

Verantwortungs- voller KI-Einsatz: Risiken verstehen, steuern und versichern

Ein Leitfaden zur Strukturierung und
Bewertung von KI-Risiken.

Inhalt

1	Einleitung: KI als Standarddimension des unternehmerischen Chancen- und Risikomanagements	4
	Abstract	4
	KI – Risiken systematisch verstehen	5
	Was wir unter KI verstehen	5
	Risikopotenziale und Grenzen der Leistungsfähigkeit von KI-Systemen	7
	Etablierung in der praktischen Anwendung	8
2	Governance, Compliance und Steuerung bei KI-Nutzung	9
	EU AI Act: Governance- und Compliance-Anforderungen	9
	Hochrisiko-KI in ausgewählten Anwendungsfeldern – einschließlich agentischer Systeme	10
	Nicht-Hochrisiko-KI mit Compliance-Pflichten	12
	Ausblick: Von Compliance zu »Good AI Governance«	15
	Wie machen es die anderen? Ein Blick ins Ausland	16
3	KI-Sicherheit ist mehr als Cybersicherheit	17
	Einordnung der Thematik	17
	Mögliche Probleme und Herausforderungen	18
	Vorbereitung auf den Einsatz von KI-Agenten	19
	Technische Sicherheitsarchitektur (strukturelle Schutzmaßnahmen)	20
	Wie wird dieser Ablauf konkret in der Praxis umgesetzt?	21
	Wann ist Human-in-the-Loop zwingend nötig?	21
	Kulturelle und mitarbeiterorientierte Anpassung (menschliche Komponente)	22
4	KI-Risiken erkennen, einordnen und bewerten	23
	KI ist ein Technologiesprung mit neuen Chance- und Risikodimensionen	23
	Schadentypen: Versuch einer strukturierten Betrachtung	23
5	Eigenschäden aus KI-Nutzung	25
	Was sind Eigenschäden?	25
	Eigenschäden an dem KI-Asset	25
	KI-induzierte Sachschäden an anderen betrieblichen Vermögenswerten	28

	KI-induzierte Vermögensschäden	29
	Fazit: Eigenschadenrisiko aus KI-Nutzung	32
6	Drittschäden aus KI-Nutzung	34
	Was sind Drittschäden?	34
	Personenschaden	34
	Sachschaden	35
	Reiner Vermögensschaden	35
	Rein vertragliche Haftung / Garantien	36
	Regulatorischer und haftungsrechtlicher Ausgangspunkt	36
	Haftung des KI- Herstellers	37
	Haftung des KI-Nutzers	39
	Betreiber- und Quasi- Herstellerrolle	39
	Deliktische Haftung	39
	Gefährdungs- und Produkthaftung	40
	Vertragliche Haftung	41
	Fazit: Drittschadenrisiko KI	42
7	Governance- und Reputationsrisiken	43
8	Viele Parteien – ein KI-Modell: gemeinsame Absicherung für ein gemeinsames KI – Projekt-Risiko?	46
9	Ausblick	48
	Glossar	50
	Anhang	55
	Literaturverzeichnis	58

1 Einleitung: KI als Standarddimension des unternehmerischen Chancen- und Risikomanagements

Abstract

Künstliche Intelligenz wird zunehmend in betriebliche Prozesse, Entscheidungsstrukturen und Produkte integriert. Inzwischen nutzen bereits 57 Prozent der Unternehmen in Deutschland KI – 2025 waren es noch 36 Prozent, vor zwei Jahren lediglich 20 Prozent.¹ Der zunehmende KI-Einsatz bringt eine veränderte Risikolandschaft mit sich. Die Publikation untersucht, welche Schadenbilder sich durch den Einsatz von KI ergeben können, wie diese systematisch eingeordnet werden können und inwieweit bestehende Versicherungslösungen geeignet sind, die daraus resultierenden Risiken abzudecken.

Ziel der Publikation ist es, ein gemeinsames Verständnis von KI-bezogenen Risiken zu schaffen und eine Struktur für deren Bewertung, Steuerung und Absicherung bereitzustellen. Im Mittelpunkt steht die Perspektive der KI-nutzenden Unternehmen. Die Publikation soll zugleich einen Beitrag zur Grundlagendiskussion in der Versicherungswirtschaft leisten.

Die Analyse zeigt, dass KI in der Regel keine vollständig neuen Schadenarten schafft, sondern bestehende Risiken verändern oder verstärken kann. Für die versicherungstechnische Einordnung ist daher weniger die Technologiebezeichnung als vielmehr das konkrete Schadenbild, die Schadenursache und der Grad der Steuerbarkeit entscheidend. Bestehende Versicherungslösungen können einzelne KI-bezogene Risiken bereits heute erfassen, soweit diese innerhalb des jeweiligen Spartenzuschnitts liegen.

Besondere Herausforderungen bestehen schon heute insbesondere bei unternehmenseigenen reinen Vermögensschäden ohne Sach- oder Personenschadenbezug, Governance- und Compliance-Risiken, agentischen KI-Systemen sowie bei der Zuordnung von Verantwortung in komplexen KI-Wertschöpfungsketten. Gute Governance, klare Verantwortlichkeiten, wirksame Kontrollmechanismen und ein belastbares Risikomanagement werden damit zu zentralen Voraussetzungen für die Beherrschbarkeit und perspektivisch auch die dauerhafte Versicherbarkeit von KI-Risiken.

57%

der Unternehmen in
Deutschland nutzen KI
und 38 Prozent planen
oder diskutieren den
Einsatz.

¹ [Dataverse \(2026\)](#)

¹ [Dataverse \(2026\)](#)

Die Publikation richtet sich insbesondere an Versicherer, InsurTechs, Makler und weitere Marktteilnehmer, die sich mit der Absicherung von KI-Risiken befassen. Angesprochen sind dabei vor allem Fach- und Führungskräfte aus dem Underwriting aller Sparten, Risikomanagement, IT-Sicherheit und Governance.

KI – Risiken systematisch verstehen

Künstliche Intelligenz (KI) entwickelt sich zunehmend zu einer Basistechnologie der digitalen Wirtschaft. Ihr Einsatz betrifft nahezu alle Branchen und Wertschöpfungsstufen. Unternehmen jeder Größenordnung integrieren KI-basierte Anwendungen in ihre Prozesse und Entscheidungsstrukturen, teils in ihre Produkte. Damit ist KI heute kein isoliert zu betrachtender Technologieaspekt mehr, sondern ein wirtschaftsübergreifendes Querschnittsthema.

Der Einsatz von KI eröffnet erhebliche Innovations- und Wachstumspotenziale, bringt zugleich aber auch neue Anforderungen an Risikomanagement, Governance und Absicherung mit sich. Zentrale Voraussetzung für einen kontrollierten und skalierbaren KI-Einsatz ist ein gemeinsames und möglichst allgemein etabliertes Verständnis der relevanten Einsatz- und Risikofelder. Ziel dieser Publikation ist es daher, zur Sensibilisierung für KI als neues Querschnittsrisiko beizutragen, wesentliche aus der Nutzung von KI resultierende Risikokategorien einzuordnen und eine erste Orientierung für die anstehende Diskussion über Risikomanagement und Versicherungsansätze zu schaffen.

Diese Sensibilisierung ist dabei kein Selbstzweck, sondern Voraussetzung für einen verantwortungsvollen, rechtssicheren und innovationsfördernden Einsatz von KI. In diesem Sinne ist die vorliegende Publikation auch als Anregung für eine Grundlagendiskussion in der Versicherungswirtschaft über den zukünftigen Umgang mit KI-bedingten Risiken und deren Versicherbarkeit gedacht.

Was wir unter KI verstehen

Für eine gemeinsame Bewertung von Chancen und Risiken ist ein einheitliches Verständnis dessen erforderlich, was unter Künstlicher Intelligenz zu verstehen ist. In dieser Publikation wird daher an die Definition des EU-KI-Rechtsrahmens (EU AI Act) angeknüpft.²

² Die Definition des EU AI Acts wird hier als Referenz genutzt, weil sie rechtlich verbindlich, technologieoffen und für Governance-, Compliance- und Risikobetrachtungen anschlussfähig ist. Sie stellt insbesondere auf softwarebasierte Systeme mit einem gewissen Grad an Autonomie ab, die Ergebnisse wie Vorhersagen, Empfehlungen, Inhalte oder Entscheidungen erzeugen können. Aufgrund ihrer weiten Fassung können in der Praxis Abgrenzungsfragen entstehen, etwa gegenüber klassischen statistischen Verfahren, regelbasierten Systemen oder einfacher Automatisierung.

»KI-System«: ein maschinengestütztes System, das so konzipiert ist, dass es mit unterschiedlichem Grad an Autonomie betrieben werden kann und nach seiner Einführung Anpassungsfähigkeit zeigt, und das für explizite oder implizite Ziele aus den Eingaben, die es erhält, ableitet, wie es Ausgaben wie Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen generieren kann, die physische oder virtuelle Umgebungen beeinflussen können.

KI entfaltet ihre Stärke insbesondere dort, wo große Datenmengen analysiert, Muster erkannt und daraus Entscheidungen oder Handlungsvorschläge abgeleitet werden müssen (traditionelle KI).

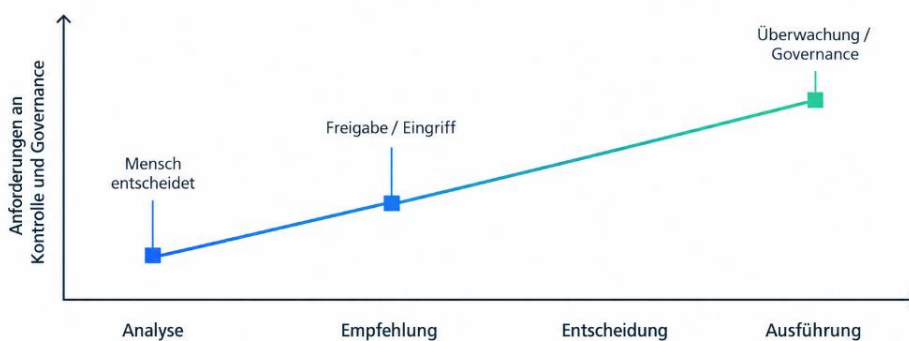
Hiervon zu unterscheiden sind KI-Systeme, die nicht primär bestehende Muster auswerten, sondern auf Grundlage der im Training aus Daten abgeleiteten statistischen Zusammenhänge neue Inhalte, insbesondere Texte, Bilder oder andere Medienformate, erzeugen (generative KI).

Eine weitere Ausprägung bilden KI-Systeme, die eigenständig Ziele verfolgen, Entscheidungen treffen und mehrstufige Aufgaben durch die koordinierte Nutzung von Informationen, Werkzeugen und Anwendungen ausführen können (agentische KI).

Gemeinsam ist diesen Ansätzen, dass sie in hohem Maße leistungsfähig, skalierbar und adaptiv sind, zugleich aber stark von Datenqualität, Modellarchitektur, Zielvorgaben und Kontrollmechanismen abhängen.

Handlungsspielraum und Kontrollanforderung

Mit zunehmendem Handlungsspielraum verändern sich die Anforderungen an die Kontrolle und Governance.



Schematische Darstellung: Mit zunehmendem Handlungsspielraum verändern sich Art und Anforderungen der menschlichen und organisatorischen Kontrolle.

Quelle: Eigene Darstellung in Anlehnung an Sheridan und Verplank (1978).

Risikopotenziale und Grenzen der Leistungsfähigkeit von KI-Systemen

KI-Systeme verfolgen keine eigenen Absichten, sondern handeln zur Erreichung der von Menschen vorgegebenen Ziele. Kommt es dabei zu unerwünschten oder unlauteren Ergebnissen, ist dies regelmäßig nicht Ausdruck eines eigenständigen Zielsystems der KI, sondern Folge unzureichender Vorgaben, Beschränkungen oder Kontrollmechanismen. Daraus ergeben sich insbesondere Risiken durch fehlerhafte Zielvorgaben, unzureichende Überwachung und eine unklare Zuordnung von Verantwortlichkeiten. Um Fehlvorstellungen zu vermeiden, sollte KI daher nicht anthropomorphisiert, also nicht mit menschlichen Eigenschaften, Absichten oder Motivationen versehen werden.

KI ist nicht im menschlichen Sinne »autonom«. Sie verfügt (noch) nicht über ein Verständnis von Verantwortung, ethischen Kontexten oder Normen. Sie ist nicht in der Lage, ethische, rechtliche oder strategische Abwägungen im menschlichen Sinne eigenständig vorzunehmen.

Darüber hinaus arbeitet KI nicht fehlerfrei. Daraus ergeben sich insbesondere folgende Risiken:

- Sie kann falsche Ergebnisse liefern, insbesondere wenn Trainingsdaten unvollständig, verzerrt oder nicht mehr aktuell sind oder wenn sich Fehler und Verzerrungen durch wiederholte Nutzung KI-generierter Daten selbst verstärken (sog. »Habsburger Effekt«).
- KI-Systeme können zudem halluzinieren, also Inhalte erzeugen, die sachlich falsch, frei erfunden oder nicht durch die zugrunde liegenden Daten gedeckt sind, aber dennoch überzeugend und plausibel erscheinen. Halluzinationen entstehen, weil KI-Modelle Wahrscheinlichkeiten berechnen und nicht zuverlässig zwischen wahr und unwahr unterscheiden können. Dies birgt das Risiko, dass fehlerhafte Informationen unbemerkt in betriebliche Prozesse und Entscheidungen einfließen.

Auch kann KI ihre Entscheidungen oftmals nicht oder nur eingeschränkt nachvollziehbar erklären.

Ein weiterer Risikofaktor ist die Skalierbarkeit von KI-Anwendungen: Möglicherweise machen Menschen im Alltag mehr Fehler als KI-Systeme. Allerdings können sich Fehler der KI aufgrund ihrer aufwandsarmen Skalierbarkeit mit hoher Geschwindigkeit und Reichweite besonders stark auswirken. Während menschliche Fehlentscheidungen oft auf Einzelfälle begrenzt bleiben, können fehlerhafte KI-gestützte Entscheidungen oder Inhalte innerhalb kurzer Zeit vielfach reproduziert und angewendet werden. Die besondere Herausforderung liegt daher nicht allein in der Vermeidung einzelner Fehler, sondern insbesondere darin, deren mögliche Auswirkungen zu begrenzen und zu kontrollieren.

Die oben skizzierten Grenzen sind für das Risikoverständnis zentral: Viele KI-bezogene Risiken entstehen nicht aus der Technologie selbst, sondern aus überhöhten Erwartungen, fehlender Kontrolle oder unklaren Verantwortlichkeiten. Ähnlich wie in der Anfangszeit von Tabellenkalkulationen, als Ergebnissen allein aufgrund der professionellen Anmutung eines Excel-Blatts vertraut wurde, besteht auch bei der

Nutzung von KI die Gefahr, Ergebnisse ungeprüft zu übernehmen, nur weil sie technisch überzeugend erscheinen.

Aus versicherungsspezifischer Sicht ist ein Aspekt besonders hervorzuheben: KI erfindet bestehende Risikokategorien in der Regel nicht neu, kann diese jedoch erheblich verstärken. Fehler, Angriffe oder Fehlentscheidungen können durch den Einsatz von KI schneller eintreten, sich leichter skalieren und im Nachhinein schwerer nachvollziehen lassen. KI wirkt in diesem Sinne weniger als eine völlig neue, eigenständige Gefahr, sondern vielmehr als Risikomultiplikator innerhalb bereits bekannter Risikofelder. Diese Perspektive ist für die Bewertung von KI-bezogenen Risiken von grundlegender Bedeutung, weil sie den Blick auf die tatsächliche Wirkung der Technologie lenkt, statt auf pauschale Zuordnungen.

Etablierung in der praktischen Anwendung

Durch ihre Leistungsfähigkeit und Potenziale ist KI heute bereits in zahlreichen betrieblichen Anwendungsfeldern etabliert; ihre Nutzung reicht von der Analyse großer Datenmengen und der Unterstützung von Entscheidungen über die Generierung von Inhalten bis hin zur zunehmend eigenständigen Bearbeitung mehrstufiger Aufgaben durch agentische Systeme. Die Weiterentwicklung insbesondere generativer und agentischer KI erweitert dabei die Bandbreite möglicher Anwendungen kontinuierlich.

Auch bei der Absicherung bestehender Unternehmensrisiken kann KI einen wichtigen Beitrag leisten. Sie kann Auffälligkeiten frühzeitig erkennen, Netzwerk- und Kommunikationsmuster analysieren, Betrugsversuche identifizieren oder Schwachstellen priorisieren und damit helfen, begrenzte Ressourcen gezielter dort einzusetzen, wo der Handlungsbedarf am größten ist. Damit kann KI auch Bestandteil moderner Risiko-, Governance- und Sicherheitsstrukturen sein.

Parallel dazu wird KI zunehmend in operative Kernprozesse integriert – etwa in der Produktionssteuerung, im Kundenservice, in der Vertragsanalyse, im Controlling oder im Zusammenspiel mit externen KI- und Cloud-Dienstleistern. Ihr Einsatz reicht dabei von Assistenz und Empfehlung über Entscheidungsunterstützung bis hin zur teilweisen oder autonomen³ Ausführung von Prozessschritten. In dem Maße, in dem KI zu einem festen Bestandteil der unternehmerischen Infrastruktur wird, gewinnt auch die Frage an Bedeutung, wie ihre Nutzung wirksam gesteuert, kontrolliert und abgesichert werden kann.

³ Siehe Glossar

2 Governance, Compliance und Steuerung bei KI-Nutzung

Das vorstehende Kapitel hat gezeigt, dass KI-Risiken nicht allein aus technischen Fehlfunktionen entstehen. Sie entstehen häufig dort, wo KI-Systeme in Geschäftsprozesse eingebettet werden, Entscheidungen beeinflussen, Verantwortung verteilen oder bestehende Kontrollmechanismen umgehen. Governance, Compliance und Steuerung bilden daher die operative Klammer zwischen Risikoverständnis, Schadenprävention und Versicherbarkeit.

Dieses Kapitel betrachtet deshalb nicht nur die rein regulatorischen Pflichten, sondern auch die Frage, wie KI im Unternehmen so organisiert, dokumentiert und überwacht werden kann, dass Risiken frühzeitig erkannt und beherrscht werden.

EU AI Act: Governance- und Compliance-Anforderungen

Der EU AI Act (Verordnung (EU) 2024/1689; im Folgenden: »EU AI Act«) ist der zentrale Rechtsrahmen für den Einsatz von KI in Europa. Er richtet sich an alle beteiligten Rollen entlang der KI-Wertschöpfungskette:

- **Anbieter** (Developer, Provider) von KI-Systemen,
- **Betreiber/Nutzer** (Deployer), die KI in ihren Geschäftsprozessen einsetzen,
- **weitere Akteure** wie Importeure, Händler und beauftragte Dritte, die das Inverkehrbringen, den Vertrieb, die Integration oder den Betrieb von KI-Systemen unterstützen.

Der EU AI Act verfolgt einen risikobasierten Ansatz:

Risikobasierter Ansatz des EU AI Act

Der EU AI Act reguliert KI-Systeme abhängig von ihrem Risikoniveau:

- **Verbotene KI-Praktiken:** KI-Anwendungen mit inakzeptablem Risiko, deren Einsatz untersagt ist.
- **Hochrisiko-KI-Systeme:** KI-Systeme in besonders sensiblen Anwendungsbereichen, für die umfassende Compliance-Anforderungen gelten.
- **KI-Systeme mit spezifischen Transparenzpflichten:** Bestimmte KI-Systeme, für die insbesondere Informations- und Kennzeichnungspflichten einzuhalten sind.
- **KI-Systeme mit geringem oder minimalem Risiko:** KI-Systeme, für die grundsätzlich keine speziellen regulatorischen Anforderungen nach dem EU AI Act gelten.

Für General-Purpose AI (GPAI)-Modelle gelten zudem eigenständige Anforderungen, die unabhängig von der konkreten Risikoeinstufung des jeweiligen Anwendungsfalls zu beachten sind.

Annex III definiert dabei konkrete Hochrisiko-Anwendungsfelder, u.a. in Kreditwürdigkeit, Beschäftigung, Gesundheitswesen und Versicherung.

Hochrisiko-KI in ausgewählten Anwendungsfeldern – einschließlich agentischer Systeme

Der EU AI Act stuft bestimmte KI-Anwendungen aufgrund ihres Einsatzkontexts und ihrer möglichen Auswirkungen auf Gesundheit, Sicherheit sowie zentrale Grundrechte – insbesondere Menschenwürde, Gleichbehandlung, Diskriminierungsverbot und Privatsphäre – als Hochrisiko ein. Für Unternehmen können insbesondere folgende in Annex III erfasste Anwendungsfelder relevant sein:

- Beschäftigung und Personalmanagement (Annex III Nr. 4) z. B. KI-gestützte Bewerberauswahl, Leistungsbewertung, Aufgabenverteilung, Beförderungs- oder Kündigungsentscheidungen sowie Formen des Algorithmic Management.
- Zugang zu wesentlichen privaten und öffentlichen Dienstleistungen (Annex III Nr. 5). Dazu zählen unter anderem bestimmte Anwendungen zur Kreditwürdigkeitsprüfung natürlicher Personen sowie zur Risikobewertung und Preisbildung in der Lebens- und Krankenversicherung.

- Biometrie (Annex III Nr. 1) z. B. bestimmte Systeme zur biometrischen Identifizierung, Kategorisierung oder Emotionserkennung, soweit sie unter die Hochrisiko-Tatbestände des AI Act fallen.
- Bildung und berufliche Ausbildung (Annex III Nr. 3) z. B. KI-Systeme zur Zulassung, Bewertung von Lernergebnissen oder Steuerung des Bildungszugangs.
- Kritische Infrastrukturen (Annex III Nr. 2) z. B. bestimmte KI-Systeme, die als sicherheitsrelevante Komponenten bei der Steuerung oder dem Betrieb kritischer Infrastruktur eingesetzt werden.
- Strafverfolgung, Migration, Asyl und Grenzkontrolle sowie Rechtspflege und demokratische Prozesse (Annex III Nr. 6–8)

Diese Bereiche sind insbesondere für Unternehmen relevant, die entsprechende Systeme entwickeln, bereitstellen oder im Auftrag öffentlicher Stellen einsetzen.

Agentische KI-Systeme sind nicht bereits aufgrund ihres hohen Autonomiegrads automatisch als Hochrisiko-KI eingestuft. Entscheidend ist, für welchen konkreten Zweck und in welchem Anwendungsfeld sie eingesetzt werden. Fällt ein agentisches System unter einen Hochrisiko-Anwendungsfall des Annex III, gelten für dieses System die einschlägigen Hochrisiko-Anforderungen des EU AI Act. Etwaige zusätzliche Pflichten⁴ für agentische Systeme als reine GPAI-Modelle sind davon rollen- und modellbezogen zu unterscheiden.

Agentische KI-Systeme

Bei agentischen KI-Systemen verschiebt sich die Risikobetrachtung von der reinen Generierung von Inhalten hin zur Ausführung von Handlungen. Entscheidend ist daher nicht die Bezeichnung als »Agent«, sondern der konkrete Autonomiegrad:

- Kann das System nur Vorschläge machen, oder
- kann es eigenständig Daten abrufen, Systeme anstoßen, Kommunikation auslösen oder Prozessschritte durchführen?

Je stärker ein System eigenständig handelt, desto wichtiger werden klare Begrenzungen des Handlungsspielraums, Freigabeprozesse, Protokollierung, menschliche Eingriffsmöglichkeiten und laufendes Monitoring. Werden agentische Systeme in Annex-III-Kontexten eingesetzt, können sie als zu Hochrisiko-KI-Systeme qualifizieren. Sie unterliegen dann sowohl den Modellpflichten für GPAI (Kapitel V) als auch den Systemspezifikationen für Hochrisiko-Systeme.

⁴ Je nach Art des Einsatzgebiet und Form der KI entstehen verschiedene Pflichten (siehe Anahg).

Nicht-Hochrisiko-KI mit Compliance-Pflichten

Auch KI-Systeme unterhalb der Hochrisiko-Schwelle können relevante operative, reputationsbezogene oder regulatorische Risiken erzeugen. Daher sollten auch für kritische Nicht-Hochrisiko-KI-Systeme angemessene Mindestkontrollen vorgesehen werden.

Dazu zählen insbesondere:

- Zweckbeschreibung,
- Verantwortlichkeiten,
- Zugriffsbeschränkungen,
- Dokumentation,
- Schulung,
- Monitoring sowie
- Eskalationswege bei auffälligen Ergebnissen oder Zweckänderungen.

Typische Beispiele:

- Chatbots im Kundenservice und Vertrieb (Transparenzpflicht nach Art. 50 EU AI Act),
- generative KI im Marketing (Kennzeichnungspflichten für synthetische Inhalte nach Art. 50 EU AI Act),
- Fraud-Detection-Modelle ohne unmittelbare Rechtsfolgen für Personen,
- interne Analytiktools, die lediglich Kennzahlen und Reports bereitstellen.

Hier stehen insbesondere folgende Artikel des EU AI Acts im Vordergrund: Art. 5 (Verbotstatbestände), Art. 4 (AI Literacy) und Art. 50 (Transparenz) sowie optional die freiwilligen Codes of Conduct (Art. 95).

Unternehmensweite KI-Governance

KI-Risiken sind in bestehende Steuerungsmechanismen (Risikomanagement, Compliance, Datenschutz, Informationssicherheit) zu integrieren:

- **KI-Verantwortlichkeiten:** Für KI-Systeme sollte eine klare Verantwortlichkeitslogik etabliert werden. Bewährt hat sich eine RACI-orientierte Zuordnung von Rollen und Pflichten entlang des Lebenszyklus eines KI-Systems. Dabei sind insbesondere Vorstand bzw. Geschäftsleitung, Risk Officer/Second Line, System Owner, Fachbereich, IT/Security sowie Legal/Compliance einzubeziehen. Die operative Verantwortlichkeit einzelner Rollen ersetzt nicht die Gesamtverantwortung der Unternehmensleitung, schafft aber Transparenz über Prüf-, Freigabe- und Überwachungspflichten.
- **KI-Policy:** konzernweite Leitlinie, die u. a. regelt
 - zulässige und unzulässige Anwendungsfelder (Art. 5),
 - verbindliche Mindestartefakte je KI-System und insbesondere je KI-Agent. Dazu zählen Zweckbeschreibung, System Owner, Nutzerkreis, Datenquellen, Modell-

/Tool-Komponenten, Freigabeentscheidung, Risikoklassifizierung, Kontrollmaßnahmen, Logging-Anforderungen sowie Monitoring-Verantwortung. Diese Informationen sollten in einem KI-Register bzw. Agenten-Steckbrief dokumentiert und fortlaufend aktualisiert werden, um Scope Creep, Zweckentfremdung und unkontrollierte Weiterentwicklung frühzeitig zu erkennen.

- Kriterien für den Einsatz eigenentwickelter und externer KI,
 - Anforderungen an Daten, Modellvalidierung und Monitoring,
 - Dokumentations- und Freigabeprozesse.
- **Maßnahmen zur Förderung der KI-Kompetenz (AI Literacy) nach Art. 4 EU AI Act**
- Unternehmen sollten angemessene Maßnahmen zur Förderung der KI-Kompetenz (AI Literacy) implementieren. Diese Maßnahmen sollten zielgruppenspezifisch ausgestaltet werden. Führungskräfte benötigen ein Grundverständnis für Einsatzpotenziale, Risiken, regulatorische Pflichten und Entscheidungsgrenzen von KI. Fachanwender benötigen praxisnahe Schulungen zu zulässiger Nutzung, Ergebnisprüfung, Bias-Risiken⁵ und Eskalationswegen. Für agentische KI sind zusätzlich Zugriffsrechte, Verantwortlichkeiten, menschliche Kontrollpunkte und die Grenzen autonomer Ausführung zu adressieren.
 - Schulung der Fachbereiche im Umgang mit KI-Outputs,
 - Sensibilisierung für Bias, Halluzinationen und den »Excel-Effekt« (unreflektiertes Vertrauen in scheinbar präzise Ergebnisse),
 - Einbindung in bestehende Weiterbildungspflichten (z. B. IDD/VersVermV).

Regulatorisch bleiben nationale und europäische Aufsichtsbehörden – etwa die BaFin und die Bundesnetzagentur – zentrale Akteure. Sie ordnen den Einsatz von KI in bestehende Aufsichts- und Risikomanagementanforderungen ein, während die inhaltlich zentralen Anforderungen an Hochrisiko-KI-Systeme und agentische KI-Systeme – Grundrechtsschutz, Transparenz und wirksame menschliche Aufsicht – primär im EU AI Act verankert sind.

Systemspezifische Anforderungen für Hochrisiko KI

Wie bereits oben ausgeführt, sind agentische KI-Systeme nicht allein aufgrund ihrer technischen Einordnung automatisch Hochrisiko-Systeme. Maßgeblich sind Einsatzkontext, Entscheidungswirkung, Autonomiegrad und die tatsächliche Möglichkeit menschlicher Kontrolle. Mit zunehmender eigenständiger Handlungsausführung steigen jedoch die Anforderungen an Risikobewertung, Freigabe, Human Oversight, Logging und Monitoring. Besonders sensibel sind Anwendungen in Bereichen wie Kranken- und Lebensversicherung, HR, Schadenregulierung oder sonstigen Entscheidungsprozessen mit erheblicher Wirkung für Betroffene.

⁵ Siehe Glossar

Ab dem 2. Dezember 2027 müssen Unternehmen, deren Systeme in den Anwendungsbereich von Annex III fallen – einschließlich agentischer Hochrisiko KI – u. a.:

- **ein Risikomanagementsystem (Art. 9)** etablieren, das KI-spezifische Risiken (Fehlentscheidungen, Diskriminierung, Governance-Versagen) laufend identifiziert, bewertet und mitigiert,
- **Daten-Governance (Art. 10)** sicherstellen: Relevanz, Repräsentativität und Bias-Kontrolle der verwendeten Daten,
- **die technische Dokumentation (Art. 11, Annex IV)** für jedes Hochrisiko-System erstellen, inkl. Modellbeschreibung, Datenquellen, Tests, Performancemetriken, Annahmen und Grenzen,
- **Logging (Art. 12)** etablieren, das nachvollziehbar dokumentiert, welche Datenquellen, Prompts, Systemausgaben, Toolaufrufe, Freigaben und Änderungen für ein KI-System relevant waren. Monitoring sollte nicht nur technische Verfügbarkeit erfassen, sondern auch Drift, Zweckabweichung, Auffälligkeiten in Nutzungsmustern, Bias-Indikatoren und Kontrollverletzungen. Nach Produktivsetzung ist ein Post-Market-Monitoring vorzusehen, das Fehlentwicklungen, regulatorische Änderungen und Anpassungsbedarf systematisch nachverfolgt.
- **Transparenz (Art. 13)** herstellen, z. B. durch klare »Instructions for Use« für Underwriter und Claims Handler,
- **Human Oversight (Art. 14)** so gestalten, dass wesentliche Entscheidungen nicht »black-box-artig« vollautomatisch erfolgen, sondern plausibilisiert und bei Bedarf übersteuert werden können,
- **Genauigkeit, Robustheit und Cybersecurity (Art. 15)** definieren, testen und überwachen,
- **ein Qualitätsmanagementsystem (Art. 17)** implementieren, das den gesamten KI-Lebenszyklus (Entwicklung, Roll-out, Änderung, Stilllegung) abdeckt,
- **Fundamental-Rights-Impact-Assessments (Art. 27)** durchführen, wo erforderlich – insbesondere in der Personenversicherung und bei HR-KI,
- **eine Konformitätsbewertung (Art. 43)** durchführen (für die meisten Annex-III-Systeme intern, für bestimmte Biometriesysteme durch eine notifizierte Stelle),
- **Post-Market-Monitoring und Incident-Reporting (Art. 72, 73)** etablieren.

Diese Anforderungen adressieren genau jene Governance- und Steuerungsrisiken, die im Schadentypenmodell zur KI verortet sind: fehlende Validierung, unklare Oversight und unzureichende Dokumentation werden zu expliziten Compliance-Risiken mit Bußgeld- und Haftungsrelevanz.

Ausblick: Von Compliance zu »Good AI Governance«

Kurzfristig wird der EU AI Act überwiegend als Compliance-Projekt in den Unternehmen ankommen. Mittel- und langfristig eröffnet er jedoch Chancen für eine systematische, versicherungstechnische Steuerung von KI-Risiken. Die Entwicklung von reiner »Pflichtenerfüllung« hin zu einer aktiven »Good AI Governance« schlägt sich dabei in drei zentralen, miteinander verzahnten Wirkungsfeldern nieder:

- **Standardisierung:** Gemeinsame Begriffe (KI-Definition), Risikokategorien (Annex III) und Dokumentationsanforderungen (Annex IV) erleichtern die Kommunikation zwischen Technologie, Risikomanagement, Underwriting und Aufsicht. Damit entsteht eine einheitliche »Sprache« für KI-Risiken, auf deren Basis Governance-Prozesse, interne Richtlinien und externe Versicherungsbedingungen konsistent ausgestaltet werden können.
- **Benchmarking:** AI Act Konformität kann zum Differenzierungsmerkmal werden – intern als Effizienz- und Qualitätstreiber, extern in der Zeichnung von KI-Risiken. Unternehmen können ihre KI-Governance im Vergleich zu Marktstandards messen, Aufsichtsanforderungen als Referenz nutzen und darauf aufbauend Entscheidungen zu Risikoappetit, Deckungskonzepten und Prämiengestaltung treffen.
- **Produktinnovation:** Klarere Haftungs- und Governance-Standards schaffen eine belastbare Grundlage für neue Deckungskonzepte – von KI-Eigenschadenlösungen bis zur viel diskutierten KI-»Bauherrenpolice« für Mehrparteien Projekte. Die regulatorisch definierten Rollen, Pflichten und Risikokategorien werden damit unmittelbar in neue Versicherungsprodukte übersetzt, die KI-Risiken nicht nur abdecken, sondern auch aktiv steuern helfen.

Zusammengenommen ermöglichen diese drei Dimensionen, KI nicht mehr nur als schwer greifbares Technologiephänomen zu betrachten, sondern als reguliertes, steuerbares Querschnittsrisiko, das methodisch erfasst und versicherungstechnisch gestaltet werden kann. Der EU AI Act bildet hierfür den verbindlichen Rahmen innerhalb des Governance- und Compliance-Kapitels.

EU AI Act-Compliance ist damit nicht nur eine formale Rechtsfrage, sondern integraler Bestandteil des Querschnittsrisikos KI. Sie beeinflusst, ob KI-bedingte Fehlentwicklungen frühzeitig erkannt, gesteuert oder zumindest haftungs- und reputationsschonend bearbeitet werden können. Als eigenständiges Governance-Risiko umfasst sie mehrere Schadendimensionen:

- **Eigenschäden** durch aufsichtsrechtliche Maßnahmen, Bußgelder von bis zu 7 Prozent des weltweiten Jahresumsatzes (Art. 99) sowie hohe Umstellungs- und Retrofit-Kosten;
- **Drittschäden** durch Haftungsansprüche bei diskriminierenden oder rechtswidrigen Entscheidungen sowie Regress gegenüber KI-Dienstleistern; und
- **Compliance- und Reputationsschäden** in Form von Vertrauensverlust, ESG-Einbußen und D&O-Risiken (etwa im Kontext von »AI Washing«).

Governance- und Reputationsrisiken sind damit keine Randthemen, sondern ein eigenständiger Schadentreiber: Sie verbinden aufsichtsrechtliche Pflichten, persönliche Organhaftung und Marktvertrauen. Ein belastbares Governance- und Dokumentationskonzept entscheidet darüber, ob KI-bedingte Fehlentwicklungen frühzeitig erkannt und begrenzt werden – oder ob sie sich als Bußgelder, Organhaftungsfälle und nachhaltige Reputationsschäden niederschlagen.

Wie machen es die anderen?

Ein Blick ins Ausland

Nachdem der EU AI Act als zentraler Referenzrahmen für KI-Governance und Risikosteuerung im europäischen Kontext beschrieben wurde, stellt sich die Frage, wie andere Rechtsräume an das Thema herangehen und welche Folgerungen sich daraus für international tätige Unternehmen und Versicherer ergeben. Ein Vergleich mit anderen Rechtsräumen - insbesondere den USA und dem Vereinigten Königreich - zeigt, dass die Herangehensweise an AI Governance unterschiedlichen Paradigmen folgt.

In den USA existiert kein einheitlicher KI-Rechtsrahmen auf Bundesebene. KI-Risiken werden innerhalb bestehender Rechtsgebiete (z. B. Wettbewerbsrecht, Verbraucherschutz) adressiert. Die Durchsetzung unterliegt dann wiederum den etablierten Aufsichtsbehörden.

Als zentrale Orientierung für die AI Governance dient das NIST AI Risk Management Framework (AI RMF), das im Einklang mit den »OECD-Prinzipien für vertrauenswürdige KI« auf Transparenz, Fairness, Robustheit und Verantwortlichkeit setzt.

Das Vereinigte Königreich verfolgt ebenfalls einen prinzipienbasierten Ansatz. Bestehende Regulierungsbehörden sollen KI-Anwendungen entlang klar formulierter Leitprinzipien steuern, statt eine eigene umfassende KI-Verordnung nach dem Vorbild des EU AI Act zu schaffen.

Im Ergebnis steht der europäische Ansatz mit dem EU AI Act für einen vergleichsweise präskriptiven, detaillierten Regulierungsrahmen, während die USA und das Vereinigte Königreich stärker prinzipien- und wirkungsorientiert regulieren. Für Unternehmen entsteht damit die Notwendigkeit, die Anforderungen des EU AI Act mit globalen Prinzipienrahmen zu verzahnen, um KI-Risiken grenzüberschreitend konsistent zu bewerten, zu bepreisen und in Deckungskonzepte zu übersetzen.

3 KI-Sicherheit ist mehr als Cybersicherheit

Einordnung der Thematik

Während Governance und Compliance den organisatorischen und regulatorischen Rahmen für den Einsatz von KI bilden, richtet sich der Blick nun auf die konkreten Sicherheitsrisiken von KI-Systemen. Mit der zunehmenden Verbreitung von KI in den

Der Begriff beschreibt die Gesamtheit aller Maßnahmen, die erforderlich sind, um KI-Systeme selbst abzusichern und zugleich die Risiken zu beherrschen, die aus ihrer Nutzung entstehen.

Unternehmen gewinnt dabei nicht nur der Schutz der Systeme vor externen Angriffen an Bedeutung, sondern auch der Umgang mit Risiken, die sich aus den spezifischen Eigenschaften und Funktionsweisen von KI ergeben. Genau an diesem Punkt setzt KI-Security an.

Es geht also sowohl um den Schutz der Technologie als auch um den verantwortungsvollen Umgang mit ihren Ergebnissen.

Der Einsatz von KI ist zunehmend integraler Bestandteil operativer und technischer Kernprozesse von Unternehmen und flächendeckend in der praktischen Anwendung etabliert. Folglich darf sich die Absicherung nicht auf die technische Ebene beschränken. Erforderlich ist vielmehr ein ganzheitlicher Ansatz, der technische, organisatorische und rechtliche Maßnahmen miteinander verbindet und aufeinander abstimmt. Nur so lässt sich sicherstellen, dass die Chancen der Technologie genutzt werden können, ohne die damit verbundenen Risiken aus dem Blick zu verlieren.

Zu einem solchen ganzheitlichen Ansatz gehören insbesondere:

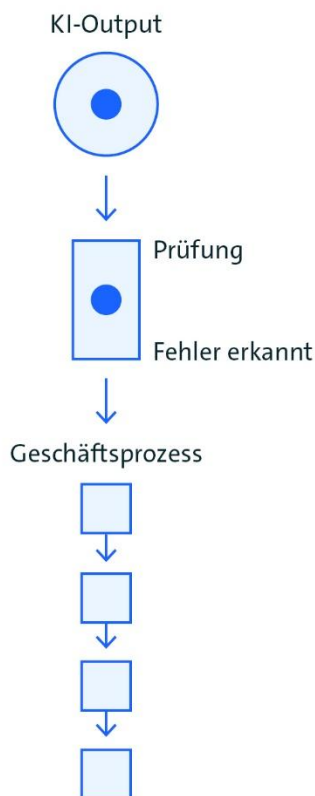
- klare Verantwortlichkeiten für KI-Anwendungen,
- dokumentierte Einsatzbereiche,
- definierte Kontroll- und Freigabeprozesse,
- der Schutz sensibler Daten,
- belastbare vertragliche Regelungen mit externen Dienstleistern,
- ein laufendes Monitoring der Systeme,
- sowie definierte Notfall- und Fallback-Prozesse.

Mögliche Probleme und Herausforderungen

So groß der Nutzen von KI ist, so vielfältig sind die Herausforderungen, die mit ihrem Einsatz einhergehen. Ein zentrales Problem liegt in der mangelnden Transparenz vieler Systeme. Die Entscheidungen oder Ergebnisse von KI-Anwendungen sind nicht immer vollständig erklärbar, weil die zugrunde liegenden Modelle komplex und in ihrer Wirkungsweise nicht ohne Weiteres nachvollziehbar sind. Das erschwert die Kontrolle, die klare Zurechnung von Verantwortlichkeiten und im Schadenfall auch die Ursachenanalyse.

Hinzu kommt das Risiko fehlerhafter oder ungeprüfter KI-Ergebnisse. Übernehmen Unternehmen die Ausgaben eines KI-Systems ungefiltert, können daraus operative Fehlentscheidungen, fehlerhafte Prozesse oder unmittelbare wirtschaftliche Schäden entstehen. Wie bereits beschrieben, verschärft sich dies bei direkter Einbindung in

A Kontrollpunkt vorhanden



B Kontrollpunkt vorhanden



● Kontrollpunkt

Schematische Darstellung: Kontrolle begrenzt die Fortpflanzung fehlerhafter KI-Ergebnisse in geschäftskritischen Prozessen. Ein Kontrollpunkt kann fehlerhafte Ergebnisse vor ihrer weiteren Verarbeitung erkennen oder zurückhalten; ohne Kontrollpunkt können sich Fehler im Prozess fortpflanzen. Die Anzahl der dargestellten Vorgänge bildet keine empirische Mengenrelation ab.

Eigene Darstellung auf Basis des Publikationstextes.

geschäftskritische Abläufe, Fehler pflanzen sich dann schnell und über ein großes Transaktionsvolumen fort, bevor sie erkannt werden.

Ein weiteres Risiko betrifft den Umgang mit Daten. KI-Systeme benötigen häufig große Datenmengen, um verlässliche Ergebnisse zu liefern. Werden dabei sensible Informationen verarbeitet oder an externe Anbieter übertragen, stellen sich unweigerlich Fragen zu Datenschutz, Vertraulichkeit, Zugriffskontrolle und Verantwortlichkeit. Der Schutz dieser Daten ist nicht nur eine rechtliche Pflicht, sondern zugleich eine Voraussetzung für das Vertrauen von Kunden, Mitarbeitenden und Geschäftspartnern.

Auch die Abhängigkeit von externen KI- oder Cloud-KI-Dienstleistern kann sich als problematisch erweisen. Sicherheitsvorfälle, Ausfälle oder Fehler beim Anbieter können sich unmittelbar auf das nutzende Unternehmen auswirken, ohne dass dieses die Ursache selbst kontrollieren oder unmittelbar beheben könnte. Die Auslagerung von KI-Funktionen verlagert damit einen Teil des operativen Risikos nach außen, ohne die Verantwortung für die eigenen Prozesse aufzuheben.

Pauschale Bewertungen greifen an dieser Stelle regelmäßig zu kurz. KI-Risiken müssen entlang der konkreten Nutzung, der Einbindung in die Prozesse, der Datenflüsse, der gewählten Sourcing-Modelle und der vorhandenen Kontrollmechanismen beurteilt werden. Erst diese differenzierte Betrachtung erlaubt eine belastbare Einschätzung des tatsächlichen Risikos und eine sachgerechte Zuordnung, z. B. zu bestehenden Versicherungslösungen und Deckungskonzepten.

Vorbereitung auf den Einsatz von KI-Agenten

Mit der zunehmenden Autonomie von KI-Systemen, insbesondere in Gestalt sogenannter KI-Agenten⁶, die eigenständig Aufgaben planen und ausführen, gewinnt die Frage der Vorbereitung des Unternehmens auf ihren Einsatz zusätzlich an Bedeutung.

Die Grundlagenarbeit für den Einsatz von KI und insbesondere von Agenten lässt sich in drei zentrale Dimensionen gliedern. Bemerkenswert ist dabei, dass sich der Fokus deutlich verschiebt: weg von einer rein technologischen Betrachtung, hin zu einer verhaltens- und prozessbasierten Strategie, die den gesamten Lebenszyklus des Systems in den Blick nimmt.

Im vorherigen Kapitel wurde auf den Governance-Aspekt eingegangen. Darüber hinaus sind die technische Sicherheitsarchitektur und die kulturelle und mitarbeiterorientierte Anpassung im Gesamtsicherheitskonzept zu betrachten. Diese zwei Dimensionen werden nachfolgend im Detail betrachtet.

⁶ Siehe Glossar

Technische Sicherheitsarchitektur (strukturelle Schutzmaßnahmen)

Im Vergleich zu klassischen IT-Schwachstellen (CVEs), die primär reaktiv durch Patches behoben werden, erfordert der Schutz vor nicht ausreichend kontrollierten KI-Agenten eine proaktive und kontinuierliche Überwachung, da das Risiko hier nicht nur im Code, sondern im unberechenbaren, autonomen Verhalten der KI liegt.

Die technischen Schutzmaßnahmen zur Absicherung von KI-Agenten lassen sich in verschiedene Ansätze gliedern, die jeweils spezifische Sicherheitsvorteile bieten, jedoch auch organisatorische Herausforderungen mit sich bringen:

Ein zentraler Baustein ist das Prinzip der minimalen Rechtevergabe (Least Privilege)⁷. Hierbei erhalten KI-Agenten nur exakt diejenigen Zugriffsrechte, die sie für ihre spezifische Aufgabe benötigen, während administrative Vollrechte strikt vermieden werden. Das minimiert den potenziellen Schadensradius bei Fehlverhalten erheblich. Allerdings erfordert dieser Ansatz einen hohen administrativen Aufwand und eine detaillierte Kenntnis darüber, welche Systeme welche Informationen beinhalten. Das heißt, dass Berechtigungsstrukturen kontinuierlich angepasst werden müssen.

Parallel dazu sichert eine robuste Authentifizierung jeden Interaktionsstrang ab. Durch standardisierte Protokolle lässt sich jede Aktion eindeutig zuordnen, was die Nachvollziehbarkeit und Compliance stärkt. Ein lückenloser Audit-Trail ist die Folge. Man muss jedoch abwägen, ob die dadurch steigende Systemkomplexität oder eine minimale Latenzerhöhung im Betrieb akzeptabel sind.

Um die Kerninfrastruktur direkt vor dem Fehlverhalten von Agenten zu schützen, bietet sich Sandboxing⁸ an. Der Agent agiert dabei in einer isolierten Umgebung ohne freien Zugriff auf das restliche Netzwerk. Das schirmt das System effektiv vor unvorhersehbarem Verhalten ab, kann aber die Performance drosseln oder notwendige Interaktionen mit anderen Systemen erschweren.

Ergänzt wird dies durch KI-spezifische Sicherheitswerkzeuge wie Firewalls⁹ und Prompt-Injection-Filter. Diese scannen Ein- und Ausgaben in Echtzeit auf bösartige Muster und unterbinden unvorhergesehene Aktionen durch spezialisierte Kontroll-Frameworks. Sie bieten aktiven Schutz gegen neuartige Angriffsvektoren. Da sich Angriffsmuster ständig weiterentwickeln, sind regelmäßige Updates zwingend erforderlich, um auch das Risiko von fälschlicherweise blockierten, legitimen Anfragen zu minimieren.

Der Mensch ist die letzte Verteidigungslinie. Die oben skizzierten Maßnahmen bilden die »harten« technischen Leitplanken. Da KI-Risiken oft dynamisch sind, empfiehlt sich ergänzend dazu der Aufbau prozessualer Kontrollschleifen¹⁰ (»Human-in-the-Loop« / [HITL]), um kritische Entscheidungen menschlich zu validieren.

⁷ Siehe Glossar

⁸ Siehe Glossar

⁹ Siehe Glossar

¹⁰ Siehe Glossar

Im Kontext von KI-Agenten ist der Mensch Teil einer Sicherheits- und Qualitätskontrollstrategie, bei der die KI nicht völlig autonom entscheidet und handelt, sondern an kritischen Punkten des Prozesses auf eine explizite menschliche Zustimmung angewiesen ist.

Dieses Vorgehen stellt sicher, dass die KI nur in vorhergesehener Art und Weise agiert und damit in ihren Handlungen jederzeit kontrollierbar bleibt. Statt einer vollautomatischen Ausführung erfordert jeder kritische Schritt eine explizite Bestätigung durch verantwortliche Mitarbeitende. Dies dient als essenzieller Prüfpunkt.

Wie wird dieser Ablauf konkret in der Praxis umgesetzt?

In einer Architektur für KI-Agenten wird »Human-in-the-Loop« meist durch sogenannte »Approval Workflows« (Freigabeschleifen) umgesetzt. Der Prozess sieht dann meistens so aus:

Zunächst erstellt der Agent in der Planungsphase einen Entwurf für seine nächste Aufgabe. Doch bevor er diesen ausführt, greift ein automatischer Interrupt und versetzt den Prozess in einen Wartemodus. Die zuständigen Mitarbeitenden werden sofort benachrichtigt und erhalten eine Zusammenfassung zur Prüfung. Sie halten nun alle Fäden in der Hand: Sie können die Aktion direkt freigeben, sie bei Bedenken komplett ablehnen oder den Entwurf manuell korrigieren, bevor der Agent fortfahren darf.

»Human-in-the-Loop« (HITL) ist die letzte Verteidigungslinie (die »Sicherheitsbremse«), falls alle technischen Maßnahmen (Sandboxing, Firewalls) versagt haben sollten.

HITL fungiert somit als entscheidende Sicherheitsbremse. Da die rechtliche Verantwortung, etwa im Sinne des EU AI Act, ohnehin beim Menschen verbleibt, sichert dieser Prozess die notwendige menschliche Verantwortlichkeit. Zudem erkennen menschliche Prüfende logische Fehler oder Kontextprobleme – wie etwa falsche Empfängerlisten – deutlich zuverlässiger als die KI selbst. Letztlich stärkt diese Form der Kontrolle auch das Vertrauen der Belegschaft in die neue Technologie.

Wann ist Human-in-the-Loop zwingend nötig?

Man unterscheidet oft zwischen verschiedenen »Freigabestufen«: Die Intensität der Kontrolle richtet sich dabei nach der jeweiligen Risikointensität. Bei risikoarmen Aufgaben wie reiner Recherche kann oft auf HITL verzichtet werden. In Bereichen mit mittlerem Risiko reicht häufig eine aktive Überwachung über Dashboards aus, bei der man im Notfall eingreifen kann. Bei kritischen Aktionen hingegen ist eine explizite Zustimmung zwingend erforderlich. Das gilt insbesondere für das Senden von Massen-E-Mails, Änderungen an Datenbanken, Finanztransaktionen oder die Veröffentlichung von Unternehmensinhalten nach außen.

Doch wie auch oben schon beschrieben, bedarf es der stetigen Überprüfung, Hinterfragung und ggf. Adaption des HITL-Gesamtkonzeptes. Genau wie die Agenten

lernen und wachsen, müssen auch die Menschen immer weiter lernen und ihre Kontrollmaßnahmen kontinuierlich an die Entwicklung der KI-Systeme anpassen.

Kulturelle und mitarbeiterorientierte Anpassung (menschliche Komponente)

Wie bereits bei der technischen Architektur und dem Human-in-the-Loop-Ansatz näher erläutert, ist die menschliche Komponente häufig der entscheidende, zugleich aber am leichtesten unterschätzte Faktor. Auch die beste technische Absicherung bleibt wirkungslos, wenn Mitarbeitende nicht in der Lage sind, KI-Systeme verantwortungsvoll zu bedienen und deren Ergebnisse kritisch einzuordnen. Daher ist es notwendig, ein kontinuierliches Training der Belegschaft einzuführen, nicht als einmalige Schulung, sondern als fortlaufender Prozess, der mit der Weiterentwicklung der Technologie Schritt hält.

Ein wesentliches Ziel muss es sein, ein tragfähiges Bewusstsein (Awareness) für den Umgang mit KI zu schaffen. Dazu gehört insbesondere eine bewusste und restriktive Rechtevergabe bei der Nutzung von Agenten: Beschäftigte sollten verstehen, welche Berechtigungen sie einem Agenten einräumen und welche Konsequenzen dies nach sich ziehen kann.

Ebenso wichtig ist die Stärkung der Eigenkompetenz, damit keine übermäßige Abhängigkeit von KI-Ergebnissen entsteht und die Fähigkeit zur eigenständigen Beurteilung erhalten bleibt. Transparenz über den Einsatz von KI sowie klare, im Arbeitsalltag verankerte Ethikrichtlinien vervollständigen diese Dimension und machen den verantwortungsvollen Umgang mit KI zu einem selbstverständlichen Bestandteil der Unternehmenskultur.

Im Ergebnis ist es stets der Mensch, der die Aussagen und Entscheidungsvorlagen der KI bewerten und einordnen können muss.

4 KI-Risiken erkennen, einordnen und bewerten

KI ist ein Technologiesprung mit neuen Chance- und Risikodimensionen

Zwar handelt es sich bei KI um eine Technologie und damit grundsätzlich um ein Technologierisiko. Gleichzeitig greift KI jedoch unmittelbar in kognitive Prozesse ein, indem sie Bewertungen, Prognosen und Empfehlungen vorbereitet oder automatisiert Entscheidungen trifft. KI wirkt damit nicht primär auf Systeme oder Infrastruktur, sondern auf die unternehmerischen Entscheidungsgrundlagen selbst.

Im Unterschied zu klassischen IT-Risiken führt ein Fehler in KI-Systemen häufig nicht zu einem technischen Ausfall, sondern zu inhaltlich fehlerhaften, verzerrten oder nicht nachvollziehbaren Aussagen, Empfehlungen oder Entscheidungen. Diese können operative Abläufe, Geschäftsstrategien oder externe Wirkungen beeinflussen, ohne dass ein klarer technischer Defekt erkennbar ist. KI ist damit zugleich »Maschine« und »Denkinstrument« – technologisch implementiert, aber kognitiv wirksam.

Gerade diese hybride Natur macht KI-Risiken besonders: Sie liegen irgendwo zwischen Technologie, Organisations- und Haftungsfragen, ohne vollständig in einer dieser Kategorien aufzugehen. KI erfordert eine eigenständige Betrachtung, die technische, menschliche und organisatorische Faktoren gleichermaßen berücksichtigt.

Schadentypen: Versuch einer strukturierten Betrachtung

Zur Einordnung KI-bezogener Risiken liegen bereits verschiedene Ansätze vor, die mögliche Schäden bestehenden Versicherungssparten zuordnen. Eine solche tabellarische Betrachtung ist beispielhaft im Anhang dargestellt. Die vorliegende Publikation wählt ergänzend einen spartenunabhängigen Ansatz, der beim Schaden selbst ansetzt und damit eine übergreifende Systematisierung ermöglicht.

Im Folgenden werden KI-bezogene Risiken systematisch identifiziert und in »Grundschatentypen« eingeordnet. Ziel ist es, eine belastbare Struktur zu schaffen, auf deren Basis Risiken vergleichbar bewertet und diskutiert werden können. Der Fokus liegt dabei zunächst auf einer analytischen Ordnung möglicher Schäden, weniger auf der Bewertung ihrer Relevanz in möglichen oder wahrscheinlichen Frequenzen und Volumina.

KI-Risiken lassen sich nicht entlang einer einzelnen Achse erklären – sie entstehen im Zusammenspiel von Verantwortungssphäre, Geschädigten und Schadensart.

Die Einordnung erfolgt entlang mehrerer Dimensionen.

■ **Geschädigte:** Die Ebene »Eigen- oder Drittschaden«

Als zentrale Struktur erscheint eine Differenzierung nach Eigen- und Drittschäden sinnvoll.

Neben der Differenzierung zwischen Eigen- und Drittschäden wird hier eine zusätzliche Kategorie für Governance, Compliance und Steuerungsrisiken eingeführt. Diese umfasst Risiken, die nicht primär durch einen klassischen Schadenseintritt, sondern durch Defizite in Organisation, Aufsicht, Verantwortungszuordnung oder regulatorischer Umsetzung entstehen. Dazu zählen insbesondere Risiken aus behördlichen Verfahren, aufsichtsrechtlichen Maßnahmen, Ordnungswidrigkeiten oder Bußgeldern. Die Ergänzung dieser Kategorie trägt dem Umstand Rechnung, dass KI-Risiken häufig nicht in formellen Schäden, sondern in Governance-Versagen, Verantwortungsdiffusion oder mangelnder Steuerbarkeit sichtbar werden und dadurch eine eigenständige risikorelevante Wirkung entfalten.

■ **Schadensart:** Die Ebene »Sach- oder Personenschaden, Vermögensfolgeschaden oder reiner Vermögensschaden«

Schließlich wird nach der Art des möglichen Schadens weiter unterschieden, insbesondere entlang der Kategorien Sach- und Personenschäden sowie Vermögensschäden. Ein besonderes Augenmerk liegt dabei auf sog. »reinen Vermögensschäden«, die im Kontext von KI-Anwendungen eine wachsende Rolle spielen und versicherungstechnisch besondere Herausforderungen mit sich bringen.

■ **Verantwortungssphäre:** Die Ebene »Nutzer oder Hersteller«

Zunächst ist zwischen Risiken auf Nutzer und Herstellerebene zu unterscheiden. In dieser Arbeit stehen bewusst die Risiken aus Sicht der Anwender von KI-Systemen im Vordergrund. Damit wird die Perspektive eingenommen, die für die meisten Unternehmen von unmittelbarer Relevanz ist, da sie KI überwiegend nutzen und nicht selbst entwickeln.

Herstellerbezogene Risiken werden, dort, wo erforderlich, kurz gesondert betrachtet, da sie anderen Verantwortlichkeiten und Risikologiken unterliegen.

5 Eigenschäden aus KI-Nutzung

Was sind Eigenschäden?

Schadenfälle, die bei einem Unternehmen unmittelbar zu einem Nachteil an ihrem eigenen Vermögen führen, werden als Eigenschäden bezeichnet. Hierzu gehören aus versicherungstechnischer Sicht

- Sachschäden, also die Beschädigung, Zerstörung und teils auch das Abhandenkommen eigener Sachwerte,
- Vermögensschäden, also schadenbedingt verringerte Einnahmen, erhöhte Aufwendungen und Kosten oder Vermögensabflüsse – und zwar als Folge eines Sach- oder Personenschadens, aber auch als sog. »reiner Vermögensschaden« ohne vorausgehenden Sach- oder Personenschaden.

Personenschäden werden im folgenden Blick auf Eigenschäden nicht näher betrachtet. Zwar sind solche Schäden auch im Zusammenhang mit dem Einsatz von KI denkbar, in der Unternehmenspraxis stehen hierzu jedoch regelmäßig arbeits- und sozialversicherungsrechtliche Fragestellungen im Vordergrund, die den Rahmen dieser Betrachtung verlassen würden.

Eigenschäden an dem KI-Asset

In der allgemeinen Diskussion um KI stehen häufig Schäden im Vordergrund, die durch den Einsatz von KI verursacht werden – also mit dem Fokus auf der KI als Schadenverursacher. Weniger beachtet wird der Teilaspekt, dass Schäden auch an einer KI-Anwendung selbst eintreten können. KI-Modelle und KI-Anwendungen können beschädigt, verändert, gelöscht oder unbrauchbar werden. Der Schaden liegt in solchen Fällen in der vollständigen oder teilweisen Beeinträchtigung des KI-Assets. Schäden an KI-Assets¹¹ können grundsätzlich zwei Varianten zugeordnet werden: erstens dem KI-Modell selbst und seinen erforderlichen Daten- und Entwicklungsbeständen sowie zweitens der Nutzbarkeit des KI-Systems.

Diese Überlegung lässt sich auf KI-Modelle und KI-Anwendungen nur eingeschränkt übertragen. KI-Modelle, ihre Parameter, Trainingsstände, Konfigurationen und Anwendungslogiken sind als solche keine Sachen im Sinne des § 90 BGB, da ihnen die für den zivilrechtlichen Sachbegriff erforderliche Körperlichkeit fehlt. Soweit im Einzelfall zugleich ein Speichermedium oder sonstige Hardware physisch beschädigt wird, kann daneben selbstverständlich ein Sachschaden vorliegen. Die Beeinträchtigung des KI-Modells selbst – etwa durch Manipulation, Löschung, Verschlechterung von Modellparametern, Verlust von Trainings- oder Wiederherstellungsdaten oder durch eine Einschränkung seiner Verfügbarkeit – ist jedoch regelmäßig als Vermögensschaden einzuordnen.

¹¹ Siehe Glossar

Im Ergebnis handelt es sich bei KI-Modellen und KI-Anwendungen um wirtschaftlich werthaltige digitale Vermögenswerte, deren Beeinträchtigung erheblichen Wiederherstellungsaufwand auslösen kann. Im Mittelpunkt steht daher nicht die Frage nach einem Sachschaden, sondern danach, ob das KI-Modell selbst, seine Wiederherstellungsgrundlagen oder seine Verfügbarkeit beeinträchtigt worden sind und welche Maßnahmen erforderlich sind, um den ursprünglichen Zustand wiederherzustellen.

■ **Variante 1: Beeinträchtigung des KI-Modells und / oder seiner Daten- und Entwicklungsstände**

Diese Kategorie umfasst Schäden am KI-Modell selbst, z. B. den Verlust, die Beschädigung oder die Manipulation trainierter Modellparameter, Fine-Tunings, Konfigurationen oder Sicherheitsmechanismen sowie Schäden an Trainings-, Validierungs- und Testdaten, Embeddings, Wissensdatenbanken, Prompt-Bibliotheken, Dokumentationen sowie sonstigen Entwicklungsartefakten. Die Beeinträchtigung kann zugleich zu einer vollständigen oder teilweisen Nichtnutzbarkeit des Modells führen.

Beispiel: Die Parameter eines speziell für die Betrugserkennung trainierten Modells gehen verloren. Das Modell muss neu trainiert werden.

Beispiel: Eine Vektordatenbank mit unternehmensspezifischem Fachwissen wird beschädigt. Das KI-System kann zwar weiterhin betrieben werden, liefert jedoch unvollständige oder fehlerhafte Antworten.

■ **Variante 2: Beeinträchtigung der Verfügbarkeit externer KI-Leistungen**

Diese Kategorie erfasst Fälle, in denen das KI-Modell oder die zugrunde liegende KI-Infrastruktur nicht im Einflussbereich des Unternehmens liegt und deshalb nicht selbst wiederhergestellt werden kann. Betroffen ist nicht die Substanz eigener KI-Assets, sondern die Verfügbarkeit einer von Dritten bereitgestellten KI-Leistung.

Beispiel: Ein KI-System basiert auf einem externen Large Language Model¹² eines Drittanbieters. Nach einem Ausfall der Modellplattform oder der Sperrung des Zugangs kann das Unternehmen die KI-Anwendung nicht mehr nutzen. Modell, Daten und Prozesse des Unternehmens bleiben unverändert, das System ist jedoch zur Nutzung nicht verfügbar.

Liegt die Behebung des Schadens innerhalb des Verantwortungs- und Kontrollbereichs des betroffenen Unternehmens, fallen Aufwendungen für Rekonstruktion, Wiederherstellung oder das erneute Training des KI-Systems an.

Cyberversicherungen können Schäden an unternehmenseigenen KI-Systemen abdecken, sofern der Schaden auf einem Cyberangriff gemäß der Definition des jeweiligen Versicherungsprodukts zurückzuführen ist. Das KI-Modell sowie seine Daten- und Entwicklungsbestände sind in diesen Fällen als Bestandteil der betroffenen

¹² Siehe Glossar

IT-Infrastruktur versichert. Ersetzt werden u. a. die Wiederherstellungskosten für Daten und Systeme und ein möglicher Betriebsunterbrechungsschaden.

Beruhet der Schaden an dem KI-Asset auf einem anderen konkreten Schadenauslöser, können tradierte Versicherungssparten (z. B. Feuerversicherung, Maschinenversicherung) greifen, sofern die Schadenursache eine dort versicherte Gefahr darstellt (z. B. Feuer, Maschinenbruch) und das in diesen Sparten typische Erfordernis des Sachschadens an einer versicherten Sache, deren Bestandteil das KI-System Bestandteil ist, erfüllt wird, es sich also z. B. nicht lediglich um einen reinen Mangel ohne physische Substanzänderung handelt. Auch die Frage, ob der Schaden »unvorhergesehen« eingetreten ist, kann hier – insbesondere in den technischen Versicherungen – eine wichtige Rolle spielen.

Schäden an KI-Assets können jedoch auch ohne Cyberangriff oder Sachschaden an einem versicherten Trägersystem entstehen. Wesentliche Ursachen sind menschliche Fehler, Prozess- und Betriebsfehler sowie Qualitätsmängel bei Datenbeständen.

Als Folge der Beeinträchtigung des KI-Modells und/oder seiner Daten- und Entwicklungsstände, aber auch direkt aufgrund einer Beeinträchtigung der Verfügbarkeit externer KI-Leistungen kann ein Betriebsunterbrechungsschaden entstehen, wenn der Betrieb vollständig unterbrochen oder teilweise gestört wird und es infolgedessen zu Ertragsausfällen oder Mehrkosten zur provisorischen Aufrechterhaltung des Betriebes kommt.

Liegt die Behebung des Schadens im Verantwortungs- und Kontrollbereich des betroffenen Unternehmens selbst, sind die Betriebsunterbrechungskomponenten bzw. -versicherungen der oben skizzierten Versicherungszweige relevant. So bietet die Cyberversicherung regelmäßig Deckung auch für den Unterbrechungsschaden, der infolge einer Netzwerksicherheitsverletzung¹³ an der versicherten IT eintritt.

Die versicherungstechnische Hürde liegt in diesen Fällen darin, dass die Beeinträchtigung des KI-Assets bei einem externen Anbieter eintritt, während sich die wirtschaftlichen Folgen beim nutzenden Unternehmen realisieren. Zwar kennen tradierte Betriebsunterbrechungsversicherungen (z. B. Feuer- oder Maschinenbetriebsunterbrechungsversicherungen) im Rahmen sog. »Rückwirkungsschäden« bereits Konstellationen, in denen Schadenereignis und Betriebsunterbrechung bei unterschiedlichen Beteiligten eintreten. Bei KI-bezogenen Ausfällen fehlt jedoch häufig der zwingend erforderliche »Sachschaden« als Anknüpfungspunkt der Deckung. Aus diesem Grund sind Rückwirkungsschäden aufgrund von KI-Fehlern, -Beeinträchtigungen oder -Ausfällen nicht in der klassischen Betriebsunterbrechungsversicherung abgedeckt.

In der Cyberversicherung ist der Bedarf zur Absicherung von Rückwirkungsschäden infolge des Ausfalls externer IT-Dienstleister seit Langem bekannt. Die hierfür entwickelten Deckungskonzepte knüpfen je nach Bedingungsmerkmal entweder konkret an eine Netzwerksicherheitsverletzung beim Dienstleister oder nur an dessen unvorhergesehene Nichtverfügbarkeit an. Ob diese Deckungserweiterungen auch Betriebsunterbrechungen infolge der Nichtverfügbarkeit externer KI-Modelle oder KI-

¹³ Siehe Glossar

Plattformen erfassen, ist vom jeweiligen Bedingungsnetzwerk abhängig und derzeit nicht immer eindeutig. Mögliche Deckungslücken sind daher im Einzelfall zu prüfen.

KI-induzierte Sachschäden an anderen betrieblichen Vermögenswerten

Greifen KI-Systeme direkt oder mittelbar in physische Wertschöpfungsprozesse ein – etwa durch Steuerung von Maschinen, Optimierung von Produktionsparametern, autonome Logistik oder intelligente Gebäudetechnik - können Sachschäden an sonstigen Vermögenswerten eines Unternehmens entstehen. Diese unterscheiden sich in ihrer äußeren Erscheinungsform häufig nicht von typischen und unter den tradierten Versicherungssparten versicherbaren Sachschäden. Sie können sich beispielsweise als Brand, Explosion, Kollision, Überhitzung oder Wasseraustritt manifestieren. Infolge des Sachschadens kann es zudem zu einer Störung des Betriebes bis hin zu einer vollständigen Betriebsunterbrechung kommen.

Die Schadensursache kann dabei sowohl in einem technischen Versagen von Softwarekomponenten als auch in einer fehlerhaften oder unangemessenen Entscheidung eines KI-Systems liegen. Gerade Letzteres stellt eine Besonderheit von KI-Anwendungen dar: Obwohl die technische Infrastruktur ordnungsgemäß funktioniert, können fehlerhafte Prognosen, Bewertungen oder Steuerungsentscheidungen der KI ein schadensauslösendes Ereignis verursachen und so zu physischen Schäden an Sachen führen.

Beispiele:

- In einem Produktionsprozess steuert ein KI-System die Fertigungsparameter fehlerhaft, insbesondere bei Temperatur und Druck. Dies führt dazu, dass Materialien thermisch geschädigt werden und Produktionschargen unbrauchbar werden. Gleichzeitig kommt es durch die ungeeigneten Betriebsbedingungen zu einer Beschädigung von Maschinenkomponenten. Der betroffene Betriebsteil steht still, es fallen Mehrkosten für die Umsetzung von Ausweichmaßnahmen an.
- Ein KI-System zur vorausschauenden Wartung prognostiziert den Zustand einer Produktionsanlage fehlerhaft und erkennt einen kritischen Verschleiß nicht rechtzeitig. In der Folge unterbleibt eine notwendige Wartung, was zu einem Ausfall der Maschine und erheblichen mechanischen Schäden führt.
- Ein KI-System trifft Steuerungsentscheidungen auf Basis fehlerhafter, nicht ausreichend validierter Sensordaten. Dadurch wird eine Anlage falsch kalibriert und über einen längeren Zeitraum außerhalb zulässiger Betriebsparameter betrieben. Dies führt zu irreversiblen Schäden an technischen Komponenten sowie zu Materialverlusten in der Produktion.

Da KI-Systeme häufig zentral entwickelt und in großem Maßstab eingesetzt werden, können einzelne Fehler weit über den ursprünglichen Anwendungsfall hinauswirken und gleichzeitig zahlreiche Anlagen, Prozesse oder Standorte betreffen.

KI-bedingte Sachschäden sind durch marktübliche sog. »Allgefahren-Sachversicherungen« derzeit überwiegend versichert. Diese Deckungskonzepte bieten umfassenden Versicherungsschutz für Sachschäden, ohne dass eine konkret benannte versicherte Gefahr (z. B. Feuer) vorliegen muss.

Auch in der Allgefahren-Dekung greifen jedoch Ausschlüsse. Gängige Einschränkungen resultieren aus dem Sachschadenbegriff (z. B. kein Versicherungsschutz bei Abhandenkommen oder reinem Verschleiß, reinem Mangel) oder aus dem Ausschluss bestimmter Gefahren (z. B. Höhere Gewalt, Krieg, Kernenergie). Die Mehrzahl der antizipierbaren KI-bedingten Sachschäden dürfte nicht unter diese typischen Ausschlussstatbestände fallen.

Ein genereller Ausschluss für Sachschäden, die durch die Nutzung von KI-Anwendungen verursacht werden, ist derzeit nicht marktüblich. Vor diesem Hintergrund hat sich analog zu »Silent Cyber« der Begriff »Silent AI« etabliert. Er beschreibt die Situation, dass bestehende Versicherungslösungen, insbesondere auf Basis von Allgefahrenkonzepten, grundsätzlich auch Sach- und Betriebsunterbrechungsschäden mit KI-Bezug erfassen können, ohne dass diese Risiken im Bedingungswerk ausdrücklich geregelt oder bei der Produktkonzeption gezielt einbezogen wurden.

Folgende Punkte sollten jedoch berücksichtigt werden:

Einige Deckungskonzepte – speziell in den Zweigen der sog. Technischen Versicherungen – setzen einen plötzlichen und unvorhergesehenen Schadeneintritt voraus. Dies kann im Einzelfall zu Deckungslücken führen. Einschränkend kann sich zudem die sog. »Serienschadenklausel« auf den Versicherungsschutz auswirken – vor allem mit Blick auf die Skaleneffekt des KI-Einsatzes werden wiederholt eintretende Sachschäden nicht mehr ohne Weiteres als »unvorhergesehen eingetreten« zu werten sein.

KI-induzierte Vermögensschäden

Als Folge eines der oben skizzierten Sachschadentypen, aber auch ohne dass ein Sachschaden überhaupt eingetreten ist, können durch den Einsatz von KI-Anwendungen Vermögensschäden bei Unternehmen entstehen.

Im Vergleich zu Sachschäden stellen sich »reine Vermögensschäden« strukturell deutlich komplexer dar. Für die weitere Betrachtung bietet sich folgende Differenzierung an:

■ Variante 1: Vermögensschäden durch Betriebsunterbrechung oder -einschränkung

Unterschiedlichste Szenarien können in einer KI-bedingten Betriebsunterbrechung resultieren. Im vorherigen Abschnitt haben wir uns Betriebsunterbrechungen infolge einer Beschädigung des KI-Assets selbst und infolge einer

unvorhergesehenen Nichtverfügbarkeit des KI-Assets bereits angesehen. Doch auch bei funktionstüchtiger und uneingeschränkter Verfügbarkeit sind Szenarien möglich, die letzten Endes zu einer Störung des Betriebes oder sogar zu einer vollständigen Betriebsunterbrechung führen: Es kommt zu Umsatzverlusten und Mehrkosten für die provisorische Aufrechterhaltung des Betriebes.

Wenn eine KI-Anwendung Fehlentscheidungen mit kritischen Folgen trifft, muss das System gestoppt werden, bis die Ursache der Fehlentscheidungen isoliert und behoben ist. Die Folgen sind umso gravierender, je größer das Volumen der durch die KI verarbeiteten Entscheidungen ist; im schlimmsten Fall wird eine schleichende Entwicklung erst spät bemerkt, wenn bereits eine Fehlentscheidungskette vorliegt.

Beispiel¹⁴:

- Beim SaaS-Anbieter Pocket OS löschte ein KI-Agent im April 2026 in neun Sekunden die gesamte Produktionsdatenbank samt Backups und legte anschließend ein detailliertes »Geständnis« über die missachteten Sicherheitsregeln ab. Die betroffenen Kunden – überwiegend Autovermietungen – verloren Reservierungen, Zahlungs- und Kundendaten und waren über 30 Stunden lang nicht arbeitsfähig. Der Schaden lag hier nicht in einer beschädigten Sache, sondern im Wegfall geschäftskritischer Daten und der daraus folgenden Betriebsunterbrechung Dritter – ein typischer reiner Vermögensschaden, der zugleich zeigt, wie ein einzelner autonomer Eingriff über eine zentral betriebene Plattform eine ganze Nutzergruppe gleichzeitig erfasst.
- Ein großer Onlinehändler hat seinen First-Level-Support und die Bestellannahme vollständig auf einen autonomen KI-Agenten umgestellt. Durch ein fehlerhaftes Modell-Update beginnt die KI plötzlich, Kunden aus Kulanz überhöhte Rabatte zu gewähren oder im Anschluss an Kundengespräche zu Lieferzeiten die betreffenden Bestellungen grundlos zu stornieren. Die Folge: Das System muss sofort abgeschaltet werden. Der Onlinehändler kann über Tage hinweg nur einen Bruchteil der Kundenanfragen bearbeiten. Es entsteht ein massiver Umsatzeinbruch (entgangener Gewinn).

¹⁴ Vgl. Cybernews, *Claude AI agent wipes firm's database in 9 seconds*, 30.04.2026.

- Variante 2: Vermögensschäden durch erhöhte Aufwendungen oder verringerte Einnahmen

Wenn KI-Technologie in preis- und abrechnungsrelevante Softwareprodukte integriert ist, können Leistungsfehler von KI-Modellen zu unerwünschten Effekten führen.

Beispiel:

- Eine KI-Anwendung erstellt fehlerhafte Nachfrageprognosen, wodurch ein Unternehmen Überbestände aufbaut und Lagerkosten entstehen.
- Eine KI-Anwendung bewertet die Bonität von Kreditnehmern falsch, so dass Darlehen mit zu günstigen Konditionen vergeben werden und teils sogar Kredite eingeräumt werden, die bei korrekter Prognose abgelehnt worden wären. Hierdurch kommt es zu einer höheren Ausfallquote, gleichzeitig wird durch die zu günstigen Konditionen auch bei insgesamt kreditwürdiger Bonität die Gewinnmarge geschmälert.
- Eine KI prognostiziert Markt- oder Preisentwicklungen fehlerhaft, wodurch finanzielle Verluste entstehen.
- Fehlerhafte Preisfestsetzung: Ein KI-System setzt Verkaufspreise aufgrund fehlerhafter Trainingsdaten deutlich unter den vorgesehenen Margen fest. Die zu niedrigen Preise führen unmittelbar zu finanziellen Verlusten.

- Variante 3: Vermögensschäden durch Vermögensabfluss

Vermögensschäden durch Vermögensabfluss entstehen, wenn ein KI-System infolge fehlerhafter Datenverarbeitung, algorithmischer Fehlentscheidungen oder der Ausnutzung von Sicherheitslücken – etwa durch Prompt Injection¹⁵ – unberechtigte oder fehlerhafte Vermögenstransfers auslöst oder den Verlust materieller oder immaterieller Vermögenswerte verursacht.

¹⁵ Siehe Glossar

Beispiele:

- Fehlerhafte Zahlungsfreigabe: Eine KI-Anwendung zur automatisierten Rechnungsbearbeitung ordnet aufgrund einer fehlerhaften Datenanalyse mehrere Rechnungen als geprüft und freigabefähig ein. Die daraufhin ausgelösten Zahlungen erfolgen zu Unrecht.
- Irrtümliche Rückerstattungen: Eine KI-Anwendung im Kundenservice genehmigt automatisiert unberechtigte Gutschriften oder Rückzahlungen an Kunden.

Reine Eigenschäden in Form von Vermögensschäden, die allein auf Fehlfunktionen, Fehlentscheidungen oder Fehlanalysen von KI-Systemen zurückzuführen sind, sind am Versicherungsmarkt noch nicht flächendeckend Gegenstand eigenständiger Deckungskonzepte. Versicherungsschutz kommt regelmäßig erst dann in Betracht, wenn der Vermögensschaden auf ein bereits etabliertes versichertes Ereignis zurückgeführt werden kann, etwa einen Cyberangriff oder eine vertrauensschadenrelevante Handlung. Reine Vermögensverluste infolge einer fehlerhaften KI-Nutzung verbleiben daher bislang weitgehend im Eigenrisiko der betroffenen Unternehmen. Bei reinen Vermögensschäden ist die Abgrenzung zwischen einem versicherbaren Technologierisiko und dem traditionell nicht versicherbaren Unternehmens- bzw. Geschäftsrisiko versicherungstechnisch anspruchsvoll. Die Entwicklung entsprechender Versicherungslösungen wird daher maßgeblich davon abhängen, Risiken hinreichend präzise zu beschreiben sowie geeignete technische und organisatorische Obliegenheiten zu definieren, um versicherbare Fehlfunktionen von KI-Systemen von unternehmerischen Fehlentscheidungen abzugrenzen.

Es existieren erste spezielle Deckungskonzepte, die gezielt Leistungsfehler von KI-Systemen und die hieraus resultierenden Vermögensschäden erfassen. Diese Versicherungslösungen richten sich an Hersteller und Betreiber großskalierter KI-Anwendungen. Sie sind in Form von Garantiever sicherungen ausgestaltet und sehen regelmäßig eine an den konkreten Leistungsparametern des KI-Systems orientierte Deckung vor. Der Versicherungsschutz knüpft an eine vertraglich definierte Leistungsabweichung des KI-Modells an, etwa wenn eine vereinbarte Prognosegenauigkeit oder ein anderer Performance-Indikator unterschritten wird. Die Deckung ist parametrisch bzw. performancebasiert ausgestaltet.

Fazit: Eigenschadenrisiko aus KI-Nutzung

Die Versicherbarkeit von Eigenschäden im Zusammenhang mit der Nutzung von KI lässt sich nicht pauschal beurteilen. Vielmehr hängt der Versicherungsschutz maßgeblich von der Art des Schadens, der zugrunde liegenden Gefahr und nicht zuletzt dem jeweiligen Bedingungswerk ab. Während einzelne Schadenkonstellationen bereits heute durch bestehende Versicherungsprodukte erfasst werden können, bestehen in anderen Bereichen noch Angebotslücken oder zumindest Abgrenzungsfragen.

Schäden an KI-Anwendungen, Modellen oder Datenbeständen können bereits heute auf unterschiedliche Weise versicherbar sein. Dies gilt zum einen, wenn sie durch eine klassische versicherte Gefahr wie Feuer, Leitungswasser oder Maschinenbruch verursacht werden. Zum anderen können sie, ebenso wie durch KI verursachte Sachschäden, im Rahmen von Allgefahrenkonzepten unter Umständen auch ohne ausdrückliche Regelung von KI-Risiken erfasst sein (»Silent AI«). Soweit die Beschädigung auf einen Cyberangriff zurückzuführen ist, kommt zudem Versicherungsschutz über bestehende Cyberversicherungen in Betracht. Daran anknüpfende Betriebsunterbrechungsschäden können regelmäßig mitversichert sein.

Eine Angebotslücke besteht aktuell noch in Bezug auf reine Vermögensschäden infolge von Fehlfunktionen oder Fehlentscheidungen von KI-Systemen. Die größte Herausforderung besteht hierzu regelmäßig in der Abgrenzung zwischen einer versicherbaren Fehlfunktion eines KI-Systems und einem nicht versicherbaren unternehmerischen Fehlentscheidungsrisiko.

6 Drittschäden aus KI-Nutzung

Was sind Drittschäden?

Ein Drittschaden ist ein Schaden, der nicht dem Versicherungsnehmer selbst entsteht, sondern einer dritten Person. Die Haftpflichtversicherung knüpft hieran an: Nach Ziff. 1 der Allgemeinen Versicherungsbedingungen für die Haftpflichtversicherung (AHB) besteht Versicherungsschutz im Rahmen des versicherten Risikos für den Fall, dass der Versicherungsnehmer wegen eines während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses (Versicherungsfall) – das einen Personen-, Sach- oder sich daraus ergebenden Vermögensschaden zur Folge hatte – aufgrund gesetzlicher Haftungsbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadensersatz in Anspruch genommen wird. Dabei sind zunächst Vermögensschäden, die weder durch Personen- noch durch Sachschäden entstanden sind (= reiner Vermögensschaden) ausgeschlossen. Je nach Deckungskonzept können diese aber explizit eingeschlossen werden.¹⁶

Mit dem zunehmenden Einsatz von KI verschiebt sich nicht die rechtliche Grundstruktur dieser Haftung, wohl aber ihre Verursachung: Drittschäden entstehen nicht mehr nur durch menschliches Fehlverhalten oder technisches Versagen, sondern durch inhaltlich fehlerhafte, voreingenommene oder unangemessene Entscheidungen automatisierter Systeme. Da KI-Anwendungen typischerweise zentral integriert und breit ausgerollt werden, kann ein einzelner Modell- oder Trainingsfehler zugleich eine Vielzahl betroffener Dritter erfassen – Bewerber, Patienten, Kreditnehmer oder Rechteinhaber gleichermaßen. Die Schadenkategorien des Haftpflichtrechts bleiben dabei dieselben; sie zeigen sich nur in neuen Erscheinungsformen.

Grundsätzlich wird in der Haftpflicht zwischen Personenschäden, Sachschäden und reinen Vermögensschäden unterschieden. Dabei zeigt sich, dass es bereits einige Praxisfälle gibt, die den jeweiligen Schadensarten zugeordnet werden können.

Personenschaden

Ein Personenschaden liegt vor, wenn ein Mensch verletzt, in seiner Gesundheit geschädigt oder getötet wird. Im KI-Kontext reicht das Spektrum von unmittelbaren körperlichen Einwirkungen autonom agierender Maschinen bis zu mittelbaren Beeinträchtigungen, die aus diskriminierenden oder fehlerhaften Entscheidungen algorithmischer Systeme folgen. Dabei ist der unechte Vermögensschaden, also der sogenannte Vermögensfolgeschaden, der Schaden, der infolge eines Personenschadens eintritt. Dies wäre z. B. der Verdienstausschlag bei einer Person, die aufgrund der körperlichen Beeinträchtigung nicht ihrer Erwerbstätigkeit nachgehen kann. Da diese Schäden keine reinen Vermögensschäden sind, werden sie den Personenschäden zugeordnet.

¹⁶ Veith, Gräfe und Gebert, Abschnitt D, § 15, Rn. 32, 33.

Leib und Körper: Hier wirkt die KI physisch auf den Menschen ein. In Südkorea wurde ein Mann von einem Industrieroboter tödlich verletzt, der ihn nicht als Person, sondern als zu bearbeitendes Objekt erfasste; in den USA wurde Tesla verklagt, nachdem ein Arbeiter durch einen Fertigungsroboter bewusstlos geschlagen wurde. Die Systeme funktionierten technisch – die schädigende Wirkung lag in der fehlerhaften Wahrnehmung und Steuerung.

Diskriminierung: Hier entsteht der Personenschaden mittelbar, durch systematische Benachteiligung. Im US-Verfahren *Mobley vs. Workday* steht der Vorwurf algorithmischer Altersdiskriminierung im Bewerbungsprozess im Raum; Wells Fargo schloss einen Vergleich über die diskriminierende Vergabe von Krediten und Stellen; gegen UnitedHealth richtet sich der Vorwurf, ein KI-System habe älteren Versicherten medizinisch gebotene Anschlussbehandlungen zu Unrecht verweigert. Gemeinsam ist diesen Fällen, dass ein skalierbares bzw. flächendeckend eingesetztes Modell eine große Zahl von Betroffenen gleichförmig benachteiligt.

Verleumdung / Defamation: Generative Systeme können unwahre, ehrverletzende Aussagen über reale Personen erzeugen und verbreiten. Das Landgericht München hatte über die Haftung für derartige KI-generierte Falschbehauptungen zu befinden – ein Schadenbild, das klassische Persönlichkeitsrechtsverletzungen in den Output automatisierter Systeme verlagert.

Sachschaden

Neben dem klassischen Sachschaden durch Zerstörung, Beschädigung oder Abhandenkommen einer Sache erweitert die KI-Nutzung potenzielle Schadenbilder. Dabei können im Zusammenhang mit dem Einsatz von KI-Systemen physische Einwirkungen durch Fehlsteuerung, falsche Konfiguration oder Entscheidungen zu fehlerhaften Produktionsprozessen oder Eingriffe in IT-Systeme zu Sachschäden führen. Der unechte Vermögensschaden ist hierbei analog wie beim Vermögensfolgeschaden bei Personenschäden der Schaden, der infolge des Sachschadens eintritt. So wäre beispielsweise die Beschädigung der Produktionsanlage durch eine fehlerhafte KI ein klassischer Sachschaden. Die daraus resultierenden Produktionsausfälle, Lieferverzögerungen sowie der entgangene Gewinn wären als unechte Vermögensschäden zu qualifizieren.

Reiner Vermögensschaden

Der reine Vermögensschaden ist haftungsrechtlich am schwierigsten zu fassen und besitzt im KI-Kontext ein besonders hohes Schadenpotenzial. Er wird daher in der klassischen Betriebshaftpflicht, anders als in Sonderkonzepten (z. B. für IKT-Dienstleister), regelmäßig nicht oder nur eingeschränkt gedeckt. In Bezug auf KI gewinnt er zudem an Bedeutung, weil autonom agierende Systeme unmittelbar in fremde Geschäftsabläufe eingreifen und wirtschaftliche Schäden, insbesondere Betriebsunterbrechungen und Datenverluste, verursachen können.

Beim SaaS-Anbieter PocketOS löschte ein KI-Agent im April 2026 in neun Sekunden die gesamte Produktionsdatenbank samt Backups und legte anschließend ein detailliertes

»Geständnis« über die missachteten Sicherheitsregeln ab.¹⁷ Die betroffenen Kunden – überwiegend Autovermietungen – verloren Reservierungen, Zahlungs- und Kundendaten und waren über 30 Stunden lang nicht arbeitsfähig. Der Schaden lag hier nicht in einer beschädigten Sache, sondern im Wegfall geschäftskritischer Daten und der daraus folgenden Betriebsunterbrechung Dritter – ein typischer reiner Vermögensschaden, der zugleich zeigt, wie ein einzelner autonomer Eingriff über eine zentral betriebene Plattform eine ganze Nutzergruppe gleichzeitig erfasst.

Rein vertragliche Haftung / Garantien

Die Versicherung von Vertragsstrafen im Rahmen der Haftpflicht zielt darauf ab, Unternehmen vor finanziellen Belastungen zu schützen, die aus vertraglich vereinbarten Strafzahlungen entstehen können. Solche Vertragsstrafen werden häufig festgelegt, um die Einhaltung bestimmter Leistungsversprechen sicherzustellen, beispielsweise Performance-Garantien von KI-Systemen. Im Falle einer Nichterfüllung können dadurch vereinbarte Strafzahlungen ganz oder teilweise fällig werden. Da diese rein vertraglichen Haftungs- und Garantieregelungen über die gesetzliche Haftpflicht hinausgehen, sind sie regelmäßig nicht vom Umfang der klassischen Haftpflichtversicherungskonzepte umfasst.

Beispielsweise kann in einer Versicherungspolice geregelt sein, dass Schäden übernommen werden, wenn die Fehlerquote eines KI-Systems einen definierten Schwellenwert (z. B. 1 Prozent) überschreitet. In diesem Kontext versichert der Versicherer das Start-up oder Scale-up, das wiederum entsprechende Leistungszusagen in seinen Lizenzverträgen macht. Dabei sind länderspezifische rechtliche Rahmenbedingungen zu beachten, insbesondere muss sichergestellt werden, dass das Unternehmen tatsächlich Garantien abgibt und nicht selbst als »Versicherer« auftritt.

Entsprechende Lösungen werden bereits von einigen spezialisierten Anbietern am Markt angeboten, vor allem für Start-ups und Scale-ups, die von der Kombination aus vertraglicher Haftung und Versicherungsschutz profitieren. Neben der finanziellen Absicherung entsteht auch ein vertrauensbildender Effekt, da die Risikobewertung durch einen Versicherer als Qualitätssignal wirkt. Gleichzeitig können potenzielle Verluste ganz oder teilweise kompensiert werden, sodass sich ein Wettbewerbsvorteil gegenüber Anbietern ohne Absicherung ergibt.

Regulatorischer und haftungsrechtlicher Ausgangspunkt

Die Versicherungsbedingungen der Drittschadendeckung stellen regelmäßig auf gesetzliche Haftpflichtbestimmungen privatrechtlichen Inhalts ab.¹⁸ Öffentlich-rechtliche Sanktionen fallen grundsätzlich nicht hierunter, sofern der Versicherungsvertrag keine ausdrücklichen Erweiterungen vorsieht. Dies gilt grundsätzlich auch für E&O- und Cyberversicherungen.¹⁹

¹⁷ Vgl. Cybernews, *Claude AI agent wipes firm's database in 9 seconds*, 30.04.2026.

¹⁸ Vgl. Veith, Gräfe, und Gebert, *Der Versicherungsprozess*, Abschn. D, § 19, Rn. 32.

¹⁹ Vgl. Höra, Teil C, § 12, Rn. 39, 40.

Für die haftungsrechtliche Einordnung des Einsatzes von KI-Agenten und automatisierten Prozessen ist zunächst zwischen aufsichtsrechtlichen und zivilrechtlichen Vorgaben zu unterscheiden. Die Verordnung (EU) 2024/1689 (EU AI Act, EU AI Act) ist als öffentlich-rechtliches Aufsichts- und Produktsicherheitsrecht ausgestaltet. Sie begründet keine eigenständigen zivilrechtlichen Schadensersatzansprüche geschädigter Dritter gegen Anbieter oder Betreiber von KI-Systemen. Verstöße werden in erster Linie durch aufsichtsrechtliche Maßnahmen und Bußgelder sanktioniert. Ob einzelne Betreiber- oder Anbieterpflichten des EU AI Act darüber hinaus als Schutzgesetze im Sinne des § 823 Abs. 2 BGB zu qualifizieren sind und damit mittelbar zivilrechtliche Ersatzansprüche begründen können, ist in der Literatur bislang nicht abschließend geklärt.²⁰

Nachdem die ursprünglich geplante KI-Haftungsrichtlinie (AI Liability Directive) im Jahr 2025 zurückgezogen wurde, richtet sich die Haftung des KI-nutzenden Unternehmens gegenüber geschädigten Dritten weiterhin nach den allgemeinen zivilrechtlichen Haftungsregimen des nationalen Rechts. Maßgeblich sind insbesondere die deliktische Haftung, die Gefährdungs- und Produkthaftung sowie die vertragliche Haftung.

Haftung des KI- Herstellers

Die Verwendung von KI-Technologie stellt Hersteller vor neue Herausforderungen. Neben dem Technologiefortschritt führen auch die Änderungen im regulatorischen Bereich zu einem erhöhten Haftungsrisiko. Gerichte in Deutschland und den USA haben bereits bestätigt, dass Unternehmen für den durch sie erzeugten KI-Output haften. Die EU-Produkthaftungsreform bezieht Software und KI-Modelle erstmals ausdrücklich in die verschuldensunabhängige Haftung mit ein, was zu einer drastischen Änderung der bestehenden Eintrittspflicht für zurechenbare Schäden führt.

Die aktuellen KI-Herstellerrisiken sind durch bestehende Versicherungsprodukte, insbesondere durch spezielle Vermögensschadenhaftpflichtprodukte, sogenannte Tech- E&O oder Tech-PI-Produkte grundsätzlich versichert; die Änderung im Bereich der Produkthaftung auf eine verschuldensunabhängige Haftung sollte keinen negativen Einfluss auf den Versicherungsschutz entfalten. Im Rahmen der Schadenabwicklung könnte die geänderte Rechtslage zu Beginn dennoch zu einem erhöhten Aufklärungsaufwand führen. Auch der Bedarf für eine gesonderte Produkthaftungsversicherung sollte von Herstellern geprüft werden.

Im Zusammenhang mit Herstellerrisiken kann in Zukunft zudem die Abgrenzung schwierig sein, wer für den erstellten Output der KI verantwortlich ist. Dies ist darauf zurückzuführen, dass die Fehlerquellen im KI-Bereich mannigfaltig sind. Sofern der Output fehlerhaft ist, kann dies entweder dem dahinterliegenden Algorithmus und daher dem Softwareunternehmen zugerechnet werden. Der mangelhafte Output kann jedoch auch darauf zurückzuführen sein, dass das die KI-Technologie nutzende Unternehmen fehlerhafte oder unvollständige Trainingsdaten zur Verfügung gestellt hat oder der zugrundeliegende Prompt nicht präzise genug war, um fehlerfreie Ergebnisse zu gewährleisten. Des Weiteren führt auch die Weiterentwicklung von KI-

²⁰ Die Richtlinie über KI-Haftung – weniger ist mehr!« von Prof. Dr. Gerhard Wagner | Sep 04, 2024.

Technologien durch Unternehmen, die grundsätzlich nicht dem Bereich der Softwareunternehmen zuzurechnen sind, zu schwierigen Abgrenzungsfragen.

Softwareunternehmen sind neben den gewöhnlichen Haftungsrisiken noch zusätzlich im Bereich der Urheberrechtsverletzungen exponiert. Dies gilt sowohl in Bezug auf das Training als auch auf die Anwendung. Werden KI-Assistenten beispielsweise zur Erstellung von Code verwendet, besteht die Gefahr, dass Code-Fragmente aus urheberrechtlich geschütztem oder unter Copyleft-Lizenzen stehendem Quellcode entstammen und dadurch ein Urheberrecht verletzt wird.

Anhand von zwei Gerichtsentscheidungen wird in der Folge auf die Reichweite der Schrankenregelung gemäß § 44b UrhG eingegangen.

Kneschke ./ LAION

Das OLG Hamburg hat mit Urteil vom 10.12.2025 (Az. 5 U 104/24) die Anwendung der Schrankenregelungen gemäß §§ 60 ff bzw 44b UrhG bejaht und somit die Nutzung derartiger Daten zur Erstellung des Trainingsmaterials als zulässig erklärt. Entgegen der Vorinstanz hat das OLG zudem festgestellt, dass ein Nutzungsvorbehalt gemäß § 44 Abs 3 UrhG nur dann wirksam ist, wenn er in maschinenlesbarer Form erfolgt.

GEMA ./ OpenAI

Eine weitere relevante Gerichtsentscheidung ist jene des LG München vom 11. November 2025, Az. 42 O 14139/24. Auch in diesem Fall wurde die Anwendbarkeit von § 44b UrhG und die dadurch bedingte gerechtfertigte Vervielfältigung von urheberrechtlich geschützten Daten gerichtlich geprüft. Dieses Urteil behandelt jedoch nicht die Verwertung von Datensätzen im Bereich der Erstellung des Trainingsdatenmaterials, sondern das Training des KI-Systems mit urheberrechtlich geschützten Daten. In seiner Entscheidung geht das LG München auch auf die Entscheidung des LG Hamburg ein und stellt klar, dass zwischen den Phasen der Vervielfältigung unterschieden werden muss. Aus Sicht des Gerichts ist zwischen drei Phasen zu differenzieren.

- **Phase 1**, das Extrahieren und die Überführung des Trainingsmaterials in ein maschinenlesbares Format, dem Erstellen des Trainingsdatenmaterials,
- **Phase 2**, die Analyse des Datenmaterials und ihre Anreicherung mit Meta-Informationen, dem Training des Modells, und
- **Phase 3**, die nachfolgende Nutzung des trainierten Modells durch Prompts und Outputs.

Gemäß dem Gericht ist lediglich die Vervielfältigung im Zuge der Phase 1 durch die Schrankenregelung nach § 44b UrhG gerechtfertigt, nicht jedoch die weitergehenden Vervielfältigungen, da diese nicht ausschließlich der Vorbereitung des Text- und Data-Mining dienen. Aus Sicht des Gerichts läuft insbesondere die Memorisierung der Liedtexte dem Rechtfertigungsgrund nach § 44b UrhG zuwider, da es sich hierbei nicht um reines Text- und Data-Mining handelt, da die Liedtexte nicht ausgewertet, sondern gesamtheitlich in das Trainingsdatenset übernommen wurden. Dies greift in die Verwertungsrechte der Urheber ein.

Des Weiteren hat das LG auch die Anwendbarkeit anderer Schrankenregelungen, im Speziellen der §§ 57 und 60d UrhG verneint. Anzumerken ist zudem noch, dass auch die

Betreiber des KI-Modells den Output der Liedtexte als »unerwünscht« angesehen haben und dies daher als eine Mangelhaftigkeit des Modells einzustufen ist.

Dem Einwand von OpenAI, das Unterlassungsgebot würde zu unverhältnismäßigen Nachteilen führen, hat das Gericht ebenfalls nicht stattgegeben.

Diese Entscheidung zeigt eindeutig auf, dass im Bereich der Vervielfältigung ein besondere Sorgfaltsmaßstab angewendet werden muss.

Abschließend ist festzustellen, dass Hersteller von KI-Systemen mit einer weitreichenden Haftungssituation konfrontiert sind und die bestehenden Versicherungslösungen noch nicht ausreichend auf diese geänderte Situation abgestimmt sind. Folgende Handlungsfelder sind zu beachten.

Haftung des KI-Nutzers

Betreiber- und Quasi- Herstellerrolle

Ausgangspunkt der Haftungs Betrachtung ist die Rolle des Unternehmens innerhalb der KI-Wertschöpfungskette. Der EU AI Act unterscheidet zwischen dem Anbieter eines KI-Systems und dessen Betreiber. Das hier betrachtete Nutzerrisiko betrifft grundsätzlich die Betreiberrolle, also den beruflichen oder betrieblichen Einsatz eines KI-Systems in eigener Verantwortung.

Von besonderer praktischer Bedeutung ist jedoch der mögliche Rollenwechsel zum Hersteller oder Quasi-Hersteller. Dies kann insbesondere dann der Fall sein, wenn ein Unternehmen eingekaufte Modelle wesentlich verändert, mit eigenen Daten nachtrainiert, mehrere Systeme zu einem eigenständigen KI-Agenten kombiniert oder die Lösung unter eigener Marke vermarktet. Dies kann auch als Teil des eigenen Produktes oder Services geschehen. Innerhalb des EU AI Acts ist dieser Rollenwechsel vom Betreiber zum Anbieter (Art. 25 EU AI Act) formal auf Hochrisiko-KI-Systeme beschränkt; das allgemeine produkthaftungsrechtliche Quasi-Hersteller-Konzept – etwa das Vermarkten unter eigener Marke oder die wesentliche Veränderung eines Produkts – ist demgegenüber älteren Ursprungs und gilt unabhängig von einer Risikoeinstufung nach dem EU AI Act. In solchen Fällen können neben den Betreiberpflichten zusätzliche Herstellerpflichten und produkthaftungsrechtliche Risiken entstehen.

Für die versicherungsrechtliche Einordnung ist diese Abgrenzung zentral, da sich hieraus unterschiedliche Haftungsregime und Deckungszuordnungen innerhalb der Betriebs-, Produkt-, E&O- oder Cyberversicherung ergeben können.

Deliktische Haftung

Gegenüber Personen, mit denen kein Schuldverhältnis besteht, ist die deliktische Haftung das zentrale Haftungsregime. Maßgebliche Anspruchsgrundlage ist § 823 BGB.

Die Haftung setzt voraus, dass durch den Einsatz eines KI-Systems ein geschütztes Rechtsgut wie Leben, Gesundheit oder Eigentum verletzt wird und dem Nutzer ein schuldhaftes Verhalten vorzuwerfen ist. In Betracht kommen beispielsweise Personen- oder Sachschäden durch fehlerhafte KI-gesteuerte Produktions-, Logistik- oder Steuerungssysteme.

Das Verschulden knüpft regelmäßig an die Verletzung von Verkehrssicherungspflichten an. Hierzu zählen insbesondere die sorgfältige Auswahl und Konfiguration des Systems, die Begrenzung von Handlungsspielräumen, die Qualitätssicherung der Eingabedaten, ein angemessenes Monitoring sowie die Möglichkeit menschlicher Eingriffe. Die Betreiberpflichten des EU AI Act dürften zukünftig zunehmend zur Konkretisierung des zivilrechtlichen Sorgfaltsmaßstabs herangezogen werden.

Von besonderer Bedeutung ist, dass reine Vermögensschäden grundsätzlich nicht vom Schutzbereich des § 823 Abs. 1 BGB erfasst werden. Gerade viele typische Fehlentscheidungen autonomer Systeme führen jedoch zu solchen Vermögensschäden. Ersatzansprüche kommen dann nur aus besonderen Schutzgesetzen oder anderen Haftungsgrundlagen in Betracht.

Werden personenbezogene Daten verarbeitet, kann daneben ein Schadensersatzanspruch nach Art. 82 DSGVO bestehen. Dieser umfasst auch immaterielle Schäden und enthält eine für Geschädigte günstigere Beweislastverteilung.

Gefährdungs- und Produkthaftung

Eine allgemeine verschuldensunabhängige Gefährdungshaftung für den Betrieb autonomer KI-Systeme besteht derzeit nicht. Der Nutzer haftet daher grundsätzlich nicht allein wegen des Einsatzes einer KI, sondern nur dann verschuldensunabhängig, wenn der Betrieb des Systems bereits in ein bestehendes Gefährdungshaftungsregime eingebettet ist, etwa im Bereich des Straßenverkehrs oder beim Betrieb bestimmter Anlagen.

Besondere Bedeutung kommt jedoch der Produkthaftung zu. Nach dem deutschen Produkthaftungsgesetz (ProdHaftG) sowie der zugrunde liegenden europäischen Produkthaftungsrichtlinie trifft die verschuldensunabhängige Haftung grundsätzlich den Hersteller eines fehlerhaften Produkts. Anders als im Deliktsrecht nach § 823 BGB kommt es dabei nicht auf ein Verschulden an, sondern allein darauf, dass ein fehlerhaftes Produkt in Verkehr gebracht wurde und hierdurch ein ersatzfähiger Schaden entstanden ist. Geschützt werden insbesondere Leben, Körper, Gesundheit und Sachgüter. Reine Vermögensschäden fallen demgegenüber nicht primär in den Anwendungsbereich der Produkthaftung.

Für Nutzer und Betreiber von KI-Systemen gewinnt die Produkthaftung zunehmend an Relevanz, weil die Herstellerstellung nicht auf den ursprünglichen Entwickler beschränkt bleibt. Bereits nach geltendem Recht kann auch ein Unternehmen als Quasi-Hersteller haften, wenn es ein Produkt unter eigenem Namen oder eigener Marke vertreibt. Im KI-Kontext können darüber hinaus Konstellationen entstehen, in denen Anwender durch erhebliche Anpassungen eines Basismodells, eigene Trainingsmaßnahmen, Fine-Tuning-Prozesse, die Integration zusätzlicher

OLG Hamm vom 12. Mai 2026 (Az. 4 UKI 3/25)

Der Chatbot einer Ästhetik-Klinik hatte den beiden tätigen Ärzten teilweise nicht-existierende Facharztartikel zugeschrieben. Der Senat ist der Argumentation, dass der Chatbot rein statistisch arbeite und den beiden Ärzten die Aussagen mangels Steuerung nicht zurechenbar seien, nicht gefolgt. Da der Betreiber den Tätigkeitsrahmen des Chatbots vorgibt, habe er ausreichenden Einfluss auf das System.

Datenquellen oder die Kombination mehrerer Systeme funktional in die Rolle eines Herstellers hineinwachsen.

Diese Entwicklung wird durch die neue Produkthaftungsrichtlinie (EU) 2024/2853 weiter verstärkt. Die Richtlinie stellt ausdrücklich klar, dass Software, KI-Systeme und digitale Dienstleistungen Produkte im produkthaftungsrechtlichen Sinne darstellen können. Zudem erweitert sie die Verantwortlichkeit auf Fälle, in denen ein Produkt nach seinem Inverkehrbringen wesentlich verändert wird. Wer ein KI-System nicht lediglich nutzt, sondern dessen Funktionsweise maßgeblich beeinflusst, weiterentwickelt oder unter eigener Verantwortung bereitstellt, kann daher künftig selbst produkthaftungsrechtlichen Risiken ausgesetzt sein.

Gerade für Unternehmen, die generative KI-Lösungen in eigene Produkte, Dienstleistungen oder Entscheidungsprozesse integrieren, entsteht hierdurch ein neues Haftungsrisiko. Die Grenze zwischen reiner Nutzung und haftungsrelevanter Produktverantwortung ist in vielen Fällen nicht eindeutig. Dies gilt insbesondere dann, wenn KI-Modelle individuell angepasst, mit unternehmenseigenen Daten trainiert oder gegenüber Kunden als eigene Lösung angeboten werden. Daher sollten die Unternehmen bei Modifizierungen und der Ausgestaltung der KI- Nutzung etwaige sich daraus ergebende Exponierungen im Blick behalten. Aus Sicht des Risikomanagements sollte daher frühzeitig geprüft werden, ob durch die konkrete Ausgestaltung des KI-Einsatzes eine Hersteller- oder Quasi-Herstellerstellung begründet werden kann.

Auch produkthaftungsrechtlich bleiben reine Vermögensschäden grundsätzlich außerhalb des Schutzbereichs. Im Vordergrund stehen Personen-, Sach- sowie nach der neuen Richtlinie bestimmte datenbezogene Schäden. Die praktische Relevanz der Produkthaftung aus Sicht eines Unternehmens ergänzt insofern die sonstige Haftung aus einem Schuldverhältnis oder deliktischer Haftung. Gleichwohl kann sie bei KI-bedingten Personen- oder Sachschäden erhebliche Bedeutung erlangen, da hier im Unterschied ein Verschulden des Herstellers nicht nachgewiesen werden muss.

Die Mitgliedstaaten haben die Produkthaftungsrichtlinie (EU) 2024/2853 bis zum 09. Dezember 2026 in nationales Recht umzusetzen. Unternehmen sollten die damit verbundenen Änderungen frühzeitig beobachten und insbesondere ihre Rolle in der KI-Wertschöpfungskette kritisch hinterfragen.

Vertragliche Haftung

Für den praktischen Einsatz von KI-Agenten in Kundenprozessen besitzt die vertragliche Haftung bzw. Haftung aus einem Schuldverhältnis regelmäßig die größte Bedeutung. Anders als das Delikts- und Produkthaftungsrecht erfasst sie auch reine Vermögensschäden, die unter anderem auch bei automatisierten Geschäftsprozessen entstehen.

Maßgeblich sind die gesetzlichen Schadensersatzansprüche wegen vertraglicher Pflichtverletzungen nach §§ 280 ff. BGB.²¹ Im Zusammenhang mit dem Einsatz von KI-

²¹ Vgl. Säcker u. a., *Münchener Kommentar zum Bürgerlichen Gesetzbuch*, Bd. 2, B. 2, § 280, Rn. 5.

Agenten können Pflichtverletzungen insbesondere in fehlerhaften automatisierten Entscheidungen, fehlerhaft ausgeführten Geschäftsvorgängen oder Verzögerungen der geschuldeten Leistung liegen.

Der KI-Agent selbst ist zwar kein Rechtssubjekt und daher kein Erfüllungsgehilfe im Sinne des § 278 BGB. Der Nutzer kann sich jedoch nicht auf die Eigenständigkeit des Systems berufen. Die Ergebnisse automatisierter Prozesse werden ihm grundsätzlich zugerechnet. Das Verschulden knüpft dabei insbesondere an Auswahl, Konfiguration, Überwachung und organisatorische Einbindung des Systems an.

Besonders bedeutsam ist die Beweislastregel des § 280 Abs. 1 Satz 2 BGB: Das Verschulden wird vermutet. Der Nutzer muss daher nachweisen können, dass er angemessene organisatorische und technische Maßnahmen getroffen hat. Eine dokumentierte KI-Governance mit klaren Verantwortlichkeiten, Kontrollmechanismen und Eskalationsprozessen gewinnt dadurch erhebliche haftungsrechtliche Bedeutung.

Fazit: Drittschadenrisiko KI

Der EU AI Act schafft derzeit kein eigenständiges zivilrechtliches Haftungsregime für Schäden durch KI-Systeme. Die Haftung gegenüber geschädigten Dritten richtet sich weiterhin nach den allgemeinen Regeln des Delikts-, Produkt- und Vertragsrechts.

Für Betreiber von KI-Agenten stehen dabei vor allem Verkehrssicherungs-, Überwachungs- und Organisationspflichten im Vordergrund. Mit zunehmender Automatisierung steigen die Anforderungen an Governance, Monitoring und menschliche Aufsicht. Gleichzeitig kann die Modifikation oder Integration fremder KI-Systeme zu einem Rollenwechsel vom Betreiber zum Hersteller oder Quasi-Hersteller führen und damit zusätzliche produkthaftungsrechtliche Risiken auslösen.

Versicherungsrechtlich ist daher stets zu prüfen, in welcher Rolle das Unternehmen handelt und welches Haftungsregime einschlägig ist. Die zentrale Risikolage liegt gegenwärtig weniger in neuartigen KI-spezifischen Haftungstatbeständen als vielmehr in der Anwendung bestehender Haftungsgrundsätze auf automatisierte und zunehmend autonome Geschäftsprozesse. Für KI-Hersteller, die Ihre Produkte an andere Unternehmen verkaufen, finden sich unabhängig von der gesetzlichen Haftung in rein vertraglich vereinbarten Leistungs- und Serviceversprechen beziehungsweise Garantien pauschalisierte Schadensersatzansprüche, die über die gesetzliche Haftung hinausgehen. Bei diesen freiwilligen Haftungsverschärfungen gegenüber Dritten versagen die gängigen Versicherungskonzepte, sodass spezialisierte und anwendungsbezogene Lösungen bei der Prüfung eines etwaigen Risikotransfers berücksichtigt werden sollten.

7 Governance- und Reputationsrisiken

Der Einsatz von KI-Lösungen in Unternehmen birgt ein hohes Risiko für die Leitungs- und Kontrollorgane. Der rasch fortschreitende Einsatz von KI-Technologien zwingt Unternehmen dazu, diese Technologien zu nutzen, um nicht den Anschluss zu verlieren. Gleichzeitig sind die Unternehmen verpflichtet, KI-Technologien gesetzeskonform zu nutzen. In diesem Spannungsfeld agieren Unternehmensorgane aktuell. Die unterlassene Auseinandersetzung mit den jeweils anwendbaren Rechtsvorschriften und daraus resultierende Schäden können zu einer persönlichen Haftung der Geschäftsführungsorgane führen. Die Argumentation, kein spezifisches Wissen im Bereich von KI-Technologien aufzuweisen, führt nicht zu einer Haftungsbefreiung der Organe, insbesondere dann nicht, wenn wesentliche Überwachungspflichten verletzt wurden. Die Bedeutung der Überwachungspflichten wurde durch eine Entscheidung des Bundesgerichtshofs erneut bekräftigt. In der Entscheidung vom 14.10.2025 (Az. II ZR 78/24) stellte der BGH fest, dass Aufsichtsratsmitglieder im Rahmen der Erfüllung ihrer Kontrollaufgaben dazu verpflichtet sind, aktiv Informationen einzuholen und eine mangelnde Berichterstattung des Vorstands nicht hinnehmen dürfen. Die dem Vorstand obliegende Bringschuld in Bezug auf eine ordnungsgemäße und vollständige Informationserteilung wird hierdurch nicht berührt; die Organe haften dem Unternehmen bzw. einem anspruchsberechtigten Dritten gegenüber solidarisch. Die Überwachung und ordnungsgemäße Dokumentation des Einsatzes von KI-Technologien sind daher zwingend notwendig.

Die mangelhafte Umsetzung der Compliance-Anforderungen des EU AI Act kann zur Verhängung von erheblichen Bußgeldern führen. Der EU AI Act sieht – ähnlich wie die DSGVO – abgestufte Bußgeldrahmen vor:

Tatbestand

Bußgeld

Besonders schwere Verstöße, etwa gegen verbotene KI-Praktiken (Art. 5 EU AI Act) oder zentrale Pflichten bei Hochrisiko bzw. GPAI-Systemen	bis zu 35 Mio. EUR oder 7 % des weltweiten Jahresumsatzes
Sonstige Verstöße gegen Pflichten des EU AI Act (z. B. Mängel im Risikomanagement, in der Datengovernance, Dokumentation, Human Oversight)	bis zu 15 Mio. EUR oder 3 % des weltweiten Jahresumsatzes

Für Unternehmen bedeutet dies:

- Eigenschäden durch Bußgelder, Aufsichtsmaßnahmen und Nutzungseinschränkungen für bestimmte Modelle oder Anwendungen,
- hohe Anpassungs- und Retrofit-Kosten, wenn Non-Compliance erst spät erkannt und nachreguliert wird,
- direkte Eingriffe ins Geschäftsmodell durch Auflagen oder Verbote spezifischer KI-Anwendungen.

Am deutschen Versicherungsmarkt sind behördliche Geldbußen regelmäßig nicht oder nur sehr eingeschränkt versicherbar. Sofern ein Unternehmen aufgrund der mangelhaften Umsetzung von Rechtsvorschriften durch seine Kontroll- und Leistungsorgane mit einem Bußgeld belastet wird, kann das Unternehmen bei seinen Organen Regress nehmen. Die Versicherbarkeit derartiger Ansprüche im Rahmen der D&O-Versicherung ist in Deutschland jedoch weiterhin strittig.

Neben den eigentlichen Strafzahlungen entstehen für Unternehmen noch weitere Schadendimensionen:

- **Compliance- und Anpassungskosten** für nachträgliche Dokumentation, Modellvalidierung, Datengovernance, Governance Boards und AI-Literacy-Programme,
- **Reputationsschäden** durch öffentlich gewordene Verstöße gegen KI-Regeln (inklusive medienwirksamer Bußgelder),
- **ESG- und Kapitalmarktrisiken** bei unzureichender oder »geschönter« KI-Governance (»AI Washing«).

Damit wird die EU AI Act-Compliance zu einem integralen Bestandteil des Querschnittsrisikos «KI». Ob KI-bedingte Fehlentwicklungen frühzeitig erkannt, gesteuert oder zumindest haftungs- und reputationsschonend bearbeitet werden, hängt maßgeblich von der Qualität der Governance ab. Ist die erforderliche Qualität nicht gegeben, führt dies zu einem persönlichen Haftungsrisiko der Leitungs- und Kontrollorgane, welches in gewissem Ausmaß auf eine D&O-Versicherung verlagert werden kann.

Neben behördlichen Sanktionen gewinnen auch vertragliche Strafzahlungen an Bedeutung:

- Leistungs- und Qualitätsgarantien für KI-Systeme gegenüber Geschäftspartnern (z. B. zugesicherte Fehlerquoten, Verfügbarkeit, Nichtdiskriminierung),
- Pönalen bei Verstößen gegen vertraglich vereinbarte KI-Governance-Standards (Audits, Dokumentations- und Oversight-Regelungen),
- Vertragsstrafen bei Verletzung KI-bezogener Transparenz- oder Kennzeichnungspflichten gegenüber Endkunden.

Diese Zahlungen sind privatrechtlich begründet, knüpfen faktisch aber häufig an Pflichten aus dem EU AI Act oder flankierender Regulierung an. Sie können als reine Vermögensschäden auftreten und damit in den Geltungsbereich von Vermögensschaden-Haftpflichtdeckungen (bspw. E&O, PI) fallen. Eine klare

vertragliche Abgrenzung (gesetzliche Haftung vs. Pönale) ist hier zentral. Anzumerken ist, dass die Versicherbarkeit von »echten« Vertragsstrafen weiterhin nur sehr eingeschränkt möglich ist und mitunter eine hohe Prämienbelastung mit sich bringt.

Das Risiko von Reputationsschäden beschäftigt Unternehmen aktuell im besonderen Maße. Dies belegt der aktuelle Allianz Risk Barometer. Die befragten Unternehmen erkennen in der rasanten Verbreitung von KI-Technologien eine hohe Gefahr für die Reputation des Unternehmens und den operativen Betrieb. Die Versicherungswirtschaft reagiert auf diese geänderte Risikolandschaft mit speziellen Deckungskonzepten. Neben den bereits existierenden Deckungen für Reputationskosten bieten einige Anbieter unter anderem im Rahmen von Cyberversicherungen Versicherungsschutz für Gewinneinbußen aufgrund eines Reputationsschadens. Trotz dieser innovativen Versicherungskonzepte muss darauf hingewiesen werden, dass nicht alle Governance- und Reputationsrisiken versicherbar sind. Zudem kann auch bei versicherten Schäden die Schadenabwicklung in der Praxis nicht zufriedenstellend verlaufen. Hier ist insbesondere auf die Schwierigkeiten bei der Erbringung des Schadensnachweises hinzuweisen.

Die fragmentierte Versicherbarkeit von KI-Risiken und die daraus resultierende Notwendigkeit, mehrere Versicherungsprodukte aufeinander abgestimmt einzukaufen, erhöhen zudem die Haftungssituation für Organe. Die Etablierung eines ordnungsgemäßen Risikomanagements, einer KI-Governance und eines umfassenden Versicherungskonzeptes ist daher unerlässlich. Governance- und Reputationsrisiken sind damit keine Randthemen, sondern ein eigenständiger Schadentreiber: Sie verbinden aufsichtsrechtliche Pflichten, persönliche Organhaftung und Marktvertrauen. Ein belastbares Governance- und Dokumentationskonzept entscheidet darüber, ob KI-bedingte Fehlentwicklungen frühzeitig erkannt und begrenzt werden – oder ob sie sich als Bußgelder, Organhaftungsfälle und nachhaltige Reputationsschäden niederschlagen.

8 Viele Parteien – ein KI-Modell: gemeinsame Absicherung für ein gemeinsames KI—Projekt-Risiko?

Eine der zentralen Herausforderungen bei der Versicherung von KI-Systemen besteht darin, dass an deren Entwicklung, Bereitstellung und Nutzung regelmäßig eine Vielzahl unterschiedlicher Parteien beteiligt ist. Schäden lassen sich daher häufig nicht eindeutig einer einzelnen verantwortlichen Partei zuordnen. So können Fehler oder Ausfälle beispielsweise auf das Basismodell eines Anbieters, die verwendeten Trainingsdaten, die Anpassung, Integration und den Betrieb durch Softwareanbieter, Systemintegratoren oder externe IT-Dienstleister, die betriebliche Konfiguration oder die konkrete Nutzung durch den Anwender zurückzuführen sein. Zwar bieten die bestehenden Haftungs- und Versicherungsregime bereits einen rechtlichen Rahmen für KI-bezogene Schäden. Aufgrund der Vielzahl beteiligter Akteure ist die Ermittlung von Schadenursachen und Verantwortlichkeiten jedoch oftmals aufwendig, sodass Schadenfälle häufig erst nach längeren Abstimmungs- und Klärungsprozessen abschließend reguliert werden können.

Eine mögliche Zukunftslösung könnte daher eine parteiübergreifende KI-Versicherung nach dem Vorbild etablierter EAR-/CAR-Deckungen (Erection All Risks / Construction All Risks) sein. Der Grundgedanke besteht darin, die verschiedenen an Entwicklung, Implementierung und Betrieb eines KI-Systems beteiligten Parteien nicht isoliert, sondern gemeinsam in einem Versicherungsmodell abzubilden. Entwickler von Basismodellen, Datenlieferanten, Integratoren, Betreiber sowie Endnutzer würden unter einer gemeinsamen Police zusammengeführt. Ein solcher Ansatz könnte die heute oftmals komplexe Zuordnung von Verantwortlichkeiten vereinfachen, die Schadenregulierung beschleunigen und die Verteilung von Prämien und Risiken auf eine breitere Basis stellen. Gleichzeitig würde die häufig schwierige Abgrenzung zwischen Produktfehler, Implementierungsfehler, Datenproblem, Fehlkonfiguration oder Fehlanwendung des Systems zumindest teilweise entschärft.

Praxisnah wäre ein solches Mehrparteien-Modell zunächst auf ein einzelnes, klar definiertes KI-System oder eine spezifische KI-Anwendung begrenzt. Dadurch ließen sich Risikoprüfung, technische Bewertung und Underwriting deutlich einfacher gestalten. Versicherer könnten die Funktionsweise, Trainings- und Governance-Prozesse sowie die Abhängigkeiten des Systems gezielt analysieren und auf dieser Grundlage eine belastbare Risikoeinschätzung vornehmen. Auch die Schadenursachenanalyse wäre in einem abgegrenzten Anwendungsfall wesentlich transparenter als in komplexen KI-Landschaften mit einer Vielzahl miteinander interagierender Modelle.

Die Skalierbarkeit dieses Ansatzes ist jedoch begrenzt. Er eignet sich ausschließlich für Konstellationen, in denen ein einzelnes dominierendes oder unternehmenskritisches

KI-Modell zentrale Geschäftsprozesse steuert und damit einen klar identifizierbaren Risikokern bildet.

9 Ausblick

Künstliche Intelligenz wird sich in den kommenden Jahren von einer innovativen Einzeltechnologie zu einem selbstverständlichen Bestandteil betrieblicher Wertschöpfung entwickeln. Mit der zunehmenden Verbreitung generativer und agentischer KI-Systeme wird ihre Integration in operative Prozesse, Entscheidungsabläufe und Produkte weiter zunehmen. Damit wächst nicht nur das Potenzial für Effizienzsteigerungen und neue Geschäftsmodelle, sondern auch die Bedeutung eines systematischen Umgangs mit den daraus resultierenden Risiken.

Die vorliegende Betrachtung legt nahe, dass die eigentliche Herausforderung künftig weniger in der Technologie selbst liegen wird als in ihrer Steuerung. KI schafft nur selten völlig neue Schadenarten, verändert jedoch bestehende Risiken durch ihre Geschwindigkeit, Skalierbarkeit und Reichweite. Fehler, Fehlentscheidungen, Haftungsfälle oder Betriebsunterbrechungen können durch KI schneller eintreten, größere Auswirkungen entfalten und sich in komplexen Prozessketten schwieriger nachvollziehen lassen. Die Risikobetrachtung wird sich daher zunehmend von einzelnen Technologien hin zu Fragen der Governance, Kontrolle und Verantwortungszuordnung verschieben.

Mit der weiteren Verbreitung autonomer und agentischer Systeme und Anwendungen mit erhöhtem Handlungsspielraum dürfte sich dieser Trend zusätzlich verstärken. Während heute vielfach die Qualität einzelner KI-Modelle im Mittelpunkt steht, werden künftig Fragen nach menschlicher Aufsicht, Eingriffsmöglichkeiten, Dokumentation, Kontrollmechanismen und organisatorischer Verantwortlichkeit zunehmend an Bedeutung gewinnen. Für Unternehmen wird damit die Fähigkeit, KI wirksam zu steuern, zu einem Wettbewerbs- und Resilienzfaktor.

Der EU AI Act bildet hierfür erstmals einen spezifisch auf künstliche Intelligenz ausgerichteten verbindlichen Regulierungsrahmen. Kurzfristig wird er vielfach als Compliance-Anforderung wahrgenommen werden. Mittelfristig dürfte er jedoch deutlich stärker als Instrument zur Standardisierung und Professionalisierung von KI-Governance wirken. Vergleichbar mit Informationssicherheits- oder Datenschutzstandards könnten sich AI-Act-konforme Prozesse, Dokumentationen und Kontrollmechanismen zu allgemein akzeptierten Qualitätsmerkmalen entwickeln.

Auch für die Versicherungswirtschaft zeichnet sich eine Weiterentwicklung bestehender Marktkonzepte ab. Bereits heute können zahlreiche KI-bezogene Schäden unter etablierten Sach-, Cyber- oder Haftpflichtdeckungen erfasst werden. Gleichzeitig werden Versicherer ihre Bedingungswerke zunehmend daraufhin überprüfen müssen, wie KI-bedingte Risiken systematisch einzuordnen sind. Die Diskussion um »Silent AI« dürfte dabei eine ähnliche Entwicklung durchlaufen wie zuvor die Diskussion um »Silent Cyber«.

Insbesondere bei Eigenschäden infolge von Fehlentscheidungen oder Leistungsdefiziten von KI-Systemen besteht noch Entwicklungsbedarf. Erste Garantie- und Performance-Konzepte zeigen, dass sich der Markt bereits mit der Frage beschäftigt, wie modell- und leistungsbezogene Risiken künftig versicherbar gemacht werden können. Die Grenze einer solchen Entwicklung bleibt jedoch das originäre Unternehmerrisiko: Wirtschaftlich nachteilige Geschäftsentscheidungen oder Fehlsteuerungen lassen sich auch dann nicht ohne Weiteres versichern, wenn sie durch

KI-Systeme vorbereitet oder beeinflusst wurden. Es ist zu erwarten, dass sich in diesem Bereich neue Deckungsansätze, Parametrien und Underwriting-Modelle herausbilden werden.

Mit zunehmender Verbreitung von KI wird sich auch die Diskussion um Haftung, Verantwortungszuordnung und Versicherbarkeit von Drittschäden weiter intensivieren. Insbesondere dort, wo KI Entscheidungen vorbereitet oder autonom ausführt, werden bestehende Haftungsregime und Versicherungskonzepte an neue technische und organisatorische Gegebenheiten angepasst werden müssen. Dabei ist durch die komplexe Verzahnung zwischen dem originären Hersteller der KI, dem KI-Nutzer, der diese für seine individuellen Zwecke adaptiert und ändert, sowie dessen Kunden, die Frage der Verantwortungszuordnung nicht einfach zu beantworten. Im Zweifel wird dies eine Einzelfrage bleiben, die im individuellen Risikomanagement geprüft und im Kontext von bestehenden Versicherungskonzepten zu bewerten sein wird. Solange dabei der Mensch als letzte Kontrollinstanz (Human in the Loop) verbleibt, sollten Tätigkeiten weiterhin unproblematisch sein, sodass die Anwendung an dieser Grenze und darüber hinaus im Fokus der weiteren Diskussion stehen wird. Gleichzeitig bleibt nicht auszuschließen, dass Versicherer die Idee für einen KI-Ausschluss einbringen werden.

Neben regulatorischen und versicherungstechnischen Fragestellungen wird verantwortungsvoller KI-Einsatz zunehmend auch zu einer Vertrauensfrage. Kunden, Mitarbeitende und Geschäftspartner erwarten, dass KI-Systeme nachvollziehbar, fair und unter angemessener menschlicher Aufsicht eingesetzt werden. Unternehmen, die Transparenz, Governance und Verantwortlichkeit frühzeitig verankern, können dadurch nicht nur regulatorische Anforderungen erfüllen und Risiken reduzieren, sondern auch Vertrauen schaffen, Akzeptanz fördern und langfristige Wettbewerbsvorteile erzielen.

Vor diesem Hintergrund spricht vieles dafür, KI künftig nicht primär als Technologie, sondern als steuerbares Unternehmensrisiko zu betrachten. Die zentrale Frage wird daher weniger lauten, ob Unternehmen KI einsetzen, sondern wie sie deren Nutzung organisieren, kontrollieren und verantworten. Mit zunehmender Verbreitung von KI werden Governance, Kontrolle und Compliance zu entscheidenden Erfolgsfaktoren für Resilienz, Versicherbarkeit und nachhaltigen wirtschaftlichen Erfolg.

Glossar

Agentic AI (AI Agents)

Agentische KI bezeichnet KI-Systeme, die Aufgaben mit einem vergleichsweise hohen Grad an Selbstständigkeit planen und ausführen können. Sie kombinieren häufig ein KI- bzw. General-Purpose-AI-Modell mit zusätzlichen Komponenten (»Scaffolding«) wie Tool-Zugriff, Speicher, Planungs- und Orchestrierungslogik. Dadurch können sie mehrstufige Aufgaben bearbeiten, externe Systeme ansteuern oder eigenständig Aktionen auslösen. Regulatorisch ist »Agentic AI« keine eigenständige Risikokategorie des EU AI Act. Entscheidend sind vielmehr Zweck, konkrete Anwendung und Rolle des jeweiligen Akteurs. Wird ein agentisches KI-System für einen in Annex III erfassten Anwendungsfall eingesetzt und erfüllt es die dortigen Voraussetzungen, kann es als Hochrisiko-KI-System einzustufen sein. Etwaige Pflichten für zugrunde liegende GPAI-Modelle und für das darauf aufbauende KI-System sind dabei rollenbezogen voneinander zu unterscheiden.

Agent Sandboxing

Sandboxing bezeichnet bei KI-Agenten die Ausführung des Agenten in einer isolierten, kontrollierten Umgebung. Diese Umgebung fungiert als Tresor, der den Zugriff des Agenten auf das restliche System, die Datenbanken oder das Netzwerk einschränkt. Sie schützt also den Inhalt nicht vor der Außenwelt, sondern die Außenwelt vor dem Inhalt. Ziel ist es, die Kerninfrastruktur vor unvorhersehbarem oder schädlichem Verhalten des Agenten zu schützen.

Diese Sicherheitsmaßnahme bringt sowohl signifikante Vorteile als auch operative Herausforderungen mit sich. Eine starke Isolierung schützt das Gesamtsystem effektiv vor Fehlfunktionen oder Manipulationen, falls der Agent etwa durch eine Prompt-Injection kompromittiert wurde. Es ist eine robuste Verteidigung. Dennoch muss man die Kehrseite beachten: Die strikte Trennung kann die Performance spürbar beeinträchtigen und den notwendigen Datenaustausch mit anderen Systemen erschweren.

Autonomie & Unvorhersehbarkeit

Agenten können Entscheidungen treffen, die außerhalb der beabsichtigten Logik liegen (»Abweichung«). Da sie eigenständig planen, ist ihr Verhalten oft schwieriger vorhersehbar als das rein textbasierte Ergebnis eines Standard-LLMs.

Bias (Verzerrung)

Bias bezeichnet systematische Verzerrungen in Daten, Modellen oder Entscheidungsprozessen, die dazu führen können, dass ein KI-System bestimmte Personen, Gruppen oder Sachverhalte wiederkehrend benachteiligt oder fehlerhaft bewertet. Ursachen können unter anderem unausgewogene oder nicht repräsentative Trainingsdaten, fehlerhafte Modellannahmen, ungeeignete Zielgrößen oder die Art der Anwendung sein. Im KI-Kontext ist Bias insbesondere relevant, wenn automatisierte Bewertungen oder Entscheidungen Auswirkungen auf Zugang, Preise, Auswahlverfahren oder Leistungen haben.

LLMs / Generative Modelle

Diese Systeme sind primär darauf optimiert, Muster in Daten zu erkennen und statistisch wahrscheinliche Inhalte (Text, Bild, Code) zu generieren. Sie sind passiv und »stateless«: Sie reagieren auf einen Eingabeprompt und beenden ihre Arbeit nach der Generierung.

KI-Agenten

Sie nutzen ein LLM als »Gehirn« für die Planung und das Verständnis, ergänzen dieses jedoch um eine Handlungsschleife (Agentic Workflow). Ein Agent kann iterativ planen, sein Vorgehen basierend auf Ergebnissen korrigieren und über APIs direkt mit Software, Datenbanken oder dem Internet interagieren, um ein übergeordnetes Ziel zu erreichen.

Das Hauptrisiko liegt in der Kombination aus Autonomie und Zugriff auf produktive Systeme. Die Hauptgefahren sind:

Privilegien & Sicherheitsrisiken

Da Agenten Zugriff auf Schnittstellen (APIs) und Datenbanken haben, können sie bei Fehlverhalten oder Manipulation (z. B. durch »Prompt Injection«) weitreichende Schäden anrichten, etwa durch unbefugte Datenzugriffe oder automatisiertes Versenden sensibler Informationen.

Dynamisches Laufzeit-Risiko

Im Gegensatz zu klassischen Sicherheitslücken (CVEs), die statisch im Code liegen und gepatcht werden können, entstehen Risiken bei KI-Agenten oft dynamisch während der Ausführung (»in-flight«). Eine reaktive Patch-Strategie greift hier oft zu kurz.

Fehlende Kontrollschleifen

Ohne installierte Sicherheitsmechanismen wie »Human-in-the-Loop« (menschliche Freigabeprozesse) können Fehler in der Logik oder Halluzinationen ungefiltert in echten Geschäftsprozessen landen, was zu finanziellen Schäden oder Compliance-Verstößen führen kann.

Hintergrund zu Claude Mythos

Bei Claude Mythos handelt es sich um ein KI-Modell des US-Unternehmens Anthropic, das als hochgradige »Dual-Use-Technologie« eingestuft wird. Es wurde gezielt darauf trainiert, versteckte Sicherheitslücken in komplexen Programmen aufzuspüren. Weil Mythos völlig autonom mehrstufige Angriffe plant, bewertet Anthropic das Modell als zu gefährlich für die Öffentlichkeit. Im Rahmen des »Project Glasswing« wird es daher nur exklusiv an ausgewählte IT-Sicherheitsfirmen und Behörden wie das BSI vergeben. Das System agiert dabei extrem agentisch. Es hat bereits tausende gefährliche Schwachstellen in Browsern gefunden. In den falschen Händen könnte diese Technologie jedoch als verheerende digitale Waffe gegen Stromnetze oder Banken eingesetzt werden.

Wirtschaftlicher Hintergrund

Experten vermuten neben reinen Sicherheitsgründen auch Infrastruktur- und Kostengründe hinter der Geheimhaltung: Die Rechenleistung für die Ausführung ist enorm und die Analyse einer einzelnen Codebasis kostet bis zu Hunderttausende Euro.

Prompt Injection

Ein Prompt Injection Angriff ist eine Methode, bei der ein Angreifer gezielt

manipulierte Eingaben (Prompts) an ein KI-Modell sendet, um dessen ursprüngliche Programmierung oder Sicherheitsanweisungen zu überschreiben.

Im Kern geht es bei diesem Angriff darum, die ursprünglichen System-Instruktionen der KI durch geschickte Formulierungen zu überlagern oder schlichtweg zu ignorieren. Die KI akzeptiert dann die Anweisungen des Angreifers als ihre neue Wahrheit. Das ist ein ernstes Problem. Während dies bei einem simplen Chatbot oft nur zu kuriosen Antworten führt, ist die Lage bei KI-Agenten kritisch. Da sie Zugriff auf externe Tools haben, könnten Angreifer sie manipulieren, um sensible Daten zu extrahieren oder E-Mails in fremdem Namen zu versenden.

Man kann es sich wie eine digitale Identitätstäuschung vorstellen: Der Angreifer verstellt seine Stimme so, dass das »Gehirn« des Agenten glaubt, der Befehl komme direkt vom Systemadministrator, obwohl er vom Benutzer stammt.

Sicherheitsexperten unterscheiden dabei primär zwischen zwei Hauptformen der Manipulation. Bei der direkten Prompt Injection, auch Jailbreaking genannt, gibt der Nutzer die schädliche Anweisung unmittelbar in den Chat ein. Die indirekte Variante ist hingegen deutlich tückischer. Hierbei platziert ein Angreifer den manipulativen Text auf einer Webseite, die der Agent während einer Recherche liest. Sobald die KI diesen Text verarbeitet, führt sie den darin versteckten Befehl unbemerkt aus.

Da dies eine der größten Schwachstellen für autonome Agenten ist, gehört die Abwehr von Prompt Injection zu den zentralen Aufgaben der Input-Validierung und des Guardrailings, die wir vorhin besprochen haben.

Guardrail

Eine Guardrail (zu Deutsch: »Leitplanke«) bezeichnet im Kontext von KI-Systemen und Agenten einen vordefinierten Sicherheits- oder Kontrollmechanismus, der sicherstellt, dass das KI-Modell innerhalb vorgegebener Grenzen agiert.

Stellen Sie sich das wie bei einer echten Leitplanke auf einer Autobahn vor: Sie verhindert, dass das Fahrzeug (die KI) von der Straße abkommt, auch wenn der Fahrer (der Nutzer oder ein bössartiger Befehl) versucht, das Fahrzeug in die Leitplanke oder den Graben zu steuern.

Was genau machen Guardrails? Guardrails fungieren als Filter- oder Prüfschicht, die zwischen dem Nutzer und dem KI-Modell (oder zwischen dem Agenten und seinem Tool/System) geschaltet ist. Sie überwachen die Kommunikation in Echtzeit und greifen ein, wenn bestimmte Regeln verletzt werden.

In der Praxis erfüllen Guardrails verschiedene kritische Funktionen zum Schutz der Infrastruktur. Input-Guardrails dienen als Eingangskontrolle und prüfen, ob die Nutzeranfrage sicher ist, indem sie beispielsweise Versuche von Prompt Injection filtern. Auf der anderen Seite stehen Output-Guardrails. Diese überwachen die Antworten der KI auf Halluzinationen oder die Preisgabe sensibler Firmendaten, bevor diese den Nutzer erreichen. Schließlich stellen Governance-Guardrails sicher, dass regulatorische Anforderungen wie die DSGVO oder der EU AI Act strikt eingehalten werden, indem sie etwa den unbefugten Datenfluss an externe Server blockieren.

KI-Firewall

Eine KI-Firewall ist ein Sicherheitsmechanismus, der Eingaben, Ausgaben und Systeminteraktionen eines KI-Systems überwacht, filtert und bei Bedarf blockiert. Sie

kann dazu beitragen, Prompt-Injection-Angriffe, unzulässige Datenabflüsse, Richtlinienverstöße oder riskante Tool-Aufrufe frühzeitig zu erkennen und zu begrenzen. Besonders relevant ist sie bei KI-Agenten mit Zugriff auf APIs, Datenbanken, E-Mail-, CRM- oder Workflow-Systeme. Eine KI-Firewall ersetzt keine Governance- oder Freigabeprozesse, kann aber als technische Schutzschicht helfen, Fehlverhalten und Manipulationen während der Ausführung zu begrenzen.

Sandbox

Eine Sandbox ist eine technisch abgegrenzte Test- oder Ausführungsumgebung, in der KI-Systeme, Modelle oder Agenten kontrolliert erprobt werden können, ohne unmittelbar auf produktive Systeme, sensible Daten oder geschäftskritische Prozesse einzuwirken. Sie dient dazu, Verhalten, Fehlerfolgen, Sicherheitsrisiken und Wechselwirkungen vor dem produktiven Einsatz zu prüfen. Im KI-Kontext ist eine Sandbox besonders relevant für Systeme mit Tool-Zugriff, Agentenfunktionen oder Prozessautomatisierung, weil sie den Handlungsspielraum begrenzt und Fehlverhalten vor der Übertragung in reale Geschäftsprozesse sichtbar machen kann.

KI-Modell

Ein KI-Modell ist der technische Kern eines KI-Systems, der auf Basis von Trainingsdaten Muster, Zusammenhänge oder Wahrscheinlichkeiten abbildet und daraus Ergebnisse wie Prognosen, Klassifikationen, Empfehlungen oder Inhalte erzeugt. Das Modell kann durch fehlerhafte Daten, Manipulation, ungeeignete Anpassungen, Verlust von Modellparametern oder technische Störungen beeinträchtigt werden. Schäden am KI-Modell können dazu führen, dass Ergebnisse unzuverlässig werden oder das Modell ganz oder teilweise neu trainiert, wiederhergestellt oder ersetzt werden muss.

KI-Anwendung

Eine KI-Anwendung ist die konkrete Einbindung eines KI-Modells in einen fachlichen oder technischen Nutzungskontext, etwa in Kundenservice, Underwriting, Schadenbearbeitung, Betrugserkennung oder Dokumentenverarbeitung. Sie umfasst neben dem Modell auch Schnittstellen, Datenquellen, Nutzeroberflächen, Berechtigungen, Workflows und Kontrollmechanismen. Wird eine KI-Anwendung beschädigt, verändert, gelöscht oder unbrauchbar, kann dies nicht nur das Modell selbst betreffen, sondern auch die Nutzung im Geschäftsprozess, die Verfügbarkeit der Anwendung oder die Qualität der ausgegebenen Ergebnisse.

KI-Asset

Ein KI-Asset ist ein wirtschaftlich werthaltiger digitaler Vermögenswert, der für Entwicklung, Betrieb oder Nutzung eines KI-Systems erforderlich ist. Dazu können KI-Modelle, Trainings-, Test- und Validierungsdaten, Embeddings, Prompt-Bibliotheken, Vektordatenbanken, Konfigurationen, Fine-Tunings, Dokumentationen oder Monitoring-Informationen gehören. Eine Beeinträchtigung eines KI-Assets liegt vor, wenn dieses beschädigt, verändert, gelöscht, manipuliert oder nicht mehr nutzbar ist und dadurch Wiederherstellungsaufwand, Betriebsunterbrechungen oder sonstige wirtschaftliche Schäden entstehen können.

Netzwerksicherheitsverletzung

Eine Netzwerksicherheitsverletzung bezeichnet einen Vorfall, bei dem die Sicherheit, Vertraulichkeit, Integrität oder Verfügbarkeit eines IT-Systems, Netzwerks oder von

darin verarbeiteten Daten beeinträchtigt wird. Typische Beispiele sind unbefugte Zugriffe, Schadsoftware, Datenmanipulation, Datenabfluss, Denial-of-Service-Angriffe oder die Kompromittierung von Zugangsdaten. Im Versicherungskontext ist der Begriff insbesondere relevant, weil Cyberversicherungen häufig an ein definiertes Cyberereignis oder eine Netzwerksicherheitsverletzung als auslösenden Versicherungsfall anknüpfen.

Anhang

A.1: Zentrale Pflichten für KI-Nutzer aus dem AI Act je nach Risikoklassifizierung:

	Zentrale Pflichten / Anforderungen	Rechtsgrundlage
Allgemeine Verbote und Querschnittspflichten	Verbot bestimmter KI-Praktiken, z. B. bestimmter manipulativer Anwendungen oder Social Scoring	Art. 5
	Anforderungen an KI-Kompetenz (AI Literacy) von Nutzern und Entscheidungsträgern	Art. 4
	Transparenzpflichten, z. B. gegenüber betroffenen Personen bei bestimmten KI-Anwendungen	Art. 50
General-Purpose AI (GPAI)	Technische Dokumentation, Zusammenfassung der Trainingsinhalte und Informationen zu urheberrechtlichen Vorgaben	Art. 51–56
	Zusätzliche Anforderungen für GPAI-Modelle mit systemischem Risiko	Art. 51–56
Hochrisiko-KI-Systeme, insbesondere nach Annex III	Risikomanagement	Art. 9
	Datengovernance	Art. 10
	Technische Dokumentation	Art. 11, Annex IV
	Logging und Nachvollziehbarkeit	Art. 12
	Transparenz und Informationen für Betreiber (»Instructions for Use«)	Art. 13

	Zentrale Pflichten / Anforderungen	Rechtsgrundlage
	Menschliche Aufsicht (Human Oversight)	Art. 14
	Genauigkeit, Robustheit und Cybersicherheit	Art. 15
	Qualitätsmanagementsystem	Art. 17
	Grundrechte- Folgenabschätzung (Fundamental Rights Impact Assessment)	Art. 27
	Konformitätsbewertung	Art. 43
	Post-Market-Monitoring und Meldung schwerwiegender Vorfälle	Art. 72, 73

A.2: KI-Risiko-Matrix

KI-Risiko	Medienhaft- pflicht	Berufs- haftpflicht	Produkt- haftpflicht	Betriebs- haftpflicht	Geistiges Eigentum	Cyber	Vertrauens- schaden	D&O
Haftung für fehlerhafte Produkte und Dienstleistungen	■	■	■	■	■	■	■	■
Urheberrecht, Marken- oder Dienstleistungsrechtsverletzung	■	■	■	■	■	■	■	■
Patentrechtsverletzung	■	■	■	■	■	■	■	■
Diskriminierung / Voreingenommenheit	■	■	■	■	■	■	■	■
Verleumdung, üble Nachrede, Beleidigung	■	■	■	■	■	■	■	■
Personenschäden	■	■	■	■	■	■	■	■
Sachschäden	■	■	■	■	■	■	■	■
Datenschutz- und Sicherheitsverletzungen	■	■	■	■	■	■	■	■
Verlust finanzieller Vermögenswerte	■	■	■	■	■	■	■	■
Marktmanipulation	■	■	■	■	■	■	■	■
Täuschung	■	■	■	■	■	■	■	■
Robotik	■	■	■	■	■	■	■	■
Produktrückruf	■	■	■	■	■	■	■	■
Betriebsunterbrechung	■	■	■	■	■	■	■	■
Verletzung der Pflichten von Führungskräften	■	■	■	■	■	■	■	■

■ Allgemein verfügbar
 ■ Limitiert
 ■ Ausgeschlossen

Literaturverzeichnis

Allianz Commercial (2026): Allianz Risk Barometer 2026.

<https://commercial.allianz.com/content/dam/onemarketing/commercial/commercial/reports/allianz-risk-barometer-2026.pdf>

beck-aktuell (2026): »Atemlos durch die Nacht« im KI-Modell kann Urheberrechte verletzen, 31. Juli 2026.

Bild-Kunst (2026): Hamburger OLG-Urteil im Fall LAION, 2026.

Bild-Kunst (2026): Text- und Data-Mining – Entscheid des Landgerichts Hamburg, Januar 2026.

Bird & Bird (2025): Higher Regional Court Hamburg Confirms AI Training was Permitted, 17. Dezember 2025.

Bitsea (2025): Wer heute noch KI-generierten Code in Umlauf bringt, hat bedingten Vorsatz für Rechtsverletzungen, 27. Oktober 2025.

Bundesgerichtshof (BGH) (2025): Urteil vom 14. Oktober 2025, Az. II ZR 78/24, abrufbar über openJur.

<https://openjur.de/u/2538374.html>

Bundesrepublik Deutschland (2026): Gesetz zur Durchführung der Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung weiterer Rechtsakte (Gesetz zur Durchführung der Verordnung über künstliche Intelligenz).

<https://www.recht.bund.de/bgbl/1/2026/223/VO.html>

CMS (2025): KI-generierter Softwarecode in der Due Diligence, 30. Oktober 2025.

Computerwoche (2026): Allianz Risk Barometer 2026: KI springt auf Platz 2, 14. Januar 2026.

<https://www.computerwoche.de/article/4116684/allianz-risk-barometer-2026-ki-springt-auf-platz-2.html>

DATEV magazin (2026): Unternehmen haften für falsche Aussagen ihres KI-Chatbot, 13. Mai 2026.

<https://www.datev-magazin.de/nachrichten-steuern-recht/recht/unternehmen-haften-fuer-falsche-aussagen-ihres-ki-chatbots-146688>

Europäische Union (2022): Verordnung (EU) 2022/2554 des Europäischen Parlaments und des Rates vom 14. Dezember 2022 über die digitale operationale Resilienz im Finanzsektor und zur Änderung der Verordnungen (EG) Nr. 1060/2009, (EU) Nr. 648/2012, (EU) Nr. 600/2014, (EU) Nr. 909/2014 und (EU) 2016/1011 (Digital Operational Resilience Act – DORA).

<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32022R2554>

Europäische Union (2024): Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202401689

Europäische Union (2026): Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI).

https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=OJ:L_202601744

GASTRO.at (2026): KI sorgt für kalte Pizza, 26. Mai 2026.

<https://gastroportal.at/pizza-hut-wegen-ki-auf-100-millionen-verklagt/96940/>

Gesamtverband der Deutschen Versicherungswirtschaft e. V. (GDV) (2020): Allgemeine Bedingungen für die Elektronikversicherung (ABE 2020).

Gesamtverband der Deutschen Versicherungswirtschaft e. V. (GDV) (2020): Allgemeine Bedingungen für die Maschinenversicherung von stationären Maschinen (AMB 2020).

Gesamtverband der Deutschen Versicherungswirtschaft e. V. (GDV) (2024): Allgemeine Versicherungsbedingungen für die Cyberrisiko-Versicherung (AVB Cyber), Stand Februar 2024.

GEMA (2025): Erstes KI-Grundsatzurteil in Europa: GEMA setzt sich gegen OpenAI durch, 11. November 2025.

Höra, Knut / Schubach, Arno (Hrsg.) (2022): Münchener Anwaltshandbuch Versicherungsrecht, 5., überarbeitete und erweiterte Auflage, München: C.H. Beck, ISBN 978-3-406-76903-0. Sack Fachmedien

IT BOLTWISE (2026): Landgericht München stoppt unbefugtes KI-Musiktraining von Suno – Urteil nicht rechtskräftig*, 3. August 2026.

<https://www.it-boltwise.de/landgericht-muenchen-stoppt-unbefugtes-ki-musiktraining-von-suno-urteil-nicht-rechtskraeftig.html>

it-daily.net (2026): Pizza Hut: Franchisenehmer fordert 100 Millionen Dollar wegen KI-Schäden, 19. Mai 2026.

<https://www.it-daily.net/shortnews/pizza-hut-franchisenehmer-klage>

Justiz Bayern (2025): Pressemitteilung des Landgerichts München I vom 11. November 2025.

Justiz Bayern (2026): Pressemitteilung des Landgerichts München I vom 31. Juli 2026.

Kanzlei Fürstenow (2026): Haftungsrisiken für Führungskräfte in 2026, 29. April 2026.

<https://www.kanzlei-fuerstenow.de/haftungsrisiken-fuer-fuehrungskraefte-in-2026/>

Kanzlei Kramarz (2025): Urheberrecht für KI-generierten Code: Eine Frage der Beweislast, 21. Juni 2025.

KPMG Law (2026): Business Judgement Rule und KI: So haften Organe für Entscheidungen*, 19. März 2026.

<https://kpmg-law.de/business-judgement-rule-und-ki-so-haften-organe-fuer-entscheidungen/>

Landgericht München I (2025): Urteil vom 11. November 2025, Az. 42 O 14139/24, abrufbar über openJur.

Legal Tribune Online (LTO) (2024): KI und Urheberrecht: Wegweisendes Urteil des LG Hamburg*, 2024.

Lime Connect / Lime Technologies (2026): Steil, Tamina: Chatbot-Haftung nach OLG Hamm: Was Unternehmen jetzt wissen müssen, 15. Juni 2026.

<https://connect.lime-technologies.com/de/blog/chatbot-haftung-olg-hamm/>

Manufaktur für Finanzplanung (2026): KI-Haftungsrichtlinien 2026: Der ultimative Guide für GmbH-Geschäftsführer zur Absicherung gegen das neue Haftungs-Vakuum, 22. Februar 2026.

<https://blog.manufaktur-fuer-finanzplanung.de/2026/02/22/ki-haftungsrichtlinien-2026-der-ultimate-guide-fuer-gmbh-geschaeftsfuehrer-zur-absicherung-gegen-das-neue-haftungs-vakuum/>

Oberlandesgericht Hamburg (2025): Urteil vom 10. Dezember 2025, Az. 5 U 104/24, abrufbar über openJur.

Rehkatsch Rechtsanwälte (2026): Kneschke gegen LAION: KI-Training & § 44b UrhG, 26. Juni 2026.

Säcker, Franz Jürgen / Rixecker, Roland / Oetker, Hartmut / Limperg, Bettina (Hrsg.) (2021–2025): Münchener Kommentar zum Bürgerlichen Gesetzbuch, 9. Auflage, München: C.H. Beck, ISBN 978-3-406-76670-1. Die 9. Auflage erschien bandweise von 2021 bis 2025. SLUB Dresden

t3n (2026): Brien, Jörn: KI sorgt für kalte Pizza: Franchise-Nehmer verklagt Pizza Hut auf 100 Millionen Dollar Schadenersatz, 25. Mai 2026.

<https://t3n.de/news/ki-pizza-hut-klage-100-millionen-1743300/>

Tabassi, Elham (2023): Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, National Institute of Standards and Technology, Gaithersburg, MD.

<https://doi.org/10.6028/NIST.AI.100-1>

Tageskarte (2026): Pizza-Hut-Franchisenehmer verklagt Kette wegen KI-System auf 100 Millionen Dollar, 18. Mai 2026.

<https://www.tageskarte.io/gastronomie/detail/pizza-hut-franchisenehmer-verklagt-kette-wegen-ki-system-auf-100-millionen-dollar.html>

Veith, Jürgen / Gräfe, Jürgen / Gebert, Yvonne (Hrsg.) (2020): Der Versicherungsprozess. Ansprüche und Verfahren, 4. Auflage, Baden-Baden: Nomos, 1.488 Seiten, ISBN 978-3-8487-5521-9.

Wagner, Gerhard (2024): »Die Richtlinie über KI-Haftung – weniger ist mehr!«, in: RDi – Recht Digital, Editorial vom 4. September 2024.

Wettbewerbszentrale (2026): OLG Hamm lässt Unternehmen für Aussagen seines Chatbots haften, Volltext verfügbar, 3. Juni 2026.

<https://www.wettbewerbszentrale.de/olg-hamm-laesst-unternehmen-fuer-aussagen-seines-chatbots-haften-volltext-verfuegbar/>

WirtschaftsWoche (2026): Tödtmann, Claudia: KI im Kundenservice: Unternehmen haften für ihre Chatbots – auch für falsche Zusagen, 21. Juni 2026.

<https://www.wiwo.de/erfolg/management/ki-im-kundenservice-unternehmen-haften-fuer-falsche-antworten/100231910.html>

Bitkom vertritt mehr als 2.200 Mitgliedsunternehmen aus der digitalen Wirtschaft. Sie generieren in Deutschland gut 200 Milliarden Euro Umsatz mit digitalen Technologien und Lösungen und beschäftigen mehr als 2 Millionen Menschen. Zu den Mitgliedern zählen mehr als 1.000 Mittelständler, über 700 Startups und nahezu alle Global Player. Sie bieten Software, IT-Services, Telekommunikations- oder Internetdienste an, stellen Geräte und Bauteile her, sind im Bereich der digitalen Medien tätig, kreieren Content, bieten Plattformen an oder sind in anderer Weise Teil der digitalen Wirtschaft. 82 Prozent der im Bitkom engagierten Unternehmen haben ihren Hauptsitz in Deutschland, weitere 8 Prozent kommen aus dem restlichen Europa und 7 Prozent aus den USA. 3 Prozent stammen aus anderen Regionen der Welt. Bitkom fördert und treibt die digitale Transformation der deutschen Wirtschaft und setzt sich für eine breite gesellschaftliche Teilhabe an den digitalen Entwicklungen ein. Ziel ist es, Deutschland zu einem leistungsfähigen und souveränen Digitalstandort zu machen.

Herausgeber

Bitkom e.V.

Albrechtstr. 10 | 10117 Berlin

Ansprechpartner/in

Lukas Spohr | Bitkom e.V.

T | +49 175 5848827 | l.spohr@bitkom.org

Verantwortliches Bitkom-Gremium

AK Digital Insurnace & InsurTech

Autorinnen und Autoren

Bastian Bahnmann, FSI Compliance & Business Development Lead Germany, Microsoft | **Dr. Peter Bärnreuther**, Artificial Intelligence Risks Senior Underwriter, Munich Re | **Christian Freudner**, VP Management Liability, Finlex | **Martin Hake**, Chief Underwriter Haftpflicht, R+V Versicherung | **Kerstin Keltner**, Managing Director, Head of Speacilty, AON Austria | **Sima Krause**, AI Governance Manager, ERGO Group AG | **Klemens Lemke**, Underwriting Manager Cyber, Hiscox Deutschland | **Sabine Pawig-Sander**, Non-Executive Director, Finlex | **Hanno Pingsmann**, Gründer & Geschäftsführer, CyberDirekt | **Axel Saß**, Chief Architect, Field CTO Organisation, Red Hat | **Stefan Schmutterer**, Berater Underwriting und Schaden CyberRisk, R+V Versicherung | **Matthias Schneider**, CRO, Markel SE | **Bogdan Werth**, Insurance Lead Germany, Microsoft

Copyright

Bitkom 2026

Diese Publikation stellt eine allgemeine unverbindliche Information dar. Die Inhalte spiegeln die Auffassung im Bitkom zum Zeitpunkt der Veröffentlichung wider. Obwohl die Informationen mit größtmöglicher Sorgfalt erstellt wurden, besteht kein Anspruch auf sachliche Richtigkeit, Vollständigkeit und/oder Aktualität, insbesondere kann diese Publikation nicht den besonderen Umständen des Einzelfalles Rechnung tragen. Eine Verwendung liegt daher in der eigenen Verantwortung des Lesers. Jegliche Haftung wird ausgeschlossen. Alle Rechte, auch der auszugsweisen Vervielfältigung, liegen beim Bitkom oder den jeweiligen Rechteinhabern.