Lernende Systeme
DIE PLATTFORM FÜR KÜNSTLICHE INTELLIGENZ

Zertifizierung von KI-Systemen

Whitepaper von Jessica Heesen,
Jörn Müller-Quade, Stefan Wrobel et al.
Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik;
Arbeitsgruppe Technologische Wegbereiter und Data Science

Kurzfassung

Die Zertifizierung von KI-Systemen kann dazu beitragen, das Vertrauen in Technologie und Anwendung zu stärken und so KI-Systeme in die Anwendung zu bringen. Ziel eines gelungenen Zertifizierungsverfahrens sollte grundsätzlich sein, Standards zu garantieren und gleichzeitig aber Überregulierung zu vermeiden und Innovationen zu ermöglichen.

Mitglieder aller Arbeitsgruppen der Plattform Lernende Systeme haben unter Federführung der beiden Arbeitsgruppen IT-Sicherheit, Privacy, Recht und Ethik sowie Technologische Wegbereiter und Data Science zusammen mit Gastautorinnen und -autoren Kriterien definiert, die zum einen zur Orientierung für die Zertifizierung von KI-Systemen herangezogen werden können und zum anderen die Entscheidung über die Notwendigkeit einer Zertifizierung unterstützen können.

Zudem geben die Expertinnen und Experten einen Überblick, wie eine effiziente Prüfinfrastruktur ausgestaltet sein könnte. Damit greifen sie zugleich den aktuellen Diskussionsstand auf und knüpfen an das bereits veröffentlichte **Impulspapier Zertifizierung von KI-Systemen** an (vgl. Heesen et al. 2020a).

Bestehende Zertifizierungsinitiativen und Verfahren für KI-Systeme

Es existieren bereits zahlreiche nationale sowie internationale Anknüpfungspunkte für gelungene Zertifizierungsinitiativen von KI-Systemen. Dazu gehören unter anderem politische Initiativen wie das Weißbuch zur Künstlichen Intelligenz der Europäischen Kommission, die Stellungnahme der Bundesregierung zu diesem Weißbuch der Europäischen Kommission, Good Practice-Beispiele aus der KI-Forschung und -Anwendung sowie weitere Initiativen zu technischen Lösungen, Standardisierung und zur Prüfung und Auditierung von KI-Systemen. Diese können einen ersten Ansatzpunkt für eine gelungene Zertifizierung von

KI-Systemen, für die bis dato in Deutschland kaum gültige und anerkannte Normen und Standards existieren, bilden.

In welchen Fällen ist eine Zertifizierung von KI-Systemen notwendig?

Nicht für jedes KI-System vor dem Hintergrund seines jeweiligen Anwendungskontextes wird eine Zertifizierung notwendig sein. Für eine gelungene Zertifizierung von KI-Systemen müssen Anwendungsfälle unterschieden werden, in denen eine Zertifizierung notwendig ist und solche, in denen es keine Konformitätsüberprüfung durch unabhängige Drittstellen braucht. Die Notwendigkeit einer Zertifizierung von KI-Systemen kann hier aus der Kritikalität des KI-Systems in einem bestimmten Anwendungskontext abgeleitet werden. Diese ist abhängig vom Anwendungskontext und setzt sich aus der Gefährdung von Menschenleben und anderen Rechtsgütern und dem Umfang der Handlungsoptionen von Menschen in bestimmten Anwendungskontexten zusammen. Anhand des Ausmaßes der Kritikalität kann auf den möglichen Regulierungsbedarf geschlossen werden. Wem die Kritikalitätsbestimmung obliegt, ist abhängig vom Anwendungskontext: Möchte der Staat für bestimmte Anwendungskontexte eine verpflichtende Zulassung oder Zertifizierung einführen, fällt ihm als regulierende Instanz die Verantwortung für die Kritikalitätseinschätzung zu. In Fällen, in denen die Zertifizierung auf freiwilliger Basis initiiert wird, kann die Kritikalitätseinschätzung von den Unternehmen selbst durchgeführt werden.

An welchen Gegenständen und Kriterien soll sich eine Zertifizierung von KI-Systemen orientieren?

Wurde die Notwendigkeit einer Zertifizierung festgestellt, so stellt sich die Frage, wie und anhand welcher Prüfkriterien KI-Systeme zertifiziert werden können. Grundsätzlich sollte bei der Zertifizierung von KI-Systemen sowohl auf allgemeine als auch auf branchenspezifische Normen, Standards, Prüfverfahren und auch gesetzliche Vorschriften zurückgegriffen beziehungsweise an diese angeschlossen werden (ISO- und DIN-Normen, geltendes (EU-)Recht). Wo nötig, müssen unter Umständen auch Lücken geschlossen oder Anpassungen vorgenommen werden. Es sollte nicht zu einer Situation kommen, in der KI-spezifische Standards und Regularien mit weitergefassten Zertifizierungen und Regularien konkurrieren.

Gegenstand der Zertifizierung

Für die Zertifizierung von KI-Systemen können verschiedene Zertifizierungsarten unterschieden werden, die sinnvoll sind. Für den Bereich der KI-Systeme empfehlen die Expertinnen und Experten entweder eine **Produkt- oder eine Mischform aus Produkt- und Prozesszertifizierung**. Die Produkt- und Prozesszertifizierung unterscheiden sich hinsichtlich Zielsetzung und Betrachtungsgegenstand, weshalb manche Prüfkriterien besser im Rahmen einer Produktzertifizierung und andere besser innerhalb einer Prozesszertifizierung abgefragt bzw. umgesetzt werden können.

Eine **Produktzertifizierung** ist eine neutrale Überprüfung der Einhaltung zugesicherter Produkteigenschaften auf der physischen Produktebene. Eine Produktzertifizierung sollte früh ansetzen (im Optimalfall bereits bei der Spezifikation des Produkts). Eine Eigenschaft der Produktzertifizierung ist, dass oft mehrere Verfahren kombiniert werden müssen, um die Kriterien zu überprüfen.

Eine Alternative oder auch Ergänzung zu einer Produktzertifizierung kann eine **Prozesszertifizierung** sein. Sie untersucht die Qualität des Herstellungs- und Entwicklungs- sowie des Einführungsprozesses im Allgemeinen und der Implementation der KI-Lösung im Besonderen. Sie dient der Reflexion der zu prüfenden Prozesse und kann unter Umständen auch durch den Hersteller oder den Betreiber selbst vorgenommen werden. Wenn ein zertifiziertes Verfahren mit speziell auf KI zugeschnittenen Instrumenten angewandt wird, kann eine Prozesszertifizierung wichtige Implikationen für Fragen der Verantwortung und Haftung geben. Gleichzeitig können gut durchgeführte Prozesse auch möglichen Fehlfunktionen vorbeugen und so zu besseren Produkten führen.

Prüfkriterien der Zertifizierung

Die anzulegenden Prüfkriterien lassen sich hinsichtlich ihrer Verbindlichkeit im Rahmen einer Zertifizierung in **Mindestkriterien**, die immer erfüllt und im jeweiligen Anwendungskontext abgeprüft werden müssen, sowie **darüber hinausgehende Kriterien** unterteilen, die abgeprüft werden können und somit eine Art „Zertifizierung Plus“ ermöglichen. Diese Kriterien sind von großer Bedeutung für eine positive und wertorientierte Entwicklung von vertrauenswürdiger KI und gehen über die Mindestanforderungen hinaus, die „vornehmlich“ der Verhinderung von evidenten und unmittelbaren Gefährdungen dienen.

Mindestkriterien, die im Rahmen einer Zertifizierung überprüft werden müssen:

Mindestkriterien

- Transparenz, Nachvollziehbarkeit, Nachprüfbarkeit und Verantwortlichkeit
- Funktionale Sicherheit/Safety/inkl. Produktsicherheit und Zuverlässigkeit
- Vermeidung von nicht-intendierten Folgewirkungen (auf andere Systeme, Menschen und die Umwelt)
- Gerechtigkeit im Sinne von Gleichheit und Diskriminierungsfreiheit
- Schutz der Privatheit und der Persönlichkeit
- Selbstbestimmung inkl. Transparenz über den Einsatz des KI-Systems und die Rolle des Menschen im Entscheidungsprozess

Darüber hinausgehende Kriterien, die im Rahmen einer Zertifizierung als „Zertifizierung Plus“ überprüft werden können:

Darüber hinausgehende Kriterien

- Offene Schnittstellen und Systemoperabilität
- Menschenzentrierung und Nutzerfreundlichkeit (Usability) inkl. Partizipation, Schutz des Einzelnen, sinnvolle Arbeitsteilung und förderliche Arbeitsbedingungen
- Nachhaltigkeit
- Kennzeichnung und Begrenzung der Systemfunktionalität

Voraussetzungen für eine gelingende Zertifizierung

KI-Systeme weisen eine besondere Dynamik auf, insbesondere weiterlernende Systeme entwickeln sich im laufenden Betrieb weiter. Dies muss bei der Wahl des Zeitpunkts und des Detailgrads der Zertifizierung berücksichtigt werden. Die Zertifizierung sollte durchgeführt werden, bevor das Produkt oder die Dienstleistung in den Verkehr gebracht wird, bei weiterlernenden Systemen

sollte die Zertifizierung regelmäßig wiederholt werden. Detailgrad und Prüftiefe bei der Zertifizierung sollten sich ebenfalls am Kritikalitätslevel eines KI-Systems in seinem Anwendungsgebiet orientieren – je höher die Kritikalität im Anwendungskontext eingeschätzt wird, desto umfangreicher sollten der Detailgrad und die Prüftiefe der Zertifizierung ausfallen.

Für eine gelingende Zertifizierung von KI-Systemen bedarf es neben der Einschätzung der Kritikalität und der Erstellung eines Prüfkatalogs zudem einer effektiven organisatorischen und technischen Infrastruktur. Damit die Konformitätsbewertung von KI-Systemen gelingt, sind etwa technische Voraussetzungen mit Blick auf Prüfwerkzeuge, Software und Testumgebungen zu erfüllen. Organisatorische Strukturen und Prozesse in Unternehmen sollten künftig zudem zu einem wichtigen, komplementären Baustein einer Zertifizierung von KI-Systemen werden. Die Kooperation zwischen Zertifizierungsstellen und Forschungsinstituten ist vor allem wichtig, um einer dynamischen Verfasstheit der Prüfstellen Rechnung zu tragen, die auf KI-Innovationen adäquat reagieren kann.

Mögliche Gestaltungsoptionen

Im Einklang mit diesen Überlegungen können konkrete Gestaltungsoptionen zur Etablierung einer gelungenen Zertifizierung von KI-Systemen abgeleitet werden, die verschiedene Akteursgruppen adressieren:

Die Forschung könnte…

- die Details der Zertifizierungsverfahren in interdisziplinären Forschungsverbünden eingehender erforschen, um dazu beizutragen, Prüfwerkzeuge zur Evaluation von KI-Systemen zu entwickeln und allgemeine Kriterien wie „Transparenz" für Wirtschaft, Nutzende und Technikentwicklung operationalisierbar zu machen. Auf dieser Basis kann die Forschung Politik, Unternehmen und Zivilgesellschaft noch eingehender zu den Chancen, Risiken und Konsequenzen der einzelnen Technologien und Anwendungsbereiche beraten.
- interdisziplinär technologische Lösungen und Methoden entwickeln, um sicherzustellen, dass KI-Systeme vertrauenswürdig sind.
- mit Unternehmen zusammen vertrauenswürdige KI-Methoden entwickeln (erklärbare, explainable AI (XAI)).
- ihre modernsten Infrastrukturen zur Verfügung stellen, damit diese Anknüpfungspunkte für erste Zertifizierungsvorhaben bilden können.
- erforschen, wo traditionelle Signalverarbeitungsmethoden aufhören und wo KI anfängt, um eine genauere Gesetzgebung zu ermöglichen.
- sich an der Entwicklung eines Konzepts zur Ausbildung von KI-Prüfingenieuren beteiligen.

Die Unternehmen könnten…

- die Bildung von Vertrauen in KI-Systeme unterstützen, indem sie **freiwillig** ethische und technische Standards ausarbeiten und offenlegen und sich verstärkt dem Einsatz von erklärbarer KI widmen. Dies stellt eine Basis für die Debatte um die Zertifizierung und Regulierung von KI-Systemen dar.
- sich an der Schaffung entsprechender Standards beteiligen und entsprechende Bedarfe identifizieren.
- sich jeweils branchenspezifisch darüber austauschen, welche Aspekte von KI-Systemen in ihrem Anwendungskontext als kritisch zu betrachten sind und wie von Seiten der Hersteller Best Practices für solche Fälle etabliert werden

könnten. Als Orientierungspunkt für einen solchen Austausch können bestehende Konzepte und Gestaltungsrichtlinien dienen. Dieser Austausch könnte eine Basis dafür darstellen, dass Unternehmen in vertrauenswürdige KI investieren und entsprechende Geschäftsmodelle entwickeln.
- ihre modernsten Infrastrukturen zur Verfügung stellen, damit diese Anknüpfungspunkte für erste Zertifizierungsvorhaben bilden können.
- Beschäftigten im Rahmen der innerbetrieblichen Weiterbildung spezielle Schulungen anbieten, die einen souveränen Umgang mit KI-Systemen zum Ziel haben.

Die Zivilgesellschaft könnte…

- Bereiche identifizieren, für die aus Sicht von Verbraucherinnen und Verbrauchern sowie Bürgerinnen und Bürgern eine Regulierung erforderlich ist. Genauso könnten Bereiche identifiziert werden, für die keine Regulierung notwendig ist und die sich für eine zivilgesellschaftliche Konformitätsprüfung etwa über ein Gütesiegel anbieten könnten.
- auf der Basis bestehender und künftig entwickelter Kriterien und Gestaltungsrichtlinien sowie der rechtlichen Rahmenbedingungen die Rolle eines „Watchdogs" einnehmen und so auf die Einhaltung der Kriterien, Richtlinien und Regeln drängen, um auf diese Weise den Einsatz von KI mitzugestalten.

Darüber hinaus bedürfen einige Aspekte in diesem Rahmen eines **gesellschaftlichen Diskurses** unter Einbeziehung aller relevanten Stakeholder aus Wirtschaft, Wissenschaft und Zivilgesellschaft. Diese betreffen:

- die Definition von Kritikalitätsstufen. Dies erfordert eine Diskussion zu tragbaren Risiken und zur gerechten Verteilung der Vorteile, die aus KI-Anwendungen hervorgehen. Ferner ist eine Aufklärung über die Funktions- und Wirkungsweise und Einsatzmöglichkeiten von KI notwendig, um zu einer realistischen Einschätzung ihrer Potenziale zu kommen.
- die Notwendigkeit der Zertifizierung von KI-Systemen: Dies gilt vor allem für die Frage, inwiefern weitere Normen und Standards über die bereits existierenden Sicherheits- und Transparenzstandards von technischen (industriellen) Systemen hinaus notwendig sind.
- die Art, wie wir zukünftig mit KI leben, lernen und arbeiten wollen: Ziel ist eine Entwicklung von KI-Systemen, die so umgesetzt sind, dass sie menschliche Kompetenzen erweitern und nicht beschneiden. Ein solcher breiter Diskurs von KI stellt die Basis dar, auf der künftig über Bewertungs- und Prüfkriterien für KI-Systeme diskutiert werden könnte.

Impressum
Herausgeber: Lernende Systeme – Die Plattform für Künstliche Intelligenz | Geschäftsstelle | c/o acatech | Karolinenplatz 4 | D-80333 München | kontakt@plattform-lernende-systeme.de | www.plattform-lernende-systeme.de |
Folgen Sie uns auf Twitter: @LernendeSysteme | Stand: November 2020 | Bildnachweis: Tierney/Adobe Stock/Titel

Diese Kurzfassung entstand auf Grundlage des Whitepapers *Zertifizierung von KI-Systemen – Kompass für die Entwicklung und Anwendung vertrauenswürdiger KI-Systeme*, München, 2020. Es wurde erstellt von der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik sowie der Arbeitsgruppe Technologische Wegbereiter und Data Science der Plattform Lernende Systeme. Die Originalfassung der Publikation ist online verfügbar unter: https://www.plattform-lernende-systeme.de/publikationen.html

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

acatech
DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN