

# Hochrisiko-KI-Systeme im Spannungsfeld von Recht und Technik:

Technische Umsetzung der regulatorischen Anforderungen



## Inhaltsverzeichnis:

|                                                                      |           |
|----------------------------------------------------------------------|-----------|
| <b>1. AI meets Regulation</b>                                        | <b>4</b>  |
| Was Europa mit der KI-Verordnung vorgibt                             |           |
| <b>2. Hochrisiko-KI-Systeme</b>                                      | <b>7</b>  |
| Was Anbieter rechtlich und technisch beachten müssen                 |           |
| <b>3. Der Stand der Technik und internationale KI-Standards</b>      | <b>12</b> |
| <b>4. Von der Theorie zur Praxis</b>                                 | <b>16</b> |
| Technische Umsetzung am Beispiel eines Hochrisiko-KI-Systems         |           |
| <b>4.1. Analyse der Datenqualität</b>                                | <b>17</b> |
| <b>4.2. Fairness und Bias des KI-Systems</b>                         | <b>19</b> |
| <b>4.3. Leistungsfähigkeit und Robustheit</b>                        | <b>21</b> |
| <b>5. KI-Compliance mit der richtigen Systemlandschaft erreichen</b> | <b>23</b> |

## Executive Summary

Die europäische KI-Verordnung (KI-VO) setzt neue Maßstäbe für den sicheren und vertrauenswürdigen Einsatz von Künstlicher Intelligenz. Besonders für Hochrisiko-KI-Systeme formuliert sie umfassende Anforderungen an Transparenz, Fairness, Datenqualität und Robustheit. Während die Verordnung klare rechtliche Pflichten definiert, bleibt die technische Umsetzung weitgehend offen – der Nachweis der Konformität orientiert sich am „Stand der Technik“.

Bei der Übersetzung regulatorischer Anforderungen in technische Lösungskonzepte kommen internationale Normen ins Spiel: ISO/IEC 42001 für KI-Managementsysteme, ISO/IEC 24029 zur Robustheitsbewertung oder ISO/IEC 5259 zur Datenqualität bieten konkrete Orientierung für die technische Ausgestaltung regulatorischer Anforderungen. Unternehmen müssen beispielsweise Maßnahmen zur Bias-Analyse, zur Absicherung der Datenintegrität und zur Leistungsbewertung ihrer KI-Systeme nachweisen. Diese Anforderungen können durch ISO-konforme Prozesse operationalisiert und dokumentiert werden. Die Anwendung auf einen konkreten Use Case zeigt, wie dies in der Praxis aussehen kann.

Technische Standards sind wichtige Instrumente, um regulatorische Sicherheit zu schaffen und die technische Qualität von KI-Anwendungen transparent und nachvollziehbar zu belegen. Um im Spannungsfeld von Recht und Technik zu navigieren, ist dementsprechend Interdisziplinarität der Schlüssel. Mitarbeiter aller involvierten Fachrichtungen sollten eingebunden werden und zusammenarbeiten.



# AI meets Regulation

## Was Europa mit der KI-Verordnung vorgibt

Die Erfolge von KI-Anwendungen wie ChatGPT oder Midjourney verdeutlichen, wie vielfältig künstliche Intelligenz heute einsetzbar ist – und wie tiefgreifend sie nahezu alle Lebens- und Arbeitsbereiche verändern wird. Damit wächst auch die Verantwortung, KI-Systeme nicht nur innovativ, sondern auch vertrauenswürdig und nachhaltig zu entwickeln und zu betreiben.

### Die KI-Verordnung als regulatorischer Meilenstein

Mit der KI-Verordnung (KI-VO) schafft die Europäische Union einen weltweit einzigartigen Rechtsrahmen, der die Entwicklung und Nutzung von KI-Systemen umfassend reguliert. Ziel ist es, Risiken frühzeitig zu erkennen und zu kontrollieren – insbesondere dort, wo KI besonders sensible Entscheidungen beeinflusst.

Die KI-VO ist am 2. August 2024 in Kraft getreten. Sie sieht gestaffelte Übergangsfristen vor:

- **Ab 2. Februar 2025:** Unternehmen müssen die allgemeinen Vorschriften (Kapitel 1) sowie die Regelungen zu verbotenen Praktiken (Kapitel 2) umsetzen.
- **Ab 2. August 2026:** Es gelten die Anforderungen für Hochrisiko-KI-Systeme gemäß Anhang III.
- **Ab 2. August 2027:** Die Pflichten für KI-Systeme nach Anhang I treten in Kraft.

### Herausforderungen für Unternehmen: Risikobewertung, Rollen und Compliance

Für Unternehmen bedeutet die KI-VO vor allem eins: Komplexität. Sie müssen nicht nur prüfen, ob und wie ihre KI-Anwendungen in den Anwendungsbereich der Verordnung fallen, sondern auch klären, in welche **Risikoklasse** diese einzuordnen sind. Das ist häufig mit erheblichen Unsicherheiten verbunden.

Hinzu kommen neue **Compliance-Anforderungen** und Prozesse, die intern etabliert werden müssen – insbesondere für Hochrisiko-KI-Systeme, die strengen Pflichten unterliegen.

Zudem verpflichtet die KI-VO Unternehmen dazu, ihre **Rolle** im Lebenszyklus eines KI-Systems gemäß **Art. 25 Abs. 1 KI-VO** zu definieren. Diese Rollen – etwa Anbieter, Inverkehrbringer oder Importeur – richten sich nach der Funktion des Unternehmens entlang der KI-Wertschöpfungskette. Entscheidend ist nicht nur, wer die KI entwickelt, sondern auch, wer sie innerhalb der EU vertreibt oder in Betrieb nimmt.

## Rollen und Risikoklassen bestimmen die regulatorischen Pflichten

Ein zentrales Element der KI-Verordnung ist die Einordnung eines Unternehmens entlang der **KI-Wertschöpfungskette**. Je nachdem, welche Rolle ein Unternehmen einnimmt, ergeben sich unterschiedliche rechtliche Verpflichtungen. Die KI-VO definiert diese Rollen präzise:



**Anbieter** ist, wer ein KI-System selbst entwickelt oder entwickeln lässt (Art. 3 Nr. 3 KI-VO).



**Betreiber** ist, wer ein KI-System in eigener Verantwortung einsetzt – allerdings nur im beruflichen Kontext (Art. 3 Nr. 4 KI-VO).



**Bevollmächtigter** handelt im Auftrag des Anbieters eines KI-Systems auf Basis einer erteilten Vollmacht (Art. 3 Nr. 5 KI-VO).



**Einführer** bringt KI-Systeme aus einem Drittstaat in die EU und bietet sie hier an (Art. 3 Nr. 6 KI-VO).



**Händler** vertreibt KI-Systeme innerhalb der EU (Art. 3 Nr. 7 KI-VO).

## Risikobasierter Ansatz: Die fünf Risikoklassen der KI-VO

Neben der Rolle ist die **Risikoklasse des KI-Systems oder -Modells** entscheidend für die Compliance-Pflichten. Die KI-VO verfolgt einen risikobasierten Ansatz und unterscheidet zwischen fünf Kategorien:

- 1. Verbotene Praktiken** (Art. 5 KI-VO): Diese KI-Anwendungen sind grundsätzlich untersagt.
- 2. Hochrisiko-KI-Systeme** (Art. 6 ff., Anhang I und III): Besonders reguliert und mit umfangreichen Anforderungen versehen.
- 3. Systemisches Risiko** (Art. 55 KI-VO): Gilt für KI-Modelle mit allgemeinem Verwendungszweck, deren Einsatz besondere Risiken birgt.
- 4. Mittleres Risiko** (Art. 50 KI-VO): Es bestehen Transparenzpflichten, aber keine Konformitätsbewertung.
- 5. Geringes Risiko** (Art. 95 KI-VO): Keine spezifischen regulatorischen Anforderungen, freiwillige Selbstverpflichtungen möglich.

**Wichtig:** Die Einordnung in Risikoklassen kann **kumulativ** erfolgen. Das bedeutet, ein KI-System kann mehreren Risikokategorien gleichzeitig zugeordnet sein – mit entsprechend gestapelten Pflichten.

## Pflichten ergeben sich aus Rolle und Risiko

Die regulatorischen Anforderungen der KI-Verordnung entstehen aus der Kombination zweier Faktoren:

1. **Welche Rolle** das Unternehmen entlang der KI-Wertschöpfungskette einnimmt.
2. **Welche Risikoklasse/n** dem KI-System oder -Modell zugeordnet ist / sind.

Pflichten entstehen jedoch **nicht für jede Kombination** aus Rolle und Risiko. Beispielsweise betreffen bestimmte Anforderungen ausschließlich Anbieter von Hochrisiko-KI-Systemen, während für Betreiber geringfügiger Systeme deutlich weniger Regelungen gelten. Ein Überblick:

- Bei **Hochrisiko-KI-Systemen** ergeben sich Pflichten für alle Rollen entlang der Wertschöpfungskette – vom Anbieter bis zum Händler.
- Systeme mit **systemischem Risiko** betreffen primär Anbieter großer, allgemein nutzbarer KI-Modelle.
- **Mittleres Risiko** bringt Pflichten für Anbieter und Betreiber mit sich.
- Bei **geringem Risiko** sind regulatorische Anforderungen optional und begrenzt auf freiwillige Maßnahmen.

## Übersicht Rollen und Risikoklassen nach der KI-Verordnung:

|                                                           | Anbieter | Betreiber | Bevollmächtigter | Einführer | Händler |
|-----------------------------------------------------------|----------|-----------|------------------|-----------|---------|
| <b>Hochrisiko-KI</b><br>(Art. 6 ff., Annex I + III KI-VO) | X        | X         | X                | X         | X       |
| <b>Systemisches Risiko</b><br>(Art. 55 KI-VO)             | X        |           |                  |           |         |
| <b>Mittleres Risiko</b><br>(Art. 50 KI-VO)                | X        | X         |                  |           |         |
| <b>Geringes Risiko</b><br>(Art. 95 KI-VO)                 | (x)      | (x)       |                  |           |         |

**Legende:** x = Anwendung (x) = Freiwillige Verpflichtung

# Hochrisiko-KI-Systeme

## Was Anbieter rechtlich und technisch beachten müssen

Im Mittelpunkt der KI-Verordnung steht die besondere Regulierung sogenannter Hochrisiko-KI-Systeme. Diese unterliegen einer Vielzahl technischer und organisatorischer Pflichten – speziell für Anbieter dieser Systeme.

### Zwei Wege zur Einstufung als Hochrisiko-KI

Die Einstufung eines KI-Systems als Hochrisiko erfolgt auf zwei verschiedenen Wegen:

#### 1. Integrierte KI in physische Produkte

Ein KI-System gilt dann als Hochrisiko-KI, wenn es - vereinfacht gesagt - in ein physisches Produkt integriert ist (z. B. bei Maschinen oder Kraftfahrzeugen).

Gemäß Art. 6 Abs. 1 i. V. m. Anhang I KI-VO ist dies der Fall, wenn das KI-System entweder eigenständig oder als Sicherheitsbauteil eines anderen Produkts aufgrund bestehender EU-Harmonisierungsvorschriften einer Konformitätsbewertung unterliegt. An dieser Stelle treffen EU-Regelungen aus dem EU-Produktsicherheitsrecht mit jenen der KI-Verordnung zusammen.



## 2. Stand-alone-KI mit Einsatz in besonders sensiblen Bereichen

Auch eigenständige KI-Systeme gelten als hochriskant, wenn sie in besonders sensiblen Bereichen eingesetzt werden. Diese sind in Anhang III der KI-VO aufgelistet und darunter fallen etwa:

- Biometrische Identifikation
- Kritische Infrastruktur (z. B. Verkehr, Energieversorgung)
- Zugang zu Bildung und staatlichen Leistungen
- Personalentscheidungen (z. B. Einstellung, Kündigung)
- Kreditwürdigkeitsprüfung oder Risikobewertung bei Versicherungen
- Strafverfolgung und Migration
- Rechtspflege und demokratische Prozesse (z. B. Wahlbeeinflussung)

### Ausnahme und Rückausnahme:

Art. 6 Abs. 3 KI-VO enthält Ausnahmen, etwa wenn das KI-System keine wesentliche Auswirkung auf das Ergebnis hat. Allerdings: Wird Profiling betrieben, greift eine Rückausnahme – das System ist dann doch als Hochrisiko-KI einzustufen.

## Diese Pflichten greifen bei Hochrisiko-KI

Wird ein KI-System nach Maßgabe von Art. 6 KI-VO und den entsprechenden Anhängen als Hochrisiko-KI-System eingestuft, entfaltet dies weitreichende **rechtliche Anforderungen**, insbesondere für **Anbieter** – aber auch **Betreiber** sind betroffen.

Die Anbieter von Hochrisiko-KI-Systemen tragen die Hauptverantwortung für die rechtskonforme Gestaltung, Bereitstellung und Überwachung ihrer Systeme. Bereits **vor der Inbetriebnahme** greifen umfassende Pflichten nach **Art. 4, 8 bis 19 KI-VO**.

### Grundvoraussetzung: Nachweis der KI-Kompetenz

Unternehmen, die Hochrisiko-KI-Systeme anbieten, müssen nach Art. 4 KI-VO sicherstellen, dass alle Personen, die KI-Systeme entwickeln, nutzen oder betreiben, über ausreichend KI-Kompetenz verfügen – etwa durch personelle, technische und prozessuale Maßnahmen.



## Anbieterpflichten vor Inbetriebnahme:

- **Technische Dokumentation**

Erstellung, fortlaufende Aktualisierung und Aufbewahrung einer vollständigen technischen Dokumentation (Art. 11 i.V.m. Art. 18). Diese muss bereits vor Inverkehrbringung vorhanden sein.

- **Gewährleistung von Genauigkeit, Robustheit und Cybersicherheit**

Das KI-System muss gemäß Art. 15 KI-VO Mindestanforderungen in den Bereichen Genauigkeit, Robustheit und Cybersicherheit erfüllen.

- **Genauigkeit** bezeichnet in diesem Zusammenhang die Widerstandsfähigkeit des Systems gegenüber Fehlern, Störungen oder Unstimmigkeiten im Betrieb.
- **Robustheit** bezieht sich auf die technische Redundanz des Systems: Die Systemarchitektur und Datenhaltung müssen so gestaltet sein, dass ein möglichst hoher Schutz vor internen und externen Störungen gewährleistet ist.
- **Cybersicherheit** schließlich umfasst alle Maßnahmen zum Schutz des Systems vor dem unbefugten Zugriff Dritter, insbesondere zur Verhinderung von Angriffen, die darauf abzielen, Schwachstellen auszunutzen und das System in seiner Funktionsweise zu manipulieren.



- **Konformitätsbewertung und -erklärung**

Ein standardisiertes Konformitätsverfahren ist durchzuführen (Art. 43, 47), inkl. **schriftlicher Erklärung** der Einhaltung der Vorschriften (Art. 16 lit. f KI-VO).

- **Betriebsanleitung und menschliche Aufsicht**

Die Nutzung muss durch eine verständliche Anleitung begleitet werden (Art. 13), und während der Anwendung muss eine **wirksame menschliche Kontrolle** ermöglicht werden (Art. 14).

- **Registrierung und CE-Kennzeichnung**

Ein Hochrisiko-KI-System muss in der EU-Datenbank eingetragen werden (Art. 49 KI-VO) und unter Umständen ist eine **CE-Kennzeichnung** erforderlich (Art. 48 KI-VO).

## **Anbieterpflichten im laufenden Betrieb**

Auch nach Markteinführung bleibt der Anbieter in der Pflicht, das System regelmäßig zu überprüfen, zu dokumentieren und weiterzuentwickeln.

- **Qualitätsmanagementsystem**

Einführung eines wirksamen internen Systems zur Sicherstellung der Einhaltung der KI-VO – inkl. kontinuierlicher Kontrolle und Verbesserung (Art. 17 KI-VO).

- **Implementierung eines Risikomanagementsystems**

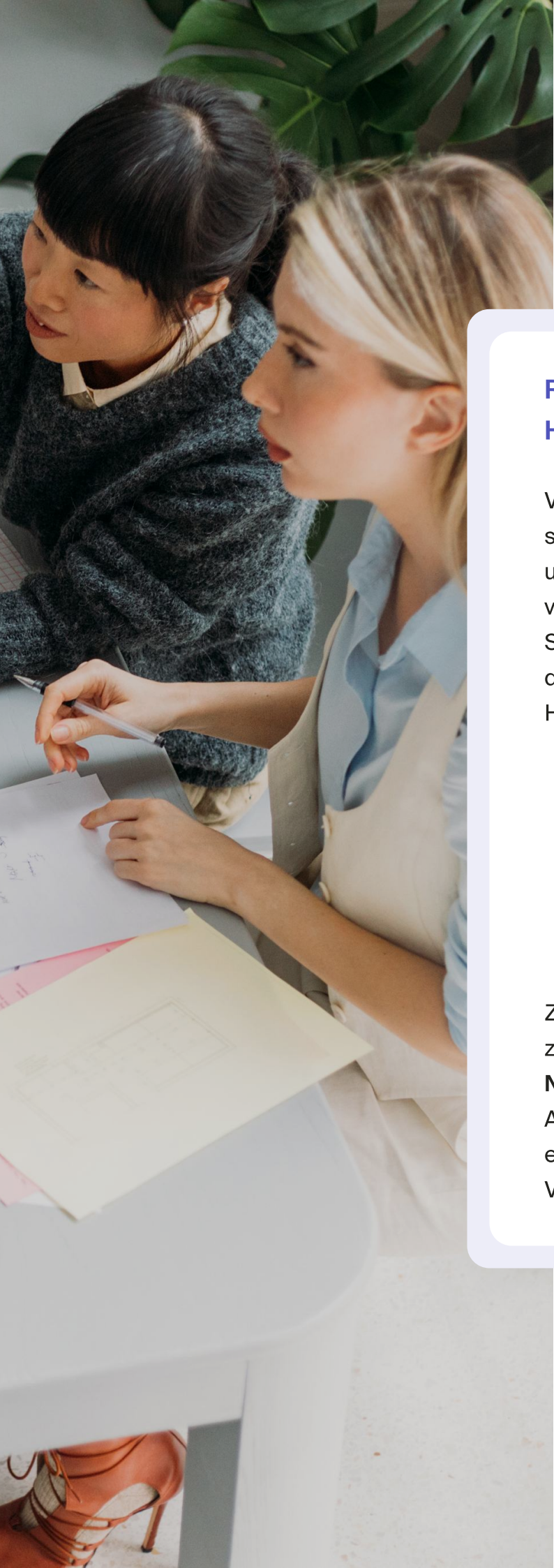
Iterative Risikoanalysen sind verpflichtend, insbesondere im Hinblick auf **Sicherheitsrisiken, Grundrechtseingriffe** und **Diskriminierungspotenzial** (Art. 9 KI-VO).

- **Verantwortungsvolle Daten- und Data-Governance-Praktiken**

Trainings-, Validierungs- und Testdaten müssen qualitativ hochwertig, zweckgebunden und repräsentativ sein (Art. 10 KI-VO). Die Verarbeitung sensibler Daten (z. B. Gesundheitsdaten) ist nur unter **strengen Voraussetzungen** zulässig – mit Verweis auf Art. 9 DSGVO.

- **Protokollierung und Aufzeichnung (Art. 12, 19)**

Automatische Logfiles technischer Abläufe müssen erstellt und revisionssicher vom Anbieter aufbewahrt werden – für Audits und ggf. aufsichtsrechtliche Prüfungen (Art. 12 i. V. m. Art. 19 KI-VO).



## Recht trifft Technik: Umsetzung bleibt Herausforderung

Viele dieser rechtlichen Anforderungen lassen sich **nicht ohne technische Expertise umsetzen** – und umgekehrt. Die KI-Verordnung bleibt bei vielen Vorgaben bewusst technologieoffen, um Spielraum für Innovation zu lassen. Doch genau das stellt Unternehmen vor praktische Herausforderungen:

- Was bedeutet „ausreichende Genauigkeit“ in einem spezifischen Kontext?
- Wie konkretisiert man Anforderungen wie Robustheit oder Nachvollziehbarkeit technisch?

Zur Beantwortung solcher Fragen bieten sich zunehmend **technische Standards (z. B. ISO/IEC-Normen)** an, die helfen, regulatorische Anforderungen operativ umzusetzen. Sie können eine Brücke schlagen zwischen abstrakten Vorschriften und konkreter technischer Realität.



## Der Stand der Technik und internationale KI-Standards

Die KI-Verordnung bleibt bei vielen technischen Anforderungen bewusst technologieoffen – etwa bei der Frage, wie Genauigkeit, Robustheit oder Cybersicherheit konkret umgesetzt werden sollen. Umso wichtiger wird daher der Rückgriff auf den „**Stand der Technik**“ als maßgebliches Referenzniveau. Dieser ist nicht nur für die technische Umsetzung zentral, sondern dient auch als **Beleg für regulatorische Konformität**.

Ein wesentliches Novum der KI-Verordnung ist, dass sie nicht nur rechtliche Anforderungen formuliert, sondern darüber hinaus den **Nachweis einer technischen Bewertung** des Systems verlangt. Insbesondere die Artikel 10, 11 und 15 KI-VO rücken dabei in den Fokus:

- *Artikel 10 Abs. 2 lit. f und g KI-VO* fordert eine umfassende **Analyse und Bewertung der Trainings- und Testdatensätze** hinsichtlich **potenzieller Verzerrungen (Bias)**, die etwa die Gesundheit und Sicherheit von Personen gefährden, Grundrechte beeinträchtigen oder zu Diskriminierung führen können. Zudem müssen Maßnahmen zur Korrektur solcher Verzerrungen dokumentiert werden.

- *Artikel 10 Abs. 2 lit. h KI-VO* verlangt eine **systematische Ermittlung etwaiger Datenlücken**, die sich negativ auf die Zweckbestimmung oder Funktionstauglichkeit des KI-Systems auswirken könnten.
- *Artikel 15 KI-VO* verpflichtet Anbieter zur **technischen Bewertung der Leistungsfähigkeit** des Systems. Es müssen geeignete Maßnahmen ergriffen werden, um die Systeme **möglichst widerstandsfähig gegenüber Fehlern, Störungen und Unstimmigkeiten** zu gestalten.
- *Artikel 11 KI-VO* schreibt schließlich die Erstellung einer umfassenden **technischen Dokumentation** vor, deren Inhalte in Anhang IV KI-VO konkretisiert sind – einschließlich der oben genannten Bewertungen.

## Fehlt: Technische Konkretisierung – gefragt: der Stand der Technik

Was die Verordnung an rechtlicher Struktur vorgibt, lässt sie an technischer Spezifikation offen. Die **konkreten technischen Maßnahmen** zur Umsetzung der Anforderungen sind in der KI-VO nicht definiert. An dieser Stelle kommt der „Stand der Technik“ ins Spiel – zusammengesetzt aus:

- aktueller wissenschaftlicher Forschung,
- anerkannten Best Practices der Industrie,
- sowie internationalen **Normen und Standards**.

Letztere sind besonders relevant, weil sie als **objektiv überprüfbare und belastbare Kriterien** gelten, um die Einhaltung technischer Anforderungen nachzuweisen.



## Orientierung durch Normung: ISO, CEN-CENELEC und Co.

Die Standardisierung im Bereich der Künstlichen Intelligenz wird auf europäischer Ebene durch das Gremium **CEN-CLC/JTC 21** der CEN-CENELEC und auf internationaler Ebene durch das Komitee **ISO/IEC JTC 1/SC 42** vorangetrieben. Die Bedeutung dieser Gremien nimmt weiter zu: Allein von der ISO/IEC gibt es aktuell **34 bereits veröffentlichte** und **weitere 40 sich in der Entwicklung befindende Standards zu KI**.

Besonders hervorzuheben sind:

- **ISO/IEC 42001:2023**

Dieser Standard beschreibt Anforderungen an ein **KI-Managementsystem** und kann als konkrete Auslegungshilfe für die Anforderungen aus **Art. 17 KI-VO** herangezogen werden.

- **ISO/IEC 24029-Reihe**

Die Reihe „Assessment of the robustness of neural networks“ enthält technische Bewertungsmethoden zur **Bestimmung der Robustheit und Leistungsfähigkeit** von KI-Systemen und unterstützt damit die Umsetzung von **Art. 15 KI-VO**.

- **ISO/IEC 5259-Reihe**

Diese Standards beschreiben **Qualitätskriterien für Daten**, die beim Training und der Evaluation von KI-Systemen verwendet werden. Sie sind eine geeignete Grundlage zur Umsetzung der Anforderungen aus **Art. 10 KI-VO**.

Die Einhaltung aktueller KI-Standards bietet nicht nur **Orientierung**, sondern auch eine **sicher dokumentierbare Grundlage** für die Bewertung und Konformität von Hochrisiko-KI-Systemen. Unternehmen sollten daher kontinuierlich den Stand der Standardisierung verfolgen und ihre technischen sowie organisatorischen Maßnahmen daran ausrichten.

Denn: Die Entwicklung technischer Normen ist dynamisch – und damit auch der „Stand der Technik“, der für die KI-Compliance nach der Verordnung entscheidend ist.

## Zuordnung der Pflichten der KI-VO zu bestehenden internationalen Normen:

| Organisatorische Anforderungen                | Normen und Standards |
|-----------------------------------------------|----------------------|
| Qualitätsmanagement-System<br>(Art. 17 KI-VO) | ISO/IEC 42001:2024   |
| Risikomanagement (Art. 9 KI-VO)               | ISO/IEC 23894:2023   |
| Data Lifecycle Management<br>(Art. 10 KI-VO)  | ISO/IEC 8183:2023    |

| Technische Anforderungen                             | Normen und Standards                               |
|------------------------------------------------------|----------------------------------------------------|
| Datenqualität (Art. 10 KI-VO)                        | ISO/IEC 5259-Reihe                                 |
| Bias-Bewertung (Art. 10 KI-VO)                       | ISO/IEC TR 24027:2021                              |
| Robustheit (Art. 15 KI-VO)                           | ISO/IEC 24029-Reihe                                |
| Leistungsfähigkeit (Art. 15 KI-VO)                   | ISO/IEC TS 4213:2022                               |
| Cybersicherheit (Art. 15 KI-VO)                      | OWASP GenAI Security Project,<br>OWASP AI Exchange |
| Vertrauensvolle Technikgestaltung<br>(Art. 15 KI-VO) | ISO/IEC TR 24028:2020                              |

# Von der Theorie zur Praxis:

## Technische Umsetzung am Beispiel eines Hochrisiko-KI-Systems

Die bisherigen Kapitel haben deutlich gemacht: Die KI-Verordnung stellt umfangreiche Anforderungen an Anbieter von Hochrisiko-KI-Systemen – sowohl rechtlich als auch technisch. Während sich einige Anforderungen durch Managementsysteme und interne Prozesse abbilden lassen, bleibt die konkrete technische Umsetzung eine der größten Herausforderungen

Wie sich diese Anforderungen in der Praxis umsetzen lassen, lässt sich am besten anhand eines konkreten Anwendungsfalls verdeutlichen. Im Folgenden zeigen wir beispielhaft, wie die regulatorischen und technischen Anforderungen der KI-Verordnung im Rahmen eines **Hochrisiko-KI-Systems zur Kreditwürdigkeitsprüfung** umgesetzt werden können. Dabei werden sowohl die Anforderungen an Datenqualität und Fairness als auch an Leistungsfähigkeit und Robustheit in den Fokus gerückt – immer mit Bezug auf die einschlägigen Artikel der KI-VO und den relevanten Stand der Technik.

*Hinweis:* Wir verzichten bei dem folgenden Beispiel bewusst auf eine datenschutzrechtliche Bewertung (u. a. i. S. v. Art. 22 DSGVO).



## Fallbeispiel: KI-System zum Kreditwürdigkeitsprüfung

Wie sich die technischen Anforderungen der KI-Verordnung in der Praxis umsetzen lassen, veranschaulichen wir im Folgenden anhand eines konkreten Anwendungsfalls:

Das sogenannte „Census Income Dataset“<sup>1</sup> enthält demografische und wirtschaftliche Merkmale einzelner Personen. Auf Basis dieser Merkmale soll ein KI-System vorhersagen, ob das Einkommen einer Person über oder unter 50.000 USD liegt.

Systeme dieser Art werden potenziell im **Finanz- und Versicherungswesen** eingesetzt – etwa zur Bewertung der Kreditwürdigkeit von Antragstellenden. Damit fallen sie explizit unter den Anwendungsbereich der KI-Verordnung: Nach Art. 6 i. V. m. Anhang III Nr. 5 lit. b KI-VO handelt es sich hierbei um ein klassisches Hochrisiko-KI-System.

Neben umfassenden organisatorischen Anforderungen – wie der Einrichtung eines Qualitätsmanagementsystems (Art. 17 KI-VO) sowie eines systematischen Risikomanagements (Art. 9 KI-VO) – greifen auch die **technischen Anforderungen**, die wir im Folgenden anhand dieses Anwendungsbeispiels im Detail erläutern.

## 4.1. Analyse der Datenqualität

Die Bewertung eines Datensatzes, der für das Training oder die Evaluation eines KI-Systems verwendet wird, hängt von mehreren Faktoren ab. Ein wesentlicher Aspekt ist die Größe des Datensatzes: Dieser muss eine ausreichende Mindestgröße aufweisen, um für das Training eines KI-Systems geeignet zu sein oder um statistisch signifikante Aussagen über dessen Leistungsfähigkeit zu ermöglichen.

Ein weiterer wichtiger Faktor ist die Verteilung der Zielwerte und Merkmale innerhalb des Datensatzes. Ein potenzieller Bias kann das Training und die Evaluation des KI-Systems unmittelbar beeinflussen. Idealerweise sollten die Zielwerte bzw. wesentliche Merkmale gleichmäßig verteilt sein, was u. a. als Kriterium für die Datenbewertung dient. Eine exemplarische Auswertung dieser Verteilungen ist in **Abbildung 1** sowie in **Abbildung 2** dargestellt.

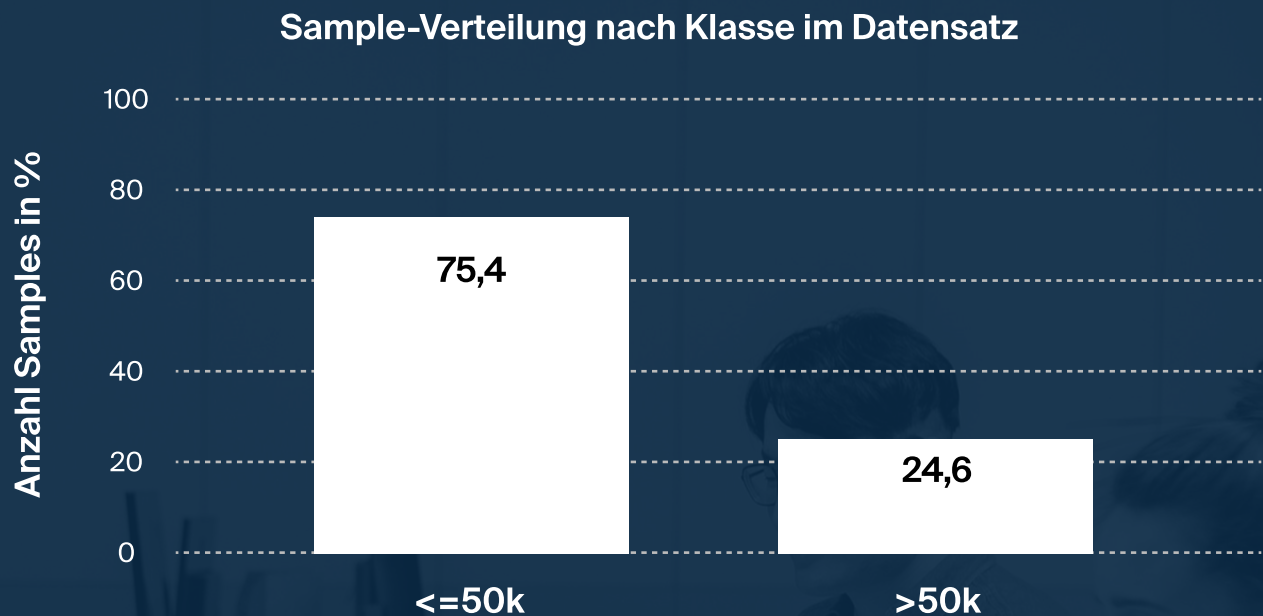
Alleine anhand dieser Kriterien kann bereits festgestellt werden, dass der zugrundeliegende Datensatz stark verzerrt ist und somit zu einem potenziellen Bias in dem KI-System führen kann. Dies schlägt sich in der finalen Bewertung der Datenqualität entsprechend nieder.

---

<sup>1</sup>R. Kohavi. "Census Income," UCI Machine Learning Repository, 1996. [Online]. Available: <https://doi.org/10.24432/C5GP7S>.

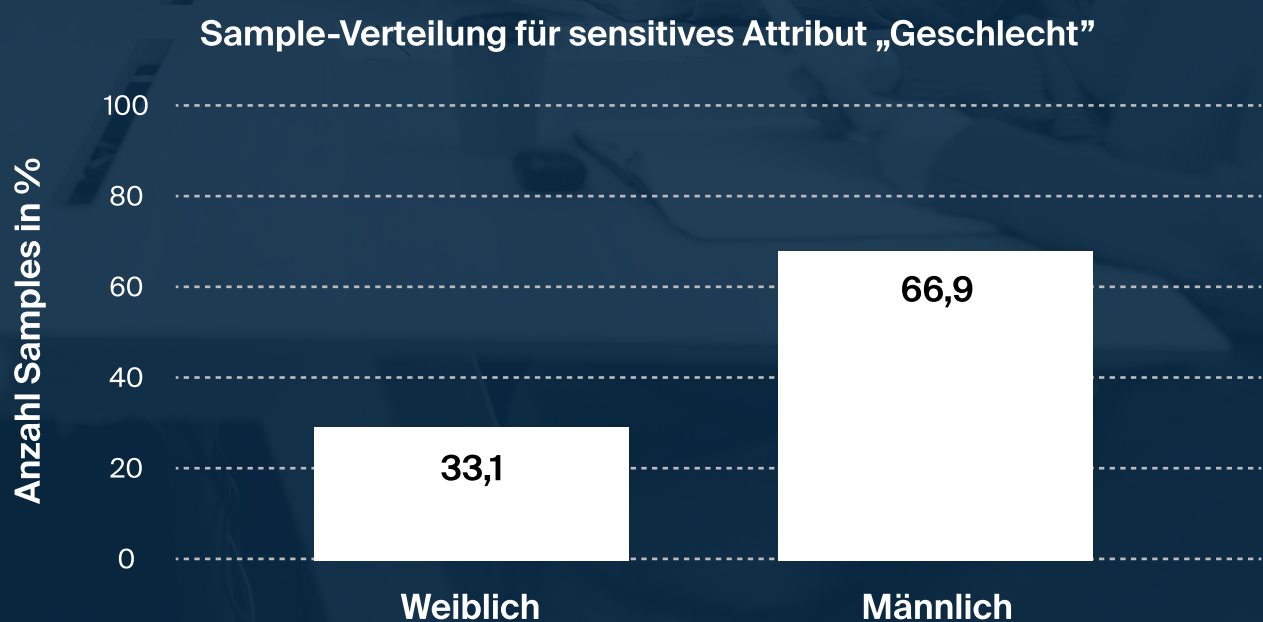
### Abbildung 1: Verteilung der Zielwerte bzw. Klassen im Datensatz

Die Personengruppe mit einem Einkommen über 50.000 USD ist deutlich unterrepräsentiert innerhalb des Datensatzes. Diese Imbalance hat großes Potenzial, einen Bias in dem KI-System zu erzeugen.



### Abbildung 2: Verteilung des Personenmerkmals „Geschlecht“ in dem Datensatz

Männliche Personen sind in dem Datensatz überrepräsentiert, was auch hier Potenzial bietet, zu einem Bias in dem KI-System zu führen.



## 4.2. Fairness und Bias des KI-Systems

Die KI-Verordnung fordert eine umfassende Analyse von Datensätzen zur Erkennung möglicher Verzerrungen (Bias) und zur Sicherstellung der Diskriminierungsfreiheit von KI-Systemen. Konkrete Vorgaben zur technischen Umsetzung macht die KI-Verordnung dabei nicht. Verschiedene technische Definitionen des Fairness-Begriffs in der Fachliteratur bieten jedoch Ansätze für eine fundierte Analyse und Bewertung.

Die Definition der „**Group Fairness**“ bspw. erfordert eine statistische Unabhängigkeit von KI-Systemen in Bezug auf sensible Merkmale wie Geschlecht oder Herkunft.

Auf diese Weise können die Gleichbehandlung unterschiedlicher Personengruppen gewährleistet und strukturelle Diskriminierung verhindert werden. Ergänzend dazu wird die Leistungsfähigkeit des Systems für verschiedene Personengruppen bewertet, um mögliche Benachteiligungen frühzeitig zu erkennen und zu vermeiden.

Für eine fundierte Bias- und Fairness-Analyse sind Annotationen zu geschützten Merkmalen in den Datensätzen erforderlich, die je nach Anwendungskontext ausgewählt werden.

Bei Hochrisiko-Systemen erlaubt die KI-Verordnung (Art. 10 Abs. 5) unter bestimmten Umständen eine Ausnahme von der DSGVO (Art. 9 Abs. 1), sodass besonders geschützte Merkmale unter zusätzlichen Sicherheitsvorkehrungen für die Erkennung und Korrektur von Verzerrungen verarbeitet werden können.

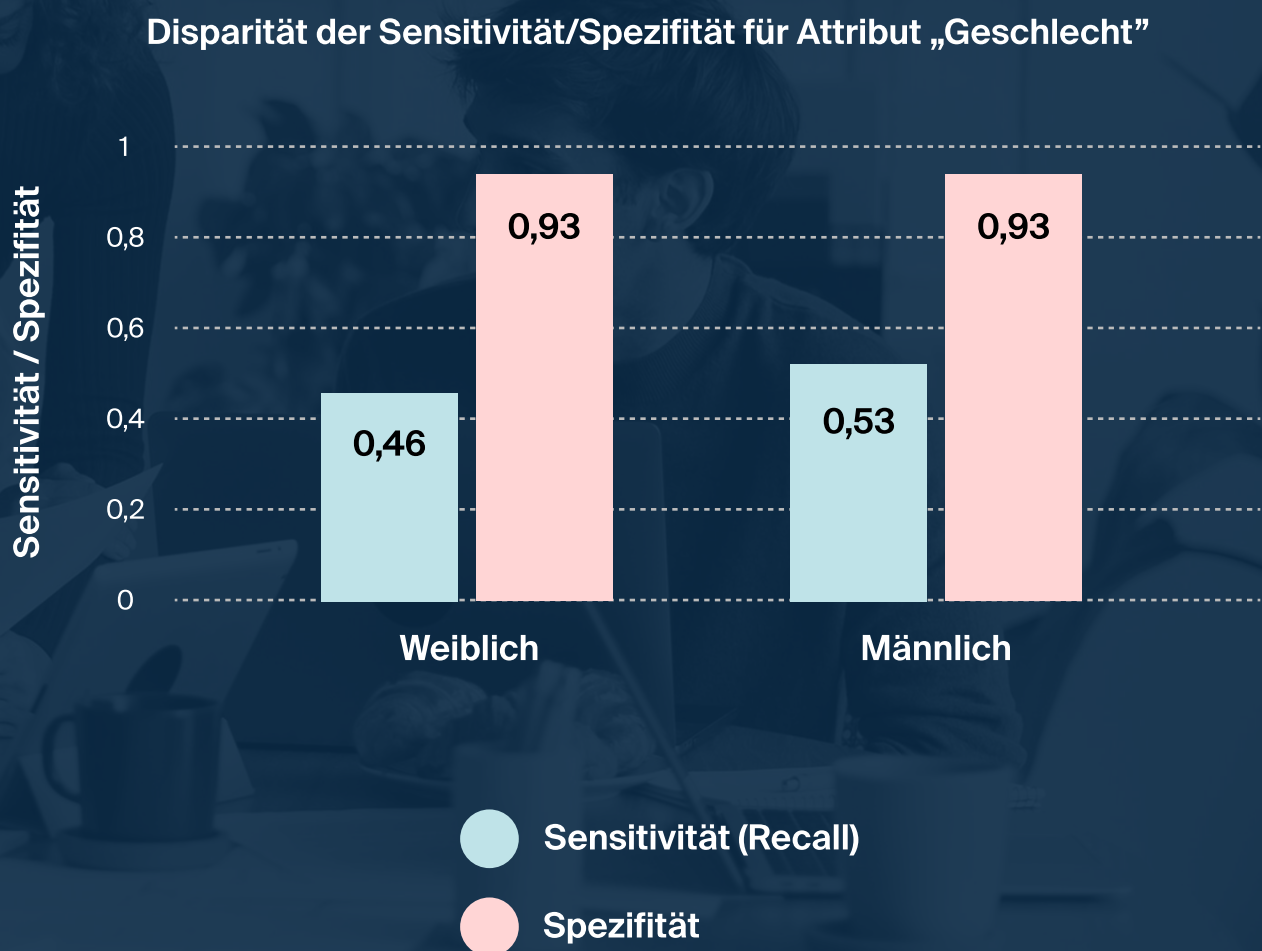
Auf die konkrete Problemstellung bezogen untersuchen wir die Leistungsfähigkeit des Systems gruppiert nach männlichen und weiblichen Personen. Dies ist im Detail in **Abbildung 3** dargestellt.

Das **Ergebnis der technischen Untersuchung auf Group Fairness ist eindeutig**: das KI-System zeigt einen strukturellen Unterschied in der Behandlung männlicher und weiblicher Personen. Dies hat im realen Betrieb erheblichen Einfluss auf die Entscheidungsfindung und verletzt somit die Anforderungen an die Fairness des Systems. Entsprechend muss hier im Entwicklungsprozess des KI-Systems nachgearbeitet werden.

### Abbildung 3: Disparität der Leistungsfähigkeit des KI-Systems gruppiert nach männlichen und weiblichen Personen

Die Sensitivität und Spezifität sind Leistungsmerkmale binärer Klassifikatoren und geben Auskunft über die Fähigkeit eines Systems, Individuen bestimmter Kategorien zu erkennen und voneinander zu differenzieren.

In diesem Fall sagt die Sensitivität aus, wie gut der Klassifikator Personen aus der Einkommensgruppe über 50.000 USD erkennt. Es fällt auf, dass die Bestimmung männlicher Personen über dieser Einkommensschwelle zuverlässiger funktioniert als für weibliche Personen.



## 4.3. Leistungsfähigkeit und Robustheit

Je nach Aufgabenstellung wird die Leistungsfähigkeit eines KI-Systems unterschiedlich bestimmt. In unserem konkreten Kontext der Klassifikation werden üblicherweise folgende Metriken berechnet:

- **Genauigkeit** („Accuracy“ - Anteil der korrekt klassifizierten Fälle)
- **Sensitivität** (Anteil der korrekt identifizierten positiven Fällen)
- **Spezifität** (Anteil der korrekt identifizierten negativen Fällen)

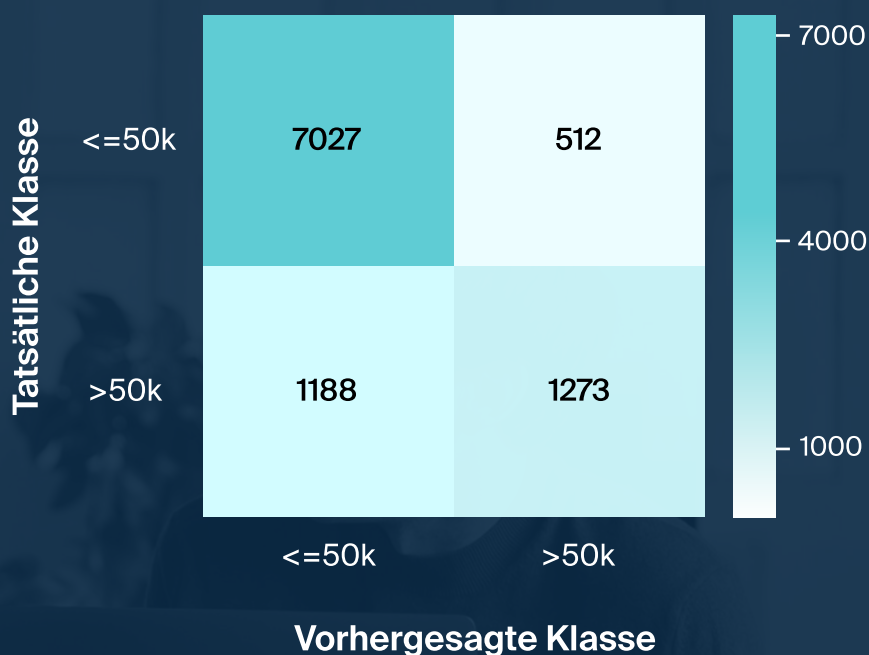
Zusammen betrachtet ermöglichen diese Metriken eine umfassende Analyse und Bewertung der Leistungsfähigkeit von Klassifikationssystemen. Die zu erzielende Leistungsfähigkeit ist allerdings von dem jeweiligen Anwendungsfall abhängig, sodass für jeden Anwendungsfall eine individuelle Bewertung erfolgen muss.

Einen ersten Überblick lässt sich mit Diagrammen wie bspw. einer **Konfusionsmatrix** gewinnen, die in **Abbildung 4** für unser Anwendungsbeispiel berechnet wurde. Hieraus lässt sich direkt ableiten, dass das KI-System tendenziell dazu neigt, Personen allgemein eher in die untere Einkommenskategorie einzusortieren. Dies kann eine direkte Folge der ungleichen Datenverteilung des Datensatzes sein, die sich letztendlich auf die Leistungsfähigkeit des KI-Systems auswirkt (vgl. Abschnitt 4.1). Solche Erkenntnisse müssen entsprechend in der technischen Dokumentation vermerkt und eingeordnet werden.



#### Abbildung 4: Konfusionsmatrix

Eine Konfusionsmatrix stellt für jeden Fall die Vorhersage des KI-Systems über die Zielklasse im Vergleich zu der im Datensatz hinterlegten tatsächlichen Klasse gegenüber. Hieraus können Schlussfolgerungen über die Leistungsfähigkeit des KI-Systems gewonnen werden. Es zeigt sich in diesem Beispiel, dass das KI-System eher zur Vorhersage der unteren Einkommensklasse (unter 50.000 USD) tendiert und somit in seiner Entscheidungsfindung verzerrt ist.



Auch die Robustheit eines KI-Systems muss gemäß Art. 15 KI-VO bewertet werden. Ein zentraler Aspekt ist dabei die Zuverlässigkeit der Unsicherheitsschätzung, die maßgeblich zur Betriebssicherheit beiträgt. Gängige KI-Systeme geben ein Konfidenzmaß an, das eine Selbsteinschätzung ihrer Vorhersagegenauigkeit darstellt.

Laut DIN SPEC 92005:2024 ist eine präzise Unsicherheitsschätzung essenziell, um bei niedriger Konfidenz geeignete Sicherheitsmaßnahmen einleiten zu können.

Zahlreiche Forschungsarbeiten haben jedoch gezeigt, dass moderne KI-Systeme ihr Konfidenzmaß häufig zu hoch einschätzen, was zu einer Abweichung von der tatsächlich beobachteten Fehlerrate führt. Liegt die tatsächliche Fehlerrate höher als die angegebene Konfidenz, spricht man von einer Fehlkalibrierung.

In unserem Beispielsystem beträgt die durchschnittliche Abweichung rund 3,6% – ein akzeptabler Wert, der auf eine zuverlässige Schätzung hinweist.

# KI-Compliance mit der richtigen Systemlandschaft erreichen

Die Umsetzung der KI-Verordnung ist kein rein technisches Projekt – und ebenso wenig eine rein rechtliche Aufgabe. Sie markiert einen umfassenden **Transformationsprozess**, der das **Zusammenspiel von Technik und Recht** erfordert. Denn die Anforderungen der KI-VO lassen sich nur dann erfüllen, wenn **juristische Vorgaben in technisch umsetzbare Maßnahmen** übersetzt werden – und technische Lösungen gleichzeitig **rechtlich abgesichert** sind.

Damit dies gelingt, braucht es ein funktionierendes System von Verantwortlichkeiten, Prozessen und Tools: **KI-Governance**.

Dieser Begriff beschreibt den organisatorischen Rahmen, in dem technische Qualität, rechtliche Konformität und unternehmerische Verantwortung koordiniert und operationalisiert werden. Dabei gilt: Wer KI verantwortungsvoll einsetzen will, muss KI-Governance als strategische Führungsaufgabe verstehen – und frühzeitig den richtigen Tech Stack etablieren.

Der Umfang der notwendigen KI-Governance ergibt sich aus vielen Faktoren: intern etwa durch die Kritikalität der eingesetzten Daten und Systeme, extern durch regulatorische Anforderungen und Erwartungen des Marktes.

## Drei Kategorien von Software-Lösungen für Ihre KI-Governance

Um die Anforderungen der KI-Verordnung effizient zu erfüllen, etablieren sich aktuell drei zentrale Kategorien von Software-Lösungen, die zusammen eine ganzheitliche Systemlandschaft bilden:

### 1. KI-Governance-Plattformen

Diese Plattformen unterstützen Organisationen bei der ganzheitlichen Steuerung von KI-Risiken und -Systemen. Sie bieten Funktionen zur Erfassung, Dokumentation und Klassifikation von KI-Anwendungen sowie zur Überwachung von Verantwortlichkeiten und zur Vorbereitung von Audits. Typische Anforderungen, die hiermit erfüllt werden können:

- Technische Dokumentation (Art. 11, 18 KI-VO)
- Risikomanagement (Art. 9 KI-VO)
- Datenschutz- und Grundrechte-Folgenabschätzung (Art. 26, 27 KI-VO)
- Betriebsanleitung (Art. 13 KI-VO)
- Qualitätsmanagement (Art. 17 KI-VO)

### 2. KI-Modell-Validierungssoftware

Diese Lösungen ermöglichen die technische Bewertung und Validierung von KI-Systemen – etwa in Bezug auf Robustheit, Genauigkeit, Bias oder Unsicherheitsschätzungen.

Hiermit können folgende regulatorische Anforderungen abgedeckt werden:

- Leistungsfähigkeit und Robustheit (Art. 15 KI-VO)
- Fairness-Analysen und Bias-Erkennung (Art. 10 Abs. 2 lit. f–h KI-VO)

### 3. AI-Literacy-Lösungen

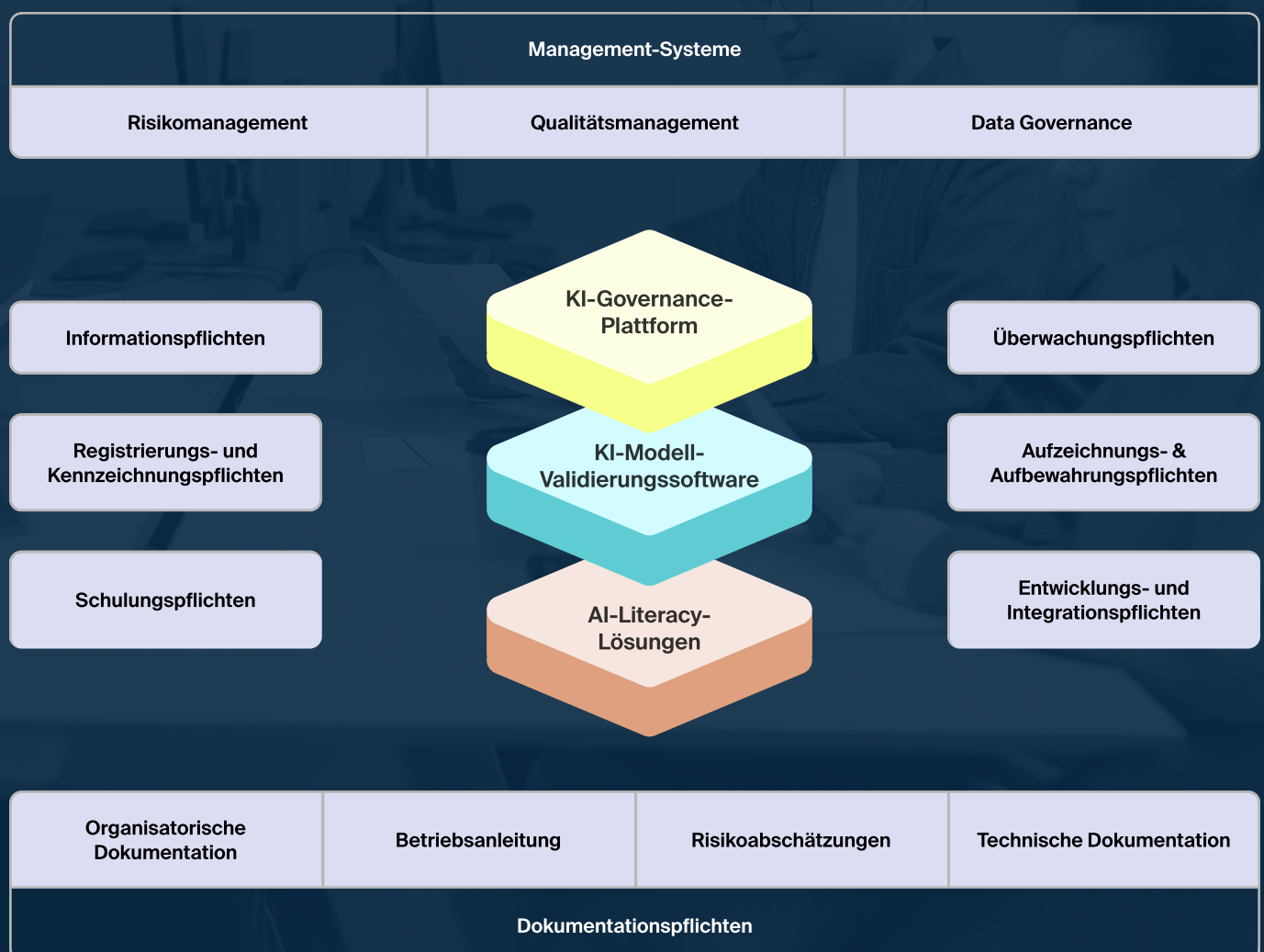
Die KI-VO verlangt in Art. 4 die Sicherstellung ausreichender KI-

Kompetenz bei allen Beteiligten entlang des Lebenszyklus. Dabei helfen AI Literacy Tools wie interaktive Schulungssysteme oder Wissensdatenbanken, diese Anforderung systematisch umzusetzen:

- KI-Kompetenz (Art. 4 KI-VO)
- Awareness-Maßnahmen für Risikobewusstsein und Governance
- Rollenspezifisches Training für Entwickler und Führungskräfte

#### Abbildung 5:

Systemlandschaft für eine effektive und rechtskonforme KI-Governance





## Warum die richtige Tools heute entscheidend sind

Angeichts der Übergangsfristen der KI-Verordnung ist **jetzt der richtige Zeitpunkt**, mit der Umsetzung der KI-Governance zu beginnen.

Wer erst mit Inkrafttreten der Pflichten reagiert, verliert wertvolle Zeit – und riskiert regulatorische wie wirtschaftliche Nachteile. Schon heute sollten Unternehmen aktiv daran arbeiten, die **technischen und organisatorischen Anforderungen der KI-VO systematisch zu operationalisieren**.

Auch wenn sich die beschriebenen Pflichten in erster Linie auf Hochrisiko-KI-Systeme beziehen, ist eine frühzeitige Umsetzung **für alle KI-Anwendungen strategisch klug**: Sie schafft Vertrauen, sichert Qualität und erhöht die Akzeptanz bei Nutzenden, Partnern und Aufsichtsbehörden gleichermaßen.

Die vorgestellten Softwarekategorien für KI-Governance helfen, Komplexität zu beherrschen und Compliance messbar zu machen. Und sie bieten die Grundlage, auf der sich zukünftige regulatorische Anforderungen – national wie international – modular und skalierbar erfüllen lassen.

**Wer KI verantwortungsvoll nutzen will, muss heute mit der Compliance von morgen beginnen.**

## Über die Autoren:

### Leah Klees

Leah Klees ist Unternehmensjuristin bei der caralegal GmbH mit Schwerpunkt auf KI-Governance und Datenschutzrecht. Sie spezialisiert sich darauf, komplexe regulatorische Anforderungen in umsetzbare, praxisnahe Maßnahmen zu übersetzen.



### Dr. Fabian Küppers

Dr. Fabian Küppers ist Experte für Künstliche Intelligenz bei e:fs TechHub GmbH, Gaimersheim. Sein Schwerpunkt liegt hierbei insbesondere auf erklärbaren, sicheren und zuverlässigen KI-Systemen.



caralegal ist ein führender Software-Anbieter für KI-Governance und Datenschutz-Compliance. Die caralegal-Plattform vereint Datenschutz, Risiko-, Audit- und KI-Management in einer intuitiven Lösung und bietet vorgefertigte Workflows zur Erfassung, Klassifizierung und Dokumentation von KI-Systemen.

#### Kontakt:

caralegal GmbH

Am Hamburger Bahnhof 4, 10557 Berlin

E-Mail: [info@caralegal.eu](mailto:info@caralegal.eu)

Website: <https://caralegal.eu>



Thetis ist die zentrale Software-Plattform zur technischen Bewertung und Absicherung von KI-Systemen. Thetis misst und bewertet Leistungsfähigkeit, Robustheit und Sicherheit in einer einzigen Lösung – für die einfache Erfüllung der technischen Anforderungen der KI-Verordnung und die effiziente Erstellung der technischen Dokumentation.

#### Kontakt:

e:fs TechHub GmbH

Dr.-Ludwig-Kraus-Straße 6, 85080 Gaimersheim

E-Mail: [info@efs-techhub.com](mailto:info@efs-techhub.com)

Website: <https://thetis.de/>