Fraunhofer

Innovation im Blick

# KI-Zertifizierung und Absicherung im Kontext des EU AI Act

## Herausforderungen und Bedürfnisse aus Sicht von Unternehmen

Janika Kutz | Vincent Philipp Göbels | Danilo Brajovic | Benjamin Fresz | Niclas Renner | Safa Omri | Jens Neuhüttler | Marco Huber | Bernd Bienzeisler

Im Auftrag des

Ipai

# Inhalt

**Executive Summary** . . . . . **4**

**1. Einleitung** . . . . . **5**

**2. Stand der Zertifizierung und Absicherung** . . . . . **8**

**3. Anforderungen** . . . . . **14**

**4. Ausblick und Handlungsempfehlungen** . . . . . **16**

**5. Zusammenfassung** . . . . . **19**

**6. Quellenverzeichnis** . . . . . **20**

# Executive Summary

**Künstliche Intelligenz als integraler Bestandteil verschiedener Lebensbereiche**
Künstliche Intelligenz (KI) transformiert zahlreiche Industrien und Bereiche unserer Gesellschaft. Die Einführung und der allgegenwärtige Einsatz von KI-basierten Systemen in verschiedenen Lebensbereichen – privat als auch beruflich – eröffnen ein enormes Potenzial zur Verbesserung von Effizienz und Qualität in zahlreichen Anwendungsfällen. Allerdings ist ihr Einsatz auch mit erheblichen Risiken und Unklarheiten verbunden, darunter algorithmische Fehler, wie z. B. fehlerhafte Prognosen, Haftungsrisiken bei Fehlern, Diskriminierungen und Datenschutzverletzungen.

**Absicherung und Zertifizierung als Chance für Künstliche Intelligenz verstehen**
Um den erfolgreichen und verantwortungsvollen Einsatz dieser Technologien zu gewährleisten, sind einheitliche Regelungen, Normen und Zertifizierungsverfahren notwendig, welche die Sicherheit und Vertrauenswürdigkeit von KI-basierten Systemen gewährleisten. In diesem Zusammenhang stellt der vorgeschlagene EU AI Act einen wichtigen Schritt hin zu einem umfassenden rechtlichen Rahmen dar, der diese Herausforderungen adressiert. Auf nationaler Ebene in Deutschland liefert die KI-Normungsroadmap eine erste Grundlage für die Absicherung von KI-basierten Systemen und Produkten nach nationalen Standards.

**Klare Anforderungen an die Umsetzung von Absicherungs- und Zertifizierungsprozessen**
In acht Interviews mit Expertinnen und Experten für KI-Anwendungen aus verschiedenen Branchen und Anwendungsbereichen haben wir über die Anforderungen seitens der Unternehmen an Absicherungs- und Zertifizierungsprozesse gesprochen. Dabei ergab sich über die verschiedenen Interviewpartnerinnen und Interviewpartner hinweg eine Reihe von klaren Anforderungen an einen Zertifizierungsprozess, welche sich in vier Kategorien einteilen lassen:

- Anforderungen an die Umsetzbarkeit aus Sicht der Unternehmen
- Anforderungen zum Erhalt der Innovationsfähigkeit
- Anforderungen an einen transparenten Prozess
- Anforderungen an die klare Rolle von Behörden und Institutionen.

Dem gegenüber stehen Anforderungen, die einen Mehrwert der Zertifizierung von KI-basierten Systemen für Endnutzende gewährleisten.

**Diverse Angebote, um Unternehmen auf dem Weg zu sicheren und zertifizierten KI-Anwendungen zu begleiten**
Unternehmen brauchen verschiedene Unterstützungsangebote, um sich künftig auf Zertifizierungsprozesse vorbereiten zu können und um sichere KI-Systeme zu entwickeln bzw. einzusetzen. Basierend auf den Interviewergebnissen konnten vier Kategorien von Angeboten identifiziert werden, die Unternehmen auf dem Weg zu einer erfolgreichen KI-Zertifizierung unterstützen können:

- Informations- und Wissenstransfer
- Netzwerk und Austausch
- individuelle Beratungsangebote
- Methoden und Tools für die praktische Anwendung.

Das Whitepaper unterstützt Unternehmen bei der Navigation durch das komplexe Feld der KI-Zertifizierung und ebnet einen klaren Pfad zur erfolgreichen Zertifizierung. Besonderer Wert wird dabei auf ein ausgewogenes Verhältnis gelegt zwischen der Sicherstellung hoher ethischer und technischer Standards einerseits und der Förderung von Innovation und Wettbewerbsfähigkeit andererseits.

# 1. Einleitung

Anwendungen der künstlichen Intelligenz (KI) haben in den vergangenen Jahren sowohl im Privat- als auch im Berufsleben an Bedeutung gewonnen. Ihre Nutzung ist dabei nicht auf eine bestimmte Zielgruppe beschränkt: So lernen Kinder bereits in spielerischen Szenarien mit Apps begleitend zum Unterricht, interaktive Bildungsplattformen gestalten den Alltag von Studierenden, aber auch ältere Menschen können dank KI-basierter Hilfsmittel, wie dem automatischen Notruf oder dem intelligenten Rollator (Bundesministerium für Bildung und Forschung o. D.), ein selbstbestimmtes Leben führen.

Wenn KI-basierte Systeme nicht nach einheitlichen Sicherheitsstandards entwickelt, betrieben und geprüft werden, sind sie mit erheblichen Risiken verbunden. Fehlerhafte KI-Systeme können die Sicherheit von Produkten oder Dienstleistungen beeinträchtigen und damit zu (lebens-)gefährlichen Situationen führen. KI-Systeme können diskriminierende oder unfaire Entscheidungen treffen, wenn sie mit voreingenommenen oder unzureichenden Daten trainiert werden. Weiterhin können Datenschutzverletzungen auftreten, bei denen personenbezogene Daten ohne Zustimmung verwendet oder weitergegeben werden. Beispiele für solche Szenarien wurden auf der Website der AI Incident Database zusammengetragen. In den Abbildungen 2, 3 und 4 wird ein Überblick über die Verteilung öffentlich bekannter KI-Vorfälle mit eingesetzter Technologie und entstandener Schadensart dargestellt. Als KI-Vorfall wird in diesem Kontext ein öffentlich bekannt gewordener Schaden oder Beinahe-Schaden bezeichnet. Die Schäden können sowohl physischer, psychischer, finanzieller, ethischer oder gesellschaftlicher Art sein, sind jedoch nicht auf diese beschränkt. (vgl. AI Incident Database o. D.)

Abbildung 1 beschreibt potenzielle Risiken im Umgang mit KI nach der Europäische Kommission (2020) und der Hochrangige Expertengruppe für Künstliche Intelligenz (2019). Es wird zwischen den Dimensionen »Ethik«, »Konformität« und

**Risikoklassen von KI, die mit Absicherung und Zertifizierung adressiert werden können**

| Ethik | Konformität | Sicherheit |
|---|---|---|
| ■ KI-Systeme können unfaire Entscheidungen basierend auf Herkunft, Ethnizität oder Glauben treffen<br>■ KI-Systeme können Vorrang über menschliches Handeln nehmen und damit Schäden verursachen | ■ Im Lebenszyklus einer KI können personenbezogene Informationen ohne Zustimmung der Besitzer verarbeitet werden (Verletzung von Personenrechten nach der DSGVO)<br>■ Bei mangelnder Transparenz eines KI-Systems können Gesetzesbrüche verschleiert und nicht festgestellt werden | ■ Autonome Maschinen können Personen oder der Umwelt schaden.<br>■ Wenn unbeteiligte Dritte Zugriff auf KI-Systeme erlangen, können essenzielle Informationen entwendet werden (Passwordleaks) oder z. B. Sicherheitssysteme deaktiviert werden. |

*Abbildung 1: Risiken von KI in Anlehnung an Hochrangige Expertengruppe für Künstliche Intelligenz 2019.*

»Sicherheit« unterschieden. Ethische Risiken treten oft im Zusammenhang mit der Fairness der KI-Ergebnisse bzw. bei Benachteiligung von bestimmten Personengruppen auf. Risiken in der Konformität entstehen u. a. bei Gesetzesbrüchen, während Sicherheitsrisiken im Einsatz von KI zu Schäden führen können. Angesichts dieser Herausforderungen sind Regelungen und Normen zur Absicherung und Zertifizierung von KI von entscheidender Bedeutung.

Um diese Herausforderungen zu adressieren, stellte die Europäische Union (EU) 2021 den EU AI Act, im Deutschen KI-Verordnung, vor (Europäische Kommission 2021). Die EU arbeitet damit an einem der international ersten Gesetzesentwürfe zur Regulierung von KI und plant, entsprechende Anwendungen in drei Risikostufen einzuteilen und, basierend auf dieser Einstufung, bestimmten Anforderungen zu unterwerfen. Dadurch soll der Einsatz von KI sicherer und vertrauenswürdiger gestaltet und die Einhaltung der Grundrechte sichergestellt werden. Ergänzend schafft in Deutschland die KI-Normungsroadmap (DIN, DKE 2022) eine erste nationale Grundlage für die Absicherung von KI-basierten Systemen.

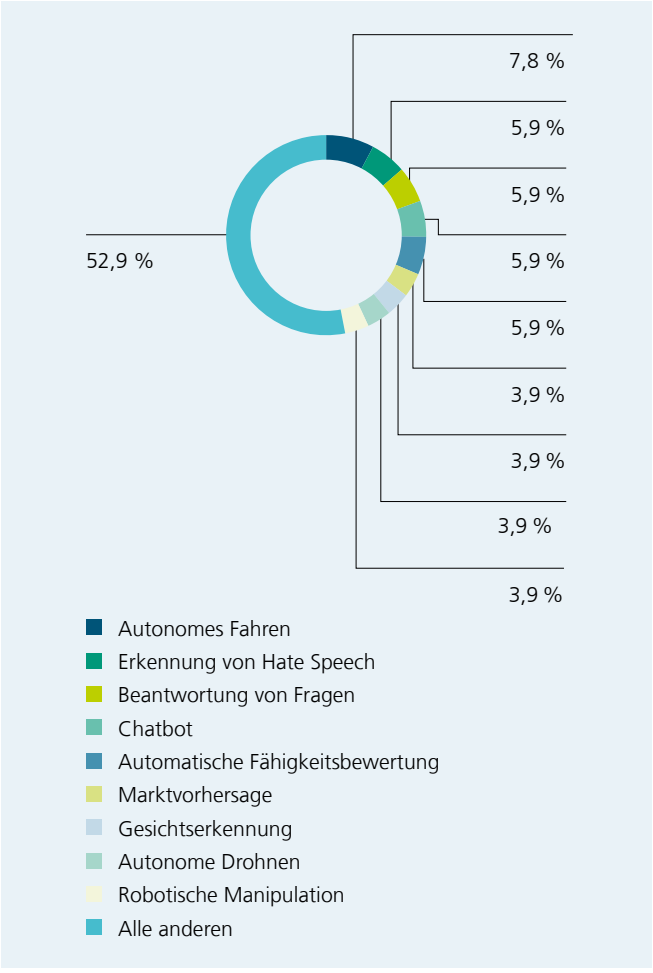


*Abbildung 2: Einsatzzweck der KI. (Quelle: AI Incident Database)*

Das Ziel all dieser Maßnahmen ist es, KI-basierte Systeme auf den Markt zu bringen, die nicht nur innovativ und effizient, sondern auch sicher und ethisch verantwortungsvoll sind. Ein solcher Ansatz ist von entscheidender Bedeutung, um das Vertrauen der Öffentlichkeit und der Stakeholder in diese Technologien zu stärken.

Abbildung 2 zeigt die relativen Häufigkeiten der Vorfälle nach Anwendungsgebiet von KI-Systemen. Es lässt sich kein dominantes Feld erkennen, in dem überdurchschnittlich viele Vorfälle entstanden. Dies unterstreicht, dass KI-Absicherung und -Zertifizierung in jeder Branche relevant ist.

Abbildung 3 veranschaulicht die Art der entstandenen Schäden. Bei mehr als 15 Prozent der Vorfälle entsteht physischer Schaden und mehr als 50 Prozent der Vorfälle zogen einen Schaden psychischer Art, an sozialen Systemen oder an Bürgerrechten nach sich. Dies zeigt einen großen Bedarf für die Sensibilisierung bezüglich derartiger Schadensquellen in der Entwicklung von KI.

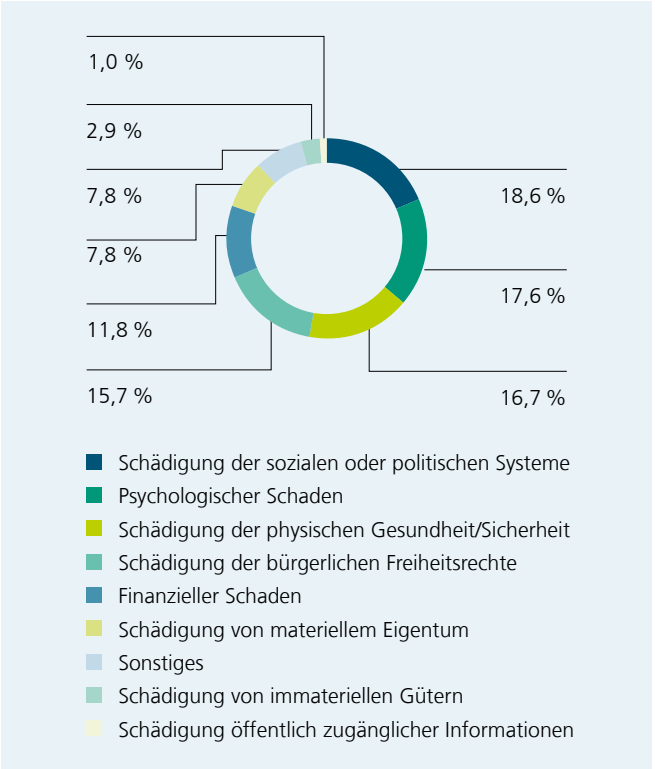


*Abbildung 3: Art der entstandenen Schäden. (Quelle: AI Incident Database)*

Abbildung 4 zeigt Grundlagen für entstandene Schäden. Fast 50 Prozent der Vorfälle sind auf Ethnizität und Geschlecht zurückzuführen. Noch zu viele Datensätze, die beim KI-Training verwendet werden, sind nicht divers und inklusiv genug und enthalten z. B. prädominant weiße Männer.

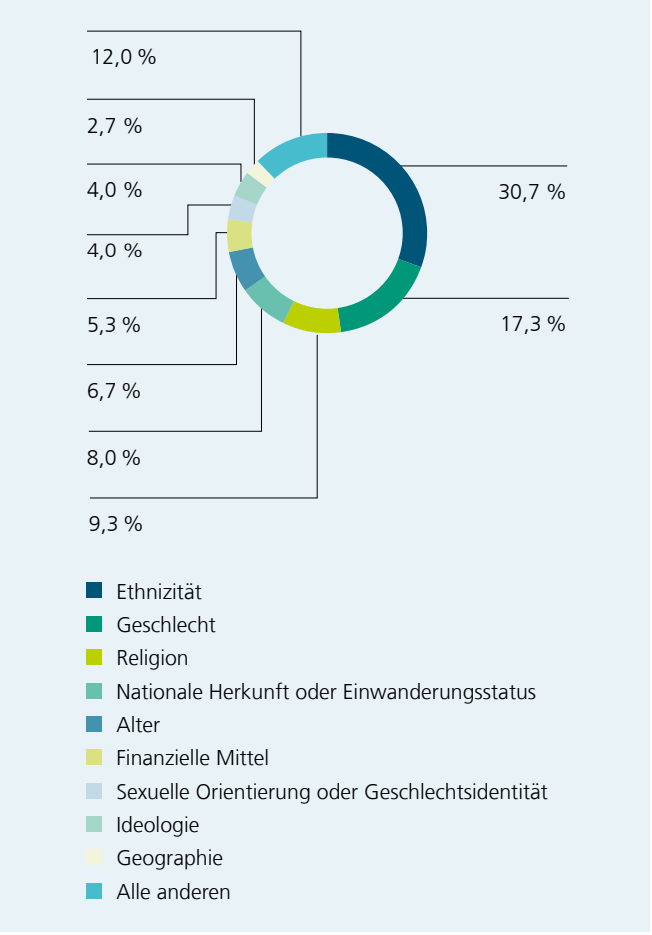


*Abbildung 4: Grundlage für den Schaden. (Quelle: AI Incident Database)*

Dieses Whitepaper liefert einen Einstieg in die Thematik. Es wird über den aktuellen Stand der gesetzlichen Regelungen und die damit verbundenen Herausforderungen informiert. Es soll für die Relevanz von Absicherungs- und Zertifizierungsprozessen beim Einsatz von KI-Systemen sensibilisiert werden. Gleichzeitig soll aufgezeigt werden, welche Angebote Unternehmen auf dem Weg zu sicheren und zertifizierten KI-Systemen unterstützen können. Ziel dieses Whitepapers ist es außerdem, einen Überblick zu geben, wo die Unternehmen bei der Absicherung und Zertifizierung von KI stehen, welche Herausforderungen in dem Zusammenhang gesehen werden und wo es Bedarf für Unterstützung gibt.

In Kapitel 2 wird der aktuelle Stand der KI-Absicherung und -Zertifizierung analysiert, um ein umfassendes Bild der gegenwärtigen Situation zu vermitteln. Anschließend werden in Kapitel 3, basierend auf Interviews mit Expertinnen und Experten, Anforderungen an KI-Systeme im Hinblick auf Sicherheit und Zertifizierung diskutiert. In Kapitel 4 wird ein Ausblick auf mögliche Entwicklungen und Handlungsempfehlungen gegeben, die dazu beitragen sollen, die Sicherheit von KI-basierten Systemen zu gewährleisten und Zertifizierungsprozesse zu unterstützen. Schließlich werden die zentralen Erkenntnisse und Ergebnisse dieses Whitepapers zusammengefasst.

# 2. Stand der Zertifizierung und Absicherung

Im folgenden Kapitel werden momentane Möglichkeiten zur Zertifizierung und Absicherung von KI vorgestellt, jeweils mit den notwendigen Begrifflichkeiten und deren Hintergründen. Dabei wird bei der Zertifizierung zwischen gesetzlicher und freiwilliger unterschieden, wovon im Moment – ohne bisher gültige Gesetzgebung – nur Letzteres möglich ist. Es ist bereits möglich, einige wichtige Aspekte von KI-Systemen zu beschreiben, die für die kommende Gesetzgebung relevant sind, insbesondere im Hinblick auf den EU AI Act. Auf diese wird am Ende des Kapitels eingegangen.

Bei der Auseinandersetzung mit dem Thema Künstliche Intelligenz ist zwischen den Begriffen Absicherung und Zertifizierung zu trennen. Im Gegensatz zu Absicherung, welche grundsätzlich alle Maßnahmen zur Minderung von Risiko beschreibt, bezeichnet Zertifizierung eine Bestätigung durch eine dritte Seite darüber, dass bestimmte Anforderungen erfüllt werden (Friederich Peter Zeuner o. D.). Dies wird als Konformitätsbewertung bezeichnet. Die zugrundeliegenden Anforderungen können dabei sowohl gesetzlich vorgeschrieben oder aber privatrechtlicher Natur und entsprechend freiwillig sein. Eine erste Übersicht findet sich in Abbildung 5, die die unterschiedlichen Zertifizierungsoptionen für KI beschreibt.

Während es bereits einige nicht-bindende Vorschläge und Arbeiten zur Absicherung und Zertifizierung von KI gibt, stellen fehlende bzw. unklare rechtliche Anforderungen für den Einsatz von KI momentan einen großen Unsicherheitsfaktor für Unternehmen dar. Um diese Lücke zu schließen und für alle Marktteilnehmenden klare Voraussetzungen zu schaffen, wird in der Europäischen Union der EU AI Act ausgearbeitet. Dieser wird zwar voraussichtlich erst in mehreren Jahren rechtskräftig, bietet aber bereits jetzt eine gute Orientierung zur Entwicklung von sicher arbeitenden und in der Europäischen Union rechtskonform einsetzbaren KI-Systemen.

### Absicherung vs. Zertifizierung

Absicherung beschreibt alle Maßnahmen zur Minderung von Risiken. Zertifizierung bezeichnet eine Bestätigung durch eine dritte Seite über das Einhalten bestimmter Vorgaben.

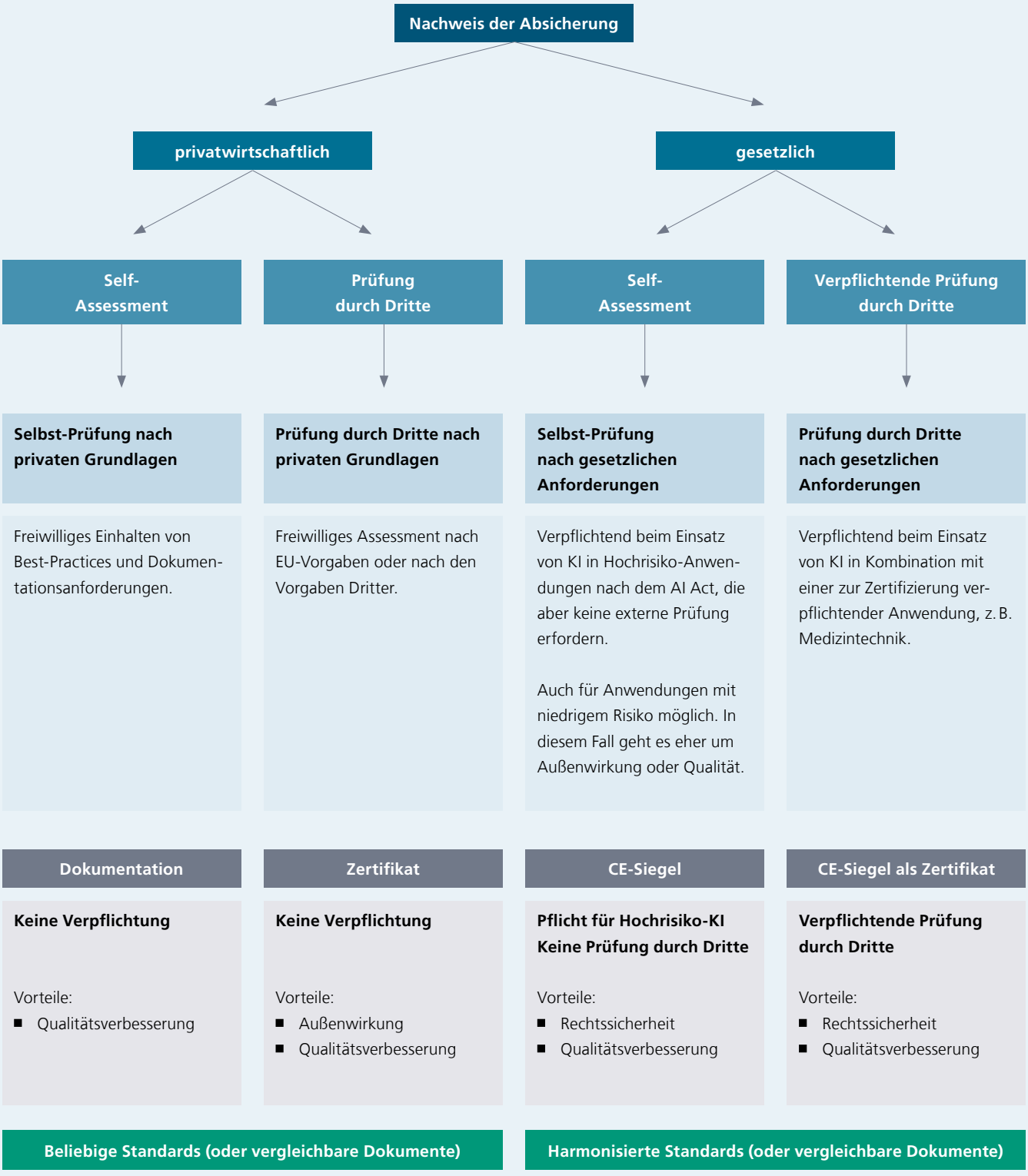


*Abbildung 5: Unterschiedliche Zertifizierungsoptionen für KI.*

**Gesetzliche Zertifizierung basierend auf Anforderungen des EU AI Act**

Da die gesetzlichen Anforderungen darüber entscheiden, ob und unter welchen Auflagen ein KI-System in der Europäischen Union eingesetzt werden darf, ist die gesetzliche Zertifizierung der Fokus dieses Kapitels. Im EU AI Act werden KI-Systeme in drei Risikoklassen eingeteilt und entsprechend dieser Einteilung mit den unten genannten Anforderungen unterlegt. Die Einteilung erfolgt dabei nach dem geschätzten Risiko des Anwendungsfalles.

Für eine Reihe von Anwendungsfällen wird der Einsatz von KI gänzlich verboten. Dazu zählen u. a. Verhaltensmanipulation von Nutzenden, wenn dadurch physische oder psychologische Schäden entstehen können, Social Scoring und biometrische Identifikation zu Strafverfolgungszwecken, wobei bei Letzterer mehrere Ausnahmen in Planung sind.

Die Bestimmungen für Hochrisikosysteme, d. h. solche, die in sicherheitskritischen Anwendungsfällen zum Einsatz kommen oder die sich wesentlich auf die Grundrechte einer oder mehrerer Personen auswirken können, bilden das Kernstück der Verordnung und könnten neben Anforderungen zu Risikomanagement, Daten-Governance, technischer Dokumentation, Transparenz, Aufzeichnungspflichten, menschlicher Aufsicht und Anforderungen an Genauigkeit eine Zertifizierung durch externe Prüfinstitute fordern. Nach aktuellem Stand wird eine solche Konformitätsbewertung durch Dritte nur dann notwendig, wenn das KI-System Teil eines Produktes ist, welches bereits auf Grund branchenspezifischer Anforderungen durch Dritte gesetzlich geprüft werden muss. Dies trifft unter anderem für Medizinprodukte zu. Alle Anwendungsfälle, die als Hochrisiko gelten, werden explizit im Anhang des EU AI Act beschrieben und umfassen unter anderem KI in kritischer Infrastruktur, bei der Bildung oder zu Strafverfolgungszwecken.

**Tabelle 1: Risikoklassen im EU AI Act**

| Risiko | Anforderungen in | Anwendungsfälle |
|---|---|---|
| Inakzeptables Risiko | Artikel 5 | (1) Verhaltensmanipulation, die zu physischen oder psychischen Schäden führen kann<br>(2) Ausnutzen von Schwächen einer Gruppe (z. B. Alter/Behinderung) zur Verhaltensmanipulation<br>(3) Social Scoring<br>(4) (Eingeschränkt) Biometrische Identifizierung |
| Foundation-Modelle | Artikel 28 | Für Foundation-Modelle, also allgemeinere KI-Modelle wie ChatGPT oder Midjourney, die als Grundlage für Anwendungen für spezifische Anwendungsfälle verwendet werden können, sollen separate Anforderungen gelten. Diese werden voraussichtlich in Artikel 28 aufgeführt und ähnlich zu den Anforderungen für Anwendungsfälle mit hohem Risiko sein. |
| Hohes Risiko | Artikel 9–15 | (1) Biometrische Identifizierung<br>(2) Betrieb kritischer Infrastruktur<br>(3) Allgemeine und berufliche Bildung<br>(4) Beschäftigung, Personalmanagement<br>(5) Zugang zu öffentlichen Diensten und Leistungen<br>(6) Strafverfolgung<br>(7) Migration, Asyl und Grenzkontrolle<br>(8) Rechtspflege und demokratische Prozesse |
| Begrenztes Risiko mit Transparenzpflicht | Artikel 52 und 69 | (1) Chatbots<br>(2) Biometrische Erkennung<br>(3) Erkennung von Emoticons<br>(4) Deep Fakes/generative KI |
| Minimales oder kein Risiko | | Alles Weitere |

Ähnliche Anforderungen wie für Hochrisiko-Systeme werden voraussichtlich für Foundation-Modelle gelten, also generelle KI-Modelle, die als Grundlage für die Entwicklung spezifischer Anwendungen genutzt werden können. Ein umfassender Überblick zu den Risikoklassen und den entsprechenden Anwendungsfällen ist in Tabelle Tabelle 1 zu finden.

Bei erfolgreicher Konformitätsbewertung eines KI-Systems ist das Ergebnis ein CE-Siegel, welches durch den Hersteller auf dem Produkt angebracht werden darf und welches signalisiert, dass das Produkt den europäischen Anforderungen entspricht. Eine Gegenüberstellung verschiedener Siegel ist in Abbildung 5 zu finden.

Für unkritische Systeme werden hingegen keine Anforderungen auferlegt, wobei es in einigen Fällen, wie z. B. Chatbots und Bildgeneratoren, die für Deep Fakes verwendet werden können, Transparenzpflichten gibt.

Auch Anbieter von KI-Systemen mit geringem Risiko sollen zu einer freiwilligen Umsetzung der im EU AI Act aufgeführten Anforderungen motiviert werden.

Die im EU AI Act beschriebenen Vorgaben sind jedoch zu ungenau, um daraus direkt ableiten zu können, was in der Entwicklung von KI-Systemen getan werden muss, um die Konformität mit dem EU AI Act sicherzustellen. So werden beispielsweise die Etablierung eines Risikomanagements, eine hohe Datenqualität sowie Transparenz gefordert, genaue Vorgaben zum Anforderungsniveau und zu dessen Erfüllung fehlen jedoch. Diese Vorgaben finden sich in Form von Normen und Standards, welche im Rahmen des Gesetzgebungsprozesses etabliert und für die Europäische Union harmonisiert werden. Die Europäische Union wird dazu einen Standardisierungsauftrag erteilen, in dessen Rahmen die entsprechenden Normen für die Harmonisierung erstellt werden. Ein großer Teil dieser Normen befindet sich derzeit noch in der Entstehung und wird voraussichtlich 2025 finalisiert werden (Hadrien Pouget 2023). Diese Standards werden u. a. auf Arbeiten der ISO aufbauen. Zu den wichtigsten zählen hierbei voraussichtlich die ISO/IEC 5259x Serie zu Datenqualität, die ISO/IEC 23894 zu Risiko-Management und die ISO/IEC 42001 zu KI-Management-Systemen. Eine gute Übersicht dazu findet sich auf der Seite der ISO/IEC JTC1/SC 42, dem bei ISO für KI zuständigen Gremium. Für Deutschland zählt die Normungsroadmap (DIN, DKE 2022) aktuell 116 teilweise branchenspezifische Normungs- und Standardisierungsbedarfe auf, die geklärt werden müssen. Brajovic et al. (2023) liefert einen Ansatz, wie die Anforderungen des EU AI Act in die Entwicklung der KI integriert werden können.

## Normungsprozess

Gesetzen wie dem EU AI Act mangelt es oft an zur Umsetzung notwendigen Details. Deshalb ergeht, sobald das Gesetz verabschiedet ist, ein Normungsauftrag an die zuständigen Normungs-Organisationen (CEN, CENELEC, ISO, …). Diese erarbeiten entsprechende Normen, die dann zur Einhaltung der gesetzlichen Vorgaben genutzt werden können. Werden diese von der Europäischen Union veröffentlicht (»harmonisiert«), müssen sie von den nationalen Normungsbehörden als nationale Normen umgesetzt werden.

**Freiwillige Zertifizierung**
Zum jetzigen Zeitpunkt ist der EU AI Act noch nicht verabschiedet, die Normen sind noch nicht final entwickelt und Prüfverfahren müssen noch ausgearbeitet werden. Damit ist eine gesetzliche Zertifizierung im Sinne einer Konformitätsprüfung für die Marktzulassung nicht möglich. Im Gegensatz zu Zertifizierungen gegenüber gesetzlich bestimmten Normen existieren bereits Label für KI-Anwendungen, meist mit Bezug zu sicherer oder vertrauenswürdiger KI. Diese Label unterscheiden sich von offiziellen Zertifizierungen in ihrem nicht standardisierten Prüfungsprozess und der Freiwilligkeit der Prüfung, können allerdings trotzdem als Zertifikate bezeichnet werden, wenn die Prüfung der zugehörigen Bedingungen von unabhängigen Dritten durchgeführt wird. Als einzige aktuelle Möglichkeit zur Zertifizierung können diese Label dazu beitragen, vertrauenswürdige und ethische KI-Systeme zu kennzeichnen, und signalisieren ein Bewusstsein für die Anforderungen an sichere KI-Systeme. Eine Gegenüberstellung zu gesetzlicher Zertifizierung ist in Abbildung 5 zu finden. Eine wichtige Arbeit in diesem Zusammenhang ist der KI-Prüfkatalog des Fraunhofer IAIS, welcher mit über 100 Seiten umfangreiche Ansätze zum Prüfen vom KI-Systemen liefert und die meisten Best-Practices abdeckt (Poretschkin et al. 2021). Einige, nicht abschließende Beispiele für bestehende freiwillige Zertifizierungen sind:

- CertAI der Münchener Rück bietet das »Trustworthy AI Seal« an, welches die Vertrauenswürdigkeit im Umgang mit KI bewertet (CertAI o. D.). Ein Teil der Bewertung basiert auf dem KI-Prüfkatalog des Fraunhofer IAIS.
- CertifAI von PricewaterhouseCoopers (PwC) und DEKRA soll ein Zertifizierungsanbieter für Künstliche Intelligenz werden, der im Dezember 2022 angekündigt wurde (Gregor Damm 2022).
- Der TÜV-IT bietet eine Reihe von Dienstleistungen im Zusammenhang mit KI an, welche u. a. mit der Ausstellung eines Zertifikats bestätigt werden können (TÜV IT o. D.).
- Das Dataset Nutrition Label ist spezifisch auf Datensätze ausgelegt und soll die Kernkomponenten eines Datensatzes strukturiert darstellen (Holland et al. 2018).

Vor allem in amerikanischen Unternehmen hat sich bezüglich des vertrauensvollen Umgangs mit KI die Dokumentation von Modellen und Datensätzen anhand sogenannter »Karten« etabliert. Die am weitesten verbreiteten Arbeiten dazu sind Model Cards (Mitchell et al. 2019), Data Cards (Pushkarna et al. 2022) und Data Sheets (Gebru et al. 2018). Diese fassen die Kerneigenschaften der KI-Modelle oder Datensätze gebündelt zusammen und sollen dabei helfen, potenzielle Fehlerquellen frühzeitig zu erkennen. Im Gegensatz zu den oben aufgeführten freiwilligen Labeln beinhalten diese Karten keine externe Prüfung und kein Zertifikat, sondern dokumentieren und unterstützen den Entwicklungsprozess. Somit können sie in einer späteren Zertifizierung unterstützen. Anwendung finden sie besonders bei IT-Unternehmen (Website zu Data Cards auf withgoogle.com; Mahima Pushkarna, Andrew Zaldivar 2022).

### Gesetzliche vs. freiwillige Zertifizierung

Gesetzliche und freiwillige Zertifizierungen sind Bestätigungen durch Dritte, dass Vorgaben eingehalten wurden. Dabei ist die gesetzliche Zertifizierung vorgeschrieben – im Fall des EU AI Act nur für Systeme, für die durch andere Gesetze bereits eine Zertifizierung notwendig ist. Die freiwillige Zertifizierung ist für alle Systeme möglich und die einzuhaltenden Vorgaben sind hier frei vom Anbieter der Zertifizierung wählbar.

**Aspekte der Absicherung**
Trotz unterschiedlicher Ursprünge von gesetzlicher und freiwilliger Zertifizierung stimmen die zu berücksichtigenden Risikodimensionen und abzusichernden Komponenten eines KI-Systems weitestgehend überein. Insbesondere die gewissenhafte Dokumentation des KI-Modells, des Entwicklungsprozesses und der ergriffenen Maßnahmen sind dabei essenziell.

Der EU AI Act stellt in den Artikeln 9–15 Anforderungen bezüglich der abzusichernden Komponenten. Diese umfassen:

- **Risikomanagement**, also Maßnahmen zur Minimierung und zum Umgang mit Gefahren
- **Daten**, wobei es insbesondere um die Qualität der Datenbasis geht
- eine umfassende **technische Dokumentation**, z. B. in Form einer detaillierten Anleitung
- **Aufzeichnungspflichten**, also das Loggen bzw. Protokollieren von Ereignissen
- **Transparenz** gegenüber Nutzenden, dass diese mit einer KI interagieren, sowie eine Interpretierbarkeit des KI-Modells an sich
- eine **menschliche Aufsicht** und damit verbunden die Möglichkeit zur menschlichen Interpretation der Ergebnisse und des Außerkraftsetzens des Systems
- Performanceanforderungen an **Genauigkeit, Robustheit und Cybersicherheit**

Außerdem wird in Artikel 61 (viii) die Beobachtung nach dem Inverkehrbringen durch die Anbieter gefordert. Darauf aufbauend gibt es allgemeine Risikodimensionen, welche für KI-Anwendungen insbesondere wichtig und bei den einzelnen oben genannten Komponenten zu berücksichtigen sind (Poretschkin et al. 2021). Diese sind:

- **Fairness**
  Schutz vulnerabler Gruppen: Alter, Geschlecht, Ethnische Gruppen, Behinderung
- **Autonomie und Kontrolle**
  Angemessener Autonomiegrad der KI-Anwendung
- **Transparenz**
  Erklärbarkeit, Nachvollziehbarkeit und Auditfähigkeit der KI-Anwendung
- **Verlässlichkeit**
  Robustheit und Konsistenz der Ausgaben
- **Sicherheit**
  Gewährleisten der funktionalen Sicherheit
- **Datenschutz**
  Schutz sensibler Daten im Kontext der KI-Anwendung

**Vorbereitung auf die Zertifizierung**
Obwohl weder EU AI Act noch Standardisierung zum jetzigen Zeitpunkt final sind, haben sich die oben genannten Risikodimensionen und abzusichernden Komponenten herauskristallisiert. Aus Unternehmenssicht sollte die Entwicklung von KI-Anwendungen daher bereits jetzt möglichst detailliert dokumentiert werden und dabei ein besonderer Fokus auf die im EU AI Act genannten Komponenten und die identifizierten Risikodimensionen gelegt werden. Obwohl es noch zu Änderungen in den konkreten Anforderungen kommen kann, stellt die frühzeitige Auseinandersetzung damit sicher, dass die nötige Expertise im Unternehmen zum Inkrafttreten des EU AI Act bereits existiert.

### Absicherung und Dokumentation

Aus Unternehmenssicht scheint es im Moment die sinnvollste Option, die Entwicklung von KI-Systemen umfangreich zu dokumentieren und dabei insbesondere auf die im EU AI Act beschriebenen Risikodimensionen achtzugeben.

Wie sehr die hier beschriebenen Lösungsstrategien bereits Einzug in deutsche Unternehmen gehalten haben, soll im folgenden Kapitel mit Hilfe von qualitativen Interviews herausgearbeitet werden.

# 3. Anforderungen

An die Gestaltung der Absicherung und Zertifizierung KI-basierter Systeme gibt es vielfältige Anforderungen. Zum einen gibt der Gesetzgeber im Rahmen des EU AI Act eine Reihe von Anforderungen vor. Andererseits gibt es Anforderungen der Nutzenden und der Anbieter von KI-basierten Systemen, die sich je nach Branche unterscheiden können. Um die unterschiedlichen Sichtweisen herauszuarbeiten, wurden acht Interviews mit Branchenexpertinnen und -experten aus Produktion, Bildung sowie beratenden Unternehmen und betroffenen (Unternehmens-)Verbänden durchgeführt. Dabei wurde eine KI-Zertifizierung im Allgemeinen, Bedarfe und potenzielle Herausforderungen diskutiert. Insbesondere wurde auch auf die oben erwähnten sechs Risikodimensionen (Fairness, Autonomie und Kontrolle, Transparenz, Verlässlichkeit, Sicherheit und Datenschutz) eingegangen und ihre Kritikalität in Bezug auf unterschiedliche Branchen diskutiert.

Dabei zeigte sich, dass die Relevanz der einzelnen Risikodimensionen in erster Linie vom Anwendungsfall und nicht von der Branche abhängt. Auch wenn es unterschiedliche Ausprägungen der Risikodimensionen in verschiedenen Branchen gibt, müssen diese für jeden Anwendungsfall neu geprüft werden und es lässt sich keine generelle branchen-basierte Betrachtungsweise ableiten. Die Interviewpartner zeigten eine unterschiedliche Sichtweise auf den Nutzen und die Risiken von KI-Zertifizierung im Sinne des EU AI Act – welche in den Kapiteln zuvor bereits ausführlich diskutiert wurde. Jedoch ergab sich über die verschiedenen Interviewpartnerinnen hinweg eine Reihe von klaren Anforderungen an einen Zertifizierungsprozess. Diese Anforderungen können in vier Kategorien eingeteilt werden. Zum einen gibt es Anforderungen, die für Unternehmen bei der Umsetzung wichtig sind, Anforderungen, die den Erhalt der Innovationsfähigkeit gewährleisten und Anforderungen an einen transparenten Prozess und eine klare Rolle von Behörden und Institutionen. Auf der anderen Seite stehen Anforderungen, die einen Mehrwert der Zertifizierung von KI-basierten Systemen für Endnutzende gewährleisten. Diese Anforderungen werden im Folgenden aufgelistet.

**Anforderungen an die Umsetzbarkeit**

- Es müssen Prozesse und Methoden geschaffen werden, die sowohl eine Zertifizierung durch Dritte als auch ein Self-Assessment nach EU-Vorgaben ermöglichen und transparent machen.
- Es müssen Methoden und Werkzeuge für eine nachvollziehbare Argumentation für die Bewertung des Anwendungsfalls und dessen Risiken erstellt werden.
- Es müssen Metriken geschaffen werden, die eine Aussage darüber zulassen, wann ein KI-basiertes System ein Risiko ausreichend adressiert und ein Zertifikat erhalten kann bzw. welche Maßnahmen getroffen werden müssen, um die Metrik zu erfüllen.
- Zertifizierung muss für Unternehmen aller Größen umsetzbar sein. Hierfür bedarf es der Unterstützung für den Zertifizierungsprozess, insbesondere für Startups und kleine und mittelständische Unternehmen.
- Es muss die Bereitschaft vorhanden sein, mit einer Restunsicherheit umzugehen.
- Das Risiko von KI-basierten Anwendungen muss quantifiziert und versichert werden können.
- Es wird genügend Zeit benötigt, um bestehende KI-basierte Systeme vor Inkrafttreten des EU AI Act zertifizieren zu können.

**Anforderungen zum Erhalt der Innovationsfähigkeit**

- Zertifizierung darf nicht zu einschränkend sein, da sonst Innovationen behindert werden.
- Der Zertifizierungsprozess muss ressourcenschonend durchführbar sein. Insbesondere sollten die geforderten Maßnahmen nicht über das technisch notwendige Minimum hinausgehen.
- Es darf keine Engpässe bei Zertifizierungsstellen geben, die eine Entwicklung verzögern.

**Anforderungen an Transparenz und Rollen**

- Es muss eindeutig sein, wann eine Zertifizierung notwendig ist. Des Weiteren muss klar geregelt sein, wann ein Self-Assessment nach EU-Vorgaben genügt und wann eine Zertifizierung und somit eine Prüfung durch Dritte erforderlich ist.
- Sowohl die Bedingungen als auch die Prozesse für eine Zertifizierung müssen verständlich und transparent sein.
- Es muss geklärt werden, für welchen Zeitraum eine Zertifizierung gültig ist und welche Änderungen an einem KI-basierten System vorgenommen werden dürfen, ohne dass eine Zertifizierung ungültig wird.
- Die Rolle der sektoralen Aufsichtsbehörden, wie z. B. Bundesnetzagentur oder Bafin, muss geregelt sein.
- Es muss sichergestellt werden, dass ein homogener Zertifizierungsmarkt innerhalb der EU entsteht, bei dem in jedem Land die gleichen Anforderungen an Unternehmen gestellt werden.

**Anforderungen für einen Mehrwert für Endnutzende von KI-Systemen**

- Zertifizierung muss dazu beitragen, Standards für Qualität und Sicherheit zu gewährleisten.
- Eine Zertifizierung muss sicherstellen, dass Nutzende von KI-basierten Systemen klar über die Funktionsweise des Systems informiert werden, um fundierte Entscheidungen treffen zu können.
- Es muss sichergestellt werden, dass bei einem Self-Assessment nach EU-Vorgaben alle Richtlinien eingehalten werden.
- Bei der Schaffung von harmonisierten Normen müssen Fachleuten aus den Bereichen Informatik, Ingenieurwissenschaften, Ethik und Fairness mitwirken.
- Es müssen Lösungen für noch offene technische Herausforderungen, wie z. B. »Emerging Properties«, bei großen Sprachmodellen geschaffen werden.
- Es muss eine Methode zur Validierung von generativen Modellen geschaffen werden.

Eine Zertifizierung muss den vielseitigen Anforderungen und Herausforderungen gerecht werden, um gleiche Chancen für Unternehmen jeglicher Größe zu bieten und Vertrauen bei allen Stakeholdern, insbesondere den Nutzenden, zu schaffen.

# 4. Ausblick und Handlungsempfehlungen

Bei der KI-Zertifizierung treffen zwei Themen – KI und der Zertifizierungsprozess – aufeinander und selten gibt es in Unternehmen für beide Themen bereits bestehende Expertise oder Prozesse. Selbst in großen Unternehmen, in denen sowohl Expertise zum Thema KI und Zertifizierung vorhanden sind, ist dieses Wissen oft an unterschiedlichen Stellen im Unternehmen vorhanden und schwer zu integrieren.

Aus den Interviews lassen sich neben den Anforderungen, die Unternehmen an Absicherungs- und Zertifizierungsprozesse stellen, auch diverse Bedürfnisse ableiten. Diese unterscheiden sich je nach Charakteristika des Unternehmens.

So gibt es Unternehmen, denen das Wissen zum Thema Absicherung und Zertifizierung KI-basierter Systeme fehlt, oder Unternehmen, denen die Ressourcen fehlen, zum einen um Zertifizierungsprozesse mit internem Personal adäquat vorzubereiten, zum anderen aber auch Zertifizierungsprozesse durch Dritte zu durchlaufen. Wieder andere Unternehmen brauchen Unterstützung bei der Übersetzung externer Anforderungen in interne Unternehmensprozesse oder bei der konkreten Gestaltung sicherer KI-Systeme, um sich auf anstehende Zertifizierungsprozesse vorzubereiten. Auch gibt es Unternehmen, die KI-basierte Lösungen aus anderen Ländern beziehen und Hilfe bei einem sicheren Betrieb und weiteren auftretenden Fragestellungen rund um die Zertifizierung dieser Systeme benötigen. Ebenfalls zeigt sich, dass es Unternehmen gibt, welche sich grundsätzlich für die Relevanz sicherer KI-Systeme weiterbilden und sensibilisieren wollen und dafür die richtigen Weiterbildungsangebote, Plattformen, Partner und Netzwerke suchen. Je nach Unternehmen und den Eigenschaften, die dieses beschreiben, ergeben sich ganz individuelle Bedürfnisse, welche adressiert werden müssen.

Es lassen sich vier Kategorien an Angeboten identifizieren, welche Unternehmen bei einer erfolgreichen KI-Zertifizierung unterstützen:

- Informations- und Wissensvermittlung
- Netzwerk und Austausch
- Individuelle Beratungsangebote
- Methoden und Tools zur Anwendung in der Praxis.

Im Folgenden wird zwischen den KI-Reifegraden »Einsteiger«, »Fortgeschrittene« und »Experten« unterschieden. Diese Kategorien dienen dazu, den Erfahrungsgrad von Unternehmen bei der Anwendung von KI grob zu klassifizieren. Einsteiger haben noch keine KI-Systeme produktiv eingesetzt, Fortgeschrittene haben einen bis wenige Anwendungsfälle umgesetzt, während Experten bereits mehrere KI-Anwendungsfälle erfolgreich realisiert haben und mit der Technologie sehr vertraut sind. Je nach KI-Reifegrad eines Unternehmens bedarf es einer unterschiedlichen Ausgestaltung dieser Angebote. Manche Angebote sind in der Regel auch erst ab einem höheren Reifegrad erfolgversprechend in Unternehmen einsetzbar. In Abbildung 6 sind die Angebote, die Unternehmen auf dem Weg zur erfolgreichen KI-Zertifizierung unterstützen, dargestellt. Diese sind in die vier aufgeführten Kategorien eingeteilt und für den Reifegrad der KI-Zertifizierung eines Unternehmens verortet. Dabei sind in Einzelfällen auch Abweichungen von dieser Einteilung denkbar.

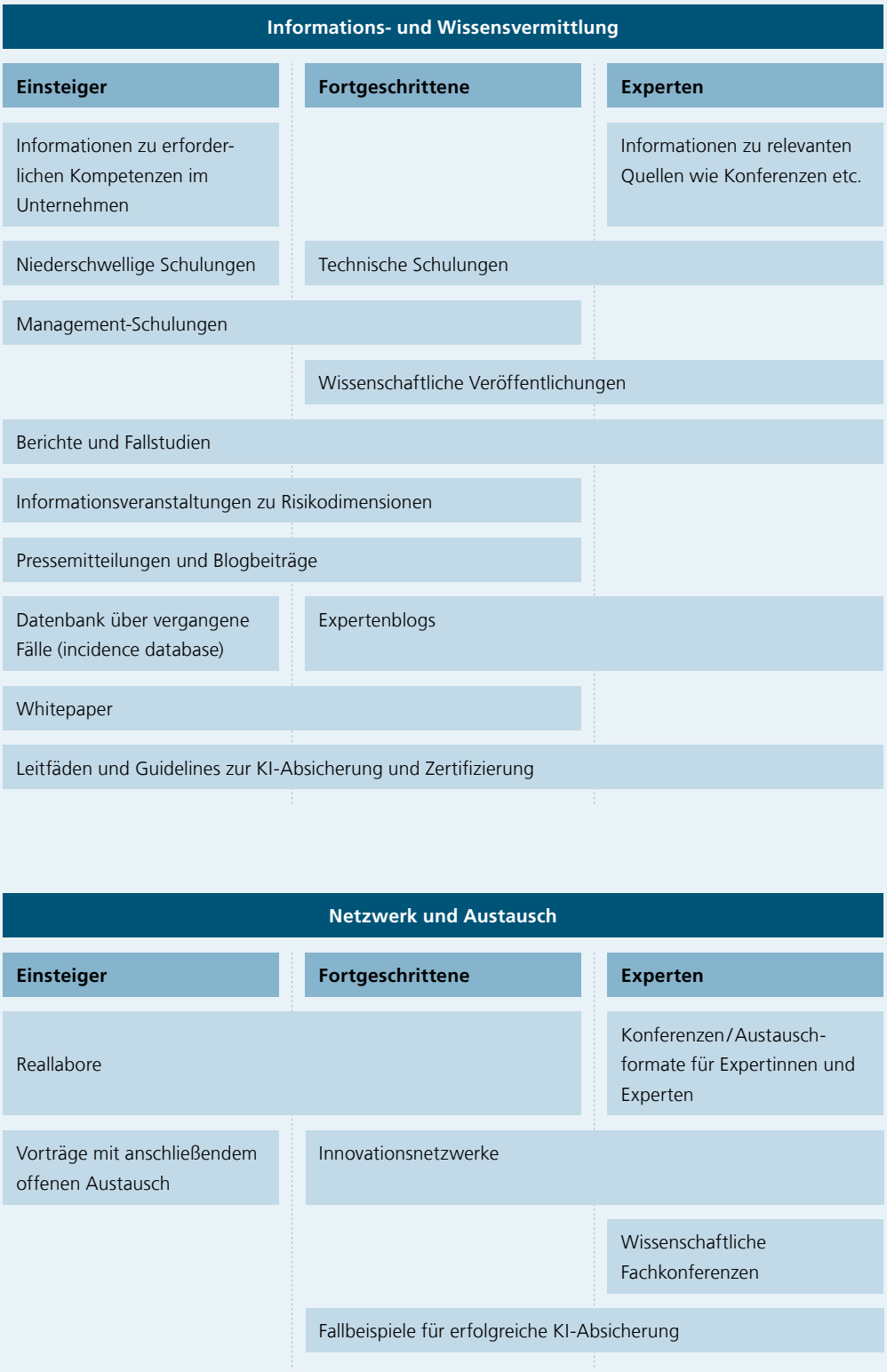


*Abbildung 6: Unterstützungsangebote für Unternehmen nach deren KI-Reifegrad. (S. 17 und 18)*

| Individuelle Beratungsangebote | | |
|---|---|---|
| **Einsteiger** | **Fortgeschrittene** | **Experten** |
| Workshop zur Einordnung des eigenen Unternehmens | Juristische Beratung zur gesetzlichen Absicherung von Eigenentwicklungen | |
| | Juristische Prüfung | |
| | Technische Beratung | |
| | Beratung zu Normen | |
| | Beratung zum Stand der Forschung | |
| | Freiwillige Zertifizierung | |
| Mentoring | | |
| Juristische Prüfung von KI-Systemen, die Unternehmen einkaufen möchten | | |
| Unterstützung bei der Erstellung einer personalisierten Zertifizierungsroadmap | | |

| Methoden und Tools zur Anwendung in der Praxis | | |
|---|---|---|
| **Einsteiger** | **Fortgeschrittene** | **Experten** |
| Self-Assessment zur Einordnung des Reifegrades des Unternehmens bei der KI-Absicherung und -Zertifizierung | Tools zur automatisierten Prüfung von KI-Eigenschaften | |
| | Dokumentations-Blaupausen für die Zertifizierung | |
| | Tools zur automatisierten Einschätzung der Kritikalität eines KI-Systems | |

*Abbildung 6: Unterstützungsangebote für Unternehmen nach deren KI-Reifegrad. (S. 17 und 18)*

# 5. Zusammenfassung

Dieses Whitepaper wirft einen umfassenden Blick auf das komplexe Thema der KI-Zertifizierung und -Absicherung im Kontext des EU AI Act. Im Fokus steht das enorme Potenzial KI-basierter Systeme, gleichzeitig wird aber auch auf die Risiken hingewiesen, die bei mangelnden Standards und unzureichenden Richtlinien auftreten können. Daher wird die Bedeutung von Vorschriften, Standards und einer umfassenden Zertifizierung für die Förderung von Vertrauen und den sicheren, rechtskonformen Einsatz von KI betont.

Unter Berücksichtigung des risikobasierten Ansatzes des EU AI Act wurden – basierend auf acht Interviews mit Expertinnen und Experten aus verschiedenen Branchen und Anwendungsbereichen – verschiedene Anforderungen und Erwartungen von Unternehmen im Kontext der KI-Zertifizierung diskutiert. Ein wesentliches Ergebnis dabei ist, dass die Relevanz der Risikodimensionen in erster Linie vom Anwendungsfall und nicht von der Branche abhängig ist. Faktoren wie die Transparenz und Machbarkeit von Zertifizierungsprozessen, klare Rollen von Behörden und Institutionen sowie die Bewahrung der Innovationsfähigkeit werden als besonders wichtig hervorgehoben. Darüber hinaus wird darauf hingewiesen, dass bei der KI-Zertifizierung stets der Mehrwert für die Endnutzerinnen und Endnutzer im Vordergrund stehen sollte.

Basierend auf den Aussagen unserer Interviewpartnerinnen und Interviewpartner scheinen die meisten Unternehmen außerdem nicht ausreichend für die kommenden Herausforderungen, die der EU AI Act mit sich bringt, vorbereitet zu sein. Abhängig vom individuellen Unternehmen wurde eine Reihe von Bedürfnissen zur optimalen Vorbereitung identifiziert. Hierzu gehören Informationsvermittlung und Wissenstransfer, Netzwerkbildung und Austausch von Erfahrungen, individuelle Beratungsangebote sowie praktische Methoden und Werkzeuge zur Unterstützung bei der Absicherung und Zertifizierung von KI-basierten Systemen.

Insgesamt unterstreicht das Whitepaper die Notwendigkeit einer effektiven Zusammenarbeit zwischen den verschiedenen Akteuren, um die Herausforderungen und Chancen, die mit der KI-Zertifizierung verbunden sind, erfolgreich zu bewältigen. Ziel ist es, einen sicheren und vertrauenswürdigen Rahmen für die Nutzung von KI in der EU zu schaffen, und Zertifizierung ist eine Möglichkeit, um dies zu erreichen. Darüber hinaus sollten jedoch auch weitere angrenzende Themen diskutiert werden, welche einen positiven Einfluss auf den sicheren Umgang mit KI-Technologien in der EU haben. Fragen rund um digitale Ethik und Verantwortung, Vertrauen in KI-Technologien beziehungsweise die Entwicklung vertrauenswürdiger KI-Anwendungen sollten weiterhin Bestandteil von Forschungs- und Entwicklungsarbeiten sein, auch unabhängig von Zertifizierungsprozessen.

Im Hinblick auf zukünftige Entwicklungen lässt sich feststellen, dass es noch erheblichen Forschungsbedarf in den Bereichen Erklärbarkeit, Verifikation und Unsicherheitsquantifizierung von KI gibt. In diesen Bereichen besteht das Potenzial, einen wesentlichen Beitrag zur Verbesserung der Transparenz und Sicherheit von KI-basierten Systemen zu leisten und dabei zu helfen, die Anforderungen des EU AI Act zu erfüllen. Darüber hinaus kann die Einbeziehung von Methoden zur Unsicherheitsquantifizierung dazu beitragen, KI-Modelle zu entwickeln, die besser in der Lage sind, ihre eigene Entscheidungsunsicherheit zu erkennen und zu kommunizieren. Neben der Modellverbesserung stellt die effiziente Sammlung und Verwendung von Trainingsdaten für KI-Modelle eine signifikante Herausforderung dar. Dabei muss sichergestellt werden, dass diese Datensätze qualitativ hochwertig, repräsentativ und frei von Verzerrungen sind. Insbesondere besteht ein dringender Bedarf an Forschung im Bereich der datenzentrierten KI, um die Effizienz und Qualität der Datensammlung und des Trainingsprozesses weiter zu verbessern.

Der Weg zur sicheren und vertrauenswürdigen Nutzung von KI erfordert die gemeinsame Anstrengung von Regulierungsbehörden, Industrie, Forschungseinrichtungen und der breiten Öffentlichkeit. Durch die Zusammenarbeit und den Austausch zwischen diesen verschiedenen Akteuren kann ein umfassendes Verständnis der Chancen und Herausforderungen der KI-Nutzung entwickelt werden, was letztlich zur Entwicklung effektiver und praxisorientierter Lösungen führt.

# 6. Quellenverzeichnis


AI Incident Database (o. D.): AI Incident Database. Online verfügbar unter https://incidentdatabase.ai/, zuletzt geprüft am 22.08.2023.

Brajovic, Danilo; Renner, Niclas; Goebels, Vincent Philipp; Wagner, Philipp; Fresz, Benjamin; Biller, Martin et al. (2023): Model Reporting for Certifiable AI: A Proposal from Merging EU Regulation into AI Development.

Bundesminsterium für Bildung und Forschung (o. D.): Intelligenter Rollator für die stationäre Pflege (RABE). Online verfügbar unter https://www.gesundheitsforschung-bmbf.de/de/intelligenter-rollator-fur-die-stationare-pflege-rabe-8613.php, zuletzt geprüft am 22.08.2023.

CertAI (Hg.) (o. D.): The CertAI Trustworthy AI seals. CertAI. Online verfügbar unter https://www.certai.com, zuletzt geprüft am 18.07.2023.

DIN, DKE (2022): Deutschen Normungsroadmap KI. Hg. v. Christoph Winterhalter Wolfgang Wahlster. DIN, DKE. Online verfügbar unter https://www.din.de/resource/blob/916792/20bf33d405710a703aa26f81362493bb/nrm-ki-deutsch-2022-final-web-250-data.pdf, zuletzt geprüft am 12.07.2023.

Europäische Kommission (2020): WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust. Online verfügbar unter https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf, zuletzt geprüft am 22.08.2023.

Europäische Kommission (2021): Artificial Intelligence Act. Online verfügbar unter https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206, zuletzt geprüft am 25.07.2023.

Friederich Peter Zeuner (o. D.): Vergabe von Zertifikaten – Rechtslage. Hg. v. Conplore Magazine. Online verfügbar unter https://conplore.com/definition-begriff-zertifikat-vergabe-von-zertifikaten-rechtslage-deutschland/, zuletzt geprüft am 22.08.2023.

Gebru, Timnit; Morgenstern, Jamie; Vecchione, Briana; Vaughan, Jennifer Wortman; Wallach, Hanna; Daumé, Hal, III; Crawford, Kate (2018): Datasheets for Datasets. Online verfügbar unter http://arxiv.org/pdf/1803.09010v8.

Gregor Damm (2022): International agierender Zertifizierungsanbieter für Künstliche Intelligenz soll entstehen / „CertifAI" soll KI-Produkte von der Entwicklung bis hin zur finalen Produktreife begleiten, prüfen und zertifizieren. PWC, Dekra. Online verfügbar unter https://www.pwc.de/de/pressemitteilungen/2022/pwc-dekra-und-der-innovationsfonds-der-stadt-hamburg-beabsichtigen-die-gruendung-von-certifai-fuer-die-pruefung-von-ki-produkten.html, zuletzt geprüft am 12.07.2023.

Hadrien Pouget (2023): Standard Setting | The Artificial Intelligence Act. Hg. v. artificialintelligenceact.eu/. Online verfügbar unter https://artificialintelligenceact.eu/standard-setting/, zuletzt geprüft am 12.07.2023.

Hochrangige Expertengruppe für künstliche Intelligenz (2019): Ethik-Leitlinien für eine vertrauenswürdige KI. Brüssel: Europäische Kommission.

Holland, Sarah; Hosny, Ahmed; Newman, Sarah; Joseph, Joshua; Chmielinski, Kasia (2018): The Dataset Nutrition Label: A Framework To Drive Higher Data Quality Standards. Online verfügbar unter https://datanutrition.org/.

ISO (o. D.): Standards by ISO/IEC JTC 1/SC 42. Artificial intelligence. Online verfügbar unter https://www.iso.org/committee/6794475/x/catalogue/p/0/u/1/w/0/d/0, zuletzt aktualisiert am 2023, zuletzt geprüft am 12.07.2023.

Mahima Pushkarna, Andrew Zaldivar (2022): The Data Cards Playbook: A Toolkit for Transparency in Dataset Documentation. Hg. v. https://ai.googleblog.com/. Google. Online verfügbar unter https://ai.googleblog.com/2022/11/the-data-cards-playbook-toolkit-for.html, zuletzt geprüft am 12.07.2023.

Mitchell, Margaret; Wu, Simone; Zaldivar, Andrew; Barnes, Parker; Vasserman, Lucy; Hutchinson, Ben et al. (2019): Model Cards for Model Reporting 81, S. 220–229. DOI: 10.1145/3287560.3287596.

Poretschkin, Maximilian; Schmitz, Anna; Akila, Maram; Adilova, Linara; Becker, Daniel; Cremers, Armin B.; Hecker, Dirk (2021): Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz. KI-Prüfkatalog. Hg. v. Fraunhofer-Institut für Intelligente Analyseund Informationssysteme IAIS. Online verfügbar unter https://www.iais.fraunhofer.de/s/ki-pruefkatalog/epaper/ausgabe.pdf, zuletzt geprüft am 22.08.2023.

Pushkarna, Mahima; Zaldivar, Andrew; Kjartansson, Oddur (2022): Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. Online verfügbar unter http://arxiv.org/pdf/2204.01075v1.

TÜV IT (Hg.) (o. D.): Künstliche Intelligenz (KI) – Prüfung & Zertifizierung von KI-Sicherheit. Online verfügbar unter https://www.tuvit.de/de/innovationen/ki/, zuletzt geprüft am 22.08.2023.

withgoogle.com (Hg.): The value of a shared understanding of AI models. Online verfügbar unter https://modelcards.withgoogle.com/about, zuletzt geprüft am 12.07.2023.

# Impressum

**Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO**
Nobelstraße 12
70569 Stuttgart
www.iao.fraunhofer.de

**Kontakt**
Dr.  Bernd Bienzeisler
Leiter Forschungs- und Innovationszentrum
Kognitive Dienstleistungssysteme
Tel. +49 711 970-2088
bernd.bienzeisler@iao.fraunhofer.de

**Satz und Layout**
Valentin Buhl

**Lektorat**
Dagmar Bartels

**Fraunhofer-Publica**
http://dx.doi.org/10.24406/publica-1875

**Alle Rechte vorbehalten**
© Fraunhofer IAO, 2023

## Kontakt

Dr.  Bernd Bienzeisler
Leiter Forschungs- und Innovationszentrum
Kognitive Dienstleistungssysteme
Tel. +49 711 970-2088
bernd.bienzeisler@iao.fraunhofer.de

Fraunhofer-Institut für Arbeitswirtschaft
und Organisation IAO
Nobelstraße 12
70569 Stuttgart

www.iao.fraunhofer.de