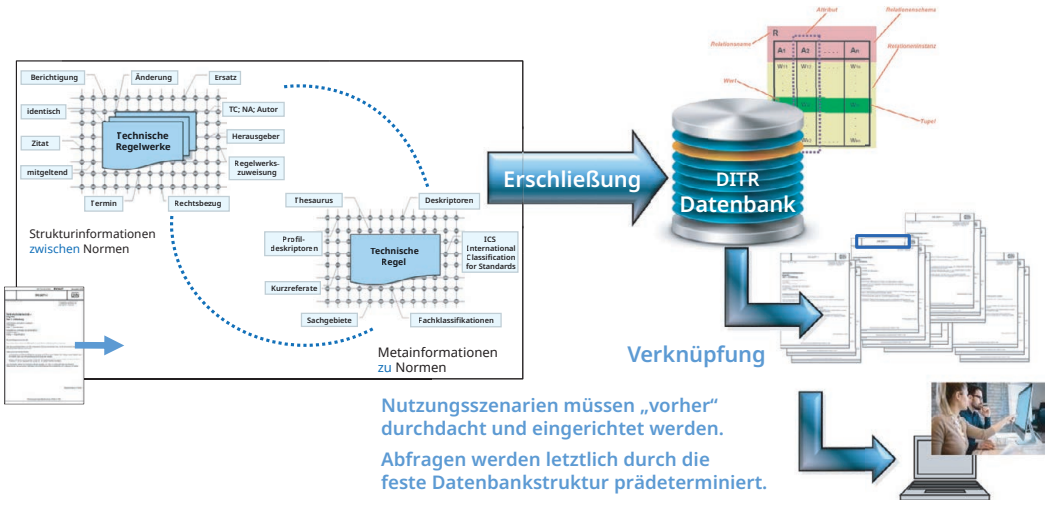
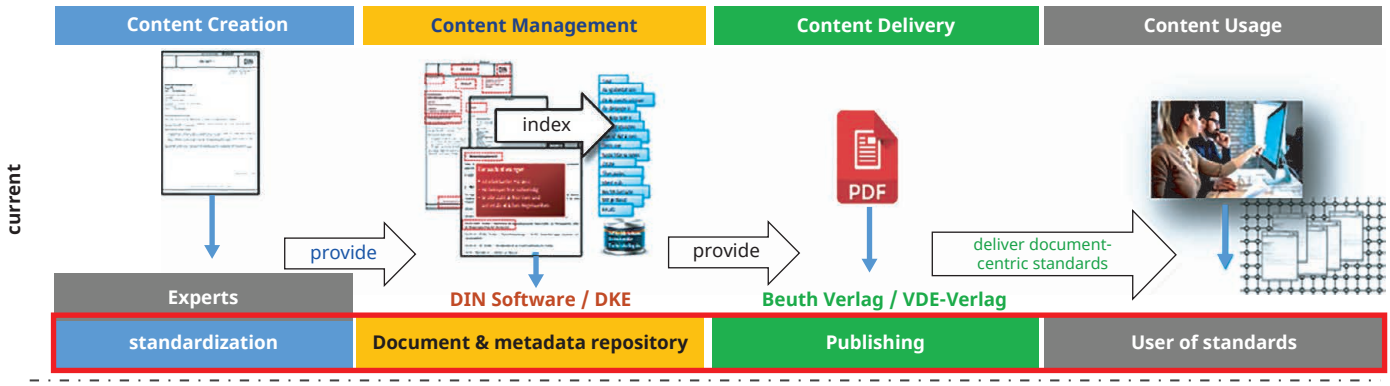


Abbildung 24: Teilautomatisierte Indexierung von Dokumenten für Kundenprozesse

LEVEL 1



Feststellung einiger relevanter Anforderungen für den bestehenden Gesamtprozess

Das „Normungsprinzip":
- Ein stabiler und detailliert beschriebener Prozess, der sich erfolgreich durchgesetzt hat, z.B. in Bezug auf Rechte an Inhalten, Beteiligung von Interessensgruppen, Entwurfsveröffentlichungen, Einspruchsverfahren, ...
- Die Normung beruht auf Prinzipien, z.B. Konsens, Einheitlichkeit, Internationalität, ...
- Basierend auf Qualitätsmerkmalen, z.B. Art der Rechtsverbindlichkeit, kartellrechtliche Freistellung, Verbraucherakzeptanz, demokratische Legitimation, Produkthaftung,

Einige Kennzahlen (in DIN e.V.):
35.000 Experten aus Industrie, Forschung, öffentlichem Dienst arbeiten zusammen mit 200 Projektleitern von DIN an 2.000 neuen Normen pro Jahr (von insgesamt 34.000 deutschen Normen)

Das „Metadaten-Prinzip":
- Zentraler Betrieb und Pflege der Metadaten-Datenbank
- Prozess (Indizierung) basierend auf seit Jahrzehnten abgestimmten standardisierten Regeln
- Professioneller Erfahrungsaustausch mit Schlüsselkunden etabliert

Einige Kennzahlen (am Beispiel der DIN Software):
Kontinuierlich weiterentwickeltes, hoch automatisiertes Verfahren mit ca. 90 Metadatenfeldern in 300 Regelwerken (national und international) führt zu ca. 60.000 Änderungen in Datensätzen p.a., die von 20 spezialisierten Mitarbeitern verwaltet werden

Das „Dienstleistungsprinzip":
- Verbreitung von nationalen und internationalen Normen und anderen technischen Vorschriften
- Entwicklung oder Aufbereitung von Expertenwissen in allen Medienformaten für Industrie, Wissenschaft, Handel, Dienstleistung, Studium und Handwerk, ...
- Anbieten von Dienstleistungen zur Prozessunterstützung des Kunden

Einige Kennzahlen (am Beispiel des Beuth Verlages):
800.000 Artikel (inkl. nationale und internationale Normen), die von 180 Mitarbeitern für 170.000 aktive Kunden angeboten werden

Das „Nutzerprinzip":
- Verwendung strukturierter und zuverlässiger Daten zur Verwaltung von (nach dem Normungsprinzip entwickelten) Dokumenten in kundenorientierten Systemen
- Zusätzlich zu den öffentlich zugänglichen Normen können Unternehmen auch interne Normen erstellen
- Die Anwendung von Normen ist freiwillig und erfolgt eigenverantwortlich

Einige Kennzahlen (im deutschen Markt):
Unternehmen nutzen ca.10 bis 20.000 Standards und Lösungen

Abbildung 25: Heutiger Workflow: Von der Normung bis zur Nutzung von Normen

Die anstehenden tief greifenden prozessualen Veränderungen im Rahmen der SMART Standards Erarbeitung des Content Managements, der Distribution und der Nutzung werden vor dem Hintergrund bestehender eingeführter und regulierter Vorgehensweisen abgegrenzt und neu definiert werden müssen. Der entscheidende Wert („asset“) eines Normungsgegenstands muss erhalten bleiben.

5.2.3 SMART Standards – Stufenmodell

Eine Herausforderung wird die Konsolidierung eines **gemeinsamen Verständnisses der Entwickler und Anwender von SMART Standards** sein: Wie wird ein zukünftiger Entwicklungsprozess von SMART Standards gestaltet werden, welche inhaltliche Struktur ist erforderlich und wie sehen die Anwendungsszenarien aus? Auf der Grundlage eines Modells, siehe Abbildung 26, wird im Folgenden eine systematische Beschreibung der Teilprozesse und von deren Teillösungen vorgenommen. In zwei ausführlichen webbasierten Seminaren werden die Modelle ebenso vorgestellt [259].

Das Modell besteht aus drei Dimensionen:

→ Darstellungsformen Level 1 bis 4: „Utility model“ – Nutzungsformate auf der Grundlage der erforderlichen Technologien, siehe auch Abbildung 26.
→ Erstellungs- und Nutzungsszenario: „Prozessmodell“ [260] – von Content Creation bis Content Usage.
→ Beispiele für die Realisierungen der Teilprozesse.

Der existierende Wertschöpfungsprozess für **Level 1** ist bereits in Kapitel 5.2 beschrieben. Er wird in Abbildung 26 zur Vollständigkeit aufgeführt, um die Transformation der einzelnen Teilprozesse auf Level 2 bis 4 besser einordnen zu können.

Die Aktivitäten in **Level 2 und 3** sind erste Schritte hin zu SMART Standards, da granulare Informationen im Rahmen bestehender Normungsprozesse entstehen werden. Weitere Ausführungen stehen im Anhang in Anhang 11.4.1.

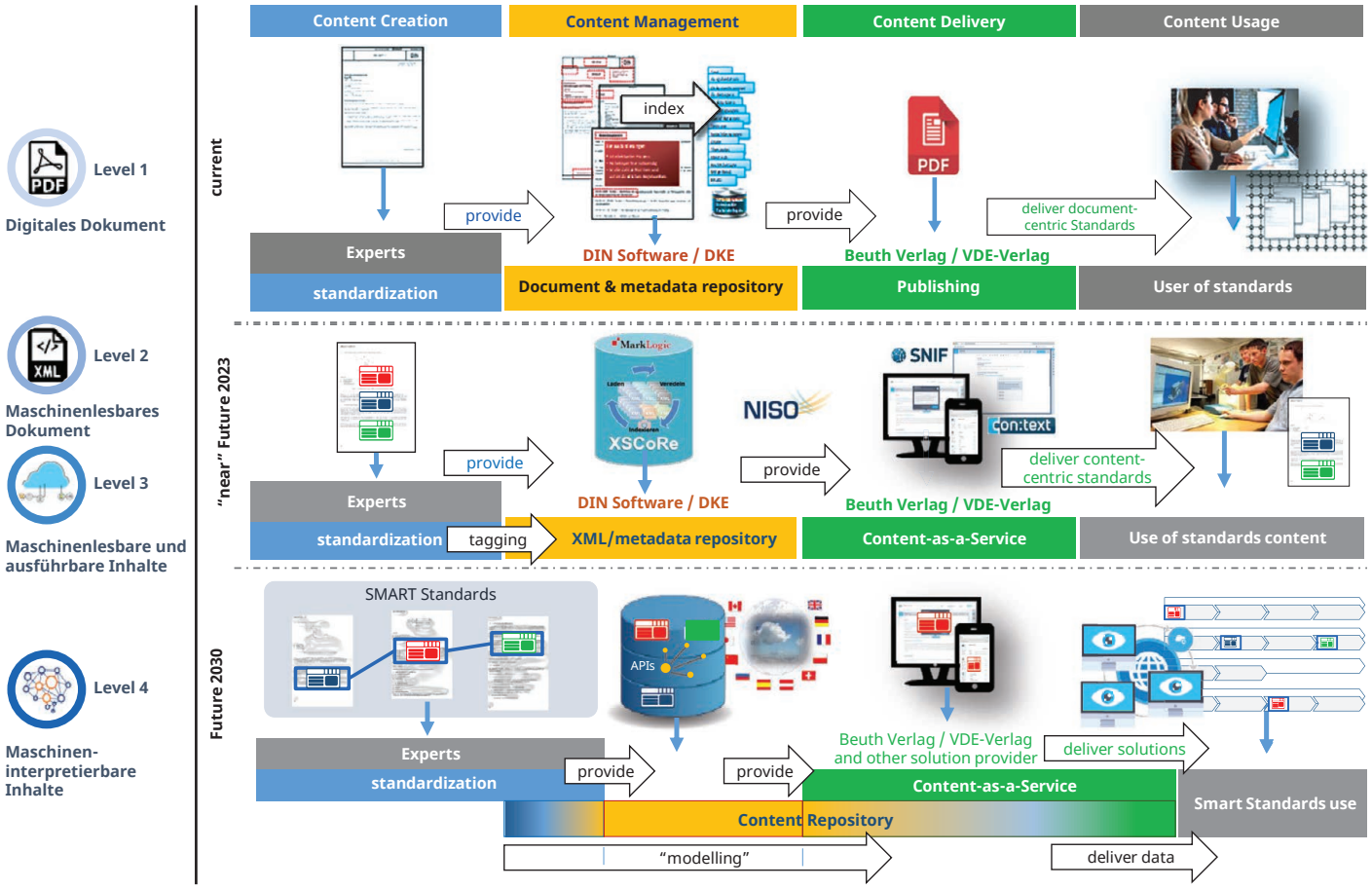


Abbildung 26: Stufenmodell SMART Standards

Level 4 repräsentiert den Endausbau einer durchgehenden SMART Standards Wertschöpfung, siehe Anhang 11.4.3.

Das Stufenmodell muss verifiziert und an andere Modelle adaptiert werden können, z. B. Referenzarchitekturmodell Industrie 4.0 (RAMI4.0) [261].

Die Schaffung einer offenen und konstruktiven Diskussionskultur bei der Einführung neuer Prozesse bedeutet auch: Man muss die Anwender (Normennutzer) und die Normungscommunity bei der Erarbeitung der Methodik frühzeitig und fortlaufend einbinden. Tatsächlich wird das Projekt SMART Standards auch „auf allen Ebenen" entsprechend national, europäisch und international begleitet und vorangetrieben. National entstanden bzw. entstehen folgende Aktivitäten rund um SMART Standards, z. B.:

→ **IDiS** (Initiative Digitale Standards von DIN/DKE):
SMART Standards of the Future (DIN):
https://din.one/site/sof
IDiS (DKE):
https://www.dke.de/de/normen-standards/digitalisierung-normung-digitalstrategie-dke-transformation

→ Zusammenarbeit mit **ANP** (Ausschuss Normenpraxis):
https://www.din.de/de/service-fuer-anwender/anp

→ Zusammenarbeit mit **BFA** (Benutzerfachausschuss der DIN Software GmbH):
https://www.dinsoftware.de/de/normen-management/benutzerfachausschuss

→ Mitarbeit im **NAGLN** (DIN-Normenausschuss Grundlagen der Normungsarbeit):
https://www.din.de/de/mitwirken/normenausschuesse/nagln

→ Zusammenarbeit mit verschiedenen Universitäten

In den o. g. Gremien wird ein wichtiger neuer Aspekt immer wieder thematisiert: Wie bereitet man die Handelnden auf die neuen Anforderungen vor?

Ein weiterer für die Zukunft systemrelevanter Aspekt betrifft die Definition der Anforderungen an die veränderten Qualifikationen der externen aber auch der DIN-internen „Akteure" im Gesamtprozess. Bestehende Konzepte müssen weiterentwickelt werden, um die in SMART-Standards-Prozessen neu entstehenden Aufgaben aller Prozessbeteiligten zu beschreiben.

5.2.4 Normen und KI

Eines der Ziele dieses Projekts ist die Ableitung von Regeln zur Formalisierung und Modellierung des Inhalts von Normen und Standards. Die daraus resultierenden Qualitätsverbesserungen der zugrunde liegenden Daten sind für die optimale Funktionalität von KI-Systemen unerlässlich. Die beabsichtigte Entwicklung eines zentralen Repositorys für die strukturierten Normdaten kann als Grundlage für qualitativ hochwertige KI-Anwendungen dienen, siehe Abbildung 27.

SMART Standards sind eine Wissensdomäne von vielen und ermöglichen es KI-Systemen grundsätzlich, die in ihnen enthaltenen Informationen automatisch und optimal in den verschiedenen Teilprozessen in einem Unternehmen zu nutzen.

Die Konzeption der notwendigen Datenmodelle und Schnittstellen wird Teil dieses Projekts sein müssen und leistet damit einen wichtigen Beitrag zur weiteren Durchdringung der KI-Anwendungen in den Teilprozessen von Unternehmen.

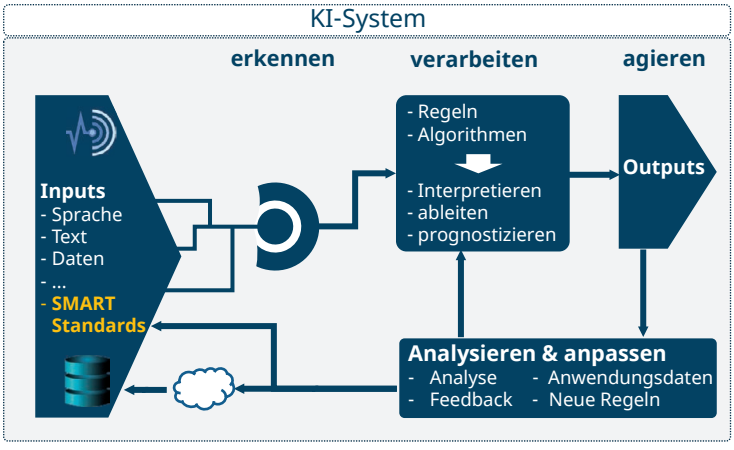


Abbildung 27: SMART Standards als Input für KI

Zahlreiche Eingangsformate für lernende/selbststeuernde Systeme, **darunter Normen**.

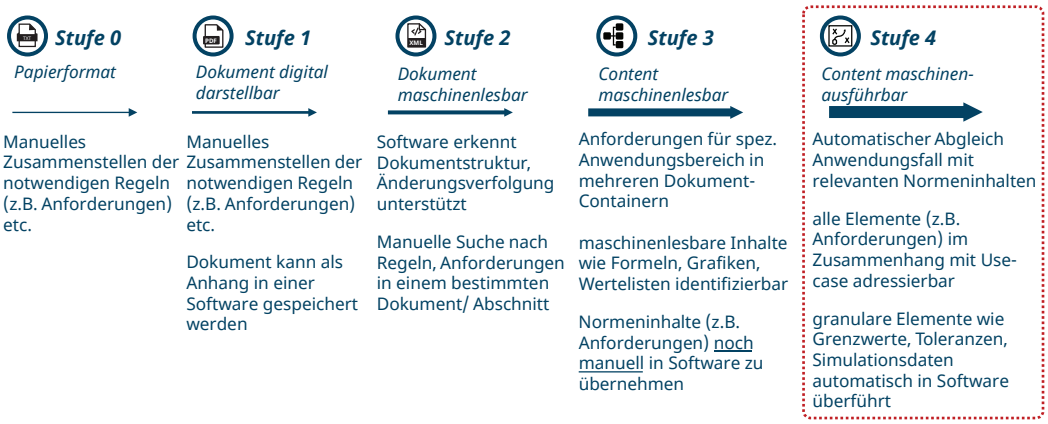
Normeninhalte müssen:

- **formalisiert/strukturiert**
- **eindeutig/fehlerfrei**
- **kontextbasiert/-sensibel**
- **granular adressierbar**

werden, um maschinell und durch Software ausführbar zu sein.

➔ **SMART Standards**

Abbildung 28: KI-Nutzungsmerkmale entsprechend „Utility Model“



Die verschiedenen KI-Nutzungsmerkmale im Stufenmodell (siehe Kapitel 5.2.3) sind in Abbildung 28 beispielhaft adaptiert. Im Projekt werden die Szenarien entsprechend Level 4 (siehe Anhang 11.4.3) anzustreben sein. Bereits Level 2 und 3 ermöglichen die Verwendung granularer Informationen, siehe Anhang 11.4.1.

5.2.5 Neugestaltung von Normen für KI-Anwendungsprozesse

Die Vorgehensweisen zur Bereitstellung granularer Normeninformationen werden unterschiedlich sein:

→ **Technologieansatz** (siehe Details im Anhang 11.4.1): Vorliegende Normendokumente werden im **Postprocessing** ohne Themen- und Anzahlbeschränkung maschinell indexiert und mittels semantischer Methoden automatisiert in granular „ansprechbaren“ Informationseinheiten bereitgestellt. Die Indexiergenauigkeit liegt derzeit bei ca. 80 Prozent gegenüber intellektuell granular aufbereiteten Dokumenten und genügt somit den Anforderungen qualifizierter Anwender, die das zerlegte Informationsangebot fachlich bewerten können. Für nachgelagerte KI-Anwendungsprozesse bedeutet das aber: Eine Validierung der Genauigkeit der Teilinformationen muss integriert werden. Der Treiber dieser Vorgehensweise sind das „Content Management“ und „Content Delivery“. Erreichbar sind **Ergebnisse im Level 3** (mit o. g. Einschränkungen) **auf der Grundlage von Level 2**.

→ **Bottom-up-Ansatz** (siehe Details im Anhang 11.4.2): Bei der Digitalisierung von Normen kann zwischen Top-down- und Bottom-up-Ansatz differenziert werden. Beide Ansätze beschäftigen sich mit Fragestellungen der Modularisierung, der Modellierung und dem Management zukünftiger Normeninhalte, allerdings aus unterschiedlichen Perspektiven. Hierbei wird der Top-down-Ansatz durch die Neugestaltung des eigentlichen Normungsprozesses und die Frage gekennzeichnet, wie zukünftige digitale Normen aufgebaut werden müssen, wohingegen sich der Bottom-up-Ansatz mit der Überführung von bereits existierenden Normeninhalten („Nachstrukturierung“) in eine maschinenausführbare Wissensrepräsentation befasst. Die Entwicklung smarter Standards bedarf sowohl einer methodischen Annäherung über einen Top-down- als auch einen Bottom-up-Ansatz. Die Treiber des Bottom-up-Ansatzes sind „Content Management und Delivery“ und das „Content Usage“. Erreichbar sind **Ergebnisse im Level 3 und Level 4** für definierte abgegrenzte Anwendungsgebiete.

→ **Top-down-Ansatz** (siehe Details im Anhang 11.4.3): Es kann nur ein Referenzdokument oder einen „Referenzinhalt“ der Norm geben und das ist der Inhalt, der vom verantwortlichen Normungsgremium geprüft und freigegeben wurde, der sogenannte Primärinhalt. Nur auf diesen beziehen sich in der Regel Gesetze oder Verträge und nur dieser Primärinhalt ist im Ernstfall relevant. Damit auch der maschinenlesbare Normeninhalt Primärinhalt sein kann, muss im **Preprocessing** (i. S. v. Normenentstehungsprozess) die Erfassung der menschengenerierten und -lesbaren sprachlichen Normeninhalte auf der Grundlage einer Struktur erfolgen, die es erlaubt, die Sprache, einschließlich der enthaltenen Semantik, eindeutig in eine maschinenlesbare Datenstruktur (z. B. Ontologie) zu überführen und umgekehrt. Die Treiber dieser Vorgehensweise sind das „Content Creation“ und das „Content Usage“. Erreichbar sind **Ergebnisse im Level 4**.

Bearbeitungsreihenfolge

Die unterschiedlichen Ansätze können und sollten parallel verfolgt werden. Der Technologieansatz liefert schnellere Erkenntnisse, die in den anderen Vorgehensweisen genutzt werden können. Darüber hinaus entstehen zügig erste – sich

wirtschaftlich rechnende – Kundenlösungen oder Prototypen und Demonstratoren, sodass praktische Erfahrungen rückgekoppelt werden können. Die Bottom-up-Vorgehensweise kann nicht geeignet sein, um den sehr großen und stetig wachsenden weltweiten Bestand an Normen nach höchsten Qualitätsansprüchen zu strukturieren. Aber auch gemäß dieser Vorgehensweise heißt es: Gezielt anfangen, um Erfahrungen zu sammeln. Das Postprocessing von Normen kann jedoch für konkrete Anwendungsfelder wirtschaftlich sein. Die „Königsklasse" für die avisierte Zielsetzung zur Erreichung von SMART Standards mit höchsten Qualitätsansprüchen für KI-Anwendungsprozesse kann nur die Verfolgung und Umsetzung einer Top-down-Methode (preprocessing) sein. Der Aufwand hierfür wird sehr hoch sein.

Wirtschaftlicher Nutzen

Der wirtschaftliche Nutzen der Standardisierung wird in einigen Ländern quantifiziert. In Deutschland erspart die Normung der Wirtschaft jährlich 17 Mrd. Euro [11]. In Frankreich trägt die Normung direkt zur Verbesserung des Bruttoinlandsprodukts bei, deren Effekt mit durchschnittlich über 5 Mrd. Euro pro Jahr beziffert wird. In Großbritannien können 28,4 Prozent des jährlichen BIP-Wachstums auf Normen zurückgeführt werden, das entspricht 9 Mrd. Euro. Die Bezifferung eines wirtschaftlichen Nutzens von SMART Standards liegt noch nicht vor und kann bisher nur qualitativ genannt werden, siehe Abbildung 29 [259].

Genau solch eine aggregierte Angabe (die zugegebenermaßen auch plakativ ist) und die Herleitung hierzu fehlen für den neuen Ansatz SMART Standards. Typische Fragestellungen sind: Welcher Nutzenanteil durch SMART Standards fällt auf die heute genannten 17 Mrd. Euro? Verlieren wir Nutzenanteile, wenn wir uns nicht mit SMART Standards beschäftigen? Oder kommt noch ein Nutzenanteil zu den 17 Mrd. Euro hinzu?

Im Rahmen des Projekts muss eine wirtschaftliche Bewertung bzgl. Aufwand, Nutzen, Realisierungszeitraum, Qualität etc. der verschiedenen Vorgehensweisen erfolgen. Danach oder projektbegleitend kann eine Priorisierung der Vorgehensweisen vorgenommen werden.

5.2.6 Zusammenfassung und Ausblick

Im vorliegenden Kapitel 5.2 „SMART Standards – Neugestaltung von Normen für KI-Anwendungsprozesse" wird eine Systematik zur Entwicklung von SMART Standards und Informationsmodellen zur Abbildung von Normen und Standards beschrieben. Beides liegt heute noch nicht vor – beides ist aber eine wichtige Voraussetzung, um KI-basierte Anwendungsmodelle mit zuverlässigen normungskonformen granularen Informationen zu versorgen.

Bei all dem, was anwenderseitig heute und zukünftig vorstellbar ist: Schlussendlich geht es darum, nachgelagert KI-Anwendungsprozesse zu entwickeln, an deren Ende Systeme entstehen, die in der Lage sind, auf Fragen des Anwenders oder der Anwendungen auf Grundlage **formalisierten und modellierten Fachwissens** aus Normen und Standards und daraus gezogener logischer Schlüsse Antworten zu liefern.

Im vorliegenden Beitrag werden zahlreiche Anforderungen mit KI-Relevanz beschrieben, die sich aus dem Projekt SMART Standards ableiten lassen. Die Umsetzung dieser Anforderungen hebt den Nutzen der Normung auf ein deutlich höheres Nutzungsniveau.

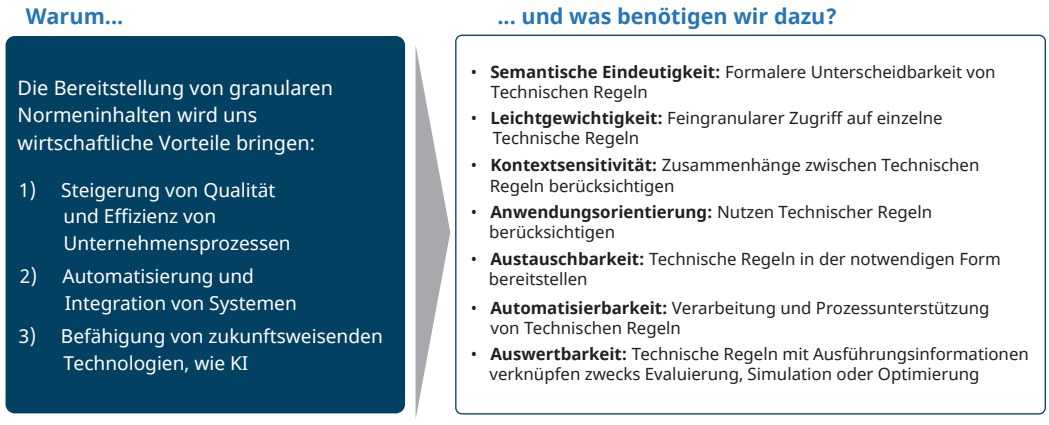


Abbildung 29: Voraussetzungen und Nutzen von SMART Standards

6 Übersicht über relevante Dokumente, Aktivitäten und Gremien zu KI

Dieses Kapitel gibt eine Übersicht über die wesentlichen Normen und Standards (6.1 und 6.2), laufende Normungs- und Standardisierungsaktivitäten (6.3) sowie Normungs- und Standardisierungsgremien (6.4). In den folgenden Tabellen sind die wichtigsten Dokumente gelistet und um zusätzliche Informationen ergänzt. Die Darstellungen erheben keinen Anspruch auf Vollständigkeit.

6.1 Veröffentlichte Normen und Standards zu KI

In Tabelle 10 werden bestehende Normen und Standards, die KI-Anwendungen explizit behandeln, aufgelistet. Weder die Tabelle insgesamt noch die Zuordnung zu den Schwerpunktthemen erheben Anspruch auf Vollständigkeit.

Tabelle 10: Bestehende Normen und Standards zu KI

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC TR 24028 [261]	KI-Grundlagen	Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence	Technischer Bericht zur Vertrauenswürdigkeit von KI-Systemen	x	x		x			x
ITU-T Y.3170 [154]	Dienstleistungsqualität bei Zukunftsnetzen	Requirements for machine learning – based quality of service assurance for the IMT-2020 Network	Anforderungen an Datensammlung, -aufbereitung und -modellierung mit Blick auf Quality of Service und Quality of Experience (siehe auch 4.3.1.3)			x				x
ITU-T Y.3173 [155]	Beurteilung der Fähigkeiten von Zukunftsnetzen	Framework for evaluating intelligence level of future networks including IMT-2020	Einschätzung von KI-Fähigkeiten bei Netzwerken (siehe auch 4.3.1.3)			x				x
ETSI TS 103 296 [152]	Emotionserkennung	Speech and Multimedia Transmission Quality (STQ); Requirements for Emotion Detectors used for Telecommunication Measurement Applications; Detectors for written text and spoken speech	Anforderungen und Eigenschaften mit Blick auf Emotionserkennung (siehe auch 4.3.1.3)		x	x				x

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ETSI TS 103 195-2 [153]	Netzwerk-architektur	Autonomic network engineering for the self-managing Future Internet (AFI); Generic Autonomic Network Architecture; Part 2: An Architectural Reference Model for Autonomic Networking, Cognitive Networking and Self-Management	Autonome Systeme: Anforderungen und Anwendungsfälle (siehe auch 4.3.1.3)			x				x
ETSI GR ENI 004 [263]	KI-Terminologie	Terminology for Main Concepts in ENI	Bericht (Group Report, GR) zur Terminologie von Netzwerken mit KI-Elementen im Kontext von Experiential Networked Intelligence (ENI)	x						
ETSI GR NFV 003 [264]	KI-Terminologie	Terminology for Main Concepts in NFV	Bericht (GR) zur Terminologie von Netzwerken mit KI-Elementen im Kontext von Network Functions Virtualisation (NFV)	x						
DIN SPEC 92001-1 [87]	Lebenszyklus von KI	Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 1: Quality Meta Model	→ konkreter Bezug zu KI : Qualitäts-Meta-Modell: Bezug zu ISO/IEC 12207 Lebenszyklusmodell → Unterscheidung in Risikoklassen: „low“ und „high risk“ KI-Module → Qualitätssäulen: Funktionalität und Performance, Robustheit und Verständlichkeit → KI-Qualität hängt ab von: Design des KI-Modells und Datenqualität → Risikomanagement entlang des gesamten Lebenszyklus wird empfohlen (siehe auch 4.1.2.3, 4.3.1.3, 4.4.2.3 und 4.7.2)		x	x	x			x

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
DIN SPEC 92001-2 [318]	Lebenszyklus von KI	Artificial Intelligence – Life Cycle Processes and Quality Requirements – Part 2: Robustness	KI-spezifische Anforderungen mit Blick auf Robustheit, insbesondere zu Adversarial Robustness und Corruption Robustness			x	x	x		
DIN SPEC 13266 [151]	Leitfaden für Deep Learning Systeme	Guideline for the development of deep learning image recognition systems	Vorgehen bei der Datensammlung über die Strukturierung der Daten zum Lernen der KI-Bilderkennung bis zur Ablaufstruktur von Lern-Experimenten und zur Qualitätssicherung (siehe auch 4.3.1.3 und 4.7.3)	x		x	x	x		x
IEEE 7010-2020 [156]	Einfluss von autonomen Systemen auf den Menschen	Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-being	Bewertungsschema von autonomen Systemen mit Blick auf Auswirkungen auf das menschliche Wohlbefinden (siehe auch 4.3.1.3)	x	x	x				x
UL 4600 [157]	Beurteilung autonomer Fahrzeuge	Standard for the Evaluation of Autonomous Products	Deckt Sicherheitsprinzipien, Risikominderung, Werkzeuge, Techniken und Life-Cycle-Prozesse für die Entwicklung und Evaluation autonomer Fahrzeuge ab. Kompatibel zu ISO/PAS 21448 und ISO 26262 (siehe auch 4.3.1.3 und 4.6.1)		x	x			x	

6.2 Veröffentlichte Normen und Standards mit Relevanz für KI

Tabelle 11 gibt einen Überblick über Normen und Standards, die noch keine detaillierten Aussagen zur Anwendung von KI-Komponenten machen, aber für die KI-Normung, -Standardisierung oder -Anwendung besonders relevant sind. Weder die Tabelle insgesamt noch die Zuordnung zu den Themen erheben Anspruch auf Vollständigkeit.

Tabelle 11: Allgemeine Normen und Standards mit Relevanz für KI-Anwendung

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC/ IEEE 12207 [58]	Lebenszyklus Software	Systems and software engineering – Software life cycle processes	→ Beschreibung von Prozessen des Lebenszyklus (Ideenfindung bis zur Stilllegung) und deren Beziehungen untereinander auf abstraktem Niveau → keine Festlegung eines Lebenszyklusmodells oder einer Entwicklungsmethode (siehe auch 4.1.2.3 und 4.3.1.2)	x	x	x		x		
ISO/IEC/ IEEE 29119 [265]–[269]	Softwaretests	Software Testing	29119-1: Konzepte und Definitionen (Concepts & Definitions) 29119-2: Testprozesse (Test Processes) 29119-3: Testdokumentation (Test Documentation) 29119-4: Testtechniken (Test Techniques) 29119-5: Keyword-Driven Testing (Keyword Driven Testing)			x		x		
ISO/IEC 15408 [48]–[50]	Sicherheitsverfahren	Information technology – Security techniques – Evaluation criteria for IT security	Definiert die Common Criteria (CC), 7 Vertrauenswürdigkeitsstufen (EAL), 11 Funktionsklassen, 7 organisatorische Klassen (siehe auch 4.1.2.2, 4.1.2.3). Teile 1 bis 3 sind veröffentlicht, Teile 3 bis 5 in Erarbeitung (siehe 6.3).		x	x	x	x		x

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC 17000ff [38]–[44]	Konformitäts-bewertung	Conformity assessment	Normenfamilie zur Konformitätsbewertung im Allgemeinen. Nicht KI-spezifisch, aber Grundlage für KI-Konformitätsbewertung (siehe auch 4.1.2.1.5 und vor allem 4.3)			x		x		x
ISO/IEC 18045 [51]	Sicherheits-verfahren	Information technology – Security techniques – Methodology for IT security evaluation	Methodik für die Bewertung der IT-Sicherheit auf Grundlage der CC („Evaluationsmethodologie“) (siehe 4.1.2.2 und 4.4.1.3)				x	x		
ISO/IEC 20546 [34]	Big Data	Information technology – Big data – Overview and vocabulary	Legt Begrifflichkeiten bzgl. Big Data fest (siehe auch 4.1.1 und 4.3.1.2)	x		x				
ISO/IEC TR 20547-1 [270]	Big Data	Information technology – Big data reference architecture – Part 1: Framework and application process	Referenzarchitektur für Big Data: Prozesse	x		x				
ISO/IEC TR 20547-2 [149]	Big Data	Information technology – Big data reference architecture – Part 2: Use cases and derived requirements	Referenzarchitektur für Big Data: Use Cases (siehe auch 4.3.1.2)	x		x				x
ISO/IEC 20547-3 [271]	Big Data	Information technology – Big data reference architecture – Part 3: Reference architecture	Referenzarchitektur für Big Data: Terminologie und Konzepte	x		x	x			
ISO/IEC TR 20547-5 [150]	Big Data	Information technology – Big data reference architecture – Part 5: Standards roadmap	Überblick über Standards, die für Big Data relevant sind (siehe auch 4.3.1.2)			x		x		x
ISO/IEC 25000 [272]	Softwarequalität	Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE	→ Leitfaden für Qualitätskriterien und die Bewertung von Softwareprodukten → Definition des SQuaRE-Modells			x	x	x		x

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC 25010 [146]	Softwarequalität	Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models	definiert Qualitätskriterien → Funktionalität: Korrektheit, Angemessenheit, Ordnungsmäßigkeit → Zuverlässigkeit: Reife, Fehlertoleranz, Wiederherstellbarkeit → Benutzbarkeit: Verständlichkeit, Bedienbarkeit, Erlernbarkeit, Robustheit → Effizienz: Wirtschaftlichkeit, Zeitverhalten, Verbrauchsverhalten → Wartungsfreundlichkeit: Analysierbarkeit, Änderbarkeit, Stabilität, Testbarkeit → Übertragbarkeit: Anpassbarkeit, Installierbarkeit, Konformität, Austauschbarkeit → Security: Integrität, Vertraulichkeit, Authentizität, Nachweisbarkeit, Haftung → Kompatibilität: Interoperabilität (können für die Erstellung der Spezifikationen und der Testfälle genutzt werden (siehe auch 4.3.1))			x	x	x		
ISO/IEC 25012 [89]	Datenqualität	Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model	Qualität des Datenprodukts: → inhärente Datenqualität (Genauigkeit, Vollständigkeit, Konsistenz, Glaubwürdigkeit, Aktualität, Zugänglichkeit,Konformität, Vertraulichkeit, Effizienz) → systemabhängige Datenqualität (Verfügbarkeit, Portabilität, Wiederherstellbarkeit, Genauigkeit, Rückverfolgbarkeit, Verständlichkeit) (siehe auch 4.1.2.3 und 4.3.1.2)			x	x			

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC 25020 [273]	Softwarequalität	Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality measurement framework	Richtlinien für Auswahl, Anwendung und Erstellung von Qualitätskennzahlen			x	x			
ISO/IEC 25021 [274]	Softwarequalität	Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality measure elements	Qualitätskriterien für Softwareentwicklung			x	x			
ISO/IEC 25024 [275]	Datenqualität	Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Measurement of data quality	Qualitätskriterien und Bewertung für sichere Softwareentwicklung			x	x			
ISO/IEC 27000ff [71]–[78], [122], [163], [210], [276]	Sicherheitsverfahren	Information technology – Security techniques	Normenfamilie zu Informationssicherheits-Managementsystemen (ISMS) mit einem Set von Unternormen zu den verschiedensten Themen – z. B. Leitfäden für ISMS Audit, Datensicherheit; ISMS für Gesundheitswesen etc. sowie u. a. → ISO/IEC 27034 sichere Softwareentwicklung (siehe auch 4.1.2.3 und 4.3.1.2) → ISO/IEC 27005 Risikomanagement (siehe auch 4.4.2.3) → ISO/IEC 27701 Ergänzung für personenbezogene Daten (privacy information management) (siehe auch 4.3.2.3.2.3)		x	x	x			

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC 29100ff [212], [277]	Sicherheits-verfahren	Information technology – Security techniques – Privacy framework	Normenfamilie zu Daten-schutz (Privacy framework), z. B. ISO/IEC 29134 Daten-schutzfolgeabschätzung (Risikobewertung)				x			
ISO/IEC 33063 [278]	Softwaretests	Information technology – Process assessment – Process assessment model for software testing	Richtlinien zur Definition und Beurteilung von Kriterien der Prozessfähigkeit in der Herstellung			x				
ISO 12100 [124], [125]	Maschinen-sicherheit	Safety of machinery – General principles for design – Risk assessment and risk reduction	Terminologie und Metho-dologie sowie allgemeine Leitsätze zur Risikobeurtei-lung und Risikominderung zur Herstellung sicherer Maschinen Im Kapitel 6: Aussagen zu Sicherheitsfunktionen, die durch programmierbare elektronische Steuerungen umgesetzt werden (siehe auch 4.2.2.4 und 4.3.1.2)		x	x	x			
ISO 13849 [126], [127]	Maschinen-sicherheit	Safety of machinery – Safety-related parts of control systems	Prinzipien der Gestaltung und Integration sicherheitsbezo-gener Teile von Steuerungen und programmierbarer elek-tronischer Systeme (siehe auch 4.2.2.4)		x	x	x			
ISO 14971 [128]	Risikomanagement	Medical devices – Application of risk management to medi-cal devices	Terminologie, Grundsätze und Prozess für das Risiko-management von Medizin-produkten, eingeschlossen Software als Medizinprodukt Beispiel für eine Norm, nach der sicherheitstechnisch relevante KI-Systeme derzeit ausgelegt werden, siehe auch Ethik (4.2.2.4)		x	x				x

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/PAS 21448 [148]	Sicherheit der Soll-Funktion	Road vehicles – Safety of the intended functionality	Sicherheit der Soll-Funktion (SOTIF) → betrachtet unangemessenes Risiko aufgrund von Gefährdungen durch Funktionsmängel der beabsichtigten Funktionalität oder durch vernünftigerweise vorhersehbaren Missbrauch durch Personen → Performance-Einschränkungen können auch der Umgebung und der Kommunikation zugeordnet werden (siehe auch 4.3.1.2)			x	x	x	x	
ISO/TR 22100 [279]–[281]	Dokumentenübersicht: Risikominderung bei Maschinen	Safety of machinery – Relationship with ISO 12100	Zielgerichtete Auswahl der verschiedenen Typen von ISO-Normen zur Maschinensicherheit 4 Teile veröffentlicht, weitere in Erarbeitung (siehe 6.3)			x	x	x		
ISO 23412 [282]	Logistik	Indirect, temperature-controlled refrigerated delivery services – Land transport of parcels with intermediate transfer	Konzentriert sich auf die technische und organisatorische Umsetzung des Transports von gekühlter Ware, kann jedoch als Grundstein für die automatische Warendistribution gesehen werden und ist deshalb relevant für KI-Anwendung (siehe auch 4.6.1)						x	
ISO 25119 [283]–[286]	Funktionale Sicherheit	Tractors and machinery for agriculture and forestry – Safety-related parts of control systems	Sicherheit bei Design, Entwicklung, Konzipierung und Produktion			x	x		x	x

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO 26262 [59]–[70]	Funktionale Sicherheit	Road vehicles – Functional safety	Management der funktionalen Sicherheit → Konzeptphase → Produktentwicklung: Systemebene → Produktentwicklung: Hardwareebene → Produktentwicklung: Softwareebene → Produktion, Betrieb und Außerbetriebnahme → Unterstützende Prozesse → ASIL- und sicherheitsorientierte Analysen (siehe auch 4.1.2.3)			x	x		x	
ISO 31000 [93]	Risikobewertung	Risk management – Guidelines	Allgemeine, nicht KI-spezifische Leitlinien zum Risikomanagement. Ansatz für das Behandeln jeglicher Art von Risiko, nicht industrie- oder sektorspezifisch. Basis für ISO/IEC 23894 zum Risikomanagement für KI. (siehe auch 4.1.3, 4.2.2.4, 4.4.2.3)		x	x	x			x
IEC 60601 1-4 [287]	Medizinprodukte	Medical electrical equipment – Part 1-4: General requirements for safety – Collateral Standard: Programmable electrical medical systems	Anforderungen an Sicherheit, Prüfungen und Richtlinien für programmierbare elektrische medizinische Systeme		x	x				x
IEC 61508 [79]–[86]	Funktionale Sicherheit von Systemen	Functional safety of electrical/electronic/programmable electronic safety-related systems	IEC 61508-3: Anforderungen an Software: Künstliche Intelligenz für die Fehlerkorrektur ab SIL 2 ausdrücklich nicht empfohlen IEC 61508-5: Beispiele zur Ermittlung der Stufe der Sicherheitsintegrität (safety integrity level) (siehe auch 4.1.2.3, 4.3.1.2, 4.4.2.3 und 4.5.2.3)	x		x	x			x

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
IEC 61511 [211], [288], [289]	Funktionale Sicherheit von Prozessleittechnik	Functional safety – Safety instrumented systems for the process industry sector	Teil 1: Allgemeines, Begriffe, Anforderungen an Systeme, Software und Hardware (siehe auch 4.4.2.3)			x	x			
IEC 61513 [290]	Anforderungen an leittechnische Systeme und Geräte	Nuclear power plants – Instrumentation and control important to safety – General requirements for systems	Konzept des Sicherheitslebenszyklus für die gesamte leittechnische Architektur sowie einzelne Systeme		x	x	x			
IEC 62061 [129]	Funktionale Sicherheit von Steuerungssystemen	Safety of machinery – Functional safety of safety-related electrical, electronic and programmable electronic control systems	Auswahl und Entwurf eines sicherheitsbezogenen elektrischen, elektronischen und programmierbaren elektronischen Steuerungssystems (SRECS) und Ansatz zu Risikoabschätzung und Festsetzung des Sicherheits-Integritätslevels (SIL) (siehe auch 4.2.2.4)		x	x	x			
IEC 62304 [291]	Software-Lebenszyklus	Medical device software – Software life cycle processes	Entwicklung und Wartung von Medizinprodukte-Software		x	x	x			x
IEC 62443 [199]–[209]	IT-Sicherheit	Industrial communication networks – Network and system security	Normenreihe in 13 Subteilen definiert u. a. Terminologie (IEC TS 62443-1-1) und Anforderungen, z. B. an das IT-Sicherheitsprogramm von Dienstleistern (IEC 62443-2-4), den Lebenszyklus für sichere Produktentwicklung (IEC 62443-4-1) und Security-Level (IEC 62443-3-3) (siehe auch 4.4.1.3)				x			
DIN EN 50128 [292]	Sicherheitsrelevante Software der Eisenbahn	Railway applications – Communication, signalling and processing systems – Software for railway control and protection systems	Verfahren, Prinzipien und Maßnahmen für Softwaresicherheit			x	x		x	

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ETSI TR 101 583 [175]	Sicherheitstests	Methods for Testing and Specification (MTS); Security Testing; Basic Terminology	Aufzählung und Erläuterung relevanter Methoden und Ansätze für Sicherheitstests wie beispielsweise Risikoanalyse und risikoadaptiertes Sicherheitstesten, funktionales Testen von Sicherheitsfunktionen, Performancetesten, Robustheitstesten und Penetrationstesten (siehe auch 4.3.2.3.2.4		x	x	x	x		
IEEE 1012-2016 [293]	Validierung von Hard- und Software	Standard for System, Software, and Hardware Verification and Validation	Überprüfung, ob im Zuge der Produktentwicklung Anforderungen erfüllt werden			x		x		

6.3 Laufende Normungs- und Standardisierungsaktivitäten zu KI

Tabelle 12 gibt Informationen zu laufenden Normungs- und Standardisierungsprojekten. Weder die Tabelle insgesamt noch die Zuordnung zu den Themen erheben Anspruch auf Vollständigkeit.

Tabelle 12: Überblick über laufende KI-relevante Normungs- und Standardisierungsprojekte

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/ IEC WD TS 4213	Bewertung von KI-Systemen	Information technology – Artificial Intelligence – Assessment of machine learning classification performance	Metriken zur Leistungsfähigkeit von KI	x		x		x		
ISO/IEC NP 5059	Softwarequalität	Software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) Quality Model for AI-based systems	Feststellung von Qualität für Systeme auf KI-Basis (siehe auch 4.1.1 und 4.3.1.4)	x		x		x		
ISO/IEC WD 5259-1	Datenqualität	Data quality for analytics and ML – Part 1: Overview, terminology, and examples	Datenqualitätsmanagement für maschinelles Lernen: Überblick, Terminologie und Beispiele	x		x		x		
ISO/IEC WD 5259-3	Datenqualität	Data quality for analytics and ML – Part 3: Data Quality Management Requirements and Guidelines	Datenqualitätsmanagement für maschinelles Lernen: Anforderungen und Richtlinien	x		x		x		
ISO/IEC WD 5259-4	Datenqualität	Data quality for analytics and ML – Part 4: Data quality process framework	Datenqualitätsmanagement für maschinelles Lernen: Prozesse	x		x		x		
ISO/IEC NP 5338	Entwicklung von KI-Systemen	Information technology – Artificial intelligence – AI system life cycle processes	Terminologiestandard zu Lebenszyklusprozessen von KI-Systemen (in Abstimmung)	x		x	x	x		
ISO/IEC NP 5339	Anwendungsrichtlinien	Information Technology – Artificial Intelligence – Guidelines for AI Applications	Richtlinien zur Anwendung von KI-Systemen (in Abstimmung)	x	x	x		x		

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC NP 5392	KI-Systeme	Information technology – Artificial intelligence – Reference Architecture of Knowledge Engineering	Referenzarchitektur für wissensbasierte Systeme	x		x	x	x		
ISO/IEC NP 5394	KI-Grundlagen	Information Technology – Artificial intelligence – Management System	Managemementsystem-standard für KI (vgl. 4.1.2.2)	x		x	x	x		
ISO/ IEC AWI TR 5469	Funktionale Sicherheit und KI	Functional Safety and AI Systems	Das Dokument soll Eigenschaften, relevante Risikofaktoren, verwendbare Methoden und Prozesse zur Kontrolle von KI-Systemen und zur Anwendung von KI in der Entwicklung von sicherheitsrelevanten Funktionen beschreiben. Es wird in Zusammenarbeit mit IEC/SC 65 A (der Standardisierungsgruppe, die für IEC 61508 verantwortlich ist) erstellt werden.			x	x	x		
ISO/IEC 15408	Sicherheitsverfahren	Information technology – Security techniques – Evaluation criteria for IT security	Definiert die Common Criteria (CC), 7 Vertrauenswürdigkeitsstufen (EAL), 11 Funktionsklassen, 7 organisatorische Klassen (siehe auch 4.1.2.2, 4.4.2.3). Teile 1 bis 3 sind veröffentlicht (siehe 6.2), Teile 4 und 5 in Erarbeitung				x	x		
ISO/ IEC FDIS 20547-4	Big Data	Information technology – Big data reference architecture – Part 4: Security and privacy	Referenzarchitektur für Big Data			x	x	x		x
ISO/IEC CD 22989	KI-Terminologie	Artificial intelligence – Concepts and terminology	Grundlagennorm, beschreibt Konzepte und Terminologie der Künstlichen Intelligenz (siehe auch 4.1.1 und 4.6.2.1)	x		x		x		

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC CD 23053	Maschinelles Lernen	Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)	Beschreibt ein begriffliches Rahmenwerk für ML (siehe auch 4.1.1)	x				x		
ISO/IEC CD 23894	KI-Risiko-management	Information Technology – Artificial Intelligence – Risk Management	Enthält Richtlinien für das Risikomanagement für die Entwicklung und Nutzung von KI-Systemen (siehe auch 4.1.1, 4.1.3, 4.2.3, 4.4.2.3)	x		x	x	x		x
ISO/ IEC AWI TR 24027	KI-Grundlagen	Information technology – Artificial Intelligence (AI) – Bias in AI systems and AI aided decision making	Technischer Bericht zur Beschreibung von „Bias“ in KI-Systemen (siehe auch 4.1.1)	x	x			x		
ISO/IEC NP 24029	KI-Robustheit	Artificial Intelligence (AI) – Assessment of the robustness of neural networks	ISO/IEC CD TR 24029-1: Overview ISO/IEC AWI 24029-2: Formal methods methodology (siehe auch 4.5.2.3)			x		x		x
ISO/ IEC CD TR 24030	Anwendungen	Information technology – Artificial Intelligence (AI) – Use cases	Sammlung von Nutzungs-szenarien für KI-Systeme	x		x	x	x	x	
ISO/ IEC AWI TR 24368	Ethik	Information technology – Artificial intelligence – Over-view of ethical and societal concerns	Technischer Bericht zu ethi-schen und gesellschaftlichen Problemstellungen der KI (siehe auch 4.1.1)	x		x				
ISO/ IEC AWI TR 24372	KI-Grundlagen	Information technology – Artificial intelligence (AI) – Overview of computational approaches for AI systems	Technischer Bericht zu den Methoden der KI	x		x		x		
ISO/ IEC WD TS 24462	Vertrauens-würdigkeit	Ontology for ICT Trustworthiness Assessment	Neues Projekt für eine Technische Spezifikation. Bearbeitet in ISO/IEC JTC 1/WG 13 „Trustworthiness“			x	x	x		

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ISO/IEC AWI 24668	Big Data	Information technology – Artificial intelligence – Process management framework for Big data analytics	Management für Datenanalysen im Bereich Big Data	x		x				
ISO/IEC CD 38507	Governance	Information technology – Governance of IT – Governance implications of the use of artificial intelligence by organizations	Behandelt organisatorische Governance im Zusammenhang mit KI (siehe auch 4.1.1)	x		x				
ISO/SAE 21434	IT-Sicherheit	Road vehicles – Cybersecurity engineering	Definiert Terminologie und wichtigste Aspekte der Informationssicherheit				x			
ISO/CD TR 4804	IT-Sicherheit	Road vehicles — Safety and cybersecurity for automated driving systems — Design, verification and validation methods	Arbeit zu KI in Straßenfahrzeugen. Wird in ISO/TC 22 „Road vehicles“ bearbeitet (siehe auch 4.6.1)				x	x	x	
ISO/CD TR 22100-5	Maschinensicherheit und KI	Safety of machinery – Relationship with ISO 12100 – Part 5: Implications of embedded Artificial Intelligence-machine learning	Der Technische Bericht beschreibt, wie Gefahren, die mit dem Einsatz von ML-Systemen in Maschinen entstehen, im Prozess der Risikobewertung berücksichtigt werden sollen			x	x	x		
ISO/AWI 24089	IT-Sicherheit	Road vehicles – Software update engineering	neuer Standard – in Bearbeitung			x	x	x		
ITU-T F.AI-DLFE	Beurteilung von Software auf Basis von tiefem Lernen	Deep Learning Software Framework Evaluation Methodology	Anforderungen an Architekturen des tiefen Lernens			x		x		
ITU-T F.AI-DLPB	Metriken und Beurteilung neuronaler Netze	Metrics and evaluation methods for deep neural network processor benchmark	Bewertungsschema für tiefes Lernen mit Blick auf Inferenz, Training, Anwendung, Netzwerk und Prozessor			x		x		

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
ITU-T F.VS-AIMC	Anforderungen für die Datenübertragung	Use cases and requirements for multimedia communication enabled vehicle systems using artificial intelligence	Netzwerkanforderungen mit Blick auf Vorhersagen, Planung, Mensch-Maschine-Interaktion und Training von Modellen			x		x		x
ITU-T Y.qos-ml-arc	Dienstleistungsqualität bei Zukunftsnetzen	Architecture of machine learning based QoS assurance for the IMT-2020 network	Dienstleistungsqualität bei Zukunftsnetzen mit Blick auf maschinelles Lernen			x		x		
ETSI DGS SAI 003	KI-Sicherheitstests	Securing Artificial Intelligence (SAI); Security Testing of AI	Richtlinien für Sicherheitstests von KI-Komponenten. Fokus liegt auf Daten für maschinelles Lernen			x	x	x		
ETSI DGR SAI 002	KI-Trainingsdatenqualität	Securing Artificial Intelligence (SAI); Data Supply Chain Report	Übersicht über vorhandene Verfahren zur Datengewinnung, Regeln für Datenhandhabung. Identifikation von Normungsbedarf			x	x	x		
ETSI DTR INT 008 (TR 103 821)	KI-Tests	Artificial Intelligence (AI) in Test Systems and Testing AI models: Testing of AI, with Definitions of Quality Metrics	Testframework für Systeme der Netzwerkautomatisierung wie z. B. ETSI GANA (Generic Autonomic Networking Architecture)			x		x		
IEEE P2801	Datenqualität	Recommended Practice for the Quality Management of Datasets for Medical Artificial Intelligence	QM-System für Datenaufbereitung bei KI-Medizinprodukten			x		x		x
IEEE P2802	Terminologie: Beurteilung von KI-Sicherheitsaspekten	Standard for the Performance and Safety Evaluation of Artificial Intelligence Based Medical Device: Terminology	Sicherheit, Risiko, Effektivität und QM			x	x	x		x
IEEE P2846	Mobilität	A Formal Model for Safety Considerations in Automated Vehicle Decision Making	Technologieneutrales mathematisches Modell und Prüfverfahren für automatische Entscheidungsfindung in Fahrzeugen (siehe 4.6.1)			x			x	

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
IEEE P3333.1.3	Beurteilung auf Basis von tiefem Lernen	Standard for the Deep Learning-Based Assessment of Visual Experience Based on Human Factors	Beurteilung der subjektiven und objektiven Nutzerzufriedenheit über tiefes Lernen	x		x		x		x
DIN SPEC 2343	Datenübertragung	Transmission of language-based data between artificial intelligences – Specification of parameters and formats	Format für Übertragung von Sprachdaten zwischen verschiedenen Ökosystemen für Industrieanwender, Open Source Communities und Privatanwender mit Fokus auf Interoperabilität und Nachvollziehbarkeit			x		x		
DIN SPEC 91426	Videoanalyse	Quality requirements for video-based methods of personnel selection	Vorgehen, um Fehler zu vermeiden, Diskriminierung zu verhindern und prognostische Validität der digitalen Einstellungsverfahren zu erhöhen			x		x		
NISTIR 8269	IT-Sicherheit	A Taxonomy and Terminology of Adversarial Machine Learning	Die Taxonomie ordnet verschiedene Arten von Angriffen, Verteidigungen und Konsequenzen. Die Terminologie definiert Schlüsselbegriffe im Zusammenhang mit der Sicherheit von ML in KI-Systemen				x	x		

Dokument	Thema	Titel	Kurzbeschreibung mit möglicher Relevanz zu KI	Relevanz für Schwerpunktthema						
				Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
VDE AR 2842-61	Vertrauens-würdigkeit	Development and trustworthiness of autonomous/cognitive systems	Legt einen allgemeinen Rahmen für die Entwicklung vertrauenswürdiger Lösungen und vertrauenswürdiger autonomer/kognitiver Systeme fest, einschließlich der Anforderungen an die nachfolgenden Phasen des Produktlebenszyklus (z. B. Produktion, Marketing & Vertrieb, Betrieb & Wartung, Ausmusterung & Reparatur). Definiert einen Referenz-Lebenszyklus in Analogie zu den wichtigsten Normen für funktionale Sicherheit (d. h. IEC 61508) als einen einheitlichen Ansatz, um die Gesamtleistung der Lösung und das beabsichtigte Verhalten und die Vertrauenswürdigkeit des autonomen/kognitiven Systems zu erreichen und aufrechtzuerhalten. Darüber hinaus könnte dies zu einer Grundlage für die Qualifizierung und Konformitätsbewertung von Lösungen führen, die auf autonomen/kognitiven Systemen einschließlich Elementen der Künstlichen Intelligenz basieren (siehe auch 4.5.1)		x	x	x	x	x	x

6.4 Gremien zu KI

Die folgende Tabelle 13 gibt einen Überblick über die wichtigsten KI-Normungs- und -Standardisierungsgremien. Weder die Tabelle insgesamt noch die Zuordnung zu den Themen der AGs (vgl. den einleitenden Absatz zu Kapitel 4) erheben Anspruch auf Vollständigkeit.

Tabelle 13: Überblick über die wichtigsten KI-Normungs- und -Standardisierungsgremien

	Gremium	Spiegelgremium[a]	Relevanz für Schwerpunktthema						
			Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
International	ISO/IEC JTC 1/SC 7 „Software and system engineering“	NA 043-01-07 AA	x		x		x		
	ISO/IEC JTC 1/SC 27 „Information security, cybersecurity and privacy protection“	NA 043-01-27 AA	x	x	x	x	x		
	ISO/IEC JTC 1/SC 29 „Coding of audio, picture, multimedia and hypermedia information“	NA 043-01-29 AA					x		x
	ISO/IEC JTC 1/SC 31 „Automatic identification and data capture“	NA 043-01-31 AA					x		
	ISO/IEC JTC 1/SC 37 „Biometrics“	NA 043-01-37 AA					x		x
	ISO/IEC JTC 1/SC 38 „Cloud manangement and distributed platforms“	NA 043-01-37 AA					x		
	ISO/IEC JTC 1/SC 40 „IT Service Management and IT Governance“	NA 043-01-40 AA	x		x		x		
	ISO/IEC JTC 1/SC 41 „Internet of things and related technologies“	NA 043-01-41 AA			x		x		
	ISO/IEC JTC 1/SC 42 „Artificial Intelligence“	NA 043-01-42 AA	x	x	x	x	x	x	
	ISO/TC 199 „Safety of machinery“	NA 095 BR	x	x	x		x		x
	ISO/TC 204 „Intelligent transport systems“/AG 1 „Big data and artificial intelligence“	NA 052-00-71 GA	x		x		x		
	ISO/TC 299 „Robotics“	NA 060-38-01 AA	x		x		x		x
	IEC SEG 10 „Ethics in Autonomous and Artificial Intelligence Applications“	DKE/TBINK AG			x		x		
	IEC/TC 62 „Electrical equipment in medical practice“/AG SNAIG „Software Network and Artificial Intelligence advisory Group“				x		x		

	Gremium	Spiegel-gremium[a]	Relevanz für Schwerpunktthema						
			Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
	IEC/TC 65 „Industrial process measurement – control and automation“/SC 65A „System aspects“ (Liaison mit ISO/IEC JTC 1/SC 42)				x	x	x		
	IEC/TC 65 JWG23 „Usage of new Technologies“						x		
	IEC/TC 65/WG 23 „Smart Manufacturing Framework and System Architecture“						x		
	ITU-T FG – AI4AD „Focus Group on AI for autonomous and assisted driving“				x		x	x	
	ITU-T FG – AI4EE „Focus Group on Environmental Efficiency for Artificial Intelligence and other Emerging Technologies“		x				x		
	ITU-T FG – AI4H „Focus Group on Artificial Intelligence for Health“		x		x		x		x
	ITU-T FG – ML5G „Focus Group on Machine Learning for Future Networks including 5G“		x		x		x		x
	ITU-T SG 2 „Operational aspects“				x		x		
	ITU-T SG 5 „Environment and circular economy“		x				x		
	ITU-T SG 13 „Future networks (& cloud)“		x				x		
	ITU-T SG 16 „Multimedia“		x				x		
	ITU-T SG 20 „IoT, smart cities & communities“		x		x		x		
Europäisch	CEN-CENELEC Focus Group on Artificial Intelligence	NA 043-01-42 AA	x		x		x		
	CEN/CLC/JTC 13 „Cybersecurity and Data Protection“	NA 043 BR-07 SO				x	x		x
	ETSI ISG ENI „Experiential Network Intelligence“				x		x		
	ETSI ISG NFV „Network Function Virtualisation“				x		x		
	ETSI ISG SAI „Securing Artificial Intelligence“		x		x	x	x		x
	ETSI ISG ZSM „Zero touch network & Service Management“		x				x		
	ETSI TC STQ „Speech and Multimedia Transmission Quality“		x		x		x		
	ETSI TC Cyber „Cybersecurity“					x	x		

	Gremium	Spiegel-gremium[a]	Relevanz für Schwerpunktthema						
			Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
National	DIN NA 043-01 FB „Fachbereich Grundnormen der Informationstechnik“ (mit Spiegelgremien zu ISO/IEC JTC 1 wie oben aufgelistet), darunter NA 04-01-42 AA „Künstliche Intelligenz“		x		x		x		
	DKE/AK 801.0.8 „Spezifikation und Entwurf autonomer/kognitiver Systeme“				x		x		
	DKE/AK 914.0.11 „Funktionale Sicherheit und künstliche Intelligenz“				x	x	x		
	DKE/TBINK AG „Ethik und Künstliche Intelligenz“				x		x		x
	DIN SPEC 2343 „Übertragung von sprachbasierten Daten zwischen Künstlichen Intelligenzen – Festlegung von Parametern und Formaten“				x		x		
	DIN SPEC 13266 „Leitfaden für die Entwicklung von Deep-Learning-Bilderkennungs-systemen“				x		x		x
	DIN SPEC 92001 „Künstliche Intelligenz – Life Cycle Prozesse und Qualitätsanforderungen“		x		x	x	x		
	DIN SPEC 91426 „Qualitätsanforderungen für video-basierte Methoden der Personalauswahl“				x		x		
Konsortien	IETF NMRG „Network Management Research Group“		x						
	IETF COIN „Computing in the Network Proposed Research Group“		x						
	IETF ICNRG „Information-Centric Networking Research Group“		x						
	IETF TSVWG „Transport Area Working Group“		x						
	IEEE AIMDWG „Artificial Intelligence Medical Device Working Group“		x		x				x
	IEEE P7006 „Standard for Personal Data Artificial Intelligence (AI) Agent“				x				
	IEEE P7010 „Well-being Metrics for Autonomous and Intelligent Systems“		x		x				
	IEEE P7014 „Emulated Empathy in Autonomous and Intelligent Systems“								
	IEEE P3652.1 „Guide for Architectural Framework and Application of Federated Machine Learning“				x				
	IEEE P2841 „Framework and Process for Deep Learning Evaluation“				x				

Gremium	Spiegel-gremium[a]	Relevanz für Schwerpunktthema						
		Grundlagen (4.1)	Ethik/Responsible AI (4.2)	Qualität, Konformitätsbewertung und Zertifizierung (4.3)	IT-Sicherheit bei KI-Systemen (4.4)	industrielle Automation (4.5)	Mobilität und Logistik (4.6)	KI in der Medizin (4.7)
IEEE P3333.1.3 „Deep Learning-Based Assessment of Visual Experience Based on Human Factors“				x				x
IEEE P2807 „Framework of Knowledge Graphs“				x				
CSA „Working Group Artificial Intelligence“				x				
OGC „Artificial Intelligence in Geoinformatics Domain Working Group“								
OMG „Artificial Intelligence Platform level Task Force“				x				
W3C AI KR „Artificial Intelligence Knowledge Representation“				x				

[a] Titel der genannten DIN-Normenausschüsse:

NA 043 „DIN-Normenausschuss Informationstechnik und Anwendungen (NIA)“

NA 052 „DIN-Normenausschuss Automobiltechnik (NAAutomobil)“

NA 060 „DIN-Normenausschuss Maschinenbau (NAM)“

NA 095 „DIN-Normenausschuss Sicherheitstechnische Grundsätze (NASG)“

NA 147 „DIN-Normenausschuss Qualitätsmanagement, Statistik und Zertifizierungsgrundlagen (NQSZ)“

NA 175 „DIN-Normenausschuss Organisationsprozesse (NAOrg)“

7 Abkürzungsverzeichnis

Abkürzung	Bedeutung
5G	Funkstandard der fünften Generation
AA	Arbeitsausschuss
ACLU	American Civil Liberties Union
ACM	Association for Computing Machinery
ADM	Automated decision making/Algorithm decision making
AG	Arbeitsgruppe
AI	Artificial intelligence
AK	Arbeitskreis
ALKS	Automated Lane Keeping System
AML	Adversarial Machine Learning
AR (VDE-)	VDE Anwendungsregel
ASIL	Automotive Savety Intergrity Level
Bitkom	Bundesverband Informationswirtschaft, Telekommunikation und neue Medien
BLEU	Bi-Lingual Evaluation Understudy
BMAS	Bundesministerium für Arbeit und Soziales
BMBF	Bundesministerium für Bildung und Forschung
BMI	Bundesministerium des Innern, für Bau und Heimat
BMVI	Bundesministerium für Verkehr und digitale Infrastruktur
BMWi	Bundesministerium für Wirtschaft und Energie
BR	Beirat
BSI	Bundesamt für Sicherheit in der Informationstechnik
BSI-KritisV	Verordnung zur Bestimmung Kritischer Infrastrukturen nach dem BSI-Gesetz
BVDW	Bundesverband Digitale Wirtschaft
CC	Common Criteria

Abkürzung	Bedeutung
CD	Committee Draft
CE	CE-Kennzeichnung (Conformité Européenne)
CEN	Comité Européen de Normalisation (Europäisches Komitee für Normung)
CENELEC	Comité Européen de Normalisation Électrotechnique (Europäisches Komitee für elektrotechnische Normung)
CLC	Abkürzung für CENELEC, z. B. bei Gremienbezeichnungen
COM	Mitteilung der EU-Kommission
CSA	Cloud Security Alliance
DARPA	Defense Advanced Research Project Agency
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz
DGR	Draft Group Report
DGS	Draft Group Specification
DIHK	Deutscher Industrie- und Handelskammertag
DIN	Deutsches Institut für Normung e. V.
DIN SPEC	DIN Spezifikation (Konsortialstandard)
DKE	DKE Deutsche Kommission Elektrotechnik Elektronik Informationstechnik in DIN und VDE
DSGVO	Datenschutz-Grundverordnung, Verordnung (EU) 2016/679
DSK	Datenschutzkonferenz
EAL	Evaluation Assurance Level
E/E/PE	Elektrisch, elektronisch, programmierbar elektronisch
ENI	Experiential Networked Intelligence
ENISA	European Network and Information Security Agency
ETSI	European Telecommunications Standards Institute

Abkürzung	Bedeutung
EU	Europäische Union
FG	Focus Group
FMECA	Failure Mode and Effects and Critical Analysis
GA	Gemeinschaftsarbeitsausschuss
GAIA-X	Projekt zum Aufbau einer leistungs- und wettbewerbsfähigen, sicheren und vertrauenswürdigen Dateninfrastruktur für Europa
GMA	VDI/VDE-Gesellschaft Mess- und Automatisierungstechnik
GR	Group Report
HLEG-KI	Hochrangige Expertengruppe für KI
HR	Human Resources
IAIS	Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme
IBM	International Business Machines Corporation
ICT	Information and Communications Technology
IDC	International Data Corporation
IACS	Industrial Automation and Control Systems
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IIRA	Industrial Internet Reference Architecture
IKT	Informations- und Kommunikationstechnik
IMT-2020	International Mobile Telecommunications-2020
IoT	Internet of Things
ISMS	Informationssicherheits-Managementsysteme
ISO	Internationale Organisation für Normung
IT	Informationstechnik
IT-SiG	IT-Sicherheitsgesetz

Abkürzung	Bedeutung
ITU	International Telecommunication Union
ITU-T	ITU Telecommunication Standardization Sector
JTC	Joint Technical Committee
JWG	Joint Working Group
KBS	Konformitätsbewertungsstelle
KI	Künstliche Intelligenz
KMU	Kleine und mittlere Unternehmen
KRITIS	Kritische Infrastruktur im Sinne der BSI-KritisV
LIME	Local Interpretable Model-Agnostic Explanations
MIP	Mixed Integer Linear Programming
ML	Machine Learning/Maschinelles Lernen
MLP	Multi-Layer Perceptron
MPG	Medizinproduktegesetz
MSS	Managementsystemstandard
MÜ	Maschinelle Übersetzung
NA	Normenaussschuss
NASA	National Aeronautics and Space Administration
NFV	Network Function Virtualisation
NIST	National Institute of Standards and Technology, U.S. Department of Commerce
NQDM	Normentwurf für qualitativ hochwertige Daten und Metadaten (Fraunhofer-Leitfaden)
NRM	Normungsroadmap
OECD	Organisation for Economic Co-operation and Development
OGC	Open Geospatial Consortium
OMG	Object Management Group
OPC	Open Platform Communications

Abkürzung	Bedeutung
OT	operational IT
OWL	Ontology Web Language
PDW	Prinzip der Doppelwirkung
PI4.0	Plattform Industrie 4.0
PLS	Plattform Lernende Systeme
PLT	Prozessleittechnik
QM	Qualitätsmanagement
QoS	Quality of Service
SAI	Securing Artificial Intelligence
SafeTRANS	Safety in Transportation Systems
SAT	Satisfiability Theories
SC	Sub Committee
SCI4.0	Standardization Council Industrie 4.0
SEDRIS	Synthetic Environment Data Representation and Interchange Specification
SEG	Standardization Evaluation Group
SG	Studiengruppe
SIL	Safety Integrity Level
SMS	Short Message Service
SMT	Satisfiability Modulo Theories
SO	Sonderausschuss
SPEC	DIN-Spezifikation (Konsortialstandard)
SQuaRE	Systems and software Quality Requirements and Evaluation
TBINK	Technischer Beirat Internationale Koordinierung
TC	Technical Committee
TKG	Telekommunikationsgesetz
TMG	Telemediengesetz
TR	Technical Report/Fachbericht
TS	Technical Specification/Technische Spezifikation
TÜV	Technischer Überprüfungsverein
UL	UL LLC (Underwriters Laboratories)
UNECE	United Nation Economic Commission for Europe
VDE	VDE Verband der Elektrotechnik Elektronik Informationstechnik e. V.
VDI	VDI Verein Deutscher Ingenieure e. V.
VDMA	Verband Deutscher Maschinen- und Anlagenbau e. V.
W3C	World Wide Web Consortium
WG	Working Group
WKIO	Werte, Kriterien, Indikatoren, Observablen
WMA	World Medical Association
ZDH	Zentralverband des Deutschen Handwerks
ZVEI	Zentralverband Elektrotechnik- und Elektronikindustrie e. V.

8 Quellen- und Literaturverzeichnis

[1] T. Heilmann, N. Schön, **NEUSTAAT: Politik und Staat müssen sich ändern. 64 Abgeordnete & Experten fangen bei sich selbst an – mit 103 Vorschlägen.** München: FinanzBuch, 2020.

[2] PWC, **Auswirkungen der Nutzung von künstlicher Intelligenz in Deutschland**. 2018 [Online]. Verfügbar unter: https://www.pwc.de/de/business-analytics/sizing-the-price-final-juni-2018.pdf [Letzter Zugriff: 17.08.2020].

[3] PwC, **Gobal Top 100 companies by market capitalization**. 2019 [Online]. Verfügbar unter: https://www.pwc.com/gx/en/audit-services/publications/assets/global-top-100-companies-2019.pdf [Letzter Zugriff: 17.08.2020].

[4] acatech (Hrsg.), **Künstliche Intelligenz in der Industrie**. München, 2020.

[5] HLEG-KI, **Ethics Guidelines for Trustworthy AI**. Brüssel: Europäische Komission, 2018 [Online]. Verfügbar unter: https://ec.europa.eu/futurium/en/ai-alliance-consultation [Letzter Zugriff: 13.08.2020].

[6] S. Palacio et al., **What do Deep Networks Like to See**. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Juni 2018.

[7] Y. Gil, B. Selman, **A 20-Year Community Roadmap for Artificial Intelligence Research in the US**. Computing Community Consortium (CCC) and Association for the Advancement of Artificial Intelligence (AAAI), 06.08.2019 [Online]. Verfügbar unter: https://cra.org/ccc/wp-content/uploads/sites/2/2019/08/Community-Roadmap-for-AI-Research.pdf [letzter Zugriff: 07.08.2020].

[8] W. Wahlster, **Künstliche Intelligenz versus menschliche Intelligenz**. Vorlesungsreihe 2017: Künstliche Intelligenz für den Menschen: Digitalisierung mit Verstand. Johannes Gutenberg Stiftungsprofessur [Online]. Verfügbar unter: http://www.dfki.de/wwdata/Gutenberg_Stiftungsprofessur_Mainz_2017/Lernende_Maschinen.pdf [Letzter Zugriff: 11.08.2020].

[9] UBA, **Künstliche Intelligenz im Umweltbereich – Anwendungsbeispiele und Zukunftsperspektiven im Sinne der Nachhaltigkeit**. Dessau-Roßlau, 2019 [Online]. Verfügbar unter: https://www.umweltbundesamt.de/publikationen/kuenstliche-intelligenz-im-umweltbereich [Letzter Zugriff: 28.07.2020].

[10] Datenethikkommission der Bundesregierung, **Gutachten der Datenethikkommission**. Berlin: BMI, 2019 [Online]. Verfügbar unter: https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/it-digitalpolitik/gutachten-datenethikkommission.pdf?__blob=publicationFile&v=6 [Letzter Zugriff: 10.08.2020].

[11] K. Blind, A. Jungmittag, A. Mangelsdorf, **Der gesamtwirtschaftliche Nutzen der Normung – Eine Aktualisierung der DIN-Studie aus dem Jahr 2000**. Berlin: DIN, 2011 [Online]. Verfügbar unter: https://www.din.de/de/ueber-normen-und-standards/nutzen-fuer-die-wirtschaft [Letzter Zugriff: 07.08.2020].

[12] BMWi, **Strategie Künstliche Intelligenz der Bundesregierung**. Berlin, 2018 [Online]. Verfügbar unter: www.ki-strategie-deutschland.de [Letzter Zugriff: 13.08.2020].

[13] BMF, **Mit Zuversicht und voller Kraft aus der Krise**. 3.6.2020 [Online]. Verfügbar unter: https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Themen/Schlaglichter/Konjunkturpaket/20200603-konjunkturpaket-beschlossen.html [Letzter Zugriff: 13.08.2020].

[14] Konrad-Adenauer-Stiftung (Hrsg.), **Vergleich nationaler Strategien zur Förderung von Künstlicher Intelligenz**. Sankt Augustin/Berlin, 2018.

[15] COM/2020/65, **Weißbuch zur Künstlichen Intelligenz: Ein europäisches Konzept für Exzellenz und Vertrauen**.

[16] S. Fouse, S. Cross und Z. Lapin, **DARPA's Impact on Artificial Intelligence**. AI Magazine, 41 (2), S. 3-8, Sommer 2020.

[17] Ethik-Kommission, **Automatisiertes und Vernetztes Fahren.** Berlin: BMVI, 2017 [Online]. Verfügbar unter: https://www.bmvi.de/SharedDocs/DE/Publikationen/DG/bericht-der-ethik-kommission.pdf?__blob=publicationFile [Letzter Zugriff: 01.07.2020].

[18] Deutscher Bundestag, **Enquete-Kommission „Künstliche Intelligenz – Gesellschaftliche Verantwortung und wirtschaftliche, soziale und ökologische Potenziale".** [Online]. Verfügbar unter: https://www.bundestag.de/ausschuesse/weitere_gremien/enquete_ki [Letzter Zugriff: 01.07.2020].

[19] Deutscher Bundestag, **Enquete-Kommission KI, „KI und Wirtschaft"**. Kommissionsdrucksache 19(27)92, 19.12.2019.

[20] Deutscher Bundestag, **Enquete-Kommission KI, „KI und Staat"**. Kommissionsdrucksache 19(27)93, 19.12.2019.

[21] Deutscher Bundestag, **Enquete-Kommission KI, „KI und Gesundheit"**. Kommissionsdrucksache 19(27)94, 19.12.2019.

[22] HLEG-KI, **Policy and investment recommendations for trustworthy Artificial Intelligence**. Brüssel: Europäische Komission, 2019 [Online]. Verfügbar unter: https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence [Letzter Zugriff: 13.08.2020].

[23] PI4.0, **Technologieszenario „Künstliche Intelligenz in der Industrie 4.0".** Berlin: BMWi, 2019 [Online]. Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/KI-industrie-40.html [Letzter Zugriff 07.08.2020].

[24] PI4.0, **KI in der Industrie 4.0: Orientierung, Anwendungsbeispiele, Handlungsempfehlungen.** Berlin: BMWi, 2019 [Online]. Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/ki-in-der-industrie-40-orientierung-anwendungsbeispiele-handlungsempfehlungen.html [Letzter Zugriff 07.08.2020].

[25] PI4.0, **Details of the Asset Administration Shell Part 1 – The exchange of information between partners in the value chain of Industrie 4.0 (Version 2.0.1)**. Berlin: BMWi, 2019 [Online]. Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/Details-of-the-Asset-Administration-Shell-Part1.html [Letzter Zugriff 07.08.2020].

[26] PI4.0, **Umgang mit Sicherheitsrisiken industrieller Anwendungen durch mangelnde Erklärbarkeit von KI-Ergebnissen.** Berlin: BMWi, 2019 [Online]. Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/Umgang-mit-Sicherheitsrisiken.html [Letzter Zugriff 07.08.2020].

[27] PI4.0, **Künstliche Intelligenz in Sicherheitsaspekten der Industrie 4.0.** Berlin: BMWi, 2019 [Online]. Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/KI-in-sicherheitsaspekten.html [Letzter Zugriff 07.08.2020].

[28] PI4.0, **Künstliche Intelligenz und Recht im Kontext von Industrie 4.0.** Berlin: BMWi, 2019 [Online]. Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/kuenstliche-intelligenz-und-recht.html [Letzter Zugriff 07.08.2020].

[29] PI4.0, **KI und Robotik im Dienste der Menschen.** Berlin: BMWi, 2019 [Online]. Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/BMWi%20KI%20und%20Robotik.html [Letzter Zugriff 07.08.2020].

[30] HLEG-KI, **A definition of AI: Main capabilities and scientific disciplines**. Brüssel: Europäische Komission, 2020 [Online]. Verfügbar unter: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341 [Letzter Zugriff: 10.06.2020].

[31] S. Russell, P. Norvig, **Artificial Intelligence: A Modern Approach**. 3. Aufl. Harrow, UK: Pearson, 2016.

[32] OECD, **Recommendation of the Council on Artificial Intelligence**. 2019 [Online]. Verfügbar unter: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL0449 [Letzter Zugriff: 10.06.2020].

[33] W. Wahlster, **„Künstliche Intelligenz als Grundlage autonomer Systeme“**. Informatik-Spektrum, Band 40, Heft 5, S. 409-418, 2017.

[34] ISO/IEC 20546:2019, **Information technology – Big data – Overview and vocabulary.**

[35] Auto Zeitung, 4/2019.

[36] W. Hildesheim, T. Holoyad, T. Schmidt, K. Schuhmacher, **Managing and Understanding Artificial Intelligence Solutions – The AI-Methods, Capabilities and Criticality Grid and its Value for Decision Makers, Developers and Regulators**. Beuth, 1. Auflage, 2020.

[37] D. R. Krathwohl, **A Revision of Bloom’s Taxonomy**. Theory into Practice, 41 (4), S. 212-218, 2002.

[38] DIN EN ISO/IEC 17000:2019-05, **Konformitätsbewertung – Begriffe und allgemeine Grundlagen (ISO/IEC DIS 17000:2019).**

[39] DIN EN ISO/IEC 17020:2012-07, **Konformitätsbewertung – Anforderungen an den Betrieb verschiedener Typen von Stellen, die Inspektionen durchführen (ISO/IEC 17020:2012).**

[40] DIN EN ISO/IEC 17021-1:2015-11, **Konformitätsbewertung – Anforderungen an Stellen, die Managementsysteme auditieren und zertifizieren – Teil 1: Anforderungen (ISO/IEC 17021-1:2015).**

[41] DIN EN ISO/IEC 17024:2012-11, Konformitätsbewertung – **Allgemeine Anforderungen an Stellen, die Personen zertifizieren (ISO/IEC 17024:2012)**.

[42] DIN EN ISO/IEC 17025:2018-03, **Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien (ISO/IEC 17025:2017).**

[43] DIN EN ISO/IEC 17029:2020-02, **Konformitätsbewertung – Allgemeine Grundsätze und Anforderungen an Validierungs- und Verifizierungsstellen (ISO/IEC 17029:2019).**

[44] DIN EN ISO/IEC 17065:2013-01, **Konformitätsbewertung – Anforderungen an Stellen, die Produkte, Prozesse und Dienstleistungen zertifizieren (ISO/IEC 17065:2012).**

[45] M. Poretschkin, F. Rostalski, J. Voosholz et al., **Vertrauenswürdiger Einsatz von Künstlicher Intelligenz.** Fraunhofer IAIS, 2019 [Online]. Verfügbar unter: https://www.iais.fraunhofer.de/content/dam/iais/KINRW/Whitepaper_KI-Zertifizierung.pdf [Letzter Zugriff: 10.08.2020].

[46] ISO, **WTO ISO Standards Information Gateway**. 2020 [Online]. Verfügbar unter: https://tbtcode.iso.org/sites/wto-tbt/home.html [Letzter Zugriff: 01.07.2020].

[47] CC 3.1:2017, **Common Criteria for Information Technology Security Evaluation.** [Online]. Verfügbar unter: https://www.commoncriteriaportal.org/cc/ [letzter Zugriff: 10.08.2020].

[48] ISO/IEC 15408-1:2009, **Information technology – Security techniques – Evaluation criteria for IT security – Part 1: Introduction and general model.**

[49] ISO/IEC 15408-2:2008, **Information technology – Security techniques – Evaluation criteria for IT security – Part 2: Security functional components.**

[50] ISO/IEC 15408-3:2008, **Information technology – Security techniques – Evaluation criteria for IT security – Part 3: Security assurance components.**

[51] ISO/IEC 18045:2008, **Information technology – Security techniques – Methodology for IT security evaluation.**

[52] CCRA, **Arrangement on the Recognition of Common Criteria Certificates in the field of Information Technology Security.** 2020 [Online]. Verfügbar unter: https://www.commoncriteriaportal.org/ccra/ [Letzter Zugriff: 10.08.2020].

[53] BSI, **Gemeinsame Kriterien für die Prüfung und Bewertung der Sicherheit von Informationstechnik**. 2020 [Online]. Verfügbar unter: https://www.bsi.bund.de/DE/Themen/ZertifizierungundAnerkennung/Produktzertifizierung/ZertifizierungnachCC/ITSicherheitskriterien/CommonCriteria/commoncriteria_node.html [Letzter Zugriff: 10.08.2020].

[54] ISO/IEC 38500:2015, **Information technology – Governance of IT for the organization.**

[55] ISO, IEC, **ISO/IEC Directives Part 1: Consolidated ISO Supplement, Procedures Specific to ISO**. 2020 [Online]. Verfügbar unter: https://www.iso.org/sites/directives/current/consolidated/index.xhtml [Letzter Zugriff: 01.07.2020] [Anhang L].

[56] Informationstechnikzentrum Bund, **V-Modell XT: Vorgehensmodell zum Planen und Durchführen von Systementwicklungs-Projekten**. Bonn, 2020 [Online]. Verfügbar unter: https://www.itzbund.de/DE/Produkte/V-Modell-XT/v-modell-xt_node.html [Letzter Zugriff: 01.07.2020].

[57] Beauftragte der Bundesregierung für Informationstechnik, **Das V-Modell XT**. Berlin: BMI, 2020 [Online]. Verfügbar unter: https://www.cio.bund.de/Web/DE/Architekturen-und-Standards/V-Modell-XT/vmodell_xt_node.html [Letzter Zugriff: 01.07.2020].

[58] ISO/IEC/IEEE 12207:2008-02, **Systems and software engineering – Software life cycle processes.**

[59] ISO 26262-1:2018, **Road vehicles – Functional safety – Part 1: Vocabulary.**

[60] ISO 26262-2:2018, **Road vehicles – Functional safety – Part 2: Management of functional safety.**

[61] ISO 26262-3:2018, **Road vehicles – Functional safety – Part 3: Concept phase.**

[62] ISO 26262-4:2018, **Road vehicles – Functional safety – Part 4: Product development at the system level.**

[63] ISO 26262-5:2018, **Road vehicles – Functional safety – Part 5: Product development at the hardware level.**

[64] ISO 26262-6:2018, **Road vehicles – Functional safety – Part 6: Product development at the software level.**

[65] ISO 26262-7:2018, **Road vehicles – Functional safety – Part 7: Production, operation, service and decommissioning.**

[66] ISO 26262-8:2018, **Road vehicles – Functional safety – Part 8: Supporting processes.**

[67] ISO 26262-9:2018, **Road vehicles – Functional safety – Part 9: Automotive safety integrity level (ASIL)-oriented and safety-oriented analyses.**

[68] ISO 26262-10:2018, **Road vehicles – Functional safety – Part 10: Guidelines on ISO 26262.**

[69] ISO 26262-11:2018, **Road vehicles – Functional safety – Part 11: Guidelines on application of ISO 26262 to semiconductors.**

[70] ISO 26262-12:2018, **Road vehicles – Functional safety – Part 12: Adaptation of ISO 26262 for motorcycles.**

[71] ISO/IEC 27034-1:2011, **Information technology – Security techniques – Application security – Part 1: Overview and concepts.**

[72] ISO/IEC 27034-1:2011/Cor 1:2014, **Information technology – Security techniques – Application security – Part 1: Overview and concepts – Technical Corrigendum 1.**

[73] ISO/IEC 27034-2:2015, **Information technology – Security techniques – Application security – Part 2: Organization normative framework.**

[74] ISO/IEC 27034-3:2018, **Information technology – Application security – Part 3: Application security management process.**

[75] ISO/IEC 27034-5:2017, **Information technology – Security techniques – Application security – Part 5: Protocols and application security controls data structure.**

[76] ISO/IEC TS 27034-5-1:2018, **Information technology – Application security – Part 5-1: Protocols and application security controls data structure, XML schemas.**

[77] ISO/IEC 27034-6:2016, **Information technology – Security techniques – Application security – Part 6: Case studies.**

[78] ISO/IEC 27034-7:2018, **Information technology – Application security – Part 7: Assurance prediction framework.**

[79] DIN EN 61508 Beiblatt 1:2005-10; VDE 0803 Beiblatt 1:2005-10, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 0: Funktionale Sicherheit und die IEC 61508 (IEC/TR 61508-0:2005).**

[80] DIN EN 61508-1:2011-02; VDE 08031:2011-02, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 1: Allgemeine Anforderungen (IEC 61508-1:2010).**

[81] DIN EN 61508-2:2011-02; VDE 0803-2:2011-02, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 2: Anforderungen an sicherheitsbezogene elektrische/elektronische/programmierbare elektronische Systeme (IEC 61508-2:2010).**

[82] DIN EN 61508-3:2011-02; VDE 0803-3:2011-02, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 3: Anforderungen an Software (IEC 61508-3:2010).**

[83] DIN EN 61508-4:2011-02; VDE 0803-4:2011-02, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 4: Begriffe und Abkürzungen (IEC 61508-4:2010).**

[84] DIN EN 61508-5:2011-02; VDE 08035:2011-02, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 5: Beispiele zur Ermittlung der Stufe der Sicherheitsintegrität (safety integrity level) (IEC 61508-5:2010).**

[85] DIN EN 61508-6:2011-02; VDE 0803-6:2011-02, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 6: Anwendungsrichtlinie für IEC 61508-2 und IEC 61508-3 (IEC 61508-6:2010).**

[86] DIN EN 61508-7:2011-02; VDE 08037:2011-02, **Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme – Teil 7: Überblick über Verfahren und Maßnahmen (IEC 61508-7:2010).**

[87] DIN SPEC 92001-1:2019-04, **Künstliche Intelligenz – Life Cycle Prozesse und Qualitätsanforderungen – Teil 1: Qualitäts-Meta-Modell.**

[88] DKE, **Referenzmodell für eine vertrauenswürdige KI: Erarbeitung einer neuen VDE Anwendungsregel**. Frankfurt a. M., 2020 [Online]. Verfügbar unter: https://www.dke.de/de/news/2019/referenzmodell-vertrauenswuerdige-ki-vde-anwendungsregel [Letzter Zugriff: 01.07.2020].

[89] ISO/IEC 25012:2008-12, **Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model.**

[90] L. Bruns, B. Dittwald, F. Meiners et al., **Leitfaden für qualitativ hochwertige Daten und Metadaten**. Berlin: Fraunhofer FOKUS, 2019 [Online]. Verfügbar unter: https://cdn0.scrvt.com/fokus/551bf951bf1982f5/0c96fbf464ef/NQDM_Leitfaden_2019.pdf [Letzter Zugriff: 01.07.2020].

[91] ISO 8601(2Teile):2019-02, **Date and time – Representations for information interchange.**

[92] D. Keim, K.U. Sattler, K., **Von Daten zu KI. Intelligentes Datenmanagement als Basis für Data Science und den Einsatz Lernender Systeme**. München, 2020 [Online]. Verfügbar unter **https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG1_Whitepaper_Von_Daten_zu_KI.pdf** [Letzter Zugriff: 30.10.2020].

[93] DIN ISO 31000:2018-10, **Risikomanagement – Leitlinien (ISO 31000:2018).**

[94] **Richtlinie 2006/42/EG** des Europäischen Parlaments und des Rates vom 17. Mai 2006 über Maschinen und zur Änderung der Richtlinie 95/16/EG (Neufassung).

[95] **Verordnung (EU) 2016/679** des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung).

[96] J. Angwin et al., **Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks"**. ProPublica, 23.03.2016 [Online]. Verfügbar unter: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Letzter Zugriff: 28.07.2020].

[97] ACLU, **Smart Reform is Possible.** New York, 2011 [Online]. Verfügbar unter: https://www.aclu.org/files/assets/smartreformispossible.pdf [Letzter Zugriff: 01.07.2020].

[98] CivilRights.org, **More than 100 Civil Rights, Digital Justice, and Community-Based Organizations Raise Concerns About Pretrial Risk Assessment, The Leadership Conference on Civil and Human Rights und The Leadership Conference Education Fund**. 2018 [Online]. Verfügbar unter: https://civilrights.org/2018/07/30/more-than100-civil-rights-digital-justice-and-community-based-organizations-raise-concerns-about-pretrial-risk-assessment/ [Letzter Zugriff: 01.07.2020].

[99] D. Collingridge, **The social control of technology**. New York: St. Martin's Press, 1980.

[100] D. Watson, **The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence, Minds & Machines**. 29, S. 417-440, 2019.

[101] E. Brethenoux, **An ode to the analytics grease monkeys**. KDnuggets, 2017 [Online]. Verfügbar unter: https://www.kdnuggets.com/2017/02/analytics-grease-monkeys.html [Letzter Zugriff: 01.07.2020].

[102] R. Berk et al., **Fairness in Criminal Justice Risk Assessments: The State of the Art, Sociological Methods & Research**. doi: 10.1177/0049124118782533, 2018. [S. 33].

[103] B. Lepri et al., **Fair, Transparent, and Accountable Algorithmic Decision-Making Processes: The Premise, the Proposed Solutions, and the Open Challenges**. Philosophy & Technology 31, 4, S. 611-27, 2018. [hier S. 624].

[104] M. Veale , M. Van Kleek, R. Binns, **Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making**. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems – CHI '18, Montreal QC, Canada, 2018, S. 1-14. ACM Press. doi: 10.1145/3173574.3174014.

[105] DIN EN ISO 9000:2015-11, **Qualitätsmanagementsysteme – Grundlagen und Begriffe (ISO 9000:2015).**

[106] K. A. Zweig et al., **Wo Maschinen irren können. Verantwortlichkeiten und Fehlerquellen in Prozessen algorithmischer Entscheidungsfindung**. Gütersloh: Bertelsmann Stiftung, 2018 [Online]. Verfügbar unter: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WoMaschinenIrrenKoennen.pdf [Letzter Zugriff: 14.07.2020].

[107] J. Walker, **Meet the New Boss: Big Data. Companies Trade In Hunch-Based Hiring for Computer Modeling**. The Wall Street Journal, 20.09.2012.

[108] S. Barocas, A. D. Selbst, **Big data's disparate impact**. California Law Review, 104, S. 671-732, 2016.

[109] S. Beck et al., **Künstliche Intelligenz und Diskriminierung – Herausforderungen und Lösungsansätze**. München: PLS, 2019 [Online]. Verfügbar unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_250619.pdf [Letzter Zugriff: 28.07.2020].

[110] G. S. Leventhal, J. Karuza, W. R.Fry, **Beyond fairness: A theory of allocation preferences, Justice and social interaction**. 3 (1), S. 167-218, 1980.

[111] L. Floridi et al., **AI4People: An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations**, Minds and Machines, 28, S. 689-707, 2018.

[112] Beijing Academy of Artificial Intelligence (BAAI), **Beijing AI Principles**. 2019 [Online]. Verfügbar unter: https://www.baai.ac.cn/blog/beijing-ai-principles [Letzter Zugriff: 03.03.2020].

[113] U. Garzcarek, D. Steuer, Approaching Ethical Guidelines for Data Scientists, in: N. Bauer et al. (Hrsg.), **Applications in Statistical Computing: From Music Data Analysis to Industrial Quality Improvement**. Cham: Springer, 2019, S. 151-169.

[114] J. Fjeld et al., **Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI**. Berkman Klein Center for Internet & Society, 2020.

[115] A. Jobin et al., **The global landscape of AI ethics guidelines**. Nature Machine Intelligence, 1, 9, S. 389-399, 2019.

[116] T. Hagendorff, **The Ethics of AI Ethics: An Evaluation of Guidelines**. Minds & Machines, 30, 1, S. 99-120, 2020.

[117] J. Heesen et al., **Ethik-Briefing. Leitfaden für eine verantwortungsvolle Entwicklung und Anwendung von KI-Systemen**, München, 2020. [Online]. Verfügbar unter: **https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_EB_200831.pdf**

[118] N. Huchler et al., **Kriterien für die Mensch-Maschine: Interaktion bei KI. Ansätze für die menschengerechte Gestaltung in der Arbeitswelt**. Plattform Lernende Systeme, München, 2020 [Online]. Verfügbar unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG2_Whitepaper2_220620.pdf [Letzter Zugriff: 28.07.2020].

[119] L. Inhoffen, **Künstliche Intelligenz: Deutsche sehen eher die Risiken als den Nutzen**. YouGov, 11.09.2018 [Online]. Verfügbar unter: https://yougov.de/news/2018/09/11/kunstliche-intelligenz-deutsche-sehen-eher-die-ris/ [Letzter Zugriff: 11.08.2020].

[120] DIN EN ISO 9001:2015-11, **Qualitätsmanagementsysteme – Anforderungen (ISO 9001:2015).**

[121] ISO/IEC Guide 51:2014-04, **Safety aspects – Guidelines for their inclusion in standards**.

[122] ISO/IEC 27001:2013, **Information technology** – **Security techniques** – **Information security management systems** – **Requirements.**

[123] S. Hallensleben et al., **From Principles to Practice – An interdisciplinary framework to operationalize AI ethics**. Gütersloh: Bertelsmann Stiftung, 2020 [Online]. Verfügbar unter: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf [Letzter Zugriff: 28.07.2020].

[124] DIN EN ISO 12100:2011-03, **Sicherheit von Maschinen – Allgemeine Gestaltungsleitsätze – Risikobeurteilung und Risikominderung (ISO 12100:2010).**

[125] DIN EN ISO 12100 Berichtigung 1:2013-08, **Sicherheit von Maschinen – Allgemeine Gestaltungsleitsätze – Risikobeurteilung und Risikominderung (ISO 12100:2010); Deutsche Fassung EN ISO 12100:2010, Berichtigung zu DIN EN ISO 12100:2011-03.**

[126] DIN EN ISO 13849-1:2016-06, **Sicherheit von Maschinen – Sicherheitsbezogene Teile von Steuerungen – Teil 1: Allgemeine Gestaltungsleitsätze (ISO 13849-1:2015).**

[127] DIN EN ISO 13849-2:2013-02, **Sicherheit von Maschinen – Sicherheitsbezogene Teile von Steuerungen – Teil 2: Validierung (ISO 13849-2:2012).**

[128] DIN EN ISO 14971:2020-07, **Medizinprodukte – Anwendung des Risikomanagements auf Medizinprodukte (ISO 14971:2019)**.

[129] DIN EN 62061:2016-05; VDE 011350:2016-05, **Sicherheit von Maschinen – Funktionale Sicherheit sicherheitsbezogener elektrischer, elektronischer und programmierbarer elektronischer Steuerungssysteme (IEC 62061:2005 + A1:2012 + A2:2015).**

[130] D. Dawson, E. Schleiger et al., **Artificial Intelligence: Australia's Ethics Framework**. Data61 CSIRO, Australia, 2019 [Online]. Verfügbar unter: https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf [Letzter Zugriff: 14.08.2020].

[131] J. Black, **Risk-based Regulation: Choices, Practices and Lessons Being Learnt**. in: OECD, **Risk and Regulatory Policy: Improving the Governance of Risk**. Paris, 2010, S. 185-236.

[132] I. MacNeil, **Risk control strategies: an assessment in the context of the credit crisis**. in: I. MacNeil, J. O'Brien (Hrsg.), The future of financial regulation. Oxford , Portland: Hart Pub, 2010, S. 141-160.

[133] J. Black, R. Baldwin, **When risk-based regulation aims low: A strategic framework**. Regulation & Governance, 6 (2), S. 131-148, 2012.

[134] F. Saurwein et al., **Governance of algorithms: options and limitations**, info, 17 (6), S. 35-49, 2015.

[135] M. Z. van Drunen et al., **Know your algorithm: what media organizations need to explain to their users about news personalization**. International Data Privacy Law, 9 (4), S. 220-235, 2019.

[136] T. D. Krafft, K. A. Zweig, **Transparenz und Nachvollziehbarkeit algorithmenbasierter Entscheidungsprozesse: Ein Regulierungsvorschlag aus sozioinformatischer Perspektive**. Berlin: Verbraucherzentrale Bundesverband e. V., 2019 [Online]. Verfügbar unter: https://www.vzbv.de/sites/default/files/downloads/2019/05/02/19-01-22_zweig_krafft_transparenz_adm-neu.pdf [Letzter Zugriff: 06.08.2020].

[137] Bundesärztekammer, **(Muster)Berufsordnung für die in Deutschland tätigen Ärztinnen und Ärzte – MBO-Ä 1997 –* in der Fassung der Beschlüsse des 121. Deutschen Ärztetages 2018 in Frankfurt am Main**. Deutsches Ärzteblatt, 01.02.2019, S. A1-A9 [hier § 9, S. A4].

[138] C. Reed, **How should we regulate artificial intelligence?**. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376, 20170360, 2018.

[139] N. Diakopoulos, **Algorithmic accountability reporting: On the investigation of black boxes**. Tow Center for Digital Journalism Publications, 2014.

[140] W3C, **Semantic Web**. [Online.] Verfügbar unter: https://www.w3.org/2001/sw/wiki/Main_Page [Letzter Zugriff: 11.08.2020].

[141] **Verordnung (EU) 2017/745** des Europäischen Parlaments und des Rates vom 5. April 2017 über Medizinprodukte, zur Änderung der Richtlinie 2001/83/EG, der Verordnung (EG) Nr. 178/2002 und der Verordnung (EG) Nr. 1223/2009 und zur Aufhebung der Richtlinien 90/385/EWG und 93/42/EWG des Rates.

[142] **Bundesdatenschutzgesetz** (BDSG).

[143] DSK, **Hambacher Erklärung zur Künstlichen Intelligenz.** Entschließung der 97. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder, Hambacher Schloss, 3. April 2019 [Online]. Verfügbar unter: https://www.datenschutzkonferenz-online.de/media/en/20190405_hambacher_erklaerung.pdf [Letzter Zugriff: 11.08.2020].

[144] JCGM 200:2012, **International vocabulary of metrology – Basic and general concepts and associated terms (VIM) [= ISO/IEC Guide 99]**.

[145] DIN EN ISO 19011:2018-10, **Leitfaden zur Auditierung von Managementsystemen (ISO 19011:2018).**

[146] ISO/IEC 25010:2011, **Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models.**

[147] S. Goericke (Hrsg.), **The Future of Software Quality Assurance**. Basel: Springer, 2020.

[148] ISO/PAS 21448:2019, **Road vehicles – Safety of the intended functionality.**

[149] ISO/IEC TR 20547-2:2018, **Information technology – Big data reference architecture – Part 2: Use cases and derived requirements.**

[150] ISO/IEC TR 20547-5:2018, **Information technology – Big data reference architecture – Part 5: Standards roadmap.**

[151] DIN SPEC 13266:2020-04, **Leitfaden für die Entwicklung von Deep-Learning-Bilderkennungssystemen.**

[152] ETSI TS 103 296 V1.1.1:2016-08, **Speech and Multimedia Transmission Quality (STQ) – Requirements for Emotion Detectors used for Telecommunication Measurement Applications – Detectors for written text and spoken speech.**

[153] ETSI TS 103 195-2 V1.1.1:2018-05, **Autonomic network engineering for the self-managing Future Internet (AFI) – Generic Autonomic Network Architecture – Part 2: An Architectural Reference Model for Autonomic Networking, Cognitive Networking and Self-Management.**

[154] ITUT Y.3170:2018-09, **Requirements for machine learning-based quality of service assurance for the IMT-2020 network.**

[155] ITUT Y.3173:2020-02, **Framework for evaluating intelligence levels of future networks including IMT-2020.**

[156] IEEE 7010:2020, **IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being.**

[157] UL 4600:2020-04, **Evaluation of Autonomous Products.**

[158] T. Weinberger et al., **Modelle Maschinellen Lernens: Symbolische und konnektionistische Ansätze**. Karlsruhe: Kernforschungszentrum Karlsruhe, 1994. [= KfK 5184].

[159] G. Kern-Isberner, **Fortgeschrittene Themen der Wissensrepräsentation**. Vorlesung TU Dortmund, 2020. [Online.] Zusammenfassung verfügbar unter: https://ls1-www.cs.tu-dortmund.de/de/lehrveranstaltungen/503-ftw-ss-2020/1897-fortgeschrittene-themen-der-wissensrepr%C3%A4sentation-ss-20 [Letzter Zugriff: 11.08.2020].

[160] C. E. Alchourròn et al., **On the logic of theory change: Partial meet contraction and revision functions**. Journal of Symbolic Logic, 50 (2), S. 510–530, 1985.

[161] D. Mackall et al., **Verification and validation of neural networks for aerospace systems**. Mofett Field (CA): NASA, 2002.

[162] N. Röttger et al., **Warum KI auch eine intelligente Qualitätssicherung braucht**. OBJEKTspektrum, 02/2020, S. 20-24.

[163] ISO/IEC 27701:2019, **Security techniques – Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management – Requirements and guidelines.**

[164] IATF 16949:2016, **Anforderungen an Qualitätsmanagementsysteme für die Serien- und Ersatzteilproduktion in der Automobilindustrie.**

[165] DIN EN 9100:2018-08, **Qualitätsmanagementsysteme – Anforderungen an Organisationen der Luftfahrt, Raumfahrt und Verteidigung**.

[166] H. W. Dörmann Osuna, **Ansatz für ein prozessintegriertes Qualitätsregelungssystem für nicht-stabile Prozesse**. Dissertation, Techn. Univ. Ilmenau, 2008.

[167] Hering et al, Qualitätslenkung mit Produkt-Regelkreis, in: E. Hering, J. Triemel, H.-P. Blank (Hrsg.), **Qualitätsmanagement für Ingenieure**. Berlin, Heidelberg: Springer, 2003, Kapitel E3.3.

[168] R. Schmitt, T. Pfeifer, **Qualitätsmanagement: Strategien, Methoden, Techniken.** München, Wien: Carl Hanser Verlag, 2015.

[169] G. Montavon et al., **Methods for Interpreting and Understanding Deep Neural Networks**. arXiv:1706.07979, 2017.

[170] E. Štrumbelj, I. Kononenko, **Explaining prediction models and individual predictions with feature contributions**. Knowledge and Information Systems, 41, S. 647-665, 2014.

[171] A. Shrikumar et al., **Learning Important Features Through Propagating Activation Differences**. arXiv:1704.02685, 2019.

[172] A. Datta et al., **Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems 2016 IEEE Symposium on Security and Privacy (SP)**. San Jose, CA, 2016, S. 598-617.

[173] T. Menzies, **Verification and Validation and Artificial Intelligence: Beyond the State of the Art**. Foundations 02: A V&V Workshop, Johns Hopkins University Laurel, Maryland USA, 2002.

[174] R. Ehlers, Formal verification of piece-wise linear feed-forward neural networks, arXiv:1705.01320v3, 2017.

[175] ETSI TR 101 583 V 1.1.1:2015-03 , **Methods for Testing and Specification (MTS) – Security Testing – Basic Terminology**.

[176] A. Odena, I. Goodfellow, **TensorFuzz: Debugging Neural Networks with Coverage-Guided Fuzzing**. arXiv:1807.10875, 2018.

[177] J. Guo, et al., **Dlfuzz: Differential fuzzing testing of deep learning systems**. Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, S. 739-743, 2018.

[178] NISTIR 8269, **A Taxonomy and Terminology of Adversarial Machine Learning (draft).**

[179] Y. Dong et al., **There is Limited Correlation between Coverage and Robustness for Deep Neural Networks**. arXiv:1911.05904, 2019.

[180] Spiegel Netzwelt, 10.06.2020 [Online]. Verfügbar unter: https://www.spiegel.de/netzwelt/web/honda-muss-produktion-nach-cyberangriff-stoppen-a-69f59803-216c-43c7-9ee9-59d3926d6314# [Letzter Zugriff: 11.08.2020].

[181] Welt, 19.06.2020 [Online.] Verfügbar unter: https://www.welt.de/politik/ausland/article209883177/Cyber-Angriffe-auf-Australien-Premier-Scott-Morrison-hat-eine-Vermutung.html [Letzter Zugriff: 11.08.2020].

[182] A. Berg, M. Niemeier, **Wirtschaftsschutz in der digitalen Welt**. Berlin: Bitkom, 6.11.2019 [Online]. Verfügbar unter: https://www.bitkom.org/sites/default/files/2019-11/bitkom_wirtschaftsschutz_2019.pdf [Letzter Zugriff: 17.08.2020].

[183] Kompass Informationssicherheit und Datenschutz, **IT-Sicherheits- und Risikomanagement**. Berlin: Bitkom, DIN, 2020 [Online]. Verfügbar unter: https://www.kompass-sicherheitsstandards.de/ISMS/Allgemeine-ISMS [Letzter Zugriff: 17.08.2020].

[184] DIN, DKE, **Deutsche Normungs-Roadmap IT-Sicherheit**. Berlin, Frankfurt (M.), Version 3.0, 2017 [Online]. Verfügbar unter: https://www.din.de/de/din-und-seine-partner/presse/mitteilungen/normungs-roadmap-it-sicherheit-aktualisiert-238508 [Letzter Zugriff: 17.08.2020].

[185] **Richtlinie (EU) 2016/1148** des Europäischen Parlaments und des Rates vom 6. Juli 2016 über Maßnahmen zur Gewährleistung eines hohen gemeinsamen Sicherheitsniveaus von Netz- und Informationssystemen in der Union.

[186] **Richtlinie (EU) 2016/680** des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung sowie zum freien Datenverkehr und zur Aufhebung des Rahmenbeschlusses 2008/977/JI des Rates.

[187] **Richtlinie 2002/58/EG** des Europäischen Parlaments und des Rates vom 12. Juli 2002 über die Verarbeitung personenbezogener Daten und den Schutz der Privatsphäre in der elektronischen Kommunikation (Datenschutzrichtlinie für elektronische Kommunikation).

[188] **Verordnung (EU) 2019/881** des Europäischen Parlaments und des Rates vom 17. April 2019 über die ENISA (Agentur der Europäischen Union für Cybersicherheit) und über die Zertifizierung der Cybersicherheit von Informations- und Kommunikationstechnik und zur Aufhebung der Verordnung (EU) Nr. 526/2013 (Rechtsakt zur Cybersicherheit).

[189] **Richtlinie 2001/95/EG** des Europäischen Parlaments und des Rates vom 3. Dezember 2001 über die allgemeine Produktsicherheit.

[190] **Gesetz zur Erhöhung der Sicherheit informationstechnischer Systeme** (IT-Sicherheitsgesetz).

[191] **Zweites Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680** (Zweites Datenschutz-Anpassungs- und Umsetzungsgesetz EU – 2. DSAnpUG-EU).

[192] **Telemediengesetz** (TMG).

[193] **Telekommunikationsgesetz** (TKG).

[194] **Gesetz über das Bundesamt für Sicherheit in der Informationstechnik** (BSI-Gesetz – BSIG).

[195] **Gesetz über die Bereitstellung von Produkten auf dem Markt** (Produktsicherheitsgesetz – ProdSG).

[196] **Gesetz über die Elektrizitäts- und Gasversorgung** (Energiewirtschaftsgesetz – EnWG).

[197] bitkom, **Regulierungsmapping IT-Sicherheit**. Berlin, 08.2019 [Online.] Verfügbar unter: https://www.bitkom.org/sites/default/files/2019-08/190816_regulierungsmapping.pdf [Letzter Zugriff: 17.08.2020].

[198] IEC TS 62443-1-1:2009, **Industrial communication networks – Network and system security – Part 1-1: Terminology, concepts and models.**

[199] IEC 62443-2-1:2010, **Industrial communication networks – Network and system security – Part 2-1: Establishing an industrial automation and control system security program.**

[200] IEC TR 62443-2-3:2015, **Security for industrial automation and control systems – Part 2-3: Patch management in the IACS environment.**

[201] DIN EN IEC 62443-2-4:2020-07; VDE 0802-2-4:2020-07, **IT-Sicherheit für industrielle Automatisierungssysteme – Teil 2-4: Anforderungen an das IT-Sicherheitsprogramm von Dienstleistern für industrielle Automatisierungssysteme (IEC 62443-2-4:2015 + Cor.:2015 + A1:2017).**

[202] IEC TR 62443-3-1:2009, **Industrial communication networks – Network and system security – Part 3-1: Security technologies for industrial automation and control systems.**

[203] IEC 62443-3-2:2020, **Security for industrial automation and control systems – Part 3-2: Security risk assessment for system design.**

[204] IEC 62443-3-3:2013, **Industrial communication networks – Network and system security – Part 3-3: System security requirements and security levels.**

[205] DIN EN IEC 62443-3-3:2020-01; VDE 0802-3-3:2020-01, **Industrielle Kommunikationsnetze – IT-Sicherheit für Netze und Systeme – Teil 3-3: Systemanforderungen zur IT-Sicherheit und Security-Level (IEC 62443-3-3:2013 + COR1:2014).**

[206] IEC 62443-4-1:2018, **Security for industrial automation and control systems – Part 4-1: Secure product development lifecycle requirements.**

[207] DIN EN IEC 62443-4-1:2018-10; VDE 0802-4-1:2018-10, **IT-Sicherheit für industrielle Automatisierungssysteme – Teil 4-1: Anforderungen an den Lebenszyklus für eine sichere Produktentwicklung (IEC 62443-4-1:2018).**

[208] IEC 62443-4-2:2019, **Security for industrial automation and control systems – Part 4-2: Technical security requirements for IACS components.**

[209] DIN EN IEC 62443-4-2:2019-12; VDE 0802-4-2:2019-12, **IT-Sicherheit für industrielle Automatisierungssysteme – Teil 4-2: Technische Sicherheitsanforderungen an Komponenten industrieller Automatisierungssysteme (IACS) (IEC 62443-4-2:2019).**

[210] ISO/IEC 27005:2018, **Information technology – Security techniques – Information security risk management.**

[211] DIN EN 61511-1:201902; VDE 0810-1:2019-02, **Funktionale Sicherheit – PLT-Sicherheitseinrichtungen für die Prozessindustrie – Teil 1: Allgemeines, Begriffe, Anforderungen an Systeme, Hardware und Anwendungsprogrammierung (IEC 61511-1:2016 + COR1:2016 + A1:2017).**

[212] DIN EN ISO/IEC 29134:2020-09, **Informationstechnik – Sicherheitsverfahren – Leitlinien für die Datenschutz-Folgenabschätzung (ISO/IEC 29134:2017).**

[213] Hirsch-Kreinsen et al., **Themenfelder Industrie 4.0: Forschungs- und Entwicklungsbedarfe zur erfolgreichen Umsetzung von Industrie 4.0**. Forschungsbeirat der PI4.0, 2019 [Online]. Verfügbar unter: https://www.acatech.de/wp-content/uploads/2019/09/Forschungsbeirat_Themenfelder-Industrie-4.0-2.pdf [Letzter Zugriff: 12.08.2020].

[214] J. Lee, **Industrial AI. Singapore:** Springer, 2020.

[215] GMA, **VDI-Statusreport Industrie 4.0 Wertschöpfungsketten. Düsseldorf**. 2014. https://www.vdi.de/ueber-uns/presse/publikationen/details/industrie-40-wertschoepfungsketten.

[216] PI4.0, **Fortschreibung der Anwendungsszenarien der Plattform Industrie 4.0. Berlin: BMWi, 2016 [Online].** Verfügbar unter: https://www.plattform-i40.de/PI40/Redaktion/DE/Downloads/Publikation/fortschreibung-anwendungsszenarien.html [Letzter Zugriff: 12.08.2020].

[217] SCI4.0, DIN/DKE, **Deutsche Normungsroadmap Industrie 4.0**. Version 4, Berlin: DIN und Frankfurt (M.): DKE [Online]. Verfügbar unter: https://www.din.de/de/forschung-und-innovation/themen/industrie40/roadmap-industrie40-62178 [Letzter Zugriff: 12.08.2020].

[218] E VDE-AR-E 2842-61-1:2020-07, **Entwicklung und Vertrauenswürdigkeit von autonom/kognitiven Systemen – Teil 61-1: Terminologie und Grundkonzepte.**

[219] GMA, **VDI-Statusreport Maschinelles Lernen**. Düsseldorf, 2019.

[220] VDI/VDE/VDMA 2632 Blatt 2:2015-10, **Industrielle Bildverarbeitung – Leitfaden für die Erstellung eines Lastenhefts und eines Pflichtenhefts**.

[221] VDI/VDE/VDMA 2632 Blatt 3:2017-10, **Industrielle Bildverarbeitung – Abnahme klassifizierender Bildverarbeitungssysteme**.

[222] VDI/VDE/VDMA 2632 Blatt 3.1:2020-08, **Industrielle Bildverarbeitung – Abnahme klassifizierender Bildverarbeitungssysteme – Prüfung der Klassifikationsleistung**.

[223] VDI/VDE/VDMA 2632 Blatt 4.1:2020-08, **Industrielle Bildverarbeitung – Oberflächeninspektionssysteme in der Flachstahlproduktion – Stabilitätsprüfung**.

[224] COM/2018/237, **Künstliche Intelligenz für Europa.**

[225] D. Schneider, M. Trapp, **Conditional safety certification of open adaptive systems**. ACM Transactions on Autonomous and Adaptive Systems (TAAS), 8, 2013.

[226] D. Schneider, et al., **WAP: Digital dependability identities**. 2015 IEEE 26th International Symposium on Software Reliability Engineering (ISSRE), Gaithersbury, MD, USA, 2015, S. 324-329.

[227] COM/2020/64, **Bericht über die Auswirkungen künstlicher Intelligenz, des Internets der Dinge und der Robotik in Hinblick auf Sicherheit und Haftung.** Brüssel: Europäische Kommission.

[228] **Straßenverkehrsgesetz** (StVG).

[229] **Verordnung (EU) 2018/858** des Europäischen Parlaments und des Rates vom 30. Mai 2018 über die Genehmigung und die Marktüberwachung von Kraftfahrzeugen und Kraftfahrzeuganhängern sowie von Systemen, Bauteilen und selbstständigen technischen Einheiten für diese Fahrzeuge, zur Änderung der Verordnungen (EG) Nr. 715/2007 und (EG) Nr. 595/2009 und zur Aufhebung der Richtlinie 2007/46/EG.

[230] **Übereinkommen über die Annahme harmonisierter technischer Regelungen der Vereinten Nationen für Radfahrzeuge, Ausrüstungsgegenstände und Teile, die in Radfahrzeuge(n) eingebaut und/oder verwendet werden können, und die Bedingungen für die gegenseitige Anerkennung von Genehmigungen, die nach diesen Regelungen der Vereinten Nationen erteilt wurden**. Amtsblatt der Europäischen Union, L 274, S. 4–30, 11.10.2016.

[231] **Richtlinie 2014/45/EU** des Europäischen Parlaments und des Rates vom 3. April 2014 über die regelmäßige technische Überwachung von Kraftfahrzeugen und Kraftfahrzeuganhägern und zur Aufhebung der Richtlinie 2009/40/EG.

[232] **UN Regulation No. 79**. Uniform provisions concerning the approval of vehicles with regard to steering equipment, Revision 4, 18.10.2018 [Online]. Verfügbar unter: https://www.unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/2018/R079r4e.pdf [Letzter Zugriff: 12.08.2020].

[233] PLS, **Intelligent vernetzt unterwegs**. [Online]. Verfügbar unter: https://www.plattform-lernende-systeme.de/umfeldszenario-intelligent-vernetzt-unterwegs.html [Letzter Zugriff: 12.08.2020].

[234] **Straßenverkehrs-Ordnung** (StVO).

[235] ISO/IEC 11179-1:2015, **Information technology – Metadata registries (MDR) – Part 1: Framework**.

[236] ISO/IEC TR 11179-2:2019, **Information technology – Metadata registries (MDR) – Part 2: Classification**.

[237] ISO/IEC 11179-3:2013, **Information technology – Metadata registries (MDR) – Part 3: Registry metamodel and basic attributes**.

[238] ISO/IEC 11179-4:2004, **Information technology – Metadata registries (MDR) – Part 4: Formulation of data definitions**.

[239] ISO/IEC 11179-5:2015, **Information technology – Metadata registries (MDR) – Part 5: Naming principles**.

[240] ISO/IEC 11179-6:2015, **Information technology – Metadata registries (MDR) – Part 6: Registration**.

[241] ISO/IEC 11179-7:2019, **Information technology – Metadata registries (MDR) – Part 7: Metamodel for data set registration**.

[242] ISO/IEC TS 11179-30:2019, **Information technology – Metadata registries (MDR) – Part 30: Basic attributes of metadata**.

[243] ISO/IEC 19763-1:2015, **Information technology – Metamodel framework for interoperability (MFI) – Part 1: Framework**.

[244] ISO/IEC 19763-3:2010, **Information technology – Metamodel framework for interoperability (MFI) – Part 3: Metamodel for ontology registration**.

[245] ISO/IEC 19763-5:2015, **Information technology – Metamodel framework for interoperability (MFI) – Part 5: Metamodel for process model registration**.

[246] ISO/IEC 19763-6:2015, **Information technology – Metamodel framework for interoperability (MFI) – Part 6: Registry Summary**.

[247] ISO/IEC 19763-7:2015, **Information technology – Metamodel framework for interoperability (MFI) – Part 7: Metamodel for service model registration**.

[248] ISO/IEC 19763-8:2015, **Information technology – Metamodel framework for interoperability (MFI) – Part 8: Metamodel for role and goal model registration**.

[249] ISO/IEC TR 19763-9:2015, **Information technology – Metamodel framework for interoperability (MFI) – Part 9: On demand model selection**.

[250] ISO/IEC 19763-10:2014, **Information technology – Metamodel framework for interoperability (MFI) – Part 10: Core model and basic mapping**.

[251] ISO/IEC 19763-12:2015, **Information technology – Metamodel framework for interoperability (MFI) – Part 12: Metamodel for information model registration**.

[252] ISO/IEC TS 19763-13:2016, **Information technology – Metamodel framework for interoperability (MFI) – Part 13: Metamodel for form design registration**.

[253] M. Salathé et al., **Focus Group on Artificial Intelligence for Health**. [Online]. Verfügbar unter: https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/default.aspx [Letzter Zugriff: 12.08.2020].

[254] J. Müller-Quade et al, **Sichere KI-Systeme für die Medizin. Datenmanagement und IT-Sicherheit in der Krebsbehandlung der Zukunft**. München, 2020 [Online] Verfügbar unter **https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_6_Whitepaper_07042020.pdf** [Letzter Zugriff: 20.10.2020].

[255] WMA, **Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects**. 2020 [Online]. Verfügbar unter: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [Letzter Zugriff: 10.06.2020].

[256] **Gesetz über Medizinprodukte** (Medizinproduktegesetz – MPG).

[257] Acatech, **Lernende Systeme: Die Plattform für Künstliche Intelligenz Deutsche Akademie der Technikwissenschaften e. V.**. 2020 [Online]. Verfügbar unter: https://www.plattform-lernende-systeme.de/startseite.html [Letzter Zugriff: 10.06.2020].

[258] C. Wischhöfer, P. Rauh, **Standards of the Future – Stand der Arbeiten zum Thema maschinenausführbarer Normeninhalte**. DIN-Mitteilungen, August 2019, S. 4-8.

[259] D. Czarny et al., **Digitale Transformation in der Normung – Ausgangssituation und Vision – Initiative Digitale Standards – Herausforderungen und Ziele**. Seminar I und II am 2.6.2020 und 10.6.2020 [Online]. Verfügbar unter: https://youtu.be/B4f09mxVHsI und https://youtu.be/laEXmB0m_PI [Letzter Zugriff: 07.08.2020].

[260] IEC SG12 TF 2, **Digital Transformation (2019) Work Package 1 – Classification Scheme and Use Cases (utility model), IEC**.

[261] R. Heidel et al., **Industrie 4.0. – Basiswissen RAMI4.0 Referenzarchitekturmodell und Industrie 4.0-Komponente**. Berlin: Beuth, 2017.

[262] ISO/IEC TR 24028:2020, **Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence.**

[263] ETSI GR ENI 004 V 2.1.1:2019-10, **Experiential Networked Intelligence (ENI) – Terminology for Main Concepts in ENI.**

[264] ETSI GR NFVSEC 003 V 1.2.1:2016-08, **Network Functions Virtualisation (NFV) – NFV Security – Security and Trust Guidance.**

[265] ISO/IEC/IEEE 29119-1:2013, **Software and systems engineering – Software testing – Part 1: Concepts and definitions.**

[266] ISO/IEC/IEEE 29119-2:2013, **Software and systems engineering – Software testing – Part 2: Test processes.**

[267] ISO/IEC/IEEE 29119-3:2013, **Software and systems engineering – Software testing – Part 3: Test documentation.**

[268] ISO/IEC/IEEE 29119-4:2015, **Software and systems engineering – Software testing – Part 4: Test techniques.**

[269] ISO/IEC/IEEE 29119-5:2016, **Software and systems engineering – Software testing – Part 5: Keyword-Driven Testing.**

[270] ISO/IEC TR 20547-1:2020, **Information technology – Big data reference architecture – Part 1: Framework and application process.**

[271] ISO/IEC 20547-3:2020, **Information technology – Big data reference architecture – Part 3: Reference architecture.**

[272] ISO/IEC 25000:2014-03, **Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Guide to SQuaRE.**

[273] ISO/IEC 25020:2019-07, **Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality measurement framework.**

[274] ISO/IEC 25021:2012-11, **Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality measure elements.**

[275] ISO/IEC 25024:2015-10, **System und Software-Engineering – Qualitätskriterien und Bewertung von System- und Softwareprodukten (SQuaRE) – Messung der Datenqualität.**

[276] ISO/IEC 27002:2013, **Information technology – Security techniques – Code of practice for information security controls.**

[277] DIN EN ISO/IEC 29100, **Informationstechnik – Sicherheitsverfahren – Rahmenwerk für Datenschutz (ISO/IEC 29100:2011, einschließlich Amd 1:2018).**

[278] ISO/IEC 33063:2015, **Information technology – Process assessment – Process assessment model for software testing.**

[279] DIN ISO/TR 22100-1:2016-07; DIN SPEC 33886:2016-07, **Sicherheit von Maschinen – Beziehung zu ISO 12100 – Teil 1: Wie ISO 12100 und Typ-B- und Typ-C-Normen zusammenhängen (ISO/TR 22100-1:2015).**

[280] DIN ISO/TR 22100-2:2014-09; DIN SPEC 33887:201409, **Sicherheit von Maschinen – Beziehung zu ISO 12100 – Teil 2: Wie ISO 12100 und ISO 138491 zusammenhängen (ISO/TR 22100-2:2013).**

[281] DIN ISO/TR 22100-3:2017-06; DIN SPEC 33888:2017-06, **Sicherheit von Maschinen – Beziehung zu ISO 12100 – Teil 3: Implementierung ergonomischer Grundsätze in Sicherheitsnormen (ISO/TR 22100-3:2016).**

[282] ISO 23412:2020, **Indirect, temperature-controlled refrigerated delivery services – Land transport of parcels with intermediate transfer.**

[283] ISO 25119-1:2018-10, **Tractors and machinery for agriculture and forestry – Safety-related parts of control systems – Part 1: General principles for design and development.**

[284] ISO 25119-2:2019-08, **Tractors and machinery for agriculture and forestry – Safety-related parts of control systems – Part 2: Concept phase.**

[285] ISO 25119-3:2018-10, **Tractors and machinery for agriculture and forestry – Safety-related parts of control systems – Part 3: Series development, hardware and software.**

[286] ISO 25119-4:2018-10, **Tractors and machinery for agriculture and forestry – Safety-related parts of control systems – Part 4: Production, operation, modification and supporting processes.**

[287] DIN EN 60601-1-4:2001-04; VDE 0750-1-4:2001-04, **Medizinische elektrische Geräte – Teil 1-4: Allgemeine Festlegungen für die Sicherheit Ergänzungsnorm: Programmierbare elektrische medizinische Systeme (IEC 606011-4:1996 + A1:1999).**

[288] DIN EN 61511-2; VDE 0810-2:2019-02, **Funktionale Sicherheit – PLT-Sicherheitseinrichtungen für die Prozessindustrie – Teil 2: Anleitungen zur Anwendung von IEC 615111 (IEC 61511-2:2016).**

[289] DIN EN 61511-3; VDE 0810-3:2019-02, **Funktionale Sicherheit – PLT-Sicherheitseinrichtungen für die Prozessindustrie – Teil 3: Anleitung für die Bestimmung der erforderlichen Sicherheits-Integritätslevel (IEC 61511-3:2016).**

[290] DIN EN 61513:2013-09; VDE 04912:2013-09, **Kernkraftwerke – Leittechnik für Systeme mit sicherheitstechnischer Bedeutung – Allgemeine Systemanforderungen (IEC 61513:2011).**

[291] DIN EN 62304:2016-10; VDE 0750101:2016-10, **Medizingeräte-Software – Software-Lebenszyklus-Prozesse (IEC 62304:2006 + A1:2015).**

[292] DIN EN 50128:2012-03; VDE 0831128:2012-03, **Bahnanwendungen – Telekommunikationstechnik, Signaltechnik und Datenverarbeitungssysteme – Software für Eisenbahnsteuerungs- und Überwachungssysteme.**

[293] IEEE 1012:2016, **IEEE Standard for System, Software, and Hardware Verification and Validation.**

[294] ISO/IEC TR 13066-2:2016-02, **Information technology – Interoperability with assistive technology (AT) – Part 2: Windows accessibility application programming interface (API)**.

[295] DIN EN ISO 13482:2014-11, **Roboter und Robotikgeräte – Sicherheitsanforderungen für persönliche Assistenzroboter (ISO 13482:2014)**.

[296] **Wissenschaftsjahr 2019 KI, „Glossar: Lernende Systeme“, Bundesministerium für Bildung und Forschung**. [Online]. Verfügbar unter: https://www.wissenschaftsjahr.de/2019/uebergreifende-informationen/glossar/detail/?tx_dpnglossary_glossarydetail%5Bcontroller%5D=Term&tx_dpnglossary_glossarydetail%5Baction%5D=show&tx_dpnglossary_glossarydetail%5Bterm%5D=27&tx_dpnglossary_glossarydetail%5BpageUid%5D=1016&cHash=2a-2f33e3e34125305328e93c376e424a [Letzter Zugriff 12.08.2020].

[297] ISO/IEC 18023-1:2006-05, **Information technology – SEDRIS language bindings – Part 1: Functional specification**.

[298] S. Jordan, C Nimtz (Hrsg.), **Lexikon Philosophie: Hundert Grundbegriffe.** Stuttgart: Reclam, 2009.

[299] D. Frey, L. K. Schmalzried, **Philosophie der Führung: Gute Führung lernen von Kant, Aristoteles, Popper & Co.** 1. Aufl. Berlin, Heidelberg: Springer-Verlag, 2013 [S. 62 f.].

[300] C. Misselhorn, **Grundfragen der Maschinenethik.** 3. Aufl. Ditzingen: Reclam, 2018.

[301] D. von der Pfordten, „Rechtsethik“, in: J. Nina-Rümelin, **Angewandte Ethik: Die Bereichsethiken und ihre theoretische Fundierung: Ein Handbuch.** 2. Aufl. Stuttgart: Alfred Kröner Verlag, 2005 [S. 207-208].

[302] M. Lutz-Bachmann, **Grundkurs Philosophie.** Bd. 7.: **Ethik**. Ditzingen: Reclam, 2013 [S. 201].

[303] W. Gründinger et al., **Mensch, Moral, Maschine.** Berlin: BVDW, 2019 [Online]. Verfügbar unter: https://www.bvdw.org/fileadmin/bvdw/upload/dokumente/BVDW_Digitale_Ethik.pdf [Letzter Zugriff: 12.08.2020] [S.20].

[304] W. Damm, P. Heidl et al., **Roadmap – Safety, Security, and Certifiability Future Man-Machine Systems, SafeTRANS-Arbeitskreises Resilient, Learning, and Evolutionary Systems**- Oldenburg: SafeTRANS, 2019 [Online]. Verfügbar unter: https://www.safetrans-de.org/de/Aktuelles/aktuelle-roadmap-%22safety%2C-security%2C-and-certifiability-of-future-man-machine-systems%22/285 [Letzter Zugriff: 12.08.2020].

[305] C. Wischhöfer, B. Oberbichler, **Normen-Management-Lösungen: Werknormenanalyse durch die DIN Software GmbH**. DIN-Mitteilungen, November 2015, S. 1014.

[306] M. Esser et al., **Digitale Content-Dienstleistungen aus dem zentralen XML Content Repository: Zentrale Ablage von Inhalten und Trennung der Inhalte von ihrer Darstellungsform**. DIN-Mitteilungen, Oktober 2017, S. 18-23.

[307] ANSI/NISO Z39.102:2017 **STS: Standards Tag Suite.**

[308] NISO, **NISO Working Group to Develop A Standards-Specific Ontology Standard (SSOS)**. [Online]. Verfügbar unter: https://www.niso.org/press-releases/2019/02/niso-working-group-develop-standards-specific-ontology-standard-ssos [Letzter Zugriff 12.08.2020].

[309] M. Schacht, **SMART Standards – Entwicklungsprozess und Contentstruktur**. DIN-Mitteilungen, Juni 2020, S. 36-42.

[310] Ehring, D.; Loibl, A.; Nagarajah, A.; Zhou, L. (2020) **Smart Standards: Automatisierungsansatz – Methodik zur Wissensrepräsentation durch Graphdatenbanken**. In: 18. Gemeinsames Kolloquium Konstruktionstechnik 2020.

[311] A. Loibl et al., **Procedure for the transfer of standards into machine-actionability**. Journal of Advanced Mechanical Design Systems and Manufacturing, 14, JAMDSM0022, 2020.

[312] VDI 2221 Blatt 1:201911, **Entwicklung technischer Produkte und Systeme – Modell der Produktentwicklung**.

[313] VDI 2221 Blatt 2:201911, **Entwicklung technischer Produkte und Systeme – Gestaltung individueller Produktentwicklungsprozesse**.

[314] DIN 8202:202003, **Normungsarbeit – Teil 2: Gestaltung von Dokumenten**.

[315] VDA, **Automotive VDA-Standardstruktur Komponentenlastenheft**. 1. Aufl., Berlin: VDA-QMC, 2007.

[316] D. Czarny et al., **Project 2 Standards of the Future**. Pilot Petroleum sector. CCMC-Report, February 2020

[317] M. Schacht, Normen-Management, in: B. Bender, K. Gericke (Hrsg.), **Pahl/Beitz Konstruktionslehre**. 9. Auflage, Wiesbaden: Springer Vieweg, 2020.

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

9 Autorenverzeichnis

Dr. Rasmus Adler, Fraunhofer IESE

Thomas Andersen, Andersen Marketing KG

Marie Anton, Bundesverband der Arzneimittel-Hersteller e. V.

Dr. Andreas Aschenbrenner, Siemens AG

Yasmeen Babar, regio iT – Gesellschaft für Informations-technologie mbH

Adam Bahlke, Motor-AI

Dr. Markus Bautsch, Stiftung Warentest

Nikolas Becker, Gesellschaft für Informatik e. V. (GI)

Justus Benning, FIR e. V. an der RWTH Aachen

Dr. Jürgen Bock, KUKA Deutschland GmbH

Bogdan Bereczki, ARGO AI

Bastian Bernhardt, IABG mbH

Dr. phil. Marija Bertovic, Bundesanstalt für Materialforschung und -prüfung (BAM)

Katharina Berwing, FIR e. V. an der RWTH Aachen

Tarek R. Besold, PHD, neurocat GmbH

Thordis Bethlehem, Berufsverband Deutscher Psychologin-nen und Psychologen e. V. (BDP)

Paul Beyer, FSD Fahrzeugsystemdaten GmbH

Jörg Bienert, Bundesverband Künstliche Intelligenz e. V.

Antonio Bikic, Ludwig-Maximilians-Universität, München

Dr. Alexander Bode, CONABO GmbH

Jürgen Bönninger, FSD Fahrzeugsystemdaten GmbH

Dr. Julia Borggräfe, Bundesministerium für Arbeit und Soziales (BMAS)

PD Dr. med. Ulrich Bork, Universitätsklinikum Carl Gustav Carus an der Technischen Universität Dresden

Matthias Brand, MBDA Deutschland GmbH

Michael Brolle, Rembe GmbH

Dr. Joachim Bühler, Verband der TÜV e. V.

Dr. Aljoscha Burchardt, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Prof. Dr. Armin Cremers, Rheinische Friedrich-Wilhelms-Universität Bonn

Stephanie Dachsberger, Plattform Lernende Systeme

Sharam Dadashnia, Scheer PAS Deutschland GmbH

Prof. Dr. Markus Dahm, IBM

Susanne Dehmel, Bitkom e. V.

Dr. Peter Deussen, Microsoft Deutschland GmbH

Verena Dietrich, imbus AG

Juergen Diller, Huawei

Alexander Dobert, Datenschutz Dobert

Jannis Dörhöfer, Verband der TÜV e. V. (VdTÜV)

Ralf Egner, Deutsche Akkreditierungsstelle GmbH (DAkkS)

Matthis Eicher, TÜV Süd

Patrik Eisenhauer, Collaborating Centre on Sustainable Consumption and Production gGmbH (CSCP)

Kentaro Ellert, PricewaterhouseCoopers GmbH

Filiz Elmas, DIN e. V.

Dr. rer. nat. Stefan Elmer, Festo SE & Co. KG

Jacques Engländer, FIR e. V. an der RWTH Aachen

Dr. Matthias Fabian, Landesärztekammer Baden-Württemberg

Dr.-Ing. Patrik Feth, SICK AG

Marc Fliehe, Verband der TÜV e. V. (VdTÜV)

Prof. Dr. Alexis Fritz, Katholische Universität Eichstätt-Ingolstadt

Dr. Martina Frost, ifaa – Institut für angewandte Arbeitswissenschaft e. V.

Andreas Fuchsberger, Microsoft Deutschland GmbH

Michael Gamer, Technische Universität Kaiserslautern

Dr. Jens Gayko, Standardization Council Industrie 4.0
Philip Gallandi, NWB e. V.

Prof. Dr. Dagmar Gesmann-Nuissl, Technische Universität Chemnitz

Wolfgang Gies, DVGW Deutscher Verein des Gas- und Wasserfaches e. V.

Marius Goebel, Spherity GmbH

Jan Götze, Airbus

Stephan Griebel, Siemens Mobility GmbH

Yvonne Gruchmann, Wirtschaftsförderung Land Brandenburg GmbH

Norman Günther, Technische Hochschule Wildau

Viktoria Hasse, Bundesverband Gesundheits-IT – bvitg e. V.

Marc Hauer, Technische Universität Kaiserslautern

Dr. Dirk Hecker, Allianz Big Data und Künstliche Intelligenz

Prof. Roland Heger, PhD, Integrata-Stiftung für humane Nutzung der Informationstechnologie

Jürgen Heiles, Siemens AG

Prof. Dr.-Ing. Michael Herdy, inpro Innovationsgesellschaft für fortgeschrittene Produktionssysteme in der Fahrzeugindustrie

Thorsten Hermann, Microsoft Deutschland GmbH

Dr. Sven Herpig, Stiftung Neue Verantwortung e. V.

Dr. Stefan Heumann, Stiftung Neue Verantwortung e. V.

Dr. Wolfgang Hildesheim, IBM Deutschland GmbH

Dr. Lukas Höhndorf, IABG mbH

Taras Holoyad, Bundesnetzagentur

Dr. Kristian Höpping, FSD Fahrzeugsystemdaten GmbH

Stephan Höppner, Atos Information Technology

Oliver Jähn, Landesärztekammer Brandenburg

Prof. Dr.-Ing. habil. Thomas Jürgensohn, HFC Human-Factors-Consult GmbH

Johannes Kalhoff, Phoenix Contact GmbH & Co. KG

Klaus Kaufmann, GS1 Germany GmbH

Michael Kayser, idox compliance

Dr. Till Klein, appliedAI Initiative for applied Artificial Intelligence by UnternehmerTUM

Klaus Kleine Büning, TÜV Nord InfraChem GmbH

Marco Knödler, YNCORIS GmbH & Co. KG

Prof. Dr. habil. Jana Koehler, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Dr. Sergii Kolomiichuk, Fraunhofer-Institut für Fabrikbetrieb und -automatisierung IFF

Roman Konertz, FernUniversität in Hagen

Roland Kossow, CyberTribe®

Tobias Krafft, Technische Universität Kaiserslautern

Dirk Kretzschmar, TÜViT GmbH

Sebastian Kriegsmann, DIN e. V.

Dr. Anna Kruspe, Deutsches Zentrum für Luft- und Raumfahrt (DLR)

Michael Krystek, Physikalisch-technische Bundesanstalt

Mark Küller, Verband der TÜV e. V. (VdTÜV)

Matthias Kuom, Europäische Kommission

Dr. Jens Lachenmaier, Universität Stuttgart

Holger Laible, Siemens AG

Philipp Lämmel, Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS

Fredi Lang, Berufsverband Deutscher Psychologinnen und Psychologen e. V.

Dr.-Ing. Christoph Legat, HEKUMA GmbH

Dr. Olga Levina, FZI Forschungszentrum Informatik

Matthias Lieske, Hitachi Europe GmbH

Georg Ludwig Lindinger, Universität Bayreuth

Daniel Loevenich, Bundesamt für Sicherheit in der Informationstechnik (BSI)

Prof. Dr. Ulrich Löwen, Siemens AG

Dr. Jackie Ma, Fraunhofer-Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, HHI

Dr.-Ing. Stefan Maack, Bundesanstalt für Materialforschung und -prüfung (BAM)

Christian Märkel, WIK GmbH

Prof. Dr. Klaus Mainzer, TUM Senior Excellence Faculty, Technische Universität München

Angelina Marko, Fraunhofer-Institut für Produktionsanlagen und Konstruktionstechnik IPK

Dr. Erik Marquardt, Verein Deutscher Ingenieure e. V. (VDI)

Jan de Meer, Hochschule für Technik und Wirtschaft Berlin (HTW Berlin)

Carsten Mehrtens, Volkswagen AG

Iris Merget, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Martin Meyer, Siemens Healthcare GmbH

Dr. Michael Mock, Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIS

Prof. Dr. Andreas Mockenhaupt, Hochschule Albstadt-Sigmaringen (Albstadt-Sigmaringen University)

Thomas Möller, Bundesverband Gesundheits-IT – bvitg e. V.

Michael Mörike, Integrata-Stiftung für humane Nutzung der Informationstechnologie

Edeltraud Mörl, Dachverband für Technologen/-innen und Analytiker/-innen in der Medizin Deutschland e. V.

Andreas Müller, Schaeffler Technologies AG & Co. KG

Tobias Nagel, Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

Jan Noelle, Rettungsdienst-Kooperation in Schleswig-Holstein gGmbH

Dr. Shane O'Sullivan, Universidade de São Paulo, Brazil

Michael Paul, Safran S.A.

Fabian Petsch, Bundesamt für Sicherheit in der Informationstechnik (BSI)

Dr. Christoph Peylo, Robert Bosch GmbH

Christoph Pogorelow, IBM Deutschland GmbH

Frank Poignée, infoteam Software AG

Dr. Maximilian Poretschkin, Fraunhofer-Institut für Intelligente Analyse und Informationssysteme IAIS

Alexander Rabe, eco – Verband der Internetwirtschaft e. V.

Golo Rademacher, Bundesministerium für Arbeit und Soziales (BMAS)

Dr. Frank Raudszus, Bundesnetzagentur

Ludwig von Reiche, NVIDIA ARC GmbH

Janis Reinelt, AICURA Medical GmbH

Axel Rennoch, Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS

Dr. Mathias Riechert, BMW Group

Dr. Patrick Riordan, Siemens AG

Marcus Röhler, Fraunhofer-Institut für Gießerei-, Composite- und Verarbeitungstechnik IGCV

Dr. Miriam Ruf, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB

Dr. Gerhard Runze, imbus AG

Ingo Sawilla, TRUMPF Werkzeugmaschinen GmbH + Co. KG

Dr.-Ing. Mario Schacht, DIN e. V.

Dr. med. Henning Schaefer, Ärztekammer Berlin

Dr. Stefan Schaffer, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Elias Schede, PricewaterhouseCoopers GmbH (PwC)

Christopher Scheel, SCHUFA Holding AG

Prof. Dr.-Ing. Ina Schieferdecker, Bundesministerium für Bildung und Forschung (BMBF)

Raoul Schönhof, Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

Dr. Thomas Schmid, Universität Leipzig

Dr. Jörg Schneider, Bundesnetzagentur

Martin A. Schneider, Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS

MinDirig Stefan Schnorr, BMWi

Dr. Kinga Schumacher, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Silvio Schwarzkopf, FSD Fahrzeugsystemdaten GmbH

Peter Seeberg, asimovero.ai

Jan Seitz, Technische Hochschule Wildau

Aydin Enes Seydanlioglu, Robert BOSCH GmbH

Annegrit Seyerlein-Klug, secunet Security Networks AG

Dr. Reiner Spallek, IABG mbH

Dr. Thomas Stauner, BMW Group

Andreas Steier, Deutscher Bundestag

Dr. Reinhard Stolle, Argo AI

Dr. Manfred Stoyke, Bundesamt für Verbraucherschutz und Lebensmittelsicherheit (BVL)

Prof. Dr. Karolina Suchowolec, Technische Hochschule Köln

Julia Szelag, AWV Arbeitsgemeinschaft für wirtschaftliche Verwaltung e. V.

Steffen Tauber, evia

Dr. Nikolay Tcholtchev, Fraunhofer-Institut für Offene Kommunikationssysteme FOKUS

Martin Tettke, Berlin Cert GmbH

Dr. Volker Treier, Deutscher Industrie- und Handelskammertag e. V.

Dr. Denise Vandeweijer, UnternehmerTUM GmbH

Dr. Silvia Vock, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)

Dr. Thomas Vögele, Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI)

Roland Vogt, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Kirsten Wagner, DVGW Deutscher Verein des Gas- und Wasserfaches e. V.

Prof. Dr. rer. nat. Dr. h.c. mult. Wolfgang Wahlster, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Dr. Thomas Waschulzik, Siemens Mobility GmbH

Prof. Dr.-Ing. Dieter Wegener, Siemens AG

Prof. Dr. Johann Wilhelm Weidringer, Bundesärztekammer (Bayer. Landesärztekammer)

Wei Wei, IBM Deutschland GmbH

Dr. Frank Werner, Software AG

Martin Westhoven, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin (BAuA)

Dr. Johannes Winter, Plattform Lernende Systeme

Christoph Winterhalter, DIN e. V.

Raoul Wintjes, DSLV Bundesverband Spedition und Logistik e. V. (DSLV)

René Wöstmann, RIF e. V. Institut für Forschung und Transfer

Yuanyuan Xiao, BTC AG

Jens Ziehn, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB

10 Weitere Mitglieder der Arbeitsgruppen

Fabian Anzmann, HKI Industrieverband Haus-, Heiz- und Küchentechnik e. V.

Eberhard Becker, DEMAG CRANES & COMPONENTS GMBH

Dr. Andreas Binder, Samson AG

Miika Blinn, Verbraucherzentrale

Thomas Boué, BSA | The Software Alliance

Gebhard Bouwer, TÜV Rheinland Industrie Service GmbH

Dr. Konstantin Böttinger, Fraunhofer-Institut für Angewandte und Integrierte Sicherheit AISEC

Dr. Alfonso Caiazzo, WIAS Berlin

Beatriz Cassoli, Technische Universität Darmstadt

Klaus Däßler, Gesellschaft für Mathematische Intelligenz

Thierry Declerck, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Alberto Diaz-Durana, HEDERA Sustainable Solutions GmbH

Dr. Markus Dicks, Bundesministerium für Arbeit und Soziales (BMAS)

Jörg Dubbert, VDI/VDE Innovation + Technik

Heiko Ehrich, TÜV NORD Mobilität GmbH & Co. KG IFM

Bernd Eisemann, Munich RE

Karl-Ludwig Elfira Blumenthal, Siemens AG

Alexandra Engelt, DIN e. V.

Prof. Dr. Stefan Evert, Friedrich-Alexander-Universität Erlangen-Nürnberg

Dr. med. Matthias Fabian, Landesärztekammer Baden-Württemberg

Dr.-Ing. Stephan Finke, DAkkS – Deutsche Akkreditierungsstelle

Robert Freund, Bundesanstalt für Materialforschung und -prüfung (BAM)

Egbert Fritzsche, VDA

Enno Garrelts, Universität Stuttgart

Marian Gläser, brighterAI

Richard Goebelt, Verband der TÜV e. V.

Benedikt Grosch, Technische Universität Darmstadt

Dr. Oliver Grün, BITMi

Dr. Thilo Hagendorff, Universität Tübingen

Andreas Hartl, Bundesministerium für Wirtschaft und Energie (BMWi)

Manfred Hefft, Domino Deutschland GmbH

Stefan Herr, innogy SE

Dr. Klaus Hesselmann, yourexpertcluster

Max Hofmann, Volkswagen AG

Dr. rer. pol. Reiner Hofmann, Universität Bayreuth

Dr.-Ing. Gerhard Imgrund, VDE

Dr. Andreas Jedlitschka, Fraunhofer-Institut für Experimentelles Software Engineering IESE

Stephan Jenzen, Airbus

Prof. Dr. Christian Johner, Johner Institut GmbH

Ninmar Lahdo, VDE

Johannes Koch, VDE

Stephan Krähnert, VDA

Danny Lubosch, Gefertec GmbH

Prof. Dr. Christoph Lütge, Technische Universität München

Dr. Oliver Maguhn, Munich RE

Johannes Melzer, Deutscher Industrie- und Handelskammertag e. V.

Dirk Michelsen, IBM Deutschland GmbH

Sebastian Micus, Deutsches Institut für Textil- und Faserforschung Denkendorf (DITF)

Stefan Mitterer, OELCHECK GmbH

Andreas Möller, ADVES GmbH & Co. KG

Prof. Dr. Jürgen Mottok, Ostbayerische Technische Hochschule Regensburg

Gert Nahler, Samson AG

Dr. med. Felix Nensa, Universitätsklinikum Essen

Thomas Niessen, Kompetenznetzwerk Trusted Cloud e. V.

Luis Oala, Fraunhofer-Institut für Nachrichtentechnik HHI

Sebastian von Oppen, Architektenkammer Berlin

Dr.-Ing. Christian Peter, BioArtProducts GmbH

Dr. Georg Plasberg, SICK AG

Christoph Preuße, Berufsgenossenschaft Holz und Metall

Dr. Gerald Quitterer, Bayerische Landesärztekammer

Dr. Martin Radtke, Bundesanstalt für Materialforschung und -prüfung (BAM)

Dr. phil. Georg Rehm, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)

Prof. Dr.-Ing. Jörg Reiff-Stephan, Technische Hochschule Wildau

Guido Reusch, eapr GmbH

Christina Rode-Schubert, trend2ability

Laurent Romary, INRIA

Jan Rösler, DIN e. V.

Nils Röttger, imbus AG

Dennis Scheuer, IBM Deutschland

Robin Schlenga, Ramboll Management Consulting GmbH

Prof. Dr. Ralf Schnieders, Hochschule für Technik und Wirtschaft Berlin

Andreas Schumann, Bundesverband der Kurier-Express-Post-Dienste

Dr.-Ing. Dennis Schüthe, Still GmbH

Philip Sperl, Fraunhofer-Institut für Angewandte und Integrierte Sicherheit AISEC

Johann-Christoph Stang, Bundesanstalt für Materialforschung und -prüfung (BAM)

Patrick Stanula, Technische Universität Darmstadt

Timo Strohmann, Technische Universität Braunschweig

Klaus Strumberger, Morgen & Morgen GmbH

Denis Suarsana, Bundesvereinigung der Deutschen Arbeitgeberverbände (BDA)

Claudia Tautorus, Verband der TÜV e. V.

Dr.-Ing. Max Ungerer, PROSTEP AG

Dr. David Urmann, VDE

Dirk Walther, Fahrzeugsystemdaten GmbH

Dr. Markus Wenzel, Fraunhofer-Institut für Nachrichtentechnik HHI

Dr. Jing Xiao, Continental Automotive

Salah Zayakh, REWE digital GmbH

Dr. Carlos Zednik, Universität Magdeburg

Dr. Thomas Zielke, Bundesministerium für Wirtschaft und Energie (BMWi)

11 Anhang

11.1 Glossar

Die in Tabelle 14 aufgeführten Begriffe sind folgenden Gebieten zugeordnet:

1 Künstliche Intelligenz (beinhaltet wegen der Beschränkung im Glossar die allgemeinen Begriffe)
2 Eigenschaften von KI-Systemen
3 Eigenschaften von Daten
4 Methoden und Techniken
5 Maschinelles Lernen (beinhaltet auch neuronale Netze)

Tabelle 14: Glossar

Gebiet	Deutsch (de)	de alternativ	Englisch (en)	Beschreibung und Quelle[30]
1	Agent	Softbot	agent	Als Agent bezeichnet man ganz allgemein eine Software- oder Hardware-Einheit, welche Informationen verarbeitet und aus einer Eingabe eine Ausgabe produziert.
3	Aktualität		Currentness	Grad der zeitlichen Gültigkeit von Daten, der in einem bestimmten Anwendungskontext relevant ist. [88]
5	Angeleitetes maschinelles Lernen	überwachtes Lernen	supervised machine learning	Technik maschinellen Lernens, die auf der Verwendung vorklassifizierter Daten beruht[31]
1	Automatisierung		automation	Ersetzung manueller Tätigkeiten durch computerisierte Methoden [294]
2	Autonomie		autonomy	Fähigkeit eines Systems, Aufgaben basierend auf seinem internen Zustand und seiner Umgebung ohne menschliche Einflussnahme auszuführen[32] [295]
1	Big Data		Big Data	Daten, die zu umfangreich, zu komplex, zu kurzlebig oder zu schwach strukturiert sind, um sie mit herkömmlichen Methoden der Datenverarbeitung auszuwerten.[33]
3	Datenqualität		data quality	Grad, in dem Charakteristiken von Daten explizit spezifizierte oder implizierte Anforderungen erfüllen. [88][34]
3	Datenschutz		Data protection	Bezieht sich auf die Erhebung und Verarbeitung personenbezogener Daten gemäß den einschlägigen Vorschriften wie die EU-Datenschutz-Grundverordnung. Beispielsweise haben betroffene Personen in Europa das Recht, dass ihre privaten Daten ausreichend vor IT-Angriffen geschützt werden.

30 Wo keine Quelle angegeben ist, stammt die Beschreibung vom Autorenteam dieser NRM. Ist die Quelle ein unveröffentlichtes Arbeitspapier, so wird diese in einer Fußnote angegeben, sonst per Literaturangabe.

31 Quelle: ISO/IEC CD 22989 (Laufendes Projekt, nicht veröffentlicht)

32 In der ISO-Beschreibung wird der Begriff „Automatik“, nicht jedoch „Autonomie“ (= „im eigenen Namen“ und somit sanktionsfähig und in eigener Verantwortung) beschrieben. Ein Gegenbeispiel ist eine Wildkamera mit vollautomatischer Einstellung und Aufnahme, die gewiss nicht als autonom angesehen wird. Im Übrigen ist ein autonomes System wohl eher mehr als nur ein kognitives System.

33 Quelle: ISO/IEC WD 20546 (Arbeitsentwurf Stand 2019-08-04, laufendes Projekt, nicht veröffentlicht.) ISO note to entry: Big data is commonly used in many different ways, for example as the name of the scalable technology used to handle big data extensive datasets.

34 Abgeleitet aus DIN EN ISO 9000:2015 [105]. Die Datenqualität wird in Charakteristiken oder auch Dimensionen beschrieben, etwa mit inhärenten Merkmalen wie Genauigkeit und Vollständigkeit oder systemabhängigen Merkmalen wie Verfügbarkeit und Wiederherstellbarkeit.

Gebiet	Deutsch (de)	de alternativ	Englisch (en)	Beschreibung und Quelle[30]
5	Deep Neural Network		Deep Neural Network	Neuronales Netzwerk, das neben der Eingabe- und Ausgabeschicht noch über weitere, sog. versteckte Schichten von Knoten verfügt (vgl. „tiefes Lernen").
2	Erklärbarkeit	Nachvollzieh-barkeit	explainability	Eigenschaft eines KI-Systems, dass Faktoren, die zu einer automatisierten Entscheidung des Systems geführt haben, durch einen Menschen verstanden werden können[35]
4	Expertensystem			Häufig regelbasiertes System, das auf symbolischer Wissensverarbeitung beruht. Beispiel: Wenn-dann-Regeln. → Symbolische, formale Repräsentation von Wissen in KI-Systemen Schlussfolgerung, mithilfe von Logik aus formalem Wissen neues Wissen herzuleiten
3	Genauigkeit		Accuracy	Grad, in dem Daten die Attribute eines Konzepts oder Ereignisses in einem gegebenen Anwendungskontext korrekt beschreiben. [88]
2	Grad der Zuverlässigkeit		dependability	Fähigkeit zur Ausführung in der geforderten Art und zum geforderten Zeitpunkt [105]
4	Inferenz	logisches Schließen	inference reasoning	Regelbasiertes Schlussfolgern; wird häufig in Expertensystemen eingesetzt
1	KI-Komponente		AI component	Systemkomponente, die Künstliche Intelligenz benutzt
1	KI-Modul		AI module	Software-Modul, das KI-Methoden implementiert [86]
1	KI-System		AI system	System, das Künstliche Intelligenz benutzt[36]
1	Kognitives System			Anpassungsfähiges System mit Schnittstellen zur digitalen Welt und zur Umwelt, das Dinge selbsttätig wahrnehmen, auf Kontexte beziehen und verstehen sowie Schlüsse daraus ziehen und lernen kann, um Aufgabenstellungen zu lösen und zu bewältigen
5	Kontinuierliches Lernen		continuous learning	Inkrementelles Training eines KI-Systems, das fortlaufend in der Produktionsumgebung des Systems stattfindet[37]
2	Kontrollier-barkeit	Steuerbarkeit	controllability	Fähigkeit eines menschlichen Operators, in zeitgerechter Weise in die Funktion eines Systems einzugreifen[38]
5	Lerndaten	Trainingsdaten	training data	Daten, die zum Training eines Modells verwendet werden[39]
1	Lernendes System		Learning system	Lernende Systeme sind Maschinen, Roboter und Softwaresysteme, die abstrakt beschriebene Aufgaben auf Basis von Daten, die ihnen als Lerngrundlage dienen, selbstständig erledigen, ohne dass jeder Schritt spezifisch vom Menschen programmiert wird. Um ihre Aufgabe zu lösen, setzen sie von Lernalgorithmen trainierte Modelle ein. Mithilfe des Lernalgorithmus können viele Systeme im laufenden Betrieb weiterlernen: Sie verbessern die vorab trainierten Modelle und erweitern ihre Wissensbasis. [296]

35 Quelle: ISO/IEC CD 22989 (Laufendes Projekt, nicht veröffentlicht).

36 Quelle: ebd.

37 Quelle: ebd.

38 Quelle: ebd.

39 Quelle: ebd.

Gebiet	Deutsch (de)	de alternativ	Englisch (en)	Beschreibung und Quelle[30]
4	Maschinelle Übersetzung		machine translation	Automatische Übersetzung von gesprochener oder geschriebener natürlicher Sprache in eine andere Sprache durch ein KI-System[40]
5	Maschinelles Lernen		machine learning	Technik, die ein System in die Lage versetzt, aus Daten und Interaktionen zu lernen
1	Modell		model	Physikalische, mathematische oder logische Repräsentation eines Systems, eines Objekts, eines Phenomäns oder eines Prozesses. [297]
5	Neuronales Netz	künstliches neuronales Netz, KNN	artificial neural net	Berechnungsnetzwerk bestehend aus einfachen Berechnungselementen und gewichteten Beziehungen zwischen diesen Elementen, dessen Ein-Ausgabe-Funktion durch das Zusammenspiel der Netzwerkelemente bestimmt wird[41]
1	Roboter		robot	Ein Roboter ist ein technisches System, das über Sensoren zur Wahrnehmung seiner Umwelt, einer zweckorientierten Verarbeitungseinheit und Effektoren zur Veränderung seiner räumlichen Relation in der Umwelt oder der Umwelt selbst verfügt.[42]
4	Robotik		robotic	Disziplin, die sich mit der Konstruktion von Robotern beschäftigt.
2	Robustheit		Robustness	Fähigkeit eines Systems, seine Funktion unter beliebigen Umständen zu erfüllen.[43]
2	Safety	Sicherheit	Safety	Bezieht sich auf die Erwartung, dass ein System unter bestimmten Umständen nicht zu einem Zustand führt, in dem menschliches Leben, Gesundheit, Eigentum oder die Umwelt gefährdet sind.
5	Schwach überwachtes maschinelles Lernen	teilüberwachtes Lernen	semi-supervised machine learning	Grad der zeitlichen Gültigkeit von Daten, der in einem bestimmten Anwendungskontext relevant ist.[44]
1	Schwache KI		narrow (or weak) AI	Zu einem bestimmten Zweck konzipiertes KI-System
2	Security	Sicherheit	Security	Hat zum Ziel, negative Auswirkungen, die ein Mensch oder eine andere Maschine auf das KI-Modul haben kann, zu verhindern. Vertraulichkeit, Integrität und Verfügbarkeit sind die wichtigsten Sicherheitsziele.

40 Quelle: ebd.

41 Quelle: ebd.

42 Quelle: ebd.

43 Quelle: ISO/IEC WD 24029-1 (Arbeitsentwurf Stand 2019-10-30, laufendes Projekt, nicht veröffentlicht)

44 Quelle: ISO/IEC WD 20546 (Arbeitsentwurf Stand 2019-08-04, laufendes Projekt, nicht veröffentlicht.) ISO note to entry: The training data for a semi-supervised learning task can include a majority of unlabelled inputs.

Gebiet	Deutsch (de)	de alternativ	Englisch (en)	Beschreibung und Quelle[30]
4	Semantische Berechnung	semantische Technologien	semantic computing	Technologien, die die Repräsentation und die Verarbeitung von Wissen zum Ziel haben[45]
1	Starke KI		general (or strong) AI	(Theoretisches Konstrukt:) allgemeine Intelligenz, die sich selbst Ziele setzen kann[46]
5	Tiefes Lernen	mehrschichtiges Lernen	deep learning	Technik maschinellen Lernens, die auf künstlichen neuronalen Netzen mit mehreren versteckten Schichten basiert
5	Trainiertes Modell	angelerntes Modell	trained model	Modell, das aus dem maschinellen Lernen resultiert[47]
5	Training		training	Prozess zur Etablierung oder Verbesserung von Modellen mithilfe maschinellen Lernens[48]
2	Transparenz		transparency	Offene, vollständige, verständliche und zugreifbare Darstellung von Informationen zu funktionalen Aspekten eines KI-Systems. Dies beinhaltet u. a. die Erklärbarkeit des KI-Systems (z. B. neuronale Netze), die Nachvollziehbarkeit des Datenschutzkonzepts sowie Informationen zu Qualitätssicherungsprozessen während der Entwicklung.
5	Unüberwachtes maschinelles Lernen	unbeaufsichtigtes maschinelles Lernen	unsupervised lerning	Technik maschinellen Lernens, die auf der Verwendung von nicht vorklassifizierten Daten beruht[49]
2	Verständlichkeit		Understandability	
5	Verstärkendes Lernen	bestärkendes Lernen	reinforcement learning	Technik maschinellen Lernens, die auf der positiven oder negativen Bewertung von Versuchen eines Systems beruht[50]
2	Vollständigkeit		Completeness	Grad, in dem Daten, die mit einer Entität assoziiert sind, Werte für all Attribute dieser Entität sowie für zu dieser in Beziehung stehenden Entitäten aufweisen. [88]
4	Wissensrepräsentation		knowledge representation	Repräsentation von Wissen, die für ein KI-System, z. B. ein Expertensystem, nutzbar ist
3	Zugänglichkeit	Verfügbarkeit		

45 Quelle: ISO/IEC CD 22989 (Laufendes Projekt, nicht veröffentlicht).

46 s. auch [43]

47 Quelle: ISO/IEC CD 22989 (Laufendes Projekt, nicht veröffentlicht).

48 Quelle: ebd.

49 Quelle: ebd.

50 Quelle: ebd.

11.2 Philosopische Grundlagen zur Ethik

Um sich mit Ethik in Bezug auf KI-Systeme auseinandersetzen zu können, sollte man sich mit den Grundlagen der Philosophie und damit ihrem Spezialgebiet, der Ethik, in unserem Kulturkreis auseinandersetzen. Im Allgemeinen versteht man unter Philosophie (altgriechisch φιλοσοφία, latinisiert **philosophia**, wörtlich „Liebe zur Weisheit“) den Versuch, die Welt und die menschliche Existenz zu ergründen, zu deuten und zu verstehen. Von anderen Wissenschaftsdisziplinen unterscheidet sich die Philosophie dadurch, dass sie sich oft nicht auf ein spezielles Gebiet oder eine bestimmte Methodologie begrenzt, sondern durch die Art ihrer Fragestellungen und ihre besondere Herangehensweise an ihre vielfältigen Gegenstandsbereiche charakterisiert ist (vgl. Einführender Artikel Wikipedia). Es gibt keine universale philosophische Methode; im Gegenteil, es existieren eine Vielzahl von ihnen, die sich wiederum an gewissen Strömungen festmachen, wie z. B. die sogenannte Hermeneutik, wobei diese als allgemein anerkanntes Verfahren in den Geisteswissenschaften zählt. Die Hermeneutik bezeichnet so etwas wie verstehende Interpretation von Dokumenten des Bewusstseins, ein Verfahren interpretierender Auslegungskunst, aber auch insgesamt eine philosophische Theorie des Verstehens in seinen Voraussetzungen, Grundlagen und Ergebnissen. Aufgrund dieser Breite ist der Begriff Hermeneutik in unterschiedlichsten Theoriezusammenhängen anzutreffen, umgekehrt wissen Kritiker der Hermeneutik oft nicht, wo sie ansetzen sollen. Als eine weitere Methode sei die Dialektik genannt; eine philosophische Methode, die die Position, von der sie ausgeht, durch gegensätzliche Behauptungen infrage stellt und in der Synthese beider Positionen eine Erkenntnis höherer Art zu gewinnen sucht.

Die philosophische Herangehensweise wurde durch Sigmund Freud (1856–1939) ergänzt, der eine nennenswerte Bedeutung für eine Veränderung in der Betrachtung des Menschen legte. Er war ein österreichischer Neurophysiologe, Tiefenpsychologe, Kulturtheoretiker und Religionskritiker. Insbesondere durch die Begründung der Psychoanalyse gilt er als einer der einflussreichsten Denker und Weltveränderer des 20. Jahrhunderts. Seine Theorien und Methoden werden bis heute diskutiert, angewendet und kritisiert. Die Psychoanalyse war deshalb so wegweisend, weil sie erstmals einen Zugang zum Unbewussten und damit zum Handeln der Menschen und der Betrachtung des Seins, einen anderen Zugang ermöglichten. Aus der Psychoanalyse haben sich später die verschiedenen Schulen der Psychologie entwickelt.

Das Lexikon Philosophie [298] liefert eine gute Ausgangsdefinition zur Ethik: „sie wird definiert als jener Teilbereich der Philosophie, der sich mit den Voraussetzungen und der Bewertung menschlichen Handelns befasst, und ist das methodische Nachdenken über die Moral. Im Zentrum der Ethik steht das spezifisch moralische Handeln, insbesondere hinsichtlich seiner Begründbarkeit und Reflexion (Ethik beschreibt und beurteilt Moral kritisch). (...) Die Ethik und ihre benachbarten Disziplinen (z. B. Rechts-, Staats- und Sozialphilosophie) werden auch als „praktische Philosophie“ zusammengefasst, da sie sich mit dem menschlichen Handeln befasst.“

„In der Ethik können die Teilgebiete der normativen Ethik, Metaethik und angewandten Ethik unterschieden werden. Die normative Ethik entwickelt wertende Theorien über wünschenswertes Handeln. Gegenstand der Metaethik ist die normative Ethik selbst – sie hinterfragt beispielsweise ihre Grundannahmen oder analysiert die Prozesse der normativen Ethik. Die angewandte Ethik fokussiert spezifische Lebensbereiche und versucht diese unter Berücksichtigung der normativen Ethik und Metaethik zu reflektieren und zu gestalten.“ [299]

Es gibt diverse neuzeitliche Ethikansätze, die man auf die künstliche Intelligenz übertragen kann, und auch Philosophen und Werke, die sich mit ethischer KI auseinandersetzen. Interessant ist hier, dass sich die Ethik erstmals auf eine Maschine, statt alleinig auf den Menschen bezieht.

Wenngleich eine KI-Ethik noch nicht klar und abschließend definiert ist, so ist sie sicherlich im Bereich der angewandten Ethik zu verorten. Sie besitzt, wo es um die ethischen Überlegungen in Bezug auf die technischen Aspekte der KI geht, starke Bezüge zu dem Querschnittsbereich Maschinenethik [300], wo es um die sozio-technischen und ökonomischen Aspekte geht, hat sie starke Bezüge zu der Wirtschaftsethik. Darüber hinaus wird sie regelmäßig Bezüge zu den vertikalen Bereichen der Angewandten Ethik besitzen, wie beispielsweise der Bio- oder Medizinethik, wann immer bereichsspezifische ethische Überlegungen unter Berücksichtigung des Phänomens KI aktualisiert werden sollen.

Zudem haben sich in vielen Anwendungsgebieten „Ethiken“ entwickelt, die sich aus den spezifischen Herausforderungen einzelner Anwendungsfelder entwickelt haben, wie z. B.

→ die Rechtsethik, sie ist sowohl ein Teil der Rechtswissenschaften als auch Teil der angewandten Philosophie.

Sie unterscheidet sich in zwei grundlegenden Aspekten von den anderen „Wissenschaftsethiken". „Zum einen treffen ethische und moralische Normen hier nicht auf einen stärker faktisch bzw. naturgesetzlich strukturierten Wirklichkeitsausschnitt – wie etwa Natur, Technik und Medizin, sondern auf das Recht, als eine Begriffs- und Normenordnung, die die Wirklichkeit grundsätzlich [...] normativ überwölbt und gestaltet und begrifflich strukturiert." Zum Zweiten beschäftigt sich die Rechtsethik mit diesen Fragen schon, seit es menschliche Gesellschaften und ihre philosophischen Betrachtungen gibt [300].

→ Die Medizinethik oder medizinische Ethik beschäftigt sich mit den sittlichen Normsetzungen, die für das Gesundheitswesen gelten sollen. Sie hat sich aus der ärztlichen Ethik entwickelt, betrifft aber alle im Gesundheitswesen tätigen Personen, Institutionen und Organisationen und nicht zuletzt die Patienten. Nahestehende Disziplinen sind die Medical Humanities und die Bioethik. Als grundlegende Werte gelten das Wohlergehen des Menschen, das Verbot zu schaden (lat.: „nihil nocere/neminem laedere!") und das Recht auf Selbstbestimmung der Patienten (Prinzip der informierten Zustimmung), allgemeiner das Prinzip der Menschenwürde [302].

Zum jetzigen Zeitpunkt geht man von einer Ethik im Sinne eines Allgemeinverständnisses aus, die gewährleistet, dass KI-Systeme letztlich in der Anwendung sowohl unseren gesetzlichen Regeln folgen als auch „verantwortungsvoll" mit menschlichen Werten unserer Gesellschaft umgehen. Diesen kann man sich auch mittels ethischer Kriterien von Fachverbänden (z. B. High-Level-Group der EU oder der Plattform Lernende Systeme) nähern.

Insbesondere aus der Presse und den Medien – und damit das Bewusstsein der Öffentlichkeit betreffend – sowie teilweise in der Forschung, werden im Zusammenhang mit dem Einsatz und der rasanten Weiterentwicklung sowie dem bereits eingesetzten, aber auch anvisierten Einsatz der KI häufig ethische Dilemmata in den Raum gestellt, wie z. B. bei einem automatisierten Fahrzeug die Frage nach dessen „Verhalten" in kritischen Verkehrssituationen. Soll das KI-System sich ggf. für eine Person gegenüber einer Gruppe bei Gefahr für Leib und Leben entscheiden?

Spricht man also über Ethik in Bezug auf KI, ist das Eingehen auf ethische/moralische Dilemmata eine Notwendigkeit, auch wenn das Ziel sein muss, sie bereits vorab zu verhindern, sprich, die autonome Maschine so zu konzipieren, dass es niemanden gefährden würde. Oder aber man „füttert" das KI-System bereits vorab mit den ethischen Werten unseres Kulturkreises.

Ein **ethisches Dilemma** ist eine Situation, bei der eine Handlungsentscheidung erforderlich ist, obwohl jede mögliche Handlungsoption, einschließlich der des Nicht-Handels, unweigerlich gegen ein Ethikpostulat verstößt. Da der Einsatz von KI-Systemen immer zu einem gewissen Maß mit Kontrollverlust einhergeht und damit Risiken schafft, findet stets eine implizite Abwägung zwischen den potenziellen Gefahren und dem potenziellen Nutzen der KI statt. Besonders kritisch ist dies, wenn die Gesundheit oder das Leben von Menschen potenziell in Gefahr gebracht werden können. Das Bundesverfassungsgericht hat die Möglichkeit des „Verrechnens" von Menschenleben bereits 2006 ausgeschlossen [303]. Die Ethik-Kommission bestätigte dies 2017 prinzipiell für automatisierte Fahrzeuge, öffnete sich aber wie folgt: „**Eine allgemeine Programmierung auf eine Minderung der Zahl von Personenschäden kann vertretbar sein**".

Das **Prinzip der Doppelwirkung (PDW)** behandelt die Frage der moralischen Verantwortung, wenn eine moralisch gute Entscheidung eine (unbeabsichtigte) ethisch schlechte Nebenwirkung hat. Ein Sonderfall des PDW ist die **Doppelverwendbarkeit (Dual Use)**. Hierbei stellt sich die Frage, ob ein Entwickler bzw. Hersteller für eine von ihm unbeabsichtigte oder untersagte schädliche Nutzung verantwortlich ist. Der Begriff stammt ursprünglich aus der Exportkontrolle von gleichzeitig zivil und militärisch nutzbaren Produkten, wird aber auch auf ethische Dilemmata angewendet.

Sowohl der Exkurs in die konkrete ethische Fragestellung der Dilemmata quasi am Ende der Kette, als auch die Betrachtung der ethisch-philosophischen Gesamtentwicklung unserer Gesellschaft machen deutlich, dass bei der Entwicklung und dem Einsatz von KI keine übergeordnete und universell einsetzbare Ethik existiert, aus der man gültige Regeln ableiten kann. Die zuvor angerissene philosophische und ethische Entwicklung unserer Gesellschaft, insbesondere bezogen auf die generellen Werte, macht hier ganz deutlich, dass der europäische Kulturkreis einen geeigneten Rahmen, der sich mit unseren Gesetzen vereinbaren lässt, aus (westlichen) Werten und Normen entwickeln und ableiten muss.

11.3 SafeTRANS Roadmap

Ein Modell (siehe Abbildung 30) zur Nachvollziehbarkeit der Komplexität von Mensch-Maschinen-Systemen und somit ein gutes Beispiel für die Verzahnung von Safety und Security liefert die Roadmap des SafeTRANS-Arbeitskreises mit dem Titel „Safety, Security, and Certifiability of Future Man-Machine Systems“. Danach sollen mit Menschen interagierende KI-Systeme unter den fünf Gesichtspunkten „Systemstärke (Strength)“, „Kontext (Context)“, „Kooperation (Cooperation)“, „Verantwortung und Reflexion (Responsibility & reflection)“ sowie „Integrität und Zertifizierung (Integrity & certification)“ charakterisiert werden können. Den Gesichtspunkten sind Zielvektoren mit Skalen für spezifische Prozesse, Methoden und Fähigkeiten zugewiesen.

Die thematische Kongruenz zu dieser Normungsroadmap KI kommt über den Vektor „Verantwortung und Reflexion“ sowie „Integrität und Zertifizierung“. Sie stellen dar, wie Entscheidungen des Systems auf rechtlicher, ethischer sowie moralischer Ebene abgewogen werden („Verantwortung und Reflexion“) sowie die Bewertung einer Entscheidung nach Konsistenz, Vertrauenswürdigkeit und Risikoeinstufung („Integrität und Zertifizierung“) erfolgen können. Die weiteren Zielvektoren sind weiterführend beschrieben. „Systemstärke“ ist durch Autonomie-, Intelligenz- sowie Evolutionsgrade repräsentiert. Unter „Kontext“ wird eine Analyse der menschlichen sowie physikalischen Umgebung des zu untersuchenden Systems vorgeschlagen. Die „Kooperation“ erwägt eine Unterstützung durch weitere Systeme oder einen menschlichen Eingriff [304].

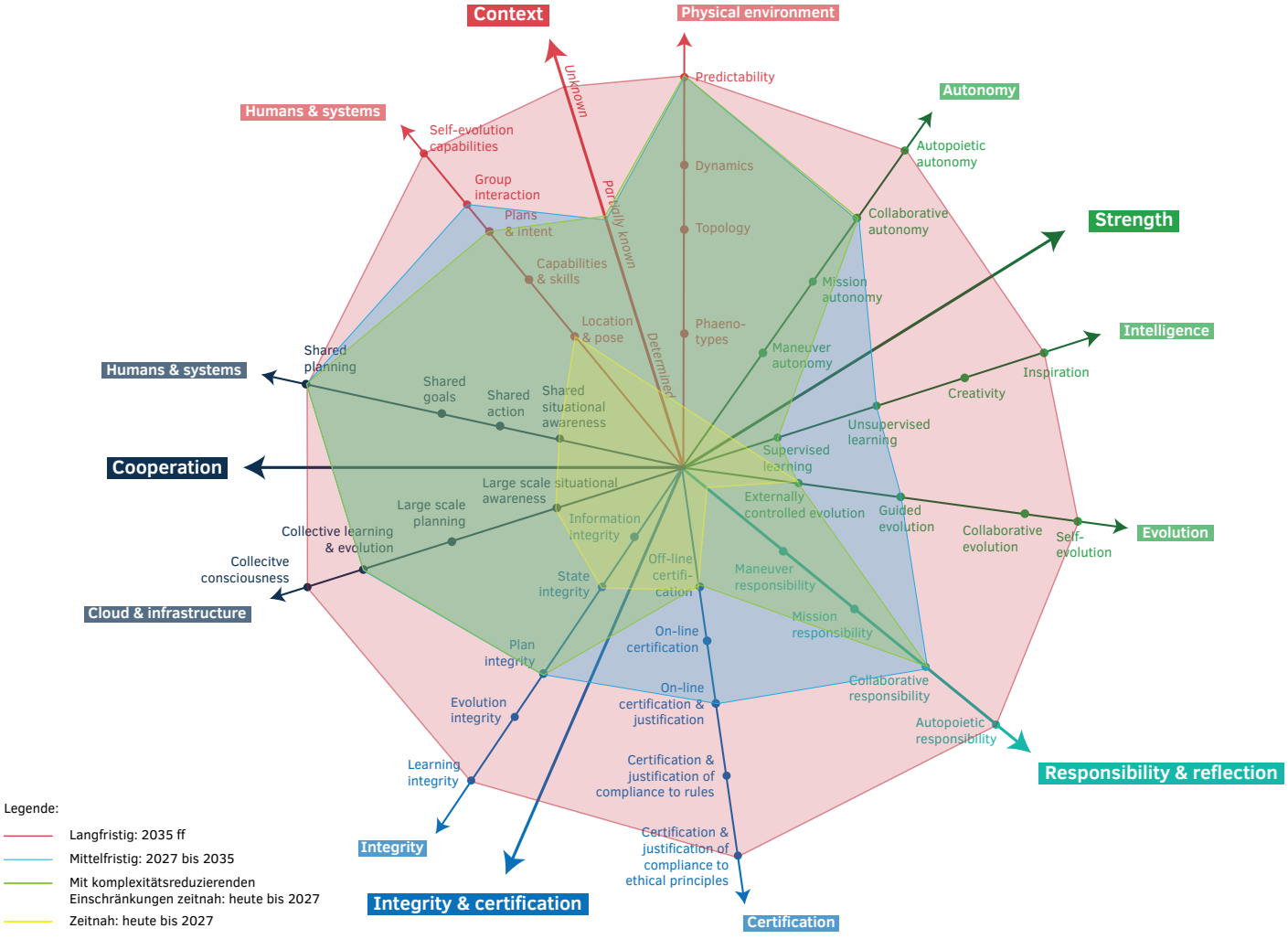


Abbildung 30: SafeTRANS Roadmap zur Nachvollziehbarkeit der Komplexität von Mensch-Maschinen-Systemen [304]

11.4 SMART Standards – Neugestaltung von Normen für KI-Anwendungsprozesse

In diesem Anhang wird die zentrale Fragestellung vertieft, wie eine Neugestaltung von Normen für KI-Anwendungprozesse aussehen und welche technologischen Ansätze hierbei zum Tragen kommen können.

Die Vorgehensweisen zur Bereitstellung granularer Normeninformationen werden unterschiedlich sein (siehe auch Kapitel 5.2.5).

→ **Technologieansatz**: Vorliegende Normendokumente werden im **Postprocessing** ohne Themen- und Anzahlbeschränkung maschinell indexiert und mittels semantischer Methoden automatisiert und in granular „ansprechbaren" Informationseinheiten bereitgestellt. Die Indexiergenauigkeit liegt derzeit bei ca. 80 % gegenüber intellektuell granular aufbereiteten Dokumenten und genügt somit den Anforderungen qualifizierter Anwender, die das zerlegte Informationsangebot fachlich bewerten können. Für nachgelagerte KI-Anwendungsprozesse bedeutet das aber: Eine Validierung der Genauigkeit der Teilinformationen muss integriert werden. Der Treiber dieser Vorgehensweise sind das „Content Management" und „Content Delivery". Erreichbar sind **Ergebnisse im Level 3** (mit o. g. Einschränkungen) **auf der Grundlage von Level 2**.

→ **Bottom-up-Ansatz**: Bei der Digitalisierung von Normen kann zwischen Top-down- und Bottom-up-Ansatz differenziert werden. Beide Ansätze beschäftigen sich mit Fragestellungen der Modularisierung, der Modellierung und dem Management zukünftiger Normeninhalte, allerdings aus unterschiedlichen Perspektiven. Hierbei wird der Top-down-Ansatz durch die Neugestaltung des eigentlichen Normungsprozesses und die Frage gekennzeichnet, wie zukünftige digitale Normen aufgebaut werden müssen, wohingegen sich der Bottom-up-Ansatz mit der Überführung von bereits existierenden Normeninhalten („Nachstrukturierung") in eine maschinenausführbare Wissensrepräsentation befasst. Die Entwicklung smarter Standards bedarf sowohl einer methodischen Annäherung über einen Top-down- als auch einen Bottom-up-Ansatz. Die Treiber des Bottom-up-Ansatzes sind „Content Management und Delivery" und das „Content Usage". Erreichbar sind **Ergebnisse im Level 3 und Level 4** für definierte abgegrenzte Anwendungsgebiete.

→ **Top-down-Ansatz**: Es kann nur ein Referenzdokument oder einen „Referenzinhalt" der Norm geben und das ist der Inhalt, der vom verantwortlichen Normungsgremium geprüft und freigegeben wurde, der sogenannte Primärinhalt. Nur auf diesen beziehen sich in der Regel Gesetze oder Verträge und nur dieser Primärinhalt ist im Ernstfall relevant. Damit auch der maschinenlesbare Normeninhalt Primärinhalt sein kann, muss im **Preprocessing** (i. S. v. Normenentstehungsprozess) die Erfassung der menschengenerierten und -lesbaren sprachlichen Normeninhalte auf der Grundlage einer Struktur erfolgen, die es erlaubt, die Sprache, einschließlich der enthaltenen Semantik, eindeutig in eine maschinenlesbare Datenstruktur (z. B. Ontologie) zu überführen und umgekehrt. Die Treiber dieser Vorgehensweise sind das „Content Creation" und das „Content Usage". Erreichbar sind **Ergebnisse im Level 4**.

Bearbeitungsreihenfolge

Die unterschiedlichen Ansätze können und sollten parallel verfolgt werden. Der Technologieansatz liefert schnellere Erkenntnisse, die in den anderen Vorgehensweisen genutzt werden können. Darüber hinaus entstehen zügig erste – sich wirtschaftlich rechnende – Kundenlösungen oder Prototypen und Demonstratoren, sodass praktische Erfahrungen rückgekoppelt werden können. Die Bottom-up-Vorgehensweise kann nicht geeignet sein, um den sehr großen und stetig wachsenden weltweiten Bestand an Normen nach höchsten Qualitätsansprüchen zu strukturieren. Aber auch gemäß dieser Vorgehensweise heißt es: Gezielt anfangen, um Erfahrungen zu sammeln. Das Postprocessing von Normen kann jedoch für konkrete Anwendungsfelder wirtschaftlich sein. Die „Königsklasse" für die avisierte Zielsetzung zur Erreichung von SMART Standards mit höchsten Qualitätsansprüchen für KI-Anwendungsprozesse kann nur die Verfolgung und Umsetzung einer Top-down-Methode (Preprocessing) sein. Der Aufwand hierfür wird sehr hoch sein.

11.4.1 Nutzung von granularen Inhalten mittels Technologieansatz

Level 2 und 3

Der Prozess für Level 2 und 3 ist – wie für Level 1 – durch abgegrenzte tradierte Verantwortungsbereiche gekennzeichnet. Dies vereinfacht zwar aus organisatorischer Sicht die Umsetzung von Lösungen, verhindert jedoch ein integriertes übergreifendes Handeln, das für SMART Standards im Level 4

zwingend erforderlich wird. Der Fokus auf IT-gestützte Verfahren und deren Weiterentwicklung im „Content Management“ und „Content Delivery“ bietet die Chance, schnell zu konkreten Lösungen zu gelangen, die für Level 4 einen wertvollen Input liefern. Zudem wird eine Basis für die notwendige IT-Infrastruktur gelegt. Die wesentlichen zu beantwortenden Fragestellungen sind in Abbildung 31 aufgezeigt.

Neue digitale Lösungen für die Normenanwendung, die auf der XML-Technologie basieren, sind entstanden oder werden derzeit entwickelt [305], [306]. Weitere Lösungen, die aufgrund strukturierter Inhalte möglich sind, werden demnächst entstehen. Anhand von Beispielen wird der Stand der Weiterentwicklungen kurz beschrieben.

Für nachgelagerte KI-Anwendungsprozesse bedeutet das: Eine Validierung der Genauigkeit der automatisiert ermittelten (Partial-)Informationen muss vorgenommen werden. Erlerntes Erfahrungswissen kann die Bewertung essenziell unterstützen.

Allgemeine Regeln zur Beschreibung der (Partial-)Informationen in Normen sowie die methodische Erarbeitung der genauen Verwendungsorte (Wirkorte, siehe Anhang 11.4.3) liegen für diese Vorgehensweise bisher nicht vor und müssen erarbeitet werden. Um KI-Anwendungsprozesse skalierbar mit (Partial-)Informationen zu versorgen, sind entsprechende Festlegungen zu vereinbaren.

Level 2 und 3

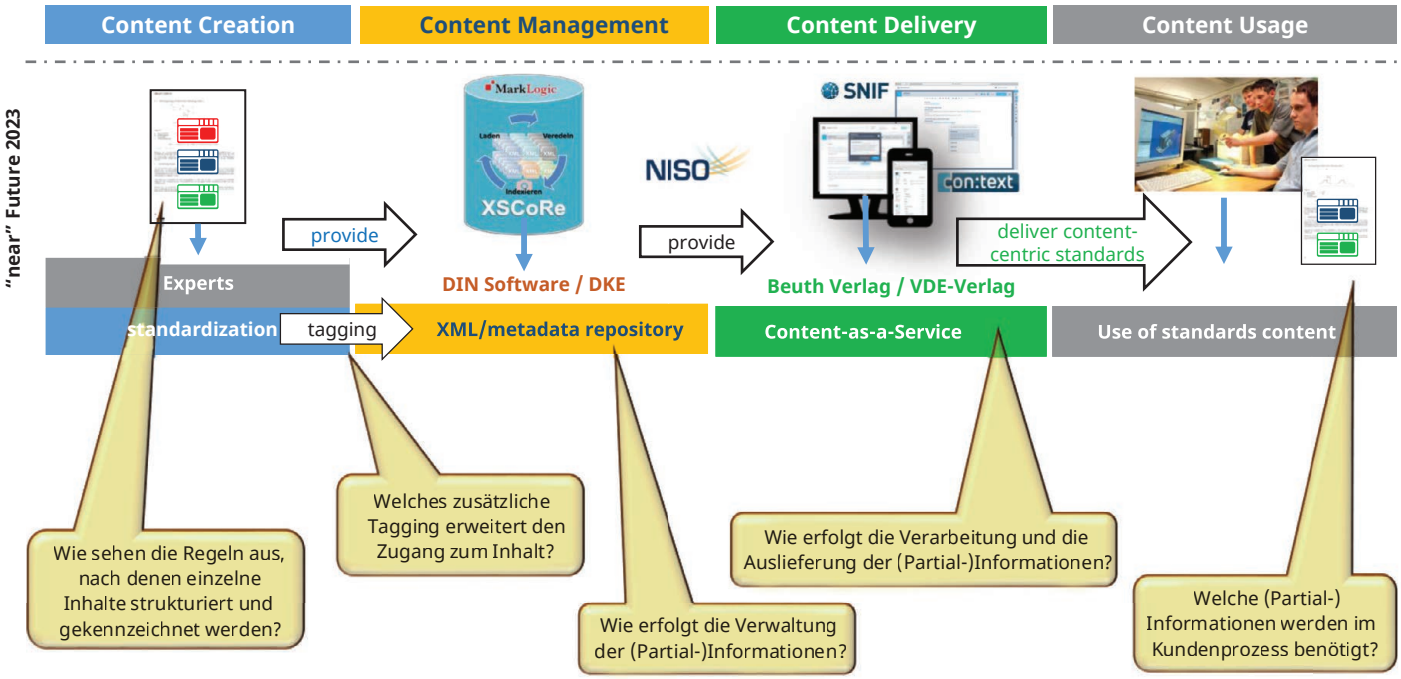


Abbildung 31: Level-2- und -3-Prozess und wesentliche Fragestellungen

BEISPIELE FÜR ENTWICKLUNGEN IM „CONTENT MANAGEMENT“ UND „CONTENT DELIVERY“

Beispiel A: Schlüsseltechnologie XML-Datenbank und Schnittstellen

Seit 2016 hat die DIN Software GmbH damit begonnen, den DIN-Normenbestand in XML zu überführen (siehe Abbildung 32). Gleichzeitig wurde eine XML-Datenbank aufgebaut [306]. Die Ziele waren und sind:

- → Entwicklung eines „XML Semantic Content-Repository“ (kurz XSCoRe) als digitale Such- und Auslieferungsplattform von Inhalten
- → Zentrale Bereitstellung und Verwaltung von Metadaten und Normeninhalten für digitale Informations- und Wissensprodukte der DIN-Gruppe
- → Bereitstellung von Schnittstellen zur Unterstützung der Content-Management-Prozesse der XML-Workflows von DIN und Beuth, um diese effektiv an das XML-Repository anzubinden
- → Basis für Entwicklung von „granularen Content-as-a-Service-basierten Plattformdiensten“ und somit Beschleunigung der digitalen Transformation der Unternehmensprozesse der DIN-Gruppe

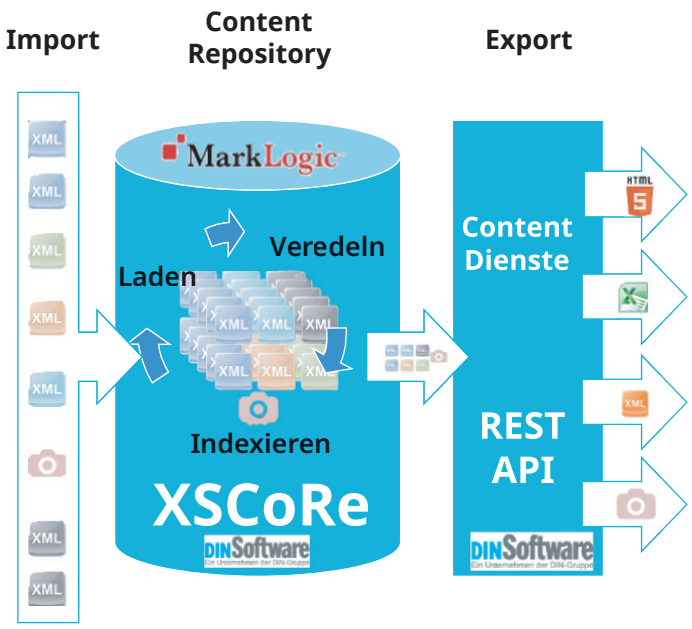


Abbildung 32: Überführung des DIN-Normenbestands in XML

XSCoRe ist insoweit weiterzuentwickeln, dass beispielsweise Anforderungen im RegIF-Format abgebildet werden können.

Beispiel B: Schlüsseltechnologie XML-Anwendungsregeln – NISO Z39.102-2017 [307]

Um sicherzustellen, dass Normen ein identisches XML-Format aufweisen, wurde eine „Standard Tag Suite“ entwickelt [308]:

- → 2017-10-09 von ANSI als US-Norm anerkannt
- → Ca. 30 Regelsetzer haben daran mitgearbeitet, inkl. BSI, SFS, DIN, CEN, ISO, IEC, IEEE, ASTM, ASME
- → Das definierte „Tag Set“ ist die Basis für Austausch und Bereitstellung von XML-Normen (ISO, IEC, DIN, ASME, ...)
- → DIN-Produkte nutzen Content im NISOSTS-Format

Gleichzeitig wurde vereinbart, 2020 in einer NISO Working Group einen „Standards-Specific Ontology Standard (SSOS)“ weiterzuentwickeln, siehe NISO-Information [308].

Die Mitglieder der National Information Standards Organization (NISO) haben ein neues Projekt zur Schaffung eines standardspezifischen Ontologiestandards (Kurztitel: NISO SSOS) genehmigt. Es wird eine Arbeitsgruppe gebildet, die eine Ontologie auf hoher Ebene entwickeln und standardisieren soll, um eine begrenzte Anzahl von Kernkonzepten und Beziehungen zu beschreiben, die sich zunächst auf den Lebenszyklus von Standards konzentrieren.

Dies wird die Verwendung von Standards erleichtern, eine konsistentere Entdeckung und Navigation innerhalb der Standards unterstützen und eine Grundlage für andere semantische Anwendungen, wie z. B. verknüpfte Daten, im Ökosystem der Standards schaffen. Die Einigung auf eine Ontologie ermöglicht es Normenherstellern und -vertreibern, bestehende Investitionen in XML weiter zu nutzen. Sie baut auf bereits bestehenden Arbeiten wie der NISO Standard Tag Suite auf, einem ANSI/NISO-Standard, bei dem es sich um einen Satz von XML-Elementen handelt, die ein gemeinsames Format für die Darstellung und den Austausch von Inhalten von Standards bereitstellen, unabhängig davon, in welcher Form die Inhalte schließlich an die Kunden geliefert werden.

Der Standards-Specific Ontology Standard (SSOS) bildet die Grundlage, die erforderlich ist, um in Schlüsselbereichen wie der verbesserten maschinellen Lesbarkeit (Level 2) voranzukommen. Die damit verbundene inhaltsbezogene Aufwertung der Dokumente wird vor allem auch für KI-basierte Anwendungen auswertbar sein. Da sich das Projekt zurzeit noch in einer Startphase befindet, sollten KI-bezogene Anforderungen benannt und – wenn möglich – einbezogen werden.

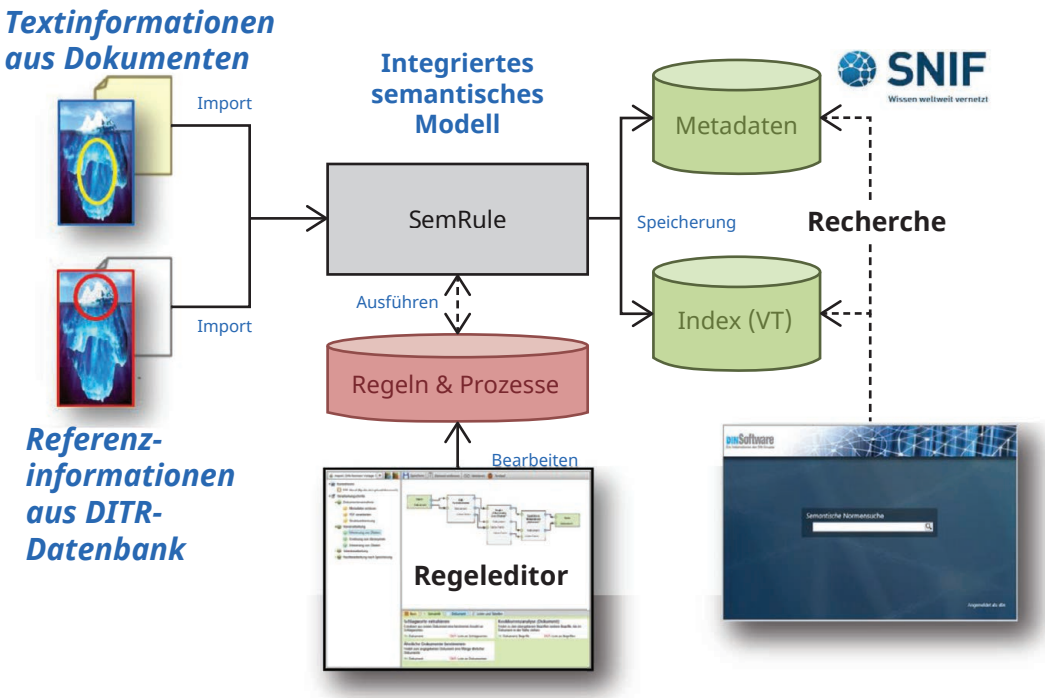
Beispiel C: Schlüsseltechnologie „Semantisches Informationsframework (SNIF)“
Mit dem „Semantischen Normeninformationsframework (SNIF)“ ist eine inhaltliche, semantische Erschließung von Normen realisiert worden [309]. Es werden in SNIF die Metadaten aus der DITR-Datenbank und die Normentexte semantisch indexiert, wodurch eine hohe Qualität der Ergebnisse erreicht wird (siehe Abbildung 33).

Die anwenderfreundliche Benutzeroberfläche zur Datenextraktion und Aktualisierung der Datenbank erlaubt es auch Mitarbeitern, die keine Programmierkenntnisse besitzen, Prozesse zur Extraktion relevanter Informationen aus DIN-Normen bedarfsgesteuert mittels eines Regeleditors einzurichten und anzupassen.

Bereits heute werden:
→ Ähnlichkeitsbeziehungen zwischen Normen systematisch analysiert,
→ vielfältige Indexierungs- und Content-Anreichungsdienstleistungen erbracht und
→ ein maßgeschneiderter Metadatenservice durch SNIF überhaupt erst ermöglicht.

SNIF ist eine Basistechnologie, deren Extraktionsmöglichkeiten noch nicht vollständig ausgeschöpft sind. Es ist im KI-Projekt zu prüfen, in welcher Weise das Framework genutzt werden kann.

Abbildung 33: SNIF-Basistechnologie



Beispiel D: Schlüsseltechnologie „con:text“, um verteilte Informationen zu finden und Zusammenhänge aufzuzeigen

Auf der Basis XML-konvertierter Dokumente und unter Einhaltung des NISO STS wurde der Service „con:text“ entwickelt, der an verschiedene Normen-Managementsysteme gekoppelt werden kann. Das Funktionsset zielt darauf ab, den Inhalt tiefergehend zu erfassen, Zusammenhänge simultan auszuspielen und anwendungsfreundlich über zahlreiche Funktionen sichtbar zu machen. In einer weiteren Ausbaustufe wird die Erstellung unternehmenseigener Dokumente (z. B. Werksnormen, technische Liefervorschriften) im bidirektionalen Zusammenspiel mit DIN-Normen unterstützt, siehe Abbildung 34.

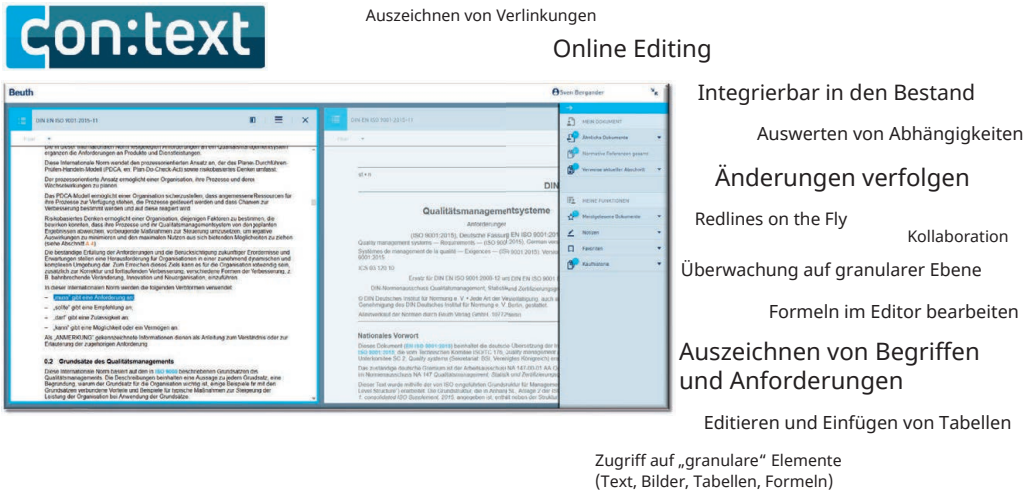
Datenbasis für „con:text:

→ Konvertierung von 35.000 DIN-EN-ISO-Dokumenten in XML

→ Ca. 4,5 Mio. aktive Verlinkungen

Das Funktionsset von con:text spiegelt die Anforderungen der Anwender wider. Somit entsteht hier ein Anwendungs-Know-how, das für die Funktionsbildung von KI-Anwendungsprozessen relevant sein kann. Gleichzeitig kann die Anwendung con:text von den Ergebnissen des KI-Projekts profitieren. Die Zusammenarbeit im KI-Projekt soll ermöglicht werden.

Abbildung 34: Erweitertes Funktionsset in Anwendungen mit dem Fokus „Inhalt“



Beispiel E: Schlüsseltechnologie „Extraktion von Normeninhalten“

Auf der Basis XML-konvertierter Dokumente und dem „Semantischen Normen-Informationsframework (SNIF)“ wird eine Recherchemethode entwickelt, die Textstellen regelbasiert aus Normen extrahieren kann (siehe Abbildung 35).

→ Der Content wird aus der XML-Datenbank MarkLogic über Schnittstellen zur Verfügung gestellt (in der Regel einzelne Sections).

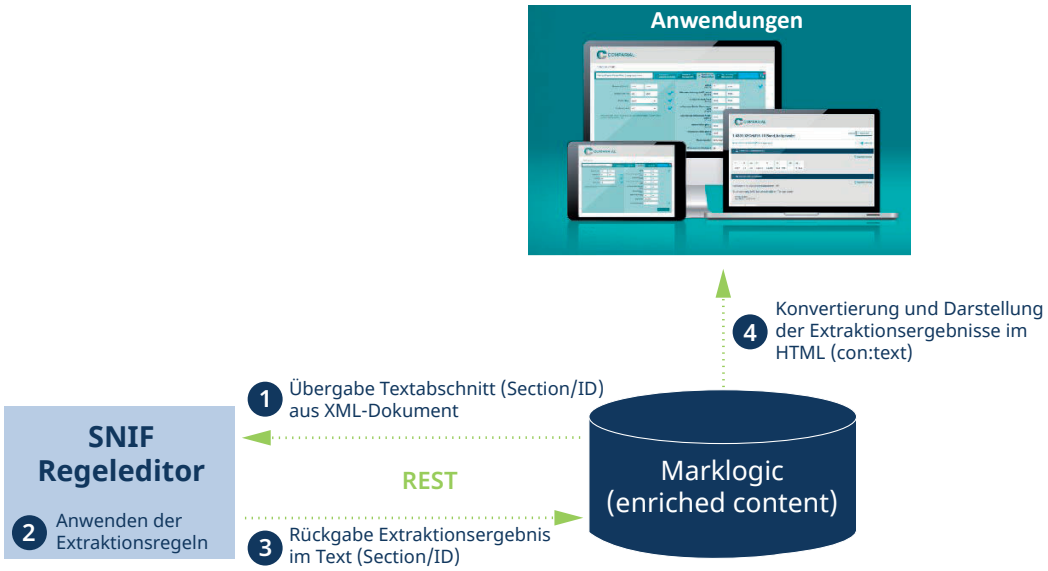
→ Das Semantische Normen-Informationsframework (SNIF) kann diesen Content im XML-Format über eigene REST-Schnittstellen entgegennehmen.

→ Die Analyse und die Verarbeitung (Erkennen von Anforderungen und Empfehlungen) finden innerhalb des SNIF-Systems statt.

→ MarkLogic erhält über eine REST-Schnittstelle den erzeugten Mehrwert zurück und liefert diesen an abnehmende Anwendungen wie z. B. con:text aus.

Die Regeln für die Informationsextraktion werden entsprechend den Kundenanforderungen formuliert. Somit entsteht insbesondere hier ein Anwendungs-Know-how, das für die Regelbildung von KI-Anwendungsprozessen relevant sein wird. Die Zusammenarbeit im KI-Projekt soll ermöglicht werden.

Abbildung 35: Schlüsseltechnologie „Extraktion von Normeninhalten“



11.4.2 Bottom-up-Methode – Nachstrukturierung von Normen

Zur Realisierung des Bottom-up-Ansatzes sind fünf Schritte – „Extraktion“, „Modellierung“, „Fusion und Speicherung“, „Bereitstellung“ sowie „Anwendung“ – notwendig (vgl. Abbildung 36) [310]. Es stellt sich die Frage, wie klassifizierte Normeninhalte ohne Informationsverlust maschinenausführbar repräsentiert werden können. Die Lösung besteht in einer automatischen Extraktion von Normeninhalten (hier am Beispiel von Formeln) und deren Überführung in eine maschinenausführbare Wissensrepräsentationsform, auf die von unterschiedlichen Autorensystemen zugegriffen werden kann. Aus den Erkenntnissen, die bei der konkreten Konzeptumsetzung gewonnen werden können, lassen sich Anforderungen und Gestaltungsregeln auf eine höhere Abstraktionsebene der „Next Generation Norm“ ableiten.

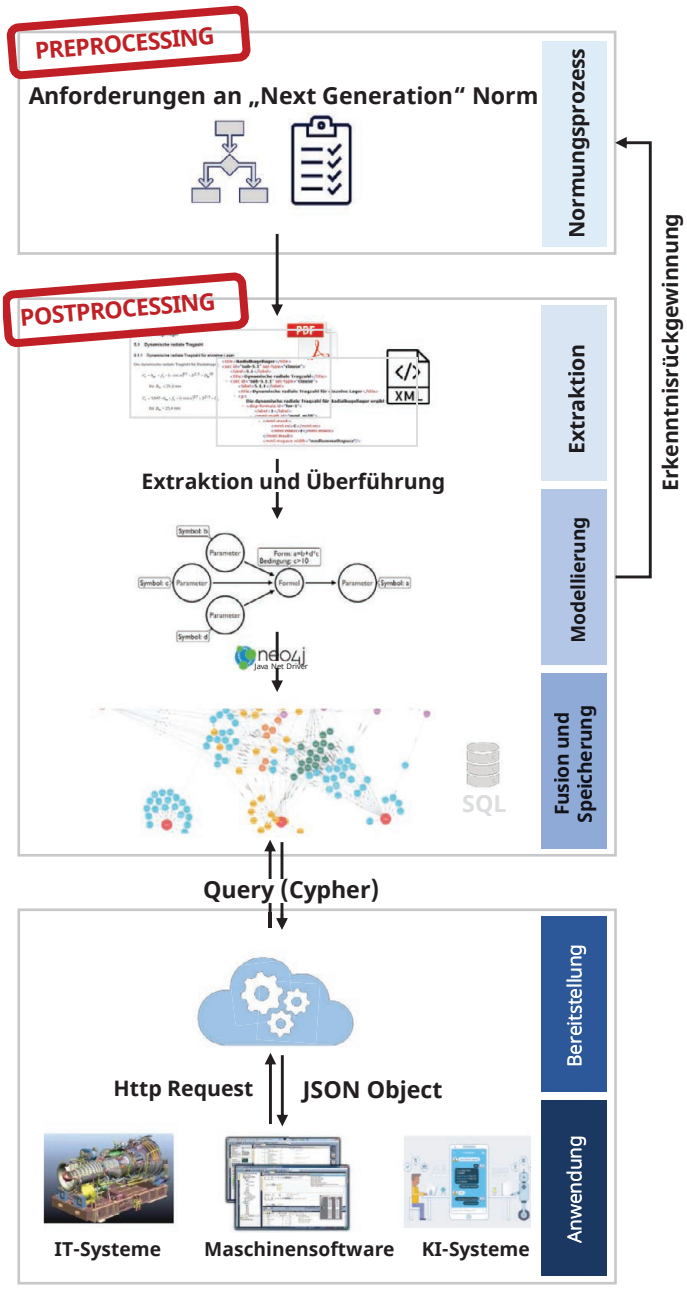


Abbildung 36: Bottom-up-Ansatz zur Nachstrukturierung von Normen

Extraktion

Normen stehen unterschiedlichen Stakeholdern nicht nur als PDF, sondern auch im XML-Format zur Verfügung, wobei Normenelemente wie beispielsweise Formeln, Tabellen, Diagramme und/oder Text im Quellcode getaggt sind. Die Extraktion beschreibt den eigentlichen Vorgang des Auslesens von relevanten Informationen. Hierzu kann ein XML-Parser verwendet werden, welcher die markierten (Formel-)Elemente einer XML-basierten Norm erkennt und in ein vordefiniertes Graphmuster transformiert.

Modellierung

Die Modellierung ermöglicht eine einfache und eindeutige Wissensrepräsentation unter Berücksichtigung der Maschinenausführbarkeit. Das Ziel besteht darin, eine automatisierte Überführung extrahierter Informationen (1:1) zu erreichen. Entsprechend der Art des Normenelements werden standardisierte Graphmuster definiert, wobei am Beispiel von Formeln ersichtlich wird, dass Parameter sowie Operatoren als Knoten und deren Beziehungen als Kanten modelliert werden können.

Fusion/Speicherung

Der Vorgehensschritt „Fusion und Speicherung“ beschreibt die Möglichkeit, alle separat generierten Graphmuster (für Formeln, Tabellen, Diagramme, Texte) zu einem erweiterbaren Wissensnetz in einer Datenbank zu aggregieren. Dies ermöglicht zum einen die Eliminierung aller redundanten Knoten und zum anderen die Wiederherstellung von Beziehungen zwischen einzelnen Normenelementen.

Bereitstellung

Die Informationsbereitstellung im Bottom-up-Ansatz dient der Entkopplung von Wissensquellen (hier: Graphdatenbank) und -verwendung (hier: Anwenderprogramm). Über einen Webservice, auf welchen unterschiedliche Autorensysteme zugreifen können, werden Abfragen in der aufgebauten Graphdatenbank durchgeführt und zurückgespielt.

Anwendung
Unter Anwendung wird hier die Nutzung von digitalisierten Normeninhalten verstanden. Die Anzahl möglicher Anwendungen, wie beispielsweise im Bereich von IT-Systemen (z. B. CAD), bei Maschinensoftware oder KI-basierten Anwendungssystemen, ist unüberschaubar. Basierend auf der Anfrage eines Autorensystems werden die relevanten Informationen in der Datenbank identifiziert und an das Autorensystem übergeben.

Letztlich lassen sich mithilfe des Bottom-up-Ansatzes Rückschlüsse für den Normungsprozess generieren. Dies ermöglicht wiederum eine Eliminierung manueller, fehleranfälliger Prozessschritte, eine signifikante Zeiteinsparung bei der Überführung von Normeninhalten in Unternehmensprozesse, die Steigerung der Qualität durch Gewährleistung einer durchgängigen Rückverfolgbarkeit von Normeninhalten sowie einen geringeren Anpassungsaufwand bei Updates von Normen.

Die Nachstrukturierung des existierenden, sehr großen Normenfundus stößt an kapazitive Grenzen und wäre nur für definierte Themenbereiche wirtschaftlich vertretbar. Hierbei ist ein Einsatz von Künstlicher Intelligenz in der Extraktionsphase des Bottom-up-Ansatzes zu untersuchen, um diesen Arbeitsschritt maschinell zu unterstützen.

11.4.3 Top-down-Methode – Entwicklung von SMART Standards

Gegenwärtig entstehen vielversprechende Konzepte bei DIN/DKE, CEN/CENELEC und ISO/IEC, bereits **während des Normungsprozesses (Preprocessing,** „Top-down-Ansatz“) entsprechende Darstellungsformen zu erarbeiten, die die Überführung in maschineninterpretierbare Formate gestatten [311], [258].

Der Gesamtprozess für Level 4 erfordert teilweise ein integriertes übergreifendes Handeln der Prozessverantwortlichen, sodass bisherige Verantwortungsgrenzen (Level 1 bis 3) überdacht und neu festgelegt werden müssen. Definitiv muss die Content-Verantwortlichkeit für die „Content Creation“ im Prozess der Entwicklung der Normen – der Primärinhalte – verortet sein. Ein Postprocessing im Sinne einer nachträglichen Interpretation bzw. „Deutung“ von Inhalten für die Weiterverarbeitung darf es nicht mehr geben.

Level 4
Die wesentlichen zu beantwortenden Fragestellungen sind in Abbildung 37 formuliert. Mit der Beantwortung der Fragestellungen wird teilweise Neuland betreten, insbesondere im Zusammenspiel mit KI-basierten Anwendungsprozessen. Eine vereinfachte Darstellung visualisiert das Zielbild eines Gesamtprozesses als Teilfunktionen SM|ART|KI, siehe Abbildung 38:

Level 4

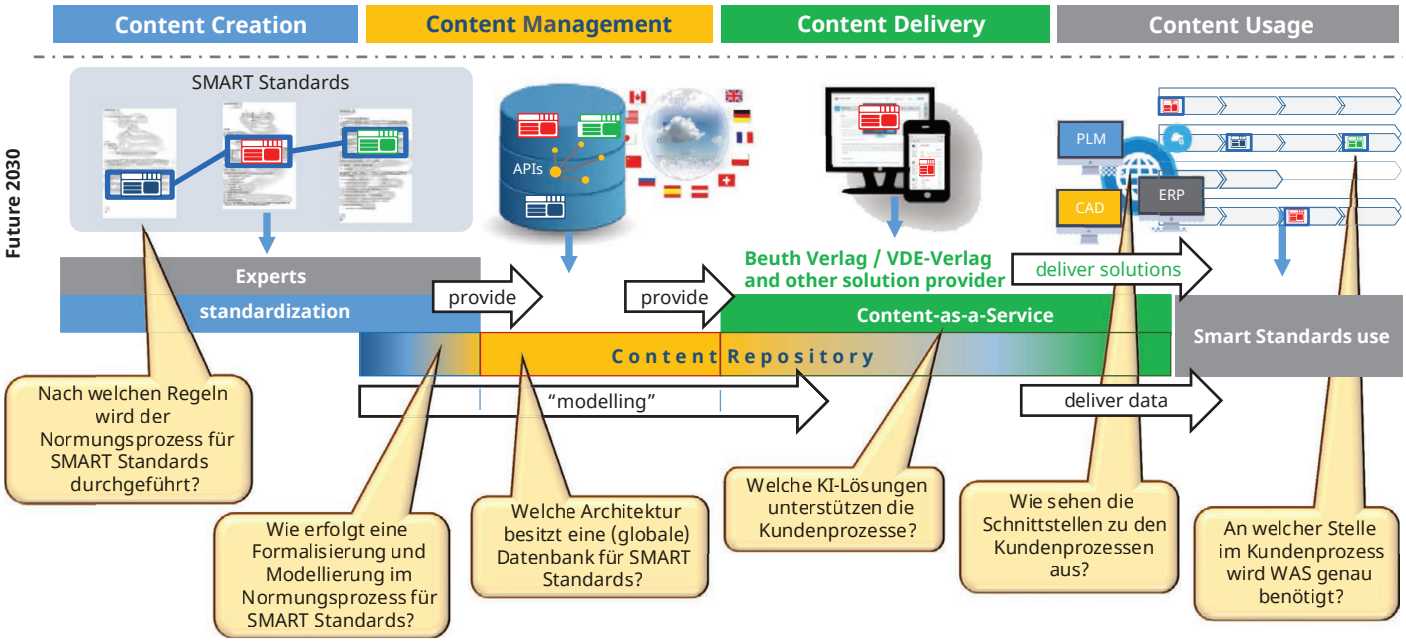


Abbildung 37: Level-4-Prozess und wesentliche Fragestellungen

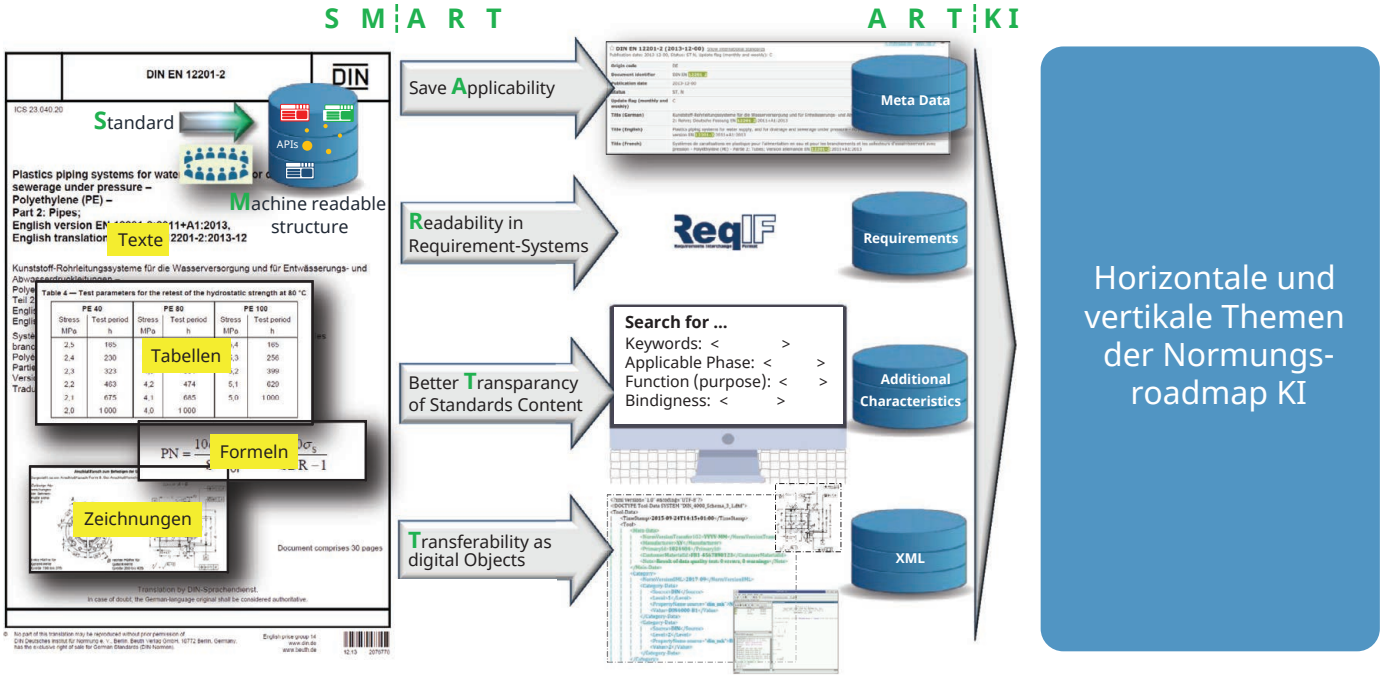


Abbildung 38: Be SMART – der Normenoutput im praktischen Anwendungskontext

SM – Transformationsfunktion (Content Creation):
Normen und Standards à maschineninterpretierbare Modelle

ART – Transportfunktion (Content Management & Delivery):
Maschineninterpretierbare Informationen à Liefermodelle

KI – Nutzungsfunktion (Content Usage):
Lieferformen à KI-basierte Anwendungsprozesse

Wie in der vorliegenden KI-Roadmap zum Ausdruck gebracht, sind „[d]ie wirtschaftlichen Einsatzfelder für KI äußerst vielfältig". Ergänzung: Sie sind **unendlich vielfältig**. Nahezu für alle Wirtschaftsbereiche und auch sonstige Anwendungsbereiche außerhalb der Wirtschaft ist KI relevant und findet sich sowohl in Form von Komponenten in Endprodukten und Dienstleistungen als auch in den produktiven Kern- und Unterstützungsprozessen innerhalb der Unternehmen. Damit ist klar, dass ein Gesamtprozess nicht von den Use Cases her aufgebaut werden kann.

Im Folgenden werden wesentliche methodische Rahmenbedingungen, in denen ein Teil der Fragestellungen erarbeitet wurde und um das Zielbild zu erreichen, zusammengefasst [309]. Sie betreffen die Erfüllung der Transportfunktion (s. o.):
→ der Entwicklungsprozess von SMART Standards und
→ die Contentstruktur (Informationsmodell).

Entwicklungsprozess von SMART Standards

Die Forderung einer Granulierung von Normen bis hin zu den kleinsten sinnvollen Informationselementen und deren Kennzeichnung ist eine der Kernaufgaben, die es zu lösen gilt. Die methodische Erarbeitung durch Fachkenner unterschiedlicher Disziplinen sowie die in nachvollziehbaren Schritten zu erfolgende Erarbeitung definierter und abgrenzbarer Normungs-Teilergebnisse sind hierbei ein wesentlicher Erfolgsfaktor. Im Folgenden wird ein Vorgehen beschrieben, das zum Gelingen der Zielstellung beitragen kann.

Der derzeitige Normungsprozess zielt darauf ab, als finales Arbeitsergebnis eines Normungsvorhabens eine abgestimmte und geprüfte Norm zu veröffentlichen. Was wird zukünftig i. S. v. SMART Standards das abzuliefernde Arbeitsergebnis sein? Man wird nicht mehr nur (aber zwingend auch) das „eine finale" Arbeitsergebnis erzeugen können. Die verschiedenen erarbeiteten Darstellungsformen des Normungsgegenstands müssen aus Produkthaftungsgründen sowie wegen einer transparenten Nachvollziehbarkeit der Teilergebnisse dokumentiert werden. Die Erarbeitung von SMART Standards ist de facto ein Entwicklungsprozess, vergleichbar mit den Systematiken der Entwicklungsprozesse von Produkten, für die die verschiedenen Entwicklungsverantwortlichkeiten analog transparent dargestellt sein müssen.

Die methodische Vorgehensweise gemäß VDI 2221 [312], [313] stellt eine Grundlage dar, die zukünftigen Normungsprozesse (Phasen, Arbeitsschritte, Arbeitsergebnisse, Akteure) zu beschreiben. In den aktuellen Vorarbeiten bei DIN/DKE und CCMC wird in Pilotprojekten die systematische Modellierung eines Normungsgegenstands erprobt und in weiterführenden Projekten konsolidiert.

Ein Ansatz hierfür sei im Folgenden zusammengefasst. Die **Phasen,** Arbeitsergebnisse und „Akteure“ sind hierin durch die gewählte Formatierung erkennbar.

Normungsantragsphase
Vor der konkreten Bearbeitung eines Normungsvorhabens wird bekanntermaßen die Relevanz und die Finanzierung geprüft, der Normenausschuss zugeordnet sowie die Zuordnung interessierter Kreise identifiziert. Zukünftig wird durch einen erweiterten **„Entscheiderkreis“,** in dem nunmehr auch der **„Normennutzer“** einen wichtigen Input einzubringen hat, die Implementierungsstufe der zu erarbeitenden SMART-Standards-Lösung festgelegt werden müssen. Die Implementierungsstufe bestimmt den **Grad der Digitalisierung eines Normungsvorhabens:**

a) Es entsteht eine Norm gemäß dem derzeitigen Entstehungs- und Nutzungsszenarios (vgl. Abbildung 26), mit Ablieferung des gesamten Contents in XML, z. B. zur weiteren Verwendung in Redaktionssystemen oder sonstigen Nutzungsumgebungen. Zusätzlich können in der Norm definierte digitale Objekte (z. B. Tabellen, Formeln, Grafiken) zur direkten Verwendung in Kundensystemen festgelegt werden.

b) Das Ziel besteht in der Entwicklung eines SMART Standards für ein zukünftiges Entstehungs- und Nutzungsszenario (vgl. Abbildung 37) in den Darstellungsformen H2H und H2M zur direkten KI-basierten Nutzung der granularen Normeninhalte in Kundenprozessen (M2M), siehe Abbildung 39.

I. Anforderungsdefinitionsphase
Die Anforderungen an den Normungsgegenstand müssen – wie bisher – formal im Einklang mit DIN 820 [314] und inhaltlich von den **„Fachexperten in den Normenausschüssen“** (im Folgenden **„Fachexperten“**) festgelegt und dokumentiert werden. Darüber hinaus gilt es, die anzustrebenden Teilergebnisse der Folgephasen, je nach entschiedener Implementierungsstufe (= Grad der Digitalisierung), zu beschreiben.

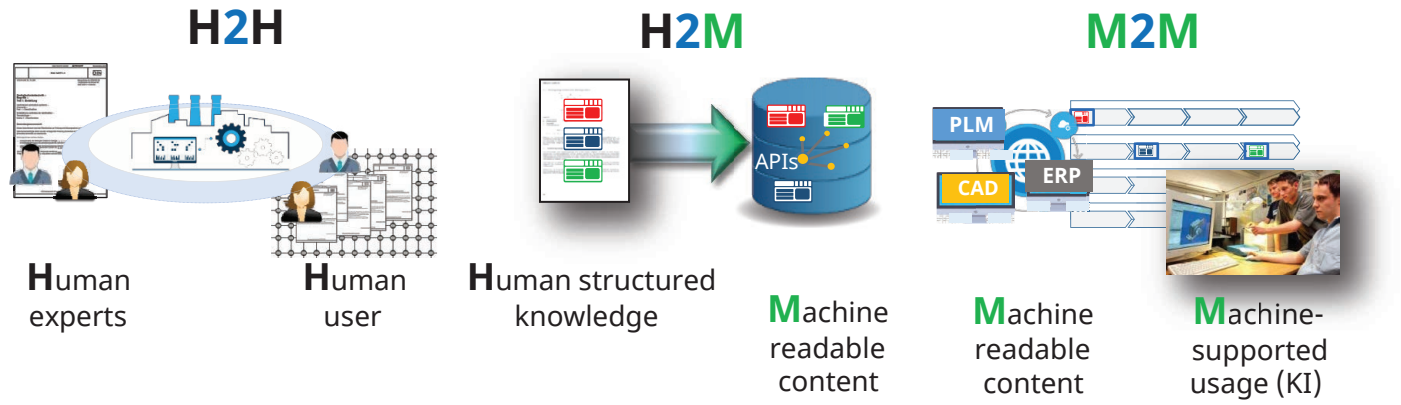


Abbildung 39: Entwicklung von SMART Standards für nachgelagerte Anwendungsprozesse

II. Normungsphase (Konzept)

Die Sprache (Prosa) der **„Fachexperten"** ist derzeit und auch in überschaubarer Zeit nicht geeignet, um daraus direkt eine maschineninterpretierbare Form im Sinne von SMART Standards zu transformieren. Diese natürlich-sprachliche Darstellungsform ist jedoch erforderlich, um Expertenwissen überhaupt zu artikulieren, zu konsolidieren und unter Berücksichtigung der DIN 820 [314] und anderer Regularien abzustimmen.

Einer methodischen Vorgehensweise folgend, wird es also erforderlich sein, dass solche Formulierungen als erarbeitetes Zwischenergebnis der Phase II. eines Normungsvorhabens erhalten und **dokumentiert** bleiben (Thema Produkthaftung) und damit den rückverfolgbaren Input für die Folgeschritte darstellen.

III. Formalisierungsphase (Ausarbeitung)

Aus II. eine formale Darstellungsform zu erzeugen, muss Aufgabe der **„Fachexperten"** bleiben. Diese Experten müssen **„erweiterte Strukturierungskompetenzen"** besitzen, um diesen Teil des Normungsprozesses nach – noch weiter festzulegenden – Regeln auszuführen.

Aktuell wird ein zielführendes Konzept verfolgt: Die Umsetzung soll auf der Grundlage des „Semantischen Tripels" erfolgen, das beispielsweise innerhalb des Resource Description Framework (RDF) als Elementarbaustein für das Semantische Web, ein globales Netzwerk der Informationen, konzipiert wurde (siehe in Phase IV.).

Die Anwendung und somit Umsetzung der Regeln kann durch „Vorgaben" beschrieben sein und durch zu entwickelnde (z. B. XML-basierte) Tools sichergestellt bzw. unterstützt werden, die eine Formalisierung erwirken bzw. fördern.

Die Tabellen oder andere Darstellungsformen sind ein Zwischenergebnis der Phase III., das **dokumentiert** werden muss (Thema Produkthaftung, Rückverfolgbarkeit). Es stellt den Input für den nächsten Schritt dar.

IV. Modellierungsphase

Im letzten Normenerarbeitungsschritt eines Normungsvorhabens müssen die **„Fachexperten mit IT-Modellierungs-Kompetenzen"** aktiv werden. Ausgehend von der eindeutigen Darstellungsform III. können diese **„Modellier-Experten"** Modelle wie z. B. Triplestores, DB-Strukturen etc. aufbauen.

Das Tripel stellt das Bindeglied zwischen der natürlichen Sprache und der Datenstruktur dar: Über eine Subjekt-Prädikat-Objekt-Beziehung wird eine Aussage aus bekannten Elementen definiert, z. B. „Gras ist grün.", welche wiederum selbst als Subjekt oder Objekt verwendet werden kann. Eine solche Verschachtelung erlaubt die Formulierung von komplexen Sätzen bzw. Festlegungen in menschlicher Sprache, eine IT-seitige Kontrolle bei der Erfassung der Inhalte sorgt dafür, dass die (verschachtelte) Tripel-Struktur eingehalten wird und gleichzeitig maschinenlesbarer Inhalt generiert wird.

Eine Normeninhaltekompetenz für den Transformationsprozess ist hierzu insofern noch förderlich, um Inkonsistenzen bei der Überführung von III. nach IV. auch inhaltlich erkennen zu können. Sollte für einzelne formale Darstellungen eine eindeutige Überführung in entsprechende Modelle nicht möglich sein, so muss hierfür in III. nachgearbeitet werden.

Im Sinne einer durchgängigen Methodik müssen die Ergebnisse der Phase IV. sorgfältig und umfassend **dokumentiert** (Thema Produkthaftung, Rückverfolgbarkeit) werden. Sie stellen den Input für weitere Implementierungen (siehe z. B. [315]) im Anwendungskontext dar.

Es sei an dieser Stelle darauf hingewiesen, dass die vier Phasen zwar nacheinander abgearbeitet werden, es sich jedoch auch immer um einen iterativen Prozess handeln wird. Darüber hinaus ist anzunehmen, dass für Normungsgegenstände, deren Darstellungsformen von Anfang an „IT-umsetzungsnah" konzipiert sind, eine integrierte Bearbeitung der Phasen II. und III. oder sogar II. bis IV. möglich sein wird. In sehr wenigen Teilbereichen der Normung, z. B. der Sachmerkmalsnormung, STLB-Bau-Normung, ist das heute schon möglich.

II. Normungsphase: Derzeit kann die Sprache (Prosa) der Fachexperten nicht direkt in eine maschineninterpretierbare Form im Sinne von SMART Standards transformiert werden. Mit zukünftig vorliegender Erfahrung und gelerntem Wissen in KI-Anwendungsprozessen ist dennoch zu konzipieren, dass eine KI-orientierte Modellierung realisiert werden kann.

III. Formalisierung und IV. Modellierung: Die Transformation mittels „Semantischer Triple" kann eine direkte Schnittstelle zu KI-Prozessen darstellen. Eine enge Zusammenarbeit ist erforderlich.

Erforderliche Contentstrukturen – das Informationsmodell
Festlegungen in Normen (Anforderungen, Empfehlungen etc.) bestehen in der Regel aus wiederkehrenden Elementen, die nach einem bestimmten Muster miteinander verknüpft werden. So beschreiben diese z. B. häufig ein System oder eine Funktion in Verbindung mit einer bestimmten Leistung bzw. Eigenschaft, die ggf. nur unter bestimmten Bedingungen einzuhalten ist.

Ein hierfür passendes Informationsmodell enthält eine möglichst allgemein verwendbare Schablone für die Formulierung von Festlegungen und definiert die Modellierung der darin enthaltenen Elemente nach dem Tripel-Konzept. Weiterhin enthält das Informationsmodell Metadaten zu den Festlegungen, durch die eine Weiterverarbeitung erleichtert wird (z. B. Verbindlichkeit oder Funktion der Festlegung). Idealerweise lassen sich diese Metadaten eindeutig aus dem Normeninhalt oder den Projektdaten ableiten und gehören dann zum Primärinhalt. Einen Ausschnitt des Informationsmodells zeigt Tabelle 15. Im Folgenden sind einige wesentliche Merkmale des Informationsmodells erläutert.

Die Normungsfunktion als gliederndes Hauptmerkmal: Der Grundgedanke früherer Überlegungen bestand darin, Inhalte gleicher Funktion zusammenzufassen, um sukzessive ein System aus vernetzten Modulen aufzubauen. Für die aktuelle Aufgabe – die Normenerarbeitung – gilt analog, die zu entwickelnden Inhalte entsprechend festzulegender Merkmale dergestalt zu strukturieren, dass eine Integration in ein zunehmend wachsendes SMART-Standards-Normenumfeld möglich ist. Für die aktuelle Aufgabenstellung hat die Normungsfunktion somit eine wichtige Bedeutung.

Definition „Normungsfunktion“: „Ein genormtes Element (kleinstes sinnvolles ‚Normengranulat‘, z. B. Satz, Abschnitt, Formel, Daten, Bild etc.) wird in einer Norm nur dann mit einer definierten Verbindlichkeit formuliert, wenn damit ein Zweck erfüllt werden soll.“

Dieser Zweck kann verschiedene beabsichtigte Teilfunktionen erfüllen:
→ Kommunikation: Verständnis herbeiführen oder fördern, Verständigung durch einheitliche Terminologie ermöglichen.
→ Qualität: Anforderungen, Sicherungsmaßnahmen und Nachhaltigkeitsprozesse festlegen.
→ Prüfung: Bedingungen, Durchführungen und Auswertungen festlegen.
→ Sicherheit: Anforderungen mittels Merkmalen an materielle (z. B. Konsumgüter) und immaterielle (z. B. Dienstleistungen) Gegenstände festlegen.
→ Grundlage: Einheitliche Maßgaben des Handelns festlegen.
→ Häufung: Materielle und immaterielle Gegenstandsvielfalt reduzieren; Vorgänge (Prozesse) vereinfachen und Aufwände (Zeit, Kosten, Material) reduzieren.
→ Recycling: Wiederverwertung/-verwendung und Weiterverwertung/-verwendung von Ressourcen regeln.
→ Zusammenhang: Normative Zusammenhänge aufzeigen und angleichen.
→ Interoperabilität: Austausch von materiellen und immateriellen Gegenständen ermöglichen; Förderung von Technologie, Warenverkehr, Anwendungen ermöglichen.

Die Entwicklung und Präzisierung weiterer Normungsfunktionen ist noch nicht abgeschlossen.

Weitere normungsbezogene Merkmale zur Strukturierung: Die formale Darstellung der unterschiedlichen Normungsgegenstände wird in der Praxis keine Hürde sein, sofern die gliedernden Merkmale verständlich und nachvollziehbar sind.

Weitere Gliederungsmerkmale und deren Ausprägungen sollen – neben dem oben beschriebenen Hauptmerkmal „Normungsfunktion“ – den einzelnen Normungsgegenstand eindeutig repräsentieren. Diese sind (derzeit):
→ Technische Merkmale des Normungsgegenstands:
- Die Darstellung kann ebenso tabellarisch erfolgen. Aktuell wird in den CCMC-Projekten ein sprachlicher Ansatz verfolgt (Subject, Action, Object). Eine Bewertung über die Nutzungsakzeptanz einer geeigneten Darstellungsform liegt noch nicht vor.
- Verbindlichkeitsmerkmal, gemäß DIN 820 [314]:
  - Verpflichtung („müssen“)
  - Empfehlung („sollten“)
  - Erlaubnis („dürfen“)
  - Möglichkeit („können“)
  - Interaktion des Normungsgegenstands, gemäß der Definition „Function of the Objective“ aus den Vorüberlegungen in den CCMC-Pilotprojekten [316]:
  - Activity
  - Constraint
  - Integral aspect
  - Interface
  - Verification

Tabelle 15: Das Informationsmodell für SMART Standards (Auszug und Arbeitsstand vom Juli 2020, DIN e. V.)

No.	property	value	occurence	data type	definition
Elements forming a provision or requirement topic					
0	title	text	optional	content	heading or title (subject + action)
1	system | subject	~	required	content	subject; product; system
2	subject-type	~	optional	metadata	connection (link)
2.1		• system	~	value	defined description of the system
2.2		• term	~	value	defined description of the system
3.1	action | modal verb	~	required	content	bindingness word; modal verb; auxiliary verb
3.2	action | main verb	~	required	content	main verb; strong verb; full verb; action
4	actor	~	required	content	effective site; inherited from scope (document) or (sub-)clause
5	performance | object	~	required	content	object; performance
6	object-type	~	optional	metadata	connection (link)
6.1		• term	~	value	defined description of the performance
6.2		• provision	~	value	provision of the defined types
6.3		• numeric value	~	value	numeric value
6.4		• unit	~	value	unit
7	condition	~	optional	content	conditions
8	margin	~	optional	content	deviations; limits; tolerances

No.	property	value	occurence	data type	definition
Attributes to the provision or requirement topic					
8	relation	relation	required		
8.1		• and	~	value	A1 and A2
…					
8.10		• xor	~	value	either A1 or A2
9	bindingness	~	required	metadata	degree of compusion
9.1		• capability	~	value	capability
…					
9.5		• requirement	~	value	requirement
10	type		required	metadata	interaction of the provision or requirement
10.1		• activity	~	value	activity (process)
…					
10.6		• verification	~	value	verification (method of proof)
11	function	~	required	metadata	requirement or standard function
11.1		• availability	≃	value	
…					
11.11		• sustainablility	≃	value	
12	smart-tag		optional	metadata	allow marking which SMART property is equivalent for this triple
12.1		• actor	≃	value	effective site; inherited from scope (document) or (sub-)clause
…					
12.5		• system	≃	value	subject; product; system

No.	property	value	occurence	data type	definition
13	classification	~	~	metadata	classification [inherited from document or topic]
14	date of activation	YYYY((-MM)?-DD)?	required	metadata	date of publication (dop) [inherited from document or topic]
15	date of creation	YYYY((-MM)?-DD)?	required	metadata	date of availability (doa) [inherited from document or topic]
16	date of deactivation	YYYY((-MM)?-DD)?	required	metadata	date of withdrawal (dow) [inherited from document or topic]
17	date of last change	YYYY((-MM)?-DD)?	required	metadata	date of availability (doa) [inherited from document or topic]
18	date of revision	YYYY((-MM)?-DD)?	required	metadata	date of publication (dop) [inherited from document or topic]
19	date of version	YYYY((-MM)?-DD)?	required	metadata	date of publication (dop) [inherited from document or topic]
20	guid	~	required	metadata	global unique identifier
21	informative	text	optional	metadata	system of interest, rationale, explanation, examples
22	keywords	~	optional	metadata	discriptors [inherited from document or topic]
23	language	ISO 639-1	required	metadata	2-letter language code in lower case
24	source-of	~	required	metadata	source; reference to standard or law text
25	status	~	required	metadata	stage-code [inherited from document or topic]
26	version number	YYYY-MM-DD hh:mm:ss	required	metadata	time stamp of last change of requirement or date of publication

- Metadaten des Gesamtdokuments, gemäß dem Stand der Indexierungsmethodik und der Erfordernisse eines Normen-Managements [317]. Es ist weiter zu prüfen, inwieweit die Erkenntnisse aus der Nutzung semantischer Methoden integriert werden können.

Merkmale mit Anwendungsbezug: Prozessuale Anwendungsaspekte („Wer nutzt welche Normungsinhalte?“) werden im Allgemeinen nicht im Normungsprozess festgelegt. Eine sinnvolle Ausnahme soll das Merkmal „Wirkort“ sein – eine zusätzliche Information, die bisher in Normen nicht durchgängig zu finden ist.

Definition „Wirkort“: „Die Stelle, an der das genormte Geschehen zur Wirkung kommt, kennzeichnet den Wirkort.“

Beispielsweise:
→ Wirkung im Prozess: z. B. Konstruktion (beispielsweise im Konzept oder der Ausarbeitung), After Sales, Reproduktion etc.
→ Wirkung bei zu erfüllenden Funktionen: z. B. Verbinden

Beispiel für zukünftige Entwicklungsprozesse unter Nutzung des Informationsmodells (siehe Abbildung 40).

I. Anforderungsdefinitionsphase

In der Normungsantragsphase wird der Grad der Digitalisierung festgelegt: Ein Normungsvorhaben entsprechend der Phasen I. bis IV. ist als SMART Standard zu entwickeln. Die Anforderungen hieran werden dokumentiert.

II. Normungsphase

Die Anforderungen werden von den Experten in textueller Form formuliert, z. B.:

„Wenn die Temperatur über 50 °C liegt oder der Druck 50 MPa überschreitet, müssen Rohr und Verbindungselemente entweder aus Material entsprechend EN 1234 bestehen oder einen Elastizitätsmodul zwischen 15.000 N/mm² und 18.000 N/mm² besitzen.“

→ Geeignet für bestehende („wortbasierte“) Normungsprozesse (z. B. Grundlage für die Entwurfsabstimmung)

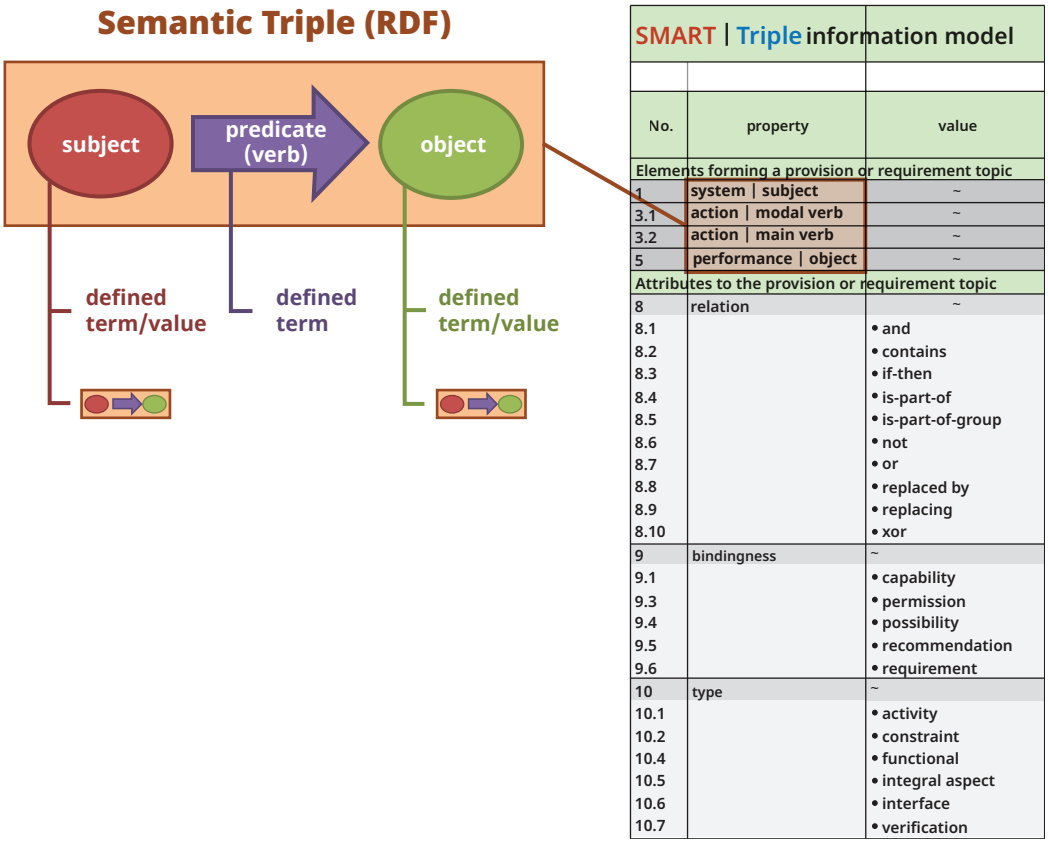


Abbildung 40: Triple-Strukturierung für zukünftige Entwicklungsprozesse unter Nutzung des Informationsmodells

III. Formalisierungsphase

Das semantische Triple ist die Grundlage für die Formalisierung:

→ Es stellt sicher, dass nur für definierte Elemente Angaben gemacht werden .
→ Es ermöglicht die Strukturierung von Norminhalten auf der Basis eindeutiger Informationselemente.
→ Verschachtelung von Triple ist möglich.
→ Die Triple-Strukturierung steigert die Qualität der Dokumente im Level 3 erheblich und legt den Grundstein für Level 4.

IV. Modellierungsphase

Als Werkzeug zur Umsetzung des Konzepts sind bei DIN derzeit speziell entwickelte grafische Benutzeroberflächen in der Erprobung, die es dem Normungsgremium erlauben, sich auf den Klartext der Norm zu konzentrieren und die beschriebenen Datenstrukturen im Hintergrund erzeugen (siehe Abbildung 41).

Experimentelle GUI zur Erfassung von Anforderungen

Die beispielsweise für ein Anforderungsmanagementsystem interessanten Metadaten können aus diesen Datenstrukturen im Anschluss eindeutig und automatisiert generiert werden. So geht etwa die Verbindlichkeit einer Festlegung aus dem (durch die Triple-Strukturierung) eindeutig identifizierbaren Modalverb hervor oder der jeweilige Normungsgegenstand kann über eine Auswertung der verknüpften Subjekt-Elemente erfolgen.

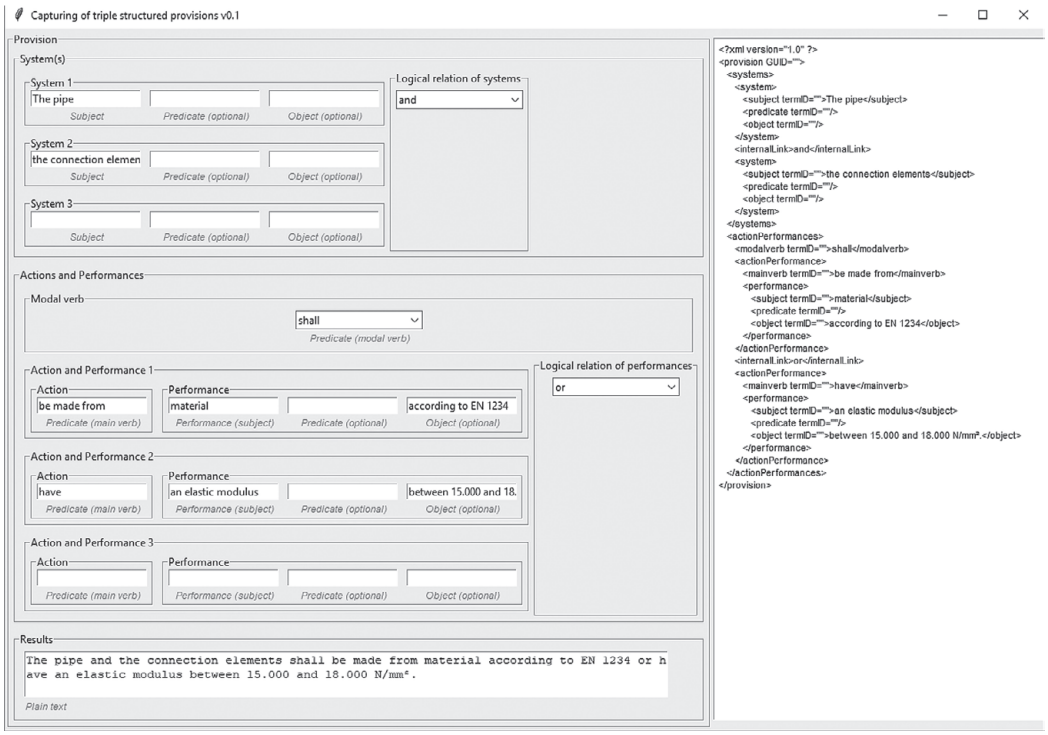


Abbildung 41: Grafische Benutzeroberfläche zur formalen Erfassung von Anforderungen

BEISPIELE FÜR POSTPROCESSING

XML-Ergebnis aus GUI (siehe Abbildung 42)

→ **Die natürliche Sprache bleibt erhalten. Alle semantischen Informationen werden in XML hinterlegt und sind für Level 4 nutzbar.**

Übersetzung nach OWL/RDF (siehe Abbildung 43)

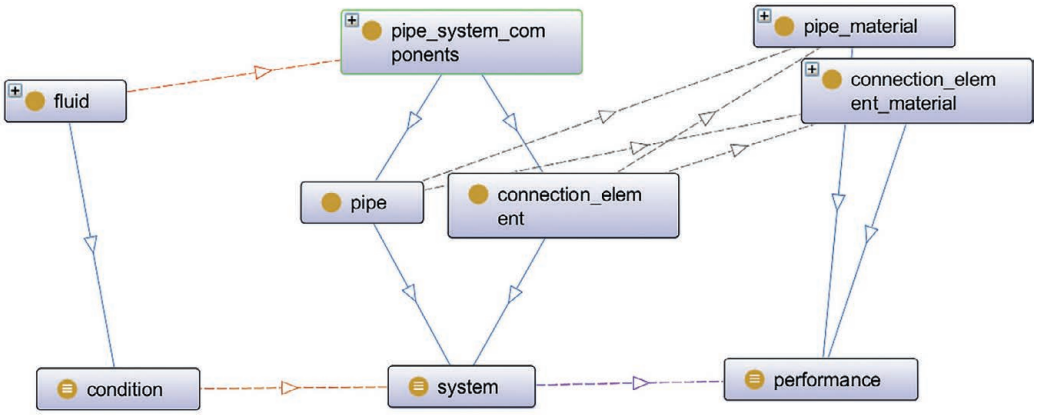
Aber auch die automatisierte Generierung von Ontologien (z. B. OWL nach W3C) für die sichere Anwendung von KI-Systemen ist möglich. Dadurch ist eine Vielzahl der denkbaren Anwendungsfälle für SMARTe Normen abgedeckt.

→ **Die Übersetzung in OWL/RDF kann in einem nachgelagerten automatisierten Prozess erfolgen.**

Abbildung 42: Beispielhafte Datenstruktur im XML, generiert aus der formalen Erfassung

```
<?xml version="1.0" encoding="UTF-8"?>
<req bindingness="requirement" reqtype="integral aspect" guid="8d53fe1a-0ca6-4e29-8327-fdb684cdb82c">
    <relations></relations>
    <functions></functions>
    <req-condition elements="2" internal-link="or" external-link="if">
        <external-link term-id="882c695b-eb80-46ae-9197-0df5080721d6">
            If
        </external-link>
        <element type="triple" number="1">
            <triple>
                <subject type="term" term-id="9d90ed81-9530-4a4c-b7c6-2d3988dc8c0f">
                    temperature
                </subject>
                <predicate type="term" term-id="9eefe750-ea1d-4e16-a5da-d42fd417c527">
                    is
                </predicate>
                <object type="value" term-id="">
                    above 50°C
                </object>
            </triple>
        </element>
        <internal-link term-id="38c68a7d-0964-43cc-89e3-efb1385896ee">
            or
        </internal-link>
```

Abbildung 43: Automatisierte Übersetzung in OWL/RDF



DIN e. V.

Burggrafenstraße 6
10787 Berlin
Tel.: +49 30 2601-0
E-Mail: presse@din.de
Internet: www.din.de

Stand: November 2020

DKE Deutsche Kommission Elektrotechnik
Elektronik Informationstechnik in DIN und VDE
Stresemannallee 15
60596 Frankfurt am Main
Tel.: +49 69 6308-0
Fax: +49 69 08-9863
E-Mail: standardisierung@vde.com
Internet: www.dke.de