André Meyer-Vitali

AI Act / KI-VO und Standardisierung

Übersicht über die Landschaft der Normierung

Bei der Einführung der KI-Verordnung müssen noch viele Aspekte definiert werden, damit eine sorgfältige Prüfung geschehen kann. Wir betrachten die Landschaft der Standardisierung, definieren Risiken und das damit einhergehende Vertrauen. Einige Methoden und Prüfmetriken werden dargestellt

1 Einleitung

Die Anwendung von Methoden und Systemen mit künstlicher Intelligenz ist heute unverzichtbar und bildet eine Voraussetzung, um große Herausforderungen in Gesellschaft und Umwelt zu bewältigen. Diese Anwendungen bringen jedoch auch Risiken mit sich, die entsprechend abgesichert werden müssen. Deshalb sind Prüfverfahren erforderlich, die hochriskante KI-Systeme bewerten können. Während die Analyse symbolischer KI-Systeme weitgehend verstanden ist, eignen sich die dabei verwendeten Methoden nur bedingt, datengetriebene subsymbolische KI-Systeme zu untersuchen. Diese Systeme operieren oft autonom, d.h., sie führen Entscheidungen und Aufgaben eigenständig und ohne direkten menschlichen Eingriff aus. Zudem sind sie nichtdeterministisch, d.h., ihr Verhalten ist aufgrund stochastischer Faktoren im Trainingsprozess und der Komplexität der Modelle nicht vollständig vorhersehbar. Diese Eigenschaften machen subsymbolische KI-Systeme besonders leistungsfähig aber auch schwer kalkulierbar, was die Bewertung und Sicherung solcher Systeme zu einer besonderen Herausforderung macht.

2 Vertrauen

Vertrauen wird definiert als die Bereitschaft eines Vertrauensgebers, sich den Handlungen eines Akteurs (Vertrauensnehmer) auszusetzen, die er nicht direkt kontrollieren kann. Es erfordert die Existenz von Unsicherheit und Risiken. Das Maß an Vertrauen steigt mit der vom Vertrauensgeber wahrgenommenen Wohlwollen, Kompetenz und Integrität des Vertrauensnehmers. Vertrauens-kalibrierung ist der Prozess der Anpassung und Angleichung der tatsächlichen und wahrgenommenen Vertrauens-würdigkeit, wobei die tatsächliche Vertrauenswürdigkeit der Grad ist, in dem ein System oder Akteur die geforderten oder versprochenen Erwartungen und Leistungen erfüllt.

Um Vertrauenswürdigkeit zu garantieren, ist es notwendig, die technischen Metriken, Methoden und Standards, sowie die entsprechenden Prozesse und Werkzeuge, zu entwickeln, die eine Prüfung und Zertifizierung von hoch-riskanten KI-Systemen erlauben.

3 Standards

Es gibt schon eine große Menge an bestehenden Prüfmethoden und Standards, um darauf auf-bauend neue, innovative Ansätze zu entwickeln. Zudem sind viele Organisationen in diesem Bereich tätig, sowohl in der Forschung als auch in der Anwendung und es ist wichtig, sowohl nationale als auch internationale Akteure zu berücksichtigen, um Best Practices zu identifizieren und Synergien zu nutzen.

Die Zusammenarbeit zwischen verschiedenen Disziplinen und Sektoren wird entscheidend sein, um ein robustes und zuverlässiges Prüf- und Zertifizierungssystem für hochriskante KI-Systeme zu etablieren. Ein solches System wird spätestens ab dem 2. August 2026 verfügbar sein müssen, wenn die europäische KI-Verordnung definitiv und für alle Anwendungen in Kraft tritt.

Ein wichtiger Schritt ist das Definieren der Begrifflichkeit. Dabei kann auf bestehende Standards zurückgegriffen werden, wie sie von nationalen und internationalen Organisationen und Instanzen wie ISO/IEC[1], OECD[2], NIST[3], ETSI[4], IEEE[5], DIN, VDE, DKE, CEN-CENELEC, AFNOR, etc. schon erarbeitet wurden.

Dr. André Meyer-Vitali

ist Informatiker und promovierte in Software Engineering, Ubiquitous Computing und verteilter KI an der Universität Zürich.
Derzeit ist er Senior Researcher am DFKI (Deutschland) mit Schwerpunkt auf der Entwicklung und Förderung von vertrauenswürdiger KI (Trusted AI). Er ist der wissenschaftliche Leiter des Centre for European Research in Trusted Artificial Intelligence (CERTAIN).
E-Mail: andre.meyer-vitali@dfki.de

1 Common Criteria of IT security evaluation (ISO/IEC 15408), Artificial intelligence concepts and terminology (ISO/IEC 22989), Guidance on risk management (ISO/IEC 23894)

2 OECD AI Principles for Trustworthy AI, https://oecd.ai/en/ai-principles

3 NIST AI Risk Management Framework (AI RMF), https://www.nist.gov/itl/ai-risk-management-framework; AI RMF Generative AI Profile, https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-generative-artificial-intelligence

4 Towards a Harmonized Documentation Scheme for Trustworthy AI (T038), https://portal.etsi.org/xtfs/#/xTF/T038/

5 IEEE CertifAIEd, https://engagestandards.ieee.org/ieeecertifaied.html

Weitere wichtige Quellen sind der Kriterienkatalog für KI-Cloud-Dienste (AIC4[6]) des BSI, der Prüfkatalog von Fraunhofer IAIS[7] und die DIN Normungsroadmap KI[8] mit dem AI Methods, Capabilities and Criticality Grid[9] (AI-MC[2]). Ebenso relevant sind, beispielsweise, Z-Inspection[10], AP4AI[11], DEKRA[12], AI Quality & Testing Hub[13], ISTQB[14] und das Fraunhofer HHI Zertifizierungs-Rahmenwerk[15]. Bei der OECD wurde ein Reporting Framework for Developing Advanced AI Systems[16],[17] entwickelt, das auf dem Hiroshima Code of Conduct[18] der G7 beruht.

4 Risikostufen

Gemäß der Verordnung (EU) 2024/1689 (AI-Act) werden KI-Systeme in verschiedene Risikostufen eingeteilt, wobei die Hochrisiko-KI-Systeme (Art. 6; Anhang III) von besonderem Interesse sind. Die zu entwickelnden Prüfmethoden sollen jedoch flexibel gestaltet werden, sodass sie auch auf andere Risikoklassen anwendbar sind. Ein KI-System gilt als vertrauenswürdig, wenn es die festgelegten Spezifikationen und Anforderungen in Bezug auf seine erwarteten und gewünschten Funktionen zuverlässig erfüllt. Dabei sind zahlreiche Charakteristiken von Bedeutung, die in ihrer Gesamtheit sicherstellen, dass diese Systeme das Vertrauen der Benutzer verdienen. Zu diesen Charakteristiken gehören unter anderem die Sicherheit, Robustheit, Transparenz und Fairness des Systems. Darüber hinaus ist es entscheidend, dass diese Systeme ethische Standards einhalten und sich kontinuierlich an neue Herausforderungen anpassen lassen, um auch langfristig vertrauenswürdig zu bleiben. Die Entwicklung solcher Prüfmethoden und -standards ist daher nicht nur eine technische, sondern auch eine gesellschaftliche Aufgabe, die maßgeblich dazu beiträgt, die Akzeptanz und das Vertrauen in KI-Systeme in der breiten Öffentlichkeit zu stärken. Diese Charakteristiken umfassen die Anforderungen der High-Level Expert Group on Artificial Intelligence (AI HLEG, ALTAI[19]).

5 Methoden

Die Überprüfung anhand der Anforderungen und Kriterien soll nicht nur einmalig erfolgen. In einem iterativen Entwicklungsprozess soll eine zirkuläre Wertschöpfung erreicht und die Erfüllung der Anforderungen kontinuierlich überprüft werden, als Bestandteil des Lebenszyklus-Managements von KI-Systemen. Da KI-Systeme ständigen Veränderungen und Anpassungen unterliegen, ist es von entscheidender Bedeutung, dass die Anforderungen und die daraus abgeleiteten

Prüfkriterien regelmäßig überprüft, aktualisiert und neu bewertet werden. Dieser fortlaufende Prozess ermöglicht es, auf neue technologische Entwicklungen und veränderte Rahmenbedingungen flexibel zu reagieren, wodurch die Qualität und Vertrauenswürdigkeit der KI-Systeme über ihre gesamte Lebensdauer hinweg gewährleistet werden. Darüber hinaus fördert dieses dynamische Prüfverfahren die kontinuierliche Verbesserung und Anpassung der Systeme, was dazu beiträgt, dass sie langfristig effektiv und sicher in ihrer jeweiligen Anwendungsumgebung bleiben.

Moderne KI-Systeme werden unter Einsatz einer Vielzahl von generischen Komponenten, Modulen und Bibliotheken entwickelt, die sowohl von den Entwicklern selbst als auch von verschiedenen Anbietern bereitgestellt werden. Diese Bausteine bilden das Fundament für die Funktionalität und Leistungsfähigkeit der KI-Systeme. Sie müssen einer kontinuierlichen Kontrolle unterzogen wer-den, um ihre Qualität und Sicherheit zu gewährleisten. Ebenso wichtig ist die fortlaufende Überprüfung der Werkzeuge und Plattformen, die zur Entwicklung und zum Betrieb dieser Systeme eingesetzt werden. Diese umfassende Evaluation stellt sicher, dass alle verwendeten Technologien den aktuellen Standards entsprechen und dass potenzielle Risiken frühzeitig erkannt und behoben werden. Durch diesen ganzheitlichen Ansatz wird die Integrität des gesamten Entwicklungs- und Betriebsprozesses gewahrt, was die Schaffung robuster und vertrauenswürdiger KI-Systeme ermöglicht.

Auch der Prüfprozess selbst kann teilweise automatisiert werden. Beispielsweise können verschiedene Prüfkriterien und Metriken berechnet werden und das Sammeln der Daten, sowie das Schreiben des Berichts kann standardisiert unterstützt werden. Es ist jedoch wichtig, dass dabei die menschliche Aufsicht bewahrt wird, damit die gewünschte Funktionalität auf Wirksamkeit und Korrektheit glaubhaft zertifiziert wird.

6 KI-Verordnung

Die KI-Verordnung (im Folgenden: KI-VO, auch „AI Act" genannt)[20] trat am 1. August 2024 in Kraft und soll als erstes um-

6 Artificial Intelligence Cloud Service Compliance Criteria Catalogue, https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/AIC4/aic4.html

7 Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz, https://www.iais.fraunhofer.de/de/forschung/kuenstliche-intelligenz/ki-pruefkatalog.html

8 Normungsroadmap Künstliche Intelligenz (Ausgabe 2), https://din.one/pages/viewpage.action?pageId=33620030

9 Schmid, T., e.a. (2021). The AI Methods, Capabilities and Criticality Grid. KI – Künstliche Intelligenz, 35(3), 425–440. https://doi.org/10.1007/s13218-021-00736-4

10 Z-Inspection®: A Process to Assess Trustworthy AI, https://z-inspection.org/

11 Accountability Principles for Artificial Intelligence (AP4AI), https://www.ap4ai.eu/

12 DEKRA Testing & Certification Services, https://www.dekra.com/en/ai-testing-certification-services/

13 AI Quality & Testing Hub, https://aiqualityhub.com/en/

14 ISTQB® AI Testing (CT-AI), https://www.istqb.org/certifications/artificial-inteligence-tester

15 Fraunhofer Auditing and Certification of AI Systems, https://www.hhi.fraunhofer.de/en/departments/ai/technologies-and-solutions/auditing-and-certification-of-ai-systems.html

16 Reporting Framework for the Hiroshima AI Process (HAIP) International Code of Conduct for Organizations Developing Advanced AI Systems, https://transparency.oecd.ai/about

17 https://www.dfki.de/web/news/oecd-stellt-in-paris-reporting-werkzeug-fuer-fortgeschrittene-ki-anwendungen-vor

18 Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems, https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document05_en.pdf

19 Assessment List for Trustworthy Artificial Intelligence (ALTAI), https://op.europa.eu/de/publication-detail/-/publication/73552fcd-f7c2-11ea-991b-01aa75ed71a1

20 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnung (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der

fassendes KI-Gesetz der Welt einen Rechtsrahmen für Künstliche Intelligenz (KI) schaffen. Die Europäische Kommission veröffentlichte im April 2021 einen Entwurf für eine neue Verordnung zur Regulierung künstlicher Intelligenz.[21] Die finale Fassung wurde am 12. Juli 2024 im Amtsblatt der Europäischen Union veröffentlicht und trat 20 Tage später, am 1. August 2024, in Kraft. Die Verordnung gilt schrittweise ab dem 2. August 2026. Die Kapitel I und II gelten ab dem 2. Februar 2025, die Kapitel III, Abschnitt 4, Kapitel V, Kapitel VII und Kapitel XII sowie Art. 78 KI-VO gelten ab dem 2. August 2025, mit Ausnahme des Art. 101 KI-VO. Art. 6 Abs. 1 KI-VO und die entsprechenden Pflichten gelten ab dem 2. August 2027.

Die KI-VO ist das Ergebnis langer Diskussionen und Verhandlungen, die durch das Spannungsfeld zwischen Regulierung und Technologieoffenheit geprägt waren.[22] Die EU-Kommission hat bereits im Jahr 2018 eine High-Level Expert Group eingesetzt, um Ethik-Leitlinien für eine vertrauenswürdige KI zu erarbeiten, die im KI-Weißbuch aufgegriffen wurden und teilweise als Grundlage für die KI-VO dienten.[23]

Ziel der KI-VO ist es, das Funktionieren des Binnenmarktes zu verbessern und die Einführung einer auf den Menschen ausgerichteten und vertrauenswürdigen KI zu fördern. Gleichzeitig soll ein hohes Schutzniveau in Bezug auf Gesundheit, Sicherheit und die in der Charta verankerten Grundrechte gewährleistet werden. Die KI-VO soll ausdrücklich die in der Charta der Grundrechte verankerten Werte der Union gewährleisten und den Schutz von Personen, Unternehmen, Demokratie und Rechtsstaatlichkeit sowie der Umwelt erleichtern. KI soll als Instrument zur Verbesserung des menschlichen Wohlergehens dienen. Als typische Gefahren des Einsatzes von KI werden beispielsweise Überwachung und Verletzung der Privatsphäre, voreingenommene und intransparente Entscheidungsfindung oder auch sicherheitskritische Anwendungen gesehen.[24] Gleichzeitig soll die KI-VO Innovation und Beschäftigung fördern und der EU eine Führungsrolle bei der Einführung vertrauenswürdiger KI verschaffen.[25] Die KI kann zur Weiterentwicklung und Optimierung von Produkten und Dienstleistungen in zahlreichen Sektoren beitragen.[26] Der Bundesminister für Justiz Dr. Marco Buschmann sprach in dem Zusammenhang von einem „Balanceakt" zwischen Innovation und Risikoschutz.[27]

Die KI-VO verpflichtet Anbieter und Betreiber von Hochrisiko-KI zu einem umfassenden Risikomanagement über den gesamten Lebenszyklus des Systems (Art. 9 KI-VO). Hochrisiko-KI-Systeme müssen mit qualitätsgeprüften Datensätzen entwickelt werden, um Bias, Fehler oder Lücken zu identifizieren (Art. 10 KI-VO). Eine detaillierte technische Dokumentation, die die Einhaltung der Vorgaben belegt, ist erforderlich (Art. 11 KI-VO), ebenso wie die Protokollierung und Überwachung der Systemfunktion (Art. 12 KI-VO). Zudem müssen Genauigkeit, Robustheit und Cybersicherheit während des gesamten Lebenszyklus sichergestellt werden (Art. 15 KI-VO). Anbieter müssen die genaue und zuverlässige Arbeit des Systems inklusive einer angemessenen Toleranz gegenüber Fehlern, Störungen und Unstimmigkeiten sicherstellen. Dazu möchte die Kommission in Zusammenarbeit mit Interessensträgern und Organisationen Benchmarks und Messmethoden erarbeiten. Das System muss eine angemessene menschliche Aufsicht und Kontrolle ermöglichen (Art. 14 KI-VO), und in bestimmten Fällen ist eine Grundrechte-Folgenabschätzung vor der Inbetriebnahme notwendig (Art. 27 Datenschutz-Grundverordnung,[28] im Folgenden: DSGVO).

Ausnahmen vom Anwendungsbereich gelten unter anderem für den Einsatz von KI zu militärischen, verteidigungspolitischen oder die nationale Sicherheit betreffende Zwecke (Art. 2 Abs. 3 KI-VO), ferner für die nicht-gewerbliche Nutzung von KI für Forschung und Entwicklung (Art. 2 Abs. 6 KI-VO). Auch Forschungs-, Test- und Entwicklungstätigkeiten zu KI-Systemen oder KI-Modellen, die vor dem Inverkehrbringen durchgeführt werden, sind ausgenommen (Art. 2 Abs. 8 KI-VO), ebenso sind Open-Source-KI-Systeme teilweise vom Anwendungsbereich ausgenommen (Art. 2 Abs. 12 KI-VO).

Die KI-VO soll durch harmonisierte Normen und Standards zur Unterstützung der Compliance ergänzt werden, wodurch auch die Nähe zum Produktsicherheitsrecht und der Technologiebezug deutlich wird. Ein einheitliches Kriterienset zur Überprüfung der Vertrauenswürdigkeit von KI-Systemen existiert bislang nicht. Vorschläge für Konformitätsprüfungen werden derzeit auf nationaler und internationaler Ebene diskutiert.

Entwürfe mit konkreten technischen Anforderungen werden auf europäischer Ebene in den Gremien Europäisches Komitee für Normung (im Folgenden: CEN) und Europäisches Komitee für elektrotechnische Normung (im Folgenden: CENELEC) erarbeitet. Harmonisierten Normen kommt im Rahmen der Konformitätsvermutung eine besondere Bedeutung zu, da bei Ausgestaltung und Kontrolle eines Hochrisiko-KI-Systems nach den harmonisierten Normen vermutet wird, dass die gesetzlichen Anforderungen eingehalten werden.[29] Die technischen Normen werden voraussichtlich erst Ende 2025 vorliegen.[30]

Die „Assessment List for Trustworthy Artificial Intelligence"[31] (im Folgenden: „ALTAI") der Hochrangigen Expertengruppe für Künstliche Intelligenz formuliert Fragen, die ganzheitliche Kriterien für KI-Systeme darstellen sollen. Sie berücksichtigen ökologische, datenschutzrechtliche, ethische und technische Aspekte. Sie sind nicht formal durch ein Audit überprüfbar.

Richtli-nien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz)

21 Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vor-schriften für Künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final, https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206, abgerufen am 08.08.2024.

22 Vgl. Möller-Klapperich, NJ 2024, 337, 338.

23 EU-Kommission, Weißbuch: Zur Künstlichen Intelligenz – ein europäisches Konzept für Exzellenz und Vertrauen, COM(2020) 65 final; Wendt, in: Wendt/Wendt, Das neue Recht der Künstlichen Intelligenz, 1. Aufl. 2024, § 3 Rn. 9 f.

24 Vgl. Von Welser, GRUR-Prax 2024, 485, 485.

25 Vgl. Erwgr. 2 KI-VO.

26 Europäisches Parlament, Künstliche Intelligenz: Chance und Risiken, https://www.europarl.europa.eu/topics/de/article/20200918STO87404/kunstliche-intelligenz-chancen-und-risiken, abgerufen am 06.08.2024.

27 https://www.bmj.de/SharedDocs/Zitate/DE/2024/0130_KI-VO.html, abgerufen am 06.08.2024.

28 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).

29 Wendt, in Wendt/Wendt, Das neue Recht der künstlichen Intelligenz, 1. Aufl. 2024, § 8 Rn. 1 f.

30 Wendt, in Wendt/Wendt, Das neue Recht der künstlichen Intelligenz, 1. Aufl. 2024, § 8 Rn. 2.

31 European Commission, Assessment List for Trustworthy Artificial Intelligence (ALTAI), https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment, abgerufen am 08.08.2024.

Die nationalen Normungsinstitute der Mitgliedsstaaten sind angehalten, ihre technische Expertise bereitzustellen und in die Arbeit der Komitees JTC 21[32], das von CEN und CENELEC gemeinsam ein-berufen wurde, einfließen zu lassen.

7 Prüflandschaft

Zu einer vertrauenswürdigen Prüflandschaft bedarf es einiger ineinandergreifenden Komponenten juristischer, technischer und organisatorischer Art. Die Umsetzung der gesetzlichen Richtlinien in Standards und die Definition von Prüfmetriken fällt teilweise darunter, aber auch zur technischen Entwicklung von Prüfmethoden und -werkzeugen. Diese können einerseits eingesetzt werden zur Validierung und Verifizierung bestehender KI-Systeme. Beispielweise können Labels definiert und vergeben werden.[33, 34, 35] Ebenso sollte schon während der Entwicklung darauf geachtet werden, dass die Richtlinien schon beim Entwurf einfließen und entsprechend implementiert werden.[36, 37, 38]

Zur Prüflandschaft gehört auch die Definition und Etablierung von nationalen und internationalen Instanzen und Organisationen, die eine entsprechende Zertifizierung abgeben können und dürfen (beispielsweise Third-Party Authorities, TPA).

Für deutsche Unternehmen ergibt sich in erster Linie die Bestrebung, Compliance mit den rechtlichen Vorgaben herzustellen. Dies kann angesichts der zahlreichen zu beachtenden Rechtsakte und der Komplexität der Materie nur mit einer KI-Strategie gelingen, die alle Aspekte der Verordnung und die erforderlichen Stakeholder innerhalb und außerhalb des Unternehmens berücksichtigt. Eine Prüflandschaft wird in Zukunft einen großen Beitrag zur KI-Compliance leisten. Standardisie-rung, Normen und Prüfkriterien kommen unter wirtschaftlichen Gesichtspunkten eine grundlegende Bedeutung zu. Als wesentliche Schlüsseltechnologie soll KI-Prozesse in der Industrie nach-haltiger, flexibler und effizienter gestalten. Angesichts großer Lieferketten gewinnt Interoperabilität in jeder Hinsicht an Relevanz.[39] Die Bemühungen für einheitliche Prüfkriterien können den Technologietransfer erleichtern und internationale Märkte erschließen. Die Schnelllebigkeit der Technologie stellt eine weitere Herausforderung dar. KI ist im Vergleich zu anderen Produkten deutlich schwerer zu generalisieren. Zudem ist zu berücksichtigen, dass viele ethische, soziale, politische oder auch klimaschutzrelevante Aspekte oft nicht konsensfähig sind und die Diskussionen auch die rechtliche und technische Entwicklung eng begleiten.

Die oben erwähnten Prüfkriterien und -methoden sind relevant für viele verschiedene Anwendungen, wie zum Beispiel in der Mobilität, Fertigung, Energie oder der Medizin[40], und soll einer-seits verallgemeinert[41, 42, 43] werden, aber auch anwendungsspezifische Aspekte umfassen. Insbesondere muss sichergestellt werden, dass diese im gesamten Lebenszyklus eingebettet werden. Dazu sollen Kriterien von Anfang an berücksichtigt werden und durch Rückkopplung aus dem Betrieb (Deployment) zur Entwicklung aktuell gehalten und ständig weiterentwickelt werden (Full Lifecycle Optimisation).

32 CEN/CENELEC, Artificial Intelligence, https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/, abgerufen am 08.08.2024.

33 Morik, K., Kotthaus, H., Heppe, L., Heinrich, D., Fischer, R., Pauly, A., & Piatkowski, N. (2021). The Care Label Con-cept: A Certification Suite for Trustworthy and Resource-Aware Machine Learning (arXiv:2106.00512). arXiv. https://doi.org/10.48550/arXiv.2106.00512

34 Scharowski, N., Benk, M., Kühne, S. J., Wettstein, L., & Brühlmann, F. (2023). Certification Labels for Trustworthy AI: Insights From an Empirical Mixed-Method Study. Proceedings of the 2023 ACM Conference on Fairness, Ac-countability, and Transparency, 248–260. https://doi.org/10.1145/3593013.3593994

35 Hauschke, A., Puntschuh, M., Hallensleben, S., & Loh, W. (2022). VDE SPEC 90012 V1.0—VCIO based description of systems for AI trustworthiness characterisation. https://www.vde.com/en/working-areas/standards/spec/vde-spec-publications

36 van Bekkum, M., de Boer, M., van Harmelen, F., Meyer-Vitali, A., & Teije, A. ten. (2021). Modular design patterns for hybrid learning and reasoning systems. Applied Intelligence, 51(9), 6528–6546. https://doi.org/10.1007/s10489-021-02394-3

37 Meyer-Vitali, A., & Mulder, W. (2024). Engineering Principles for Building Trusted Human-AI Systems. In K. Arai (Ed.), Intelligent Systems and Applications (pp. 468–485). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-66428-1_30

38 Hermanns, H., Lauber-Rönsberg, A., Meinel, P., Sterz, S., & Zhang, H. (2024). AI Act for the Working Programmer (arXiv:2408.01449). arXiv. https://doi.org/10.48550/arXiv.2408.01449

39 DIN, Deutsche Normungsroadmap Künstliche Intelligenz, 2022, S. 179.

40 D Schwabe, K Becker, M Seyferth, A Klaß, T Schaeffter (2024). The METRIC-framework for assessing data quality for trustworthy AI in medicine: A systematic review. Npj Digital Medicine, 7(1), 1–30. https://doi.org/10.1038/s41746-024-01196-4

41 Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 99, 101896.

42 Awadid, A., Amokrane-Ferka, K., Sohier, H., Mattioli, J., Adjed, F., Gonzalez, M., & Khalfaoui, S. (2024). AI Systems Trustworthiness Assessment: State of the Art. Workshop on Model-Based System Engineering and Artificial Intel-ligence – MBSE-AI Integration 2024. https://hal.science/hal-04400795

43 Rutinowski, J., Klüttermann, S., Endendyk, J., Reining, C., & Müller, E. (2024). Benchmarking Trust: A Metric for Trustworthy Machine Learning. In L. Longo, S. Lapuschkin, & C. Seifert (Eds.), Explainable Artificial Intelligence (pp. 287–307). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63787-2_15