



Whitepaper Zertifizierung, 2023

# Welche Zertifizierung ist hilfreich für Verbraucher\*innen?

Acht Thesen zu wirksamen freiwilligen Zertifikaten  
von KI-Produkten

Dr. Gergana Baeva, Robert Brunner, Filiz Elmas, Benjamin Fresz, Christine Fuß,  
Prof. Dr. Claudia Heß, JProf. Dr. Philipp Kellmeyer, Malin Mahner,  
Prof. Dr. Heiko Paulheim, Dr. Maximilian Poretschkin, Michael Puntschuh,  
Annegrit Seyerlein-Klug

## Welchen Mehrwert bringen KI-Zertifikate für Verbraucher\*innen?

Ob bei Lebensmitteln, beim Fahrstuhl oder dem Auto: Zertifikate sollen anzeigen, dass Produkte gut und sicher sind. Dafür müssen sie bestimmte überprüfbare Anforderungen erfüllen oder in einem strukturierten Prozess mit wirksamen Maßnahmen zur Qualitätssicherung und Risikokontrolle entwickelt worden sein.<sup>1</sup> Zertifikate schaffen daher Orientierung für Verbraucher\*innen in Bezug auf die Qualität von Produkten und Diensten, die sie selbst kaum bewerten können. Welche\*r Verbraucher\*in ist schon fähig, die Sicherheit eines Gläschens Babybrei oder eines Fahrstuhls eigenständig zu bewerten?

Dies gilt auch für komplexe KI-Systeme, bei denen oftmals kaum nachvollziehbar ist, wie sie zu konkreten Ergebnissen kommen.<sup>2</sup> Dies gilt insbesondere für Lai\*innen. Zertifikate und Gütesiegel können wertvolle Indikatoren sein, um die Vertrauenswürdigkeit einer Anwendung zu beurteilen. Sie haben zudem die Funktion, verschiedene Produkte und Dienste vergleichbar darzustellen, und sind daher ein mögliches Instrument für Verbraucher\*innen, um solche Systeme souveräner und selbstbestimmter zu nutzen.

Wie nützlich Zertifikate für Verbraucher\*innen tatsächlich sind, hängt jedoch von verschiedenen Faktoren und nicht zuletzt von der Frage, was unter Zertifizierung verstanden wird. In der Alltagssprache werden Begriffe wie Zertifikate, Gütesiegel und Labels oft synonym genutzt. Prüfende Organisationen verstehen unter einer Zertifizierung, dass ein „unparteiischer Dritter“ prüft, ob ein Erzeugnis, ein Verfahren oder eine Dienstleistung mit bestimmten Standards und Normen übereinstimmt.<sup>3</sup> Im Ergebnis kennzeichnen Zertifikate auf diese Weise geprüfte Produkte oder Prozesse.<sup>4</sup> Solche unabhängigen Zertifizierungen von KI-Systemen stehen im Mittelpunkt der vorliegenden Arbeit.

## KI-Zertifikate jenseits gesetzlicher Pflichten

Die von der *Europäischen Union* geplante KI-Verordnung<sup>5</sup> soll erstmals einen gesetzlichen Rahmen für KI-Systeme festsetzen, der aber viele Umsetzungsprobleme aufwirft. So sind Prüfungen gesetzlich nur für den Bereich der Hochrisiko-KI-Systeme vorgesehen. Das Ziel solcher Prüfungen ist, die Konformität hinsichtlich der gesetzlichen Anforderungen zu bewerten. Die meisten Hochrisiko-Systeme sollen lediglich im Rahmen einer Selbstbewertung und nicht durch einen akkreditierten, unabhängigen Dritten geprüft werden. Ergebnisse solcher Selbstbewertungen sind keine Zertifikate im eigentlichen Sinne. Die Möglichkeit, KI-Systeme für nicht vorgesehene Zwecke zu nutzen, kann ein weiteres Problem darstellen. Zum Beispiel gelten herkömmliche Sprachassistenzsysteme oder Smartwatches mit ihren üblichen Einsatzzwecken laut Verordnungsentwurf nicht als Hochrisiko-Systeme. Jedoch ist es möglich, Sprachassistenten als Notfallmeldesysteme oder eine Smartwatch zur Aufzeichnung und Analyse von medizinisch relevanten Patient\*innen-daten zu nutzen, die zur Diagnose und Therapie von Erkrankungen eingesetzt werden können. In diesem neuen Anwendungskontext würden diese Systeme in den Hochrisikobereich fallen und müssten entsprechende Anforderungen umsetzen. Im Zusammenhang mit KI-Systemen können also Graubereiche entstehen, die zu anderen Prüfungsanforderungen führen als bislang üblich. Die Gefahr des Missbrauchs sollte bei KI-Systemen sorgfältig berücksichtigt werden, insbesondere wenn sie für risikoreiche Zwecke genutzt werden könnten.

Jenseits des Hochrisikobereichs sind auch freiwillige Zertifizierungen sinnvoll. Diese bieten im Vergleich zu gesetzlichen Verpflichtungen mehr Gestaltungsraum und können weitere Anforderungen abdecken. So könnten freiwillige Zertifikate als Nachweis dafür dienen, dass KI-Systeme mit geringen Risiken bestimmten ethischen Prinzipien entsprechen, wie im neuen Vorschlag zu KI-Verordnung des *Europäischen Parlaments* gefordert.<sup>6</sup> Denn auch bei nicht risikoreichen Anwendungen haben Betroffene ein Interesse daran, dass diese zuverlässig, robust und sicher sind. Freiwillige Zertifikate werden aus Sicht von Unternehmen zudem als Instrumente betrachtet, um das Vertrauen der Verbraucher\*innen in KI-Anwendungen zu fördern.<sup>7</sup>

1 Glatzner 2016, S. 290f.  
2 Schmid 2022, S. 291.  
3 DIN EN ISO/IEC 17000:2020 „Konformitätsbewertung – Begriffe und allgemeine Grundlagen“.  
4 Kneip 2022, S. 731.  
5 Europäische Kommission 2021.  
6 Europäisches Parlament 2023, Art. 4a, S. 142f.  
7 Kneip 2022, S. 731.

## Mehr Orientierung für Verbraucher\*innen durch Zertifikate?

Entsprechen freiwillige Zertifizierungen allerdings keinen klaren Anforderungen, drohen sie schnell an Aussagekraft und Relevanz zu verlieren und unübersichtlich für Verbraucher\*innen zu werden. Das Beispiel Nachhaltigkeitssiegel aus der Textil- und Lebensmittelindustrie zeigt eine solche Fehlentwicklung eindrücklich. Eine Studie zur Verwendung von Nachhaltigkeitslabeln auf großen E-Commerce-Plattformen in Deutschland stellte fest, dass eine große Anzahl der Artikel als nachhaltig gekennzeichnet wird.<sup>8</sup> Nur 14 Prozent der gekennzeichneten Produkte trugen jedoch glaubwürdige, von Dritten geprüfte Nachhaltigkeitssiegel. Zwei Drittel der Artikel wurden mit einem eigenen Label ausgestattet, das von den Hersteller\*innen oder Händler\*innen selbst definiert und ausgestellt worden war.<sup>9</sup> Das Problem der eigenen Labels ist nicht nur, dass die Anforderungen unklar sind, sondern auch, dass Unternehmen sich dabei häufig nur auf bestimmte Aspekte hinsichtlich der Nachhaltigkeit der Produkte beschränken. Zudem erfolgt die Prüfung nicht durch unabhängige Dritte. Die Zuverlässigkeit des Labels kann durch Interessenkonflikte abgeschwächt werden. Außerdem entsteht so eine Vielzahl unterschiedlicher Anforderungskataloge. Da Verbraucher\*innen keinen Überblick über diese vielfältigen Labels haben können, tragen diese häufig eher zu Verwirrung und Skepsis bei.<sup>10</sup>

Zentrale Fragen mit Blick auf Verbraucher\*innen sind: Welche Anforderungen an die Vertrauenswürdigkeit stellen sie? Inwieweit weichen diese Anforderungen von gesetzlichen Auflagen ab? In welchen Bereichen gehen sie über diese hinaus? Umfragestudien zeigen, dass sich Verbraucher\*innen ein besseres Verständnis und mehr Transparenz beim KI-Einsatz wünschen.<sup>11</sup> Gleichzeitig zeigen Befragungen einen starken Wunsch nach externer Überprüfung von Zuverlässigkeit und Vertrauenswürdigkeit der Systeme.<sup>12</sup> Zudem werden oft Lösungen präferiert, die eine menschliche Aufsicht über die KI-Ergebnisse vorsehen.<sup>13</sup> Solche Wünsche lassen sich in den verschiedenen Einsatzkontexten nicht immer gleich umsetzen. Denn Marktsituation und Kontext sind für die Wirkung freiwilliger Zertifikate entscheidend: Handelt es sich um eine Situation, in der die Verbraucher\*innen die KI-Nutzung ohne Nachteile ablehnen können? Können sie ihr Wissen um Zertifizierung und

Vertrauenswürdigkeit von KI-Systemen in ihre Handlungsspielräume einbeziehen? Insgesamt ist bisher nicht ausreichend erforscht, welche Bedenken Verbraucher\*innen haben und wie ihre Handlungssouveränität durch begleitende Informationen und Zertifikate gesteigert werden kann.

## Ziele und Grenzen des vorliegenden Whitepapers

In diesem Whitepaper fokussieren wir uns auf die durch die KI-Verordnung nicht erfasste freiwillige Zertifizierung. Dabei stehen die Interessen und Sichtweisen der Verbraucher\*innen im Vordergrund. Als Verbraucher\*innen betrachten wir nicht nur die unmittelbaren Nutzer\*innen von KI-Systemen, sondern auch Privatpersonen, die indirekt von KI-Ergebnissen betroffen sein können. Zudem berücksichtigen wir Menschen in der Rolle als souveräne Datensubjekte, deren personenbezogene Daten in die Entwicklung von KI-Systemen einfließen. Dieses breite Verständnis des Verbraucher\*innenbegriffs entspricht dem Selbstverständnis des *Zentrums für vertrauenswürdige Künstliche Intelligenz (ZVKI)*.<sup>14</sup>

Wir stellen Thesen auf, wie freiwillige Zertifizierungen von KI-Systemen für Verbraucher\*innen nützlich sein können. Die im Verlauf vorgestellten und begründeten acht Thesen wurden in einer Reihe von interdisziplinären Workshops erarbeitet. Dabei identifizierten wir Herausforderungen einer KI-Zertifizierung aus Verbraucher\*innensicht und präzisierten diese iterativ. Die Gruppendiskussionen dienten als Grundlage dieses Whitepapers und bauten auf Erkenntnissen der Fach-Arbeitsgruppe „KI Zertifizierung“ des ZVKI auf. Alle Co-Autor\*innen waren an der Erstellung des finalen Texts beteiligt.

Die aufgestellten Thesen thematisieren grundlegende Anforderungen an eine KI-Zertifizierung aus einer übergeordneten Perspektive, unabhängig von den spezifischen Details einer KI-Anwendung. Somit liefern sie keine Aussagen zu konkreten Inhalten der Zertifizierung. Detaillierte Empfehlungen setzen eine umfassende Analyse von Einsatzfällen voraus, die in diesem Rahmen nicht geleistet werden kann.

8 Gossen et al. 2022.

9 Ebd., S. 5.

10 Ebd., S. 2f.

11 MeMo\*KI 2022.

12 TÜV 2021, S. 30ff.; Busse/ Baeva 2022.

13 Overdiek/Petersen 2022, S. 21.

14 Vgl. dazu auch Baeva 2022.

## Wie sollen KI-Systeme zertifiziert und KI-Zertifikate vermittelt werden?

Um durch KI-Zertifikate Sicherheit und Transparenz für Verbraucher\*innen zu schaffen, müssen eine Reihe von Herausforderungen bewältigt werden. Diese betreffen den Umfang von Zertifizierungsprozessen, die Rahmenbedingungen und die konkrete Infrastruktur für ihre Umsetzung. Zudem geht es auch um Fragen der Kommunikation

von Zertifikaten, die auf diese Weise erstellt werden. Sie müssten den Verbraucher\*innen erst bekannt sein, um eine Wirkung entfalten zu können.

Im Fokus stehen vor allem die besonderen Anforderungen an freiwillige Ansätze. Die aufgezeigten Herausforderungen gelten jedoch auch für verpflichtende Zertifizierungen, da beide aus Verbraucher\*innensicht nur dann einen Mehrwert schaffen, wenn sie relevante Risiken zuverlässig und wirksam erfassen und durch Tests überprüfen können.

**THESE 1:** Freiwillige Zertifizierungen brauchen klare Rahmenbedingungen. Dazu gehören die Ausarbeitung von Standards mit Mindestanforderungen sowie eine Durchführung durch unabhängige Prüforganisationen.

Damit freiwillige Zertifikate aussagekräftig sind, müssen sie sich an etablierten Konformitätsprüfverfahren orientieren. Das umfasst die Aushandlung von Mindeststandards für verschiedene Einsatzkontexte sowie geeignete Prüfverfahren, um relevante Risiken erkennen und bewerten zu können. Eine unabhängige Prüfung durch Dritte steigert dabei die Relevanz der Prüfung aus Sicht der Verbraucher\*innen.

Für den Bereich der verpflichtenden Zertifizierungen sind konkrete Anforderungen festgeschrieben und werden in geeigneten Standards und Normen operationalisiert. Bei freiwilligen Zertifizierungen müssen die Anforderungen erst ausgehandelt werden. Diese müssen nicht zwangsläufig mit den Auflagen an Hochrisiko-KI-Systeme übereinstimmen, jedoch sollten sie relevante Risiken abdecken. Die allgemeinen Prinzipien, die im Vorschlag des *Europäischen Parlaments* (Art. 4a) die Entwicklung und den Einsatz von vertrauenswürdigen KI-Systemen leiten sollen, bieten hierfür einen geeigneten Rahmen.

Mindestanforderungen müssen zudem einheitlich dokumentiert sein, damit sie nachvollziehbar sind. Welche Anforderungen es genau sind, hängt vom Einsatzkontext ab und sollte in gesellschaftlichen Debatten sowie im Rahmen von Standardisierungsverfahren weiter präzisiert werden. Dabei sollen Eigenschaften von KI-Systemen wie eine direkte Interaktion mit Menschen, eine effektive menschliche Aufsicht oder ein Risiko für Grundrechte wie die Meinungsfreiheit berücksichtigt werden. Insbesondere Auswirkungen auf Grundrechte sind schwer zu bewerten, da beim KI-Einsatz oft verschiedene Rechte gegeneinander abgewogen werden müssen, beispielsweise effektiven Datenschutz versus mögliche KI-vermittelte Diskriminierungen im Bewerbungsmanagement. Gleichzeitig ist es nötig, solche Abwägungen zu diskutieren – besonders deshalb, weil sich viele Risiken erst ab einem bestimmten Verbreitungsgrad von Technologien manifestieren. Gleichzeitig sollten die Anforderungen so gestaltet sein, dass sie der rasanten Entwicklung der KI-Technologie standhalten können.

Eine Zertifizierung durch externe Prüforganisationen erhöht die Zuverlässigkeit und Akzeptanz von freiwilligen Zertifikaten. Selbstbewertungen bergen die Gefahr zu niedriger Bewertungsschwellen, einer selektiven Auswahl von Bewertungskriterien, einer unzureichenden Prüfung, eines lückenhaften Monitorings und somit von unzureichendem Schutz. Der Aufbau einer Infrastruktur ist wichtig, um gegen unzureichende Zertifikate vorgehen zu können. Dazu gehören Testdatensätze, validierte Prüfwerkzeuge, aber auch ausgebildete Prüfer\*innen.

Damit Zertifikate zuverlässige Garantien für Verbraucher\*innen darstellen, müssen Verfahren und Anforderungen auch bei freiwilligen Zertifizierungen klar geregelt sein. Wünschenswert ist zudem eine gezielte Förderung von qualitativ hochwertigen freiwilligen Zertifizierungsansätzen, die mittels etablierter und unabhängiger Prüfverfahren erstellt werden.

**THESE 2:** Um die Vertrauenswürdigkeit von KI-Systemen zu fördern, braucht es eine Zertifizierung von Produkten und Prozessen. Produkt- und Prozessprüfungen ergänzen sich dabei gegenseitig. Insbesondere für KI-Systeme in kritischen Anwendungskontexten kann auf eine Produktprüfung nicht verzichtet werden.

Aus Sicht von Verbraucher\*innen zeichnen Zertifikate sichere und vertrauenswürdige Produkte aus. Das setzt umfassende Prüfungen von Produkteigenschaften voraus, um beispielsweise nachzuweisen, dass KI-Systeme Kriterien der Robustheit, Genauigkeit, Transparenz oder Sicherheit entsprechen. Einige dieser Kriterien sind jedoch bei KI-Systemen kaum technisch zu erfassen. Die Zertifizierung von KI-Anwendungen setzt daher verschiedene Schritte voraus, die sowohl Produkt- als auch Prozesseigenschaften erfassen und die in das finale Zertifikat einfließen.

## Gegenstände der Zertifizierung

Für eine Bewertung der Vertrauenswürdigkeit können verschiedene Prüfgegenstände relevant sein:<sup>15</sup>

*Produktzertifizierung:* Das KI-System als Produkt wird in Bezug auf verschiedene Anforderungen getestet. Dabei wird überprüft, ob es bestimmten Standards entspricht, z. B. in Bezug auf Transparenz und Nachvollziehbarkeit, Robustheit, Genauigkeit und Sicherheit oder Datenschutz, und die geplante Funktion erfüllt.

*Prozesszertifizierung:* Hierbei werden Prozesse der Entwicklung, des Qualitäts- und des Risiko-Managements in Bezug auf das konkrete System strukturiert bewertet. Dabei wird beispielsweise herausgearbeitet, wie Risiken erkannt oder wie Datensätze in Bezug auf Datenschutz und Repräsentativität verwaltet werden.

*Organisationszertifizierung:* Im Rahmen einer Gesamtbewertung von Unternehmen kann die grundsätzliche Fähigkeit zertifiziert werden, vertrauenswürdige KI-Systeme zu entwickeln. Dabei können Strukturen bewertet werden, beispielsweise Art, Umfang und Qualität von KI-Governance-Mechanismen.<sup>16</sup>

Produktzertifizierungen können einzelne Eigenschaften und Leistungen des KI-Systems technisch präzise bewerten, sind aber nicht hinsichtlich aller relevanten Bewertungskriterien gleichermaßen umsetzbar. Prozesszertifizierungen sind unter anderem dann relevant, wenn technisch schwer messbare Kriterien wie Fragen des Grundrechtsschutzes aufgefangen werden sollen oder eine Vielzahl von (unkritischen) KI-Systemen und Services durch ein Unternehmen bereitgestellt wird und eine Prüfung jedes einzelnen Produkts nicht marktfähig wäre. Eine genaue Dokumentation des Entwicklungsprozesses und eine Begründung von Entscheidungen nach festgelegten Kriterien sollten Teil solcher Zertifizierungen sein. Zudem sollten Zertifizierungen auch für kleinere Unternehmen bezüglich des Zeitaufwands und des methodischen Wissens machbar sein.

Verschiedene Formen und Ebenen der Zertifizierung können dazu beitragen, das Vertrauen der Verbraucher\*innen in die KI-Systeme zu fördern. Wichtig ist dabei, Prüfungen von Produkt- und Prozesseigenschaften sinnvoll zu kombinieren,<sup>17</sup> um aussagekräftige Bewertungen zu konkreten KI-Systemen erzielen zu können.

<sup>15</sup> Vgl. dazu auch Heese et al. 2020, S. 22f., sowie Schmitz et al. 2022.

<sup>16</sup> Beispielsweise das KI-Managementsystemstandard ISO 42001, analysiert in Mock 2021.

<sup>17</sup> Siehe etwa Poretschkin et al. 2022.

**THESE 3:** Eine KI-Zertifizierung soll drei Ebenen umfassen: die im Produkt eingebauten KI-Komponenten, das KI-Gesamtsystem als Produkt sowie die Umwelt, mit der es interagiert.

Bei der Zertifizierung von KI-Systemen steht meist die Kontrolle von Risiken für Nutzer\*innen, Verbraucher\*innen und Betroffene im Fokus.<sup>18</sup> KI-getriebene Produkte beinhalten eine oder mehrere KI-Komponenten, unterliegen einer oftmals komplexen Wertschöpfungskette und interagieren als Gesamtsystem mit ihrer Umwelt. Risiken können sowohl durch Wechselwirkungen zwischen KI-Komponenten und dem Gesamtsystem als auch durch solche zwischen dem Gesamtsystem und der Umwelt entstehen. Diese sind bei der Entwicklung des Systems nicht immer vorhersehbar. Eine Zertifizierung auf allen drei Ebenen kann helfen, verschiedene Risiken zu erfassen.

Produktzertifizierungen betreffen üblicherweise ein Gesamtsystem, das für einen definierten Einsatzzweck entwickelt wurde – beispielsweise ein Fahrzeug mit diversen Fahrerassistenzsystemen. Gleichzeitig können von den unterschiedlichen Komponenten des Gesamtsystems unterschiedliche Risiken ausgehen, z. B. mechanische Ausfallrisiken bei einem Fahrzeug. Welche Risiken von einer KI-Komponente im Vergleich zu anderen Komponenten ausgehen, hängt vom konkreten Produkt ab. Insbesondere der Betrieb weiterlernender KI-Komponenten kann zusätzliche Unsicherheiten beinhalten, die das Gesamtrisiko des Systems beeinflussen. Bei einer Zertifizierung sollten diese verschiedenen Risiken auf den drei Ebenen differenziert bewertet werden.

Zertifikate zielen darauf ab, ein System in einem bestimmten Anwendungsszenario zu bewerten. Das lenkt den Blick auf das Zusammenspiel von KI-System und Umwelt, in die es eingebettet ist. Der breite Begriff Umwelt umfasst allgemeine sowie nicht technische Charakteristika des Einsatzortes, beispielsweise die sozialen Normen und Strukturen, die eine Nutzungssituation definieren. Bei autonomen Fahrzeugen wären dies beispielsweise die bestehenden Regeln im Straßenverkehr, das Verhalten anderer Verkehrsteilnehmer\*innen oder spezifische Eigenschaften der physischen Umgebung. Daraus ergeben sich sektorspezifische Anforderungen an die Zertifizierung.

Die Durchführung einer differenzierten Risikoanalyse und einer darauf aufbauenden Zertifizierung stellen daher eine komplexe Aufgabe dar, die eine Beteiligung von Expert\*innen verschiedener Bereiche (Rechtswissenschaft, Ethik, Zivilgesellschaft, Domänenexpert\*innen etc.) erfordert. Solche Multi-Stakeholder-Ansätze können Entwickler\*innen dabei unterstützen, die komplexen Auswirkungen sozio-technischer KI-basierter Systeme, etwa auch auf grundrechtliche Aspekte, im Entwicklungsprozess angemessen zu berücksichtigen.

Für Verbraucher\*innen ist besonders relevant, dass reale Risiken beim KI-Einsatz zuverlässig erkannt und effektiv entschärft werden. Das umfasst sowohl eine differenzierte Betrachtung der verschiedenen Komponenten als auch eine Prüfung, die Umwelteigenschaften ausreichend berücksichtigt.

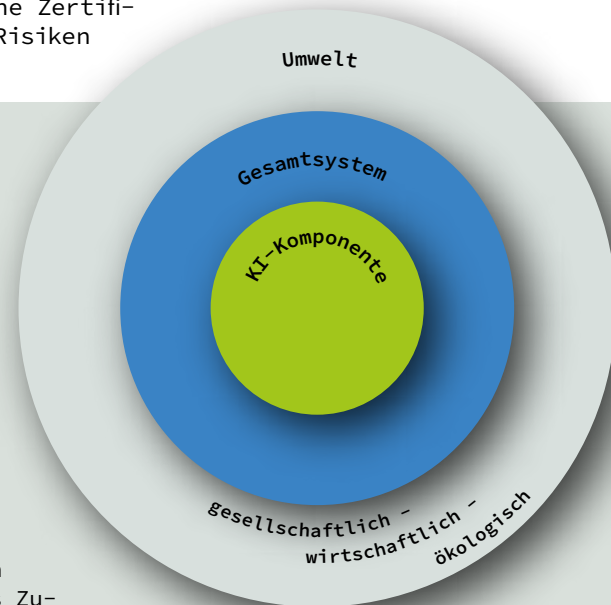


Abb. 1:  
Ebenen der  
Zertifizierung  
von KI-Systemen

<sup>18</sup> Mangelsdorf et al. 2022.

**THESE 4:** Effektive Zertifizierungen erfordern klar definierte Rollen und Kompetenzen der verantwortlichen Expert\*innen. Um die Interessen der Verbraucher\*innen und Betroffenen zu berücksichtigen, sollen Vertreter\*innen verschiedener gesellschaftlicher Gruppen an der Entwicklung entsprechender Verfahren beteiligt werden.

Aussagekräftige Zertifizierungen brauchen neben verbindlichen Testverfahren auch präzise Anforderungen an die prüfenden Expert\*innen. Dabei stellt sich die Frage, welches Kompetenzprofil Fachexpert\*innen benötigen, die an einer Prüfung beteiligt sind. Neben geeigneten Qualifikationen ist es aus Verbraucher\*innensicht erforderlich, dass Prüfer\*innen für die Interessen von Betroffenen und Verbraucher\*innen sensibilisiert sind und ihre Interessen im Rahmen der Prüfung effektiv wahren können. Die Einbindung gesellschaftlicher Gruppen wie Verbraucher\*innenschutzorganisationen in die Entwicklung von Prüfverfahren kann Prüfer\*innen bei Bedarf unterstützen.

Für die meisten Verbraucher\*innen ist die Expertise von zertifizierenden Unternehmen oft sichtbarer als die der individuellen Prüfer\*innen. Akkreditierungen für die Prüforganisationen stellen daher auch bei freiwilligen Zertifizierungen einen Mehrwert dar. Sie stärken das Vertrauen der Verbraucher\*innen in die Kompetenz und Zuverlässigkeit der Prüfer\*innen und damit in die Zertifizierung an sich.

Eine Beteiligung zivilgesellschaftlicher Perspektiven und Akteur\*innen an Prüfverfahren kann Prüfer\*innen unterstützen. Zum einen können zivilgesellschaftliche Akteur\*innen an der Entwicklung von Zertifizierungen beteiligt werden. Hierfür existieren bereits Strukturen, auf die zurückgegriffen werden kann, z. B. die Beteiligung von zivilgesellschaftlichen Organisationen an der Erstellung von Normen, die Basis einer Zertifizierung sein können. Zum anderen können Verbraucher\*innen und Betroffene punktuell an konkreten Zertifizierungen beteiligt werden. So können Verbraucher\*innenbedarfe besser abgebildet werden. Eine solche Beteiligung ermöglicht insbesondere bei Werte- und Zielkonflikten neue Einblicke.

Belange von Verbraucher\*innen und Betroffenen drohen angesichts der hohen Komplexität von KI-Prüfungen vernachlässigt zu werden. Um sie angemessen zu berücksichtigen, sollten die Interessen der Verbraucher\*innen bei der Zertifizierung von KI-Systemen abgebildet werden – durch entsprechende Expertise und Beteiligungsstrukturen.

**THESE 5:** Die Risiken unterschiedlicher Einsatzkontexte erfordern eine separate Zertifizierung. Gleichzeitig braucht es Ansätze, um Zertifizierungen durch einen schonenden Einsatz von Ressourcen attraktiv zu gestalten.

Die Anforderungen an eine Zertifizierung für unterschiedliche Einsatzzwecke festzulegen, ist eine komplexe, aber sinnvolle Aufgabe. Denn unterschiedliche Einsatzgebiete bedingen unterschiedliche (ggf. bereits existierende) technische Standards und rechtliche Anforderungen. Gleichzeitig sind Zertifizierungen mit einem hohen Ressourceneinsatz verbunden. Könnten mehrere Einsatzszenarien im Rahmen eines einzigen Zertifizierungsprozesses bearbeitet werden, hätte das Vorteile für die Anbieter\*innen der Systeme. Eine Voraussetzung dafür wäre, dass die betrachteten Einsatzkontexte hinreichend ähnlich sind und mit vergleichbaren Risiken einhergehen.

Je nach Kontext kann eine Zertifizierung von Teilsystemen nützlich sein. Da KI-Komponenten oft in Lieferketten entwickelt und in Produkten von Dritten eingebaut werden, wäre es für die Hersteller\*innen der finalen Produkte von Vorteil, wenn generelle Aspekte wie die Robustheit oder die Sicherheit von KI-Komponenten zertifiziert werden könnten. Beachtet werden sollte hierbei, dass solche generellen Aspekte kaum losgelöst vom Einsatzszenario bewertet werden können. Zudem ist das Einsatzgebiet für viele KI-Modelle entlang solcher Lieferketten nicht a priori bestimmbar, da es sich erst aus der finalen Produktgestaltung seitens der Anbieter\*innen ergibt. Große Sprachmodelle haben zum Beispiel kein bestimmtes Einsatzgebiet, sondern können für eine Vielzahl an Zwecken genutzt werden. Solche KI-Komponenten können ohne die Betrachtung des Einsatzkontexts nicht ausreichend aussagekräftig zertifiziert werden.<sup>19</sup>

Eine Zertifizierung, die kontextbezogene Risiken umfassend abdeckt, kann die Vertrauenswürdigkeit von KI-Systemen fördern. Gleichzeitig können auch Verbraucher\*innen ein Interesse daran haben, Zertifizierungsschritte ressourcenschonend zu gestalten, wenn trotzdem aussagekräftige KI-Zertifikate zustande kommen.

<sup>19</sup> Zu Zertifizierung von großen Sprachmodellen und ihren Grenzen vgl. den Ansatz von Mökander et al. 2023.

**THESE 6:** Die Vertrauenswürdigkeit von KI-Systemen sollte angesichts dynamischer Veränderungen bei weiterlernenden Systemen gewahrt werden. Zeitliche Dynamiken sind daher bei der Zertifizierung zu berücksichtigen. Zudem werden klare Kriterien für eine Re-Zertifizierung benötigt.

Bei allen Formen der Prüfung und Zertifizierung von KI-Systemen stellt sich die Frage, wie die Dynamik des Feldes angemessen erfasst werden kann: Schließlich entwickeln sich KI-Anwendungen teilweise rasant weiter und KI-Systeme können sich im Laufe ihres Einsatzes kontinuierlich verändern. Gleichzeitig kann sich das Nutzungsverhalten ändern. Bei sich weiterentwickelnden Systemen braucht es demnach ein kontinuierliches Monitoring sowie eindeutige Kriterien, wann eine Re-Zertifizierung nötig ist. Dabei sollten sowohl Änderungen im System durch Betreiber\*innen als auch mögliche Veränderungen des Einsatzkontextes berücksichtigt werden.

Der adaptive Charakter weiterlernender Systeme erfordert zukünftig ein kontinuierliches unternehmensinternes Monitoring, das z. B. erkennt, ob und unter welchen Bedingungen ein KI-System Auswirkungen hat, die außerhalb von als sicher eingestuften Grenzen liegen. Das Monitoring sollte regelmäßigen Audits unterliegen. Dokumentations- und Aufzeichnungspflichten sind Voraussetzung dafür, dass solche Audits nachvollziehbar durchgeführt werden. Gleichzeitig ist anzumerken, dass Qualitäts- und Risiko-Management-Systeme bereits jetzt kontinuierlich ausgelegt werden und den gesamten Lebenszyklus eines KI-Systems abdecken. Je nach Einsatzkontext des KI-Systems kann auch ein externes Monitoring angemessen sein.

Re-Zertifizierungen können zusätzlich zum internen Monitoring die Veränderungsdynamik eines KI-Systems erfassen. Eine Re-Zertifizierung ist dann erforderlich, wenn Veränderungen zu neuen Risiken führen.

Die Betrachtung von Veränderungen über verschiedene Ebenen hinweg kann dabei helfen, den Zeitpunkt einer sinnvollen Re-Zertifizierung zu bestimmen. Der Zyklus von Re-Zertifizierungen sollte daher mögliche Veränderungen der KI-Komponente angemessen berücksichtigen. Auch Veränderungen im Einsatzkontext selbst sollten betrachtet werden. Zum Beispiel kann die Einführung neuer Fahrzeugmodelle dazu führen, dass diese neuen Modelle von Fahrassistenzsystemen, die Kollisionen vermeiden sollen, nicht als Fahrzeuge erkannt werden.

Einschränkend lässt sich an dieser Stelle anmerken, dass die meisten Systeme zurzeit nicht kontinuierlich lernend sind, sondern strukturiert angepasst werden. Die Bewertung und die Beobachtung von weiterlernenden Systemen beinhalten technisch herausfordernde Fragen, die noch nicht gelöst sind. An diesem Punkt besteht ein großer Bedarf an Forschung und Entwicklung, um eine dynamische Regulierung zu ermöglichen, die in Zukunft immer wichtiger werden wird.

Für Verbraucher\*innen ist eine kontinuierliche Kontrolle insbesondere von weiterlernenden KI-Systemen wichtig, um möglichst schnell und angemessen neue Risiken identifizieren und ihnen begegnen zu können.

**THESE 7:** KI-Zertifikate sollen für Verbraucher\*innen und Betroffene verständlich und leicht zugänglich kommuniziert werden. Abgestufte Vermittlungsformate sollen dabei die unterschiedlichen Informationsbedarfe und Wissensstände berücksichtigen.

Damit sie Vertrauen stiften, müssen die Ergebnisse gut umgesetzter Zertifizierungen breit und angemessen kommuniziert werden. Das gilt für sämtliche Zielgruppen: Zertifikate müssen daher sowohl leicht zugänglich als auch bedarfsgerecht und verständlich gestaltet werden.

Verbraucher\*innen können verschiedene Wissensstände und Informationsbedürfnisse aufweisen. So haben beispielsweise Eltern hohe Ansprüche an den Umfang und die Vollständigkeit von Informationen zu Produkten, die ihre Kinder nutzen. Im Vergleich dazu sind die Informationsansprüche von Kund\*innen in einem Online-Shopping-Portal vergleichsweise niedriger, denn die möglichen negativen Auswirkungen bleiben überschaubar. Betroffenheit und Relevanz der KI-Ergebnisse prägen demnach die Informationsbedürfnisse. Aber auch vorhandene Kompetenzen und Ressourcen unterscheiden sich: Technikaffine Verbraucher\*innen oder Vertreter\*innen von Verbraucher\*innenorganisationen brauchen eher Zugang zu detaillierteren Informationen als andere.

Zertifikate müssen so gestaltet werden, dass sie sowohl für Verbraucher\*innen verständlich als auch für Expert\*innen im Detail nachvollziehbar sind. Eine mehrstufige Vermittlung mit unterschiedlichen Abstraktionsebenen kann geeignete Formate für Lai\*innen sowie für informierte Fachleute umfassen. Hierbei können auch Multiplikator\*innen einbezogen werden, die als Vermittler\*innen wenig informierte Verbraucher\*innen erreichen.

Um die unterschiedlichen Bedarfe an Informationen zu erfassen, ist eine Untersuchung nötig, die sich an den Anwendungsfällen orientiert. So können verschiedene Gruppen systematisiert und detailliert definiert werden, ähnlich wie bei Taxonomiestandards (vgl. z. B. die zurzeit entstehende ISO-Norm ISO/IEC CD 12792 Information technology – Artificial intelligence – Transparency taxonomy of AI systems).<sup>20</sup> Eine genaue Aufschlüsselung möglicher Verbraucher\*innengruppen bedarf weitere Untersuchungen zu den Wahrnehmungen und Bedürfnissen der Verbraucher\*innen.

Bei der Vermittlung von Zertifikaten sowie grundsätzlich von Informationen zu den eingesetzten KI-Systemen geraten manche Gruppen von Verbraucher\*innen in Vergessenheit. Betroffene von KI-Entscheidungen interagieren nicht immer direkt mit dem System und sind oft schwer zu identifizieren. Verbraucher\*innen als Datensubjekte sind auch nicht immer sichtbar, wenn es darum geht, relevante Informationen zu vermitteln.

Gerade weil sämtliche betroffene Gruppen für KI-Entwickler\*innen wenig greifbar sind, ist es notwendig, Verbraucher\*innen in ihren Rollen als Betroffene und Datensubjekte zu erreichen. Hier braucht es breit angelegte und differenzierte Vermittlungsansätze, bei denen jede\*r die Möglichkeit haben sollte, auf Inhalte des Zertifikats zuzugreifen und diese zu verstehen.

<sup>20</sup> <https://www.iso.org/standard/84111.html>.

**THESE 8:** Verbraucher\*innen, zivilgesellschaftliche Organisationen und Forscher\*innen sollen nachvollziehbare Angaben dazu erhalten, welche Organisationen an einer Zertifizierung beteiligt sind und welche Prüfkriterien für ein Zertifikat gelten.

Labels für Nachhaltigkeit, Fairness oder artgerechte Tierhaltung zeigen als Beispiele, wie schnell es unübersichtlich werden kann. Um KI-Zertifikate als Orientierungswissen bei Entscheidungen berücksichtigen zu können, brauchen Verbraucher\*innen und ihre Vertreter\*innen weitere Angaben zu Prüfverfahren, Bewertungskriterien sowie prüfenden Organisationen. Solche Informationen helfen dabei, Zertifikate vergleichbar zu machen, und geben Auskunft über die Vertrauenswürdigkeit des einzelnen KI-Zertifikats.

Um den Verbraucher\*innen Unterschiede zwischen verschiedenen Zertifikaten und Labels aufzuzeigen, ist es insbesondere nötig, die Testverfahren und Prüfkriterien transparent zu machen. Dadurch können grundlegende Unterschiede verschiedener Zertifikate nachvollziehbar werden. Adressat\*innen wären sowohl Verbraucher\*innen als auch zivilgesellschaftliche Organisationen oder Forscher\*innen, die solche Labels vergleichen könnten. Gleichzeitig ist es für viele Verbraucher\*innen kaum möglich, Kriterien und Verfahren einer KI-Prüfung nachzuvollziehen. Hierzu sind abgestufte Kommunikationsformate erforderlich.

Die Aussagekraft von freiwilligen Zertifikaten und Labels hängt in hohem Maße von den Zertifizierungsstellen ab, die sie ausstellen. Handelt es sich dabei um Organisationen mit passender Akkreditierung wie etwa durch die *Deutsche Akkreditierungsstelle*, können Verbraucher\*innen davon ausgehen, dass die Prüfungen gängigen Standards und dem aktuellen technischen Stand entsprechen. Es wäre möglich, mittels eines gemeinsamen, potenziell geschützten Vergabeprogramms für verschiedene akkreditierte Zertifizierungsorganisationen die Klarheit und Aussagekraft eines KI-Zertifikats zu steigern.

Aussagekräftige und vertrauenswürdige Zertifikate brauchen geeignete Vermittlungsformen. Zentrale Bestandteile einer guten Kommunikation sind Angaben zum Prüfverfahren, zu Prüfkriterien und zu den prüfenden Organisationen.

## Zusammenfassung: Welche Anforderungen müssen freiwillige KI- Zertifikate erfüllen, um aussagekräftig zu sein?

KI-Zertifikate haben großes Potenzial, zu mehr Wissen, Orientierung und einer größeren Handlungssouveränität bei Verbraucher\*innen beizutragen. Sie können Verbraucher\*innen bei ihren Entscheidungen unterstützen und einen Vergleich zwischen verschiedenen Diensten und Produkten ermöglichen. Sie vermitteln, für welche Einsatzzwecke und unter welchen Bedingungen die Systeme zertifiziert sind. Darüber hinaus tragen sie zur Haftungssicherheit für Unternehmen bei.

Neben gesetzlich vorgeschriebenen Zertifizierungen und CE-Kennzeichen für selbstgeprüfte Systeme können auch freiwillige Zertifikate einen Mehrwert für Verbraucher\*innen schaffen. Sie können es ihnen erleichtern, die Vertrauenswürdigkeit eines KI-Systems zu beurteilen. Freiwillige Zertifikate können auch andere Anforderungen abbilden und Schwerpunkte setzen, jenseits der gesetzlichen Anforderungen an Hochrisiko-KI-Systeme in der geplanten KI-Verordnung. Ob sich dieses Potenzial aber real entfalten kann, hängt von mehreren Voraussetzungen ab.

An erster Stelle müssen freiwillige Zertifikate auf klaren Regeln sowie Standards und Normen aufbauen. Diese sollten so entwickelt werden, dass sie relevante Anforderungen an die Vertrauenswürdigkeit von KI-Systemen abbilden. Eine Orientierung bietet der neu vom *Europäischen Parlament* vorgeschlagene Artikel 4a der KI-Verordnung, der allgemeine Prinzipien für den ethischen und vertrauenswürdigen Einsatz sämtlicher KI-Systeme beschreibt. Für die Umsetzung von KI-Zertifizierungen sind verschiedene Prüfgegenstände von Bedeutung – in Bezug auf die Zertifizierung von Produkten, aber auch von Qualitäts- und Risikomanagementprozessen. Solche Prüfschritte sollten sinnvoll aufeinander abgestimmt werden, um die relevanten Anforderungen an KI-Systeme zu erfassen.

Risiken bei KI-Produkten können aus den Interaktionen zwischen Gesamtsystem und Teilkomponenten sowie zwischen Gesamtsystem und dessen Umgebung mit ihren technischen sowie sozialen Eigenschaften entstehen. Um alle relevanten Risiken abzubilden, sollten daher an zweiter Stelle alle drei Ebenen bei der Zertifizierung

von KI-Systemen berücksichtigt werden. Zudem können sich sowohl das System selbst als auch die Umgebung über die Zeit verändern und somit die Leistungsfähigkeit des KI-Produkts oder der KI-Dienstleistung beeinflussen. Neue Risiken und Auswirkungen wären denkbar. Die Analyse solcher Veränderungen ist komplexer, wenn es sich um weiterlernende Systeme handelt. Eine weitere Anforderung ist es daher, solche zeitlichen Dynamiken sinnvoll abzubilden.

Beteiligungsverfahren können dazu beitragen, die Interessen der Verbraucher\*innen bei der Entwicklung und Umsetzung von Zertifikaten zu berücksichtigen, und auch für Unternehmen einen Mehrwert darstellen. Aus Sicht der Verbraucher\*innen ist es zudem notwendig, dass aussagekräftige Zertifikate durch unabhängige Dritte erstellt werden. Akkreditierte Prüfer\*innen mit geeigneter Expertise sowie vertrauenswürdige Zertifizierungsorganisationen sind eine zentrale Voraussetzung dafür, auch in Bezug auf Befürchtungen einer unübersichtlichen zukünftigen KI-Zertifizierungslandschaft.

Aussagekräftige und vertrauenswürdige Zertifikate benötigen schließlich geeignete Vermittlungsformen. Neben Anforderungen an Verständlichkeit und Zugang ist es hierbei wichtig, verschiedene Gruppen von Verbraucher\*innen mit unterschiedlichen Wissensständen und Motivationen zu berücksichtigen. Die ausgewählten Vermittlungsformate sollten imstande sein, sowohl Verbraucher\*innen jeweils angemessen zu informieren als auch die Transparenz der Prüfverfahren zu steigern. Wesentliche Informationen sind dabei Angaben zu Prüfverfahren, Prüfkriterien und umsetzenden Organisationen.

Zertifizierungsprogramme, die diese Anforderungen umsetzen, können auch jenseits gesetzlicher Verpflichtungen Vertrauen bei den Verbraucher\*innen stiften und somit wertvoll für die Hersteller\*innen und Anbieter\*innen von KI-Systemen sein.

## Quellen

Alle Quellen zuletzt am 25.08.2023 abgerufen.

- Baeva, Gergana: Wer sind die Verbraucher\*innen im Kontext von KI-Systemen? 2022. Online unter: [https://www.zvki.de/storage/publications/Verbraucher\\_innen\\_ZVKI\\_GB.pdf](https://www.zvki.de/storage/publications/Verbraucher_innen_ZVKI_GB.pdf).
- Busse, Franziska/ Baeva, Gergana: Was sind die richtigen Zutaten für vertrauenswürdige Künstliche Intelligenz? Ergebnisse der ZVKI-Online-Befragung: Wissen, Nachvollziehbarkeit und bewertbare Erfahrungen – Zutaten für vertrauenswürdige Künstliche Intelligenz (KI). Hg. v. Zentrum für vertrauenswürdige Künstliche Intelligenz (ZVKI). 2022. Online unter: <https://www.zvki.de/zvki-exklusiv/fachinformationen/online-befragung>.
- DIN e. V.: DIN EN ISO/IEC 17000:2020 „Konformitätsbewertung – Begriffe und allgemeine Grundlagen“. 2020. Online unter: <https://www.beuth.de/de/norm/din-en-iso-iec-17000/319777862>.
- Europäische Kommission: Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, COM(2021) 206 final. 2021. Online unter: <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX%3A52021PC0206>.
- Europäisches Parlament: Report on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence. 2023. Online unter: [https://88fe0205-a2ac-4895-ba2e-a260b-7a7b33d.usrfiles.com/ugd/88fe02\\_d1ead08d-b89d4b259552ec545415b1d4.pdf](https://88fe0205-a2ac-4895-ba2e-a260b-7a7b33d.usrfiles.com/ugd/88fe02_d1ead08d-b89d4b259552ec545415b1d4.pdf).
- Glatzner, Ludwig: Theorie und Funktion der Zertifizierung am Beispiel ISO 14001 „Umweltmanagementsysteme“. In: Friedel, Rainer/ Spindler, Edmund (Hrsg.): Zertifizierung als Erfolgsfaktor, S. 289-304. 2016. Online unter: [https://doi.org/10.1007/978-3-658-09701-1\\_21](https://doi.org/10.1007/978-3-658-09701-1_21).
- Gossen, Maike et al.: Nudging Sustainable Consumption: A Large-Scale Data Analysis of Sustainability Labels for Fashion in German Online Retail. In: Frontiers in Sustainability, Vol. 3. 2022. Online unter: <https://doi.org/10.3389/frsus.2022.922984>.
- Heese, Jessica et al.: Zertifizierung von KI-Systemen. Kompass für die Entwicklung und Anwendung vertrauenswürdiger KI-Systeme. 2020. Online unter: [https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG1\\_3\\_Whitepaper\\_Zertifizierung\\_KI\\_Systemen.pdf](https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG1_3_Whitepaper_Zertifizierung_KI_Systemen.pdf).
- ISO: ISO/IEC CD 12792 Information technology – Artificial intelligence – Transparency taxonomy of AI systems. 2023 (z. Z. in Entwicklung). Online unter: <https://www.iso.org/standard/84111.html>.
- Kneip, Veronika: Gütesiegel. In: Aßländer, Michael S. (Hrsg.): Handbuch Wirtschaftsethik, S. 731-734. 2022. Online unter: [https://doi.org/10.1007/978-3-476-05806-5\\_78](https://doi.org/10.1007/978-3-476-05806-5_78).
- Mangelsdorf, Axel et al.: Nationale und internationale Standardisierung und Zertifizierung von Daten-diensten. In: Rohde, Marieke et al. (Hrsg.): Datenwirtschaft und Datentechnologie, S. 275-288. 2022. Online unter: [https://doi.org/10.1007/978-3-662-65232-9\\_19](https://doi.org/10.1007/978-3-662-65232-9_19).
- MeMo:KI: Dashboard des Meinungsmonitor Künstliche Intelligenz. 2022. Online unter <https://www.cais-research.de/forschung/memoki/>.
- Mock, Michale et al.: Management System Support for Trustworthy Artificial Intelligence. A comparative study. 2021. Online unter: [https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche\\_intelligenz/Fraunhofer\\_IAIS\\_Study\\_%20MSS.pdf](https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/Fraunhofer_IAIS_Study_%20MSS.pdf).
- Mökander, Jakob et al.: Auditing large language models: a three-layered approach. In: AI and Ethics. 2023. Online unter: <https://doi.org/10.1007/s43681-023-00289-2>.
- Overdiek, Markus/ Petersen, Thomas: Was Deutschland über Algorithmen und Künstliche Intelligenz weiß und denkt. Ergebnisse einer repräsentativen Bevölkerungsumfrage. Hg. v. Bertelsmann Stiftung. 2022. Online unter: <https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/was-deutschland-ueber-algorithmen-weiss-und-denkt>.
- Poretschkin, Maximilian et al.: Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz – KI-Prüfkatalog. Fraunhofer IAIS. 2021. Online unter: [https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche\\_intelligenz/ki-pruefkatalog/202107\\_KI-Pruefkatalog.pdf](https://www.iais.fraunhofer.de/content/dam/iais/fb/Kuenstliche_intelligenz/ki-pruefkatalog/202107_KI-Pruefkatalog.pdf).
- Schmid, Ute: Vertrauenswürdige Künstliche Intelligenz. In: BMUV/ Rostalski, Frauke (Hrsg.): Künstliche Intelligenz. Wie gelingt eine vertrauenswürdige Verwendung in Deutschland und Europa?, S. 287-298. 2022. Online unter: <https://www.mohrsiebeck.com/buch/kuenstliche-intelligenz-9783161612992>.
- Schmitz, Anna et al.: The why and how of trustworthy AI. An approach for systematic quality assurance when working with ML components. In: Automatisierungstechnik 70, Heft 9, S. 793-804. 2022. Online unter: <https://doi.org/10.1515/auto-2022-0012>.
- TÜV-Verband e. V.: Sicherheit und Künstliche Intelligenz. Erwartungen, Hoffnungen, Risiken. Eine repräsentative Befragung der Bevölkerung in Deutschland im Auftrag des TÜV-Verbands. 2021. Online unter: [https://www.tuev-verband.de/?tx\\_epxelo\\_file%5Bid%5D=856779&cHash=1af8a-3f0e6c845423fdd637c8dbcd080](https://www.tuev-verband.de/?tx_epxelo_file%5Bid%5D=856779&cHash=1af8a-3f0e6c845423fdd637c8dbcd080).

## Autorinnen und Autoren

Das vorliegende Whitepaper ist das Ergebnis mehrerer zwischen Januar und Mai 2023 durchgeführter Workshops und baut auf den Diskussionen der ZVKI-Fach-Arbeitsgruppe „KI Zertifizierung“, die sich zwischen Mai und Oktober 2022 zu Fragen der Prüfung und der Zertifizierung von KI-Systemen beriet. Geleitet wurden die ZVKI-Fach-AG und die anschließenden Workshops durch Gergana Baeva und Maximilan Poretschkin.

### An der Erstellung des Whitepaper beteiligt waren:

**Dr. Gergana Baeva**

Zentrum für vertrauenswürdige Künstliche Intelligenz (ZVKI)

**Robert Brunner**

B'IMPRESS – impressive communication

**Filiz Elmas**

Deutsches Institut für Normung e. V. (DIN)

**Benjamin Fresz**

Fraunhofer-Institut für Produktionstechnik und Automatisierung IPA

**Christine Fuß**

Deutsches Institut für Normung e. V. (DIN)

**Prof. Dr. Claudia Heß**

IU Internationale Hochschule

**Prof. Dr. Philipp Kellmeyer**

Universität Mannheim/ Universitätsklinikum Freiburg

**Malin Mahner**

Kommissariat der deutschen Bischöfe

**Prof. Dr. Heiko Paulheim**

Universität Mannheim

**Dr. Maximilan Poretschkin**

Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS

**Michael Puntschuh**

Beyond AI Collective

**Annegrit Seyerlein-Klug**

Technische Hochschule Brandenburg

## Impressum

**Autoren\*innen:** Dr. Gergana Baeva, Robert Brunner, Filiz Elmas, Benjamin Fresz, Christine Fuß, Dr. Claudia Heß, Dr. Philipp Kellmeyer, Malin Mahner, Dr. Heiko Paulheim, Dr. Maximilian Poretschkin, Michael Puntschuh, Annegrit Seyerlein-Klug

**Inhaltliche Mitarbeit:** Raphael Haddadi

**Redaktion:** Jaana Müller-Brehm

**Lektorat:** Hannah Willing (text | struktur)

**Gestaltung und Satz:** Christoph Löffler

**Verantwortlich im Sinne des Presserechts:** Philipp Otto

**Veröffentlichung:** Oktober 2023

**Lizenz:** Dieses Whitepaper steht unter der Lizenz Creative Commons CC-BY-SA-Lizenz 4.0 International.

### Das Zentrum für vertrauenswürdige Künstliche Intelligenz (ZVKI)

Das ZVKI macht als zentraler Ort der Debatte in Deutschland die Entwicklungen rund um gesellschaftliche Fragen zu Künstlicher Intelligenz und algorithmischen Systemen greifbar. Als nationale und unparteiische Schnittstelle zwischen Wissenschaft, Wirtschaft, Politik und Zivilgesellschaft informiert das ZVKI über viele verbraucher\*innenrelevante Aspekte, ermöglicht öffentliche Diskussionen und entwickelt Instrumente zur Bewertung und Zertifizierung von vertrauenswürdiger KI.

Weitere Informationen zum ZVKI: [www.zvki.de](http://www.zvki.de)

**Kontakt:** [zvki@irights-lab.de](mailto:zvki@irights-lab.de)

Das Projekt ZVKI wird vom unabhängigen Think Tank iRights.Lab verantwortet und durchgeführt. Das iRights.Lab entwickelt Strategien und praktische Lösungen, um die Veränderungen in der digitalen Welt vorteilhaft zu gestalten. Wir unterstützen öffentliche Einrichtungen, Stiftungen, Unternehmen, Wissenschaft und Politik dabei, die Herausforderungen der Digitalisierung zu meistern und die vielschichtigen Potenziale effektiv und positiv zu nutzen. Dazu verknüpft es rechtliche, technische, ökonomische und gesellschaftspolitische Expertise.

Weitere Informationen über das iRights.Lab finden Sie unter [www.irights-lab.de](http://www.irights-lab.de).

**Verbundpartner\*innen:** Fraunhofer AISEC, Fraunhofer IAIS, Freie Universität Berlin

**Gefördert durch:** Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz (BMUV)



Gefördert durch:



Bundesministerium  
für Umwelt, Naturschutz, nukleare Sicherheit  
und Verbraucherschutz

aufgrund eines Beschlusses  
des Deutschen Bundestages