protectifAI

Die Risiken hinter dem größten Produktivitätssprung unserer Zeit

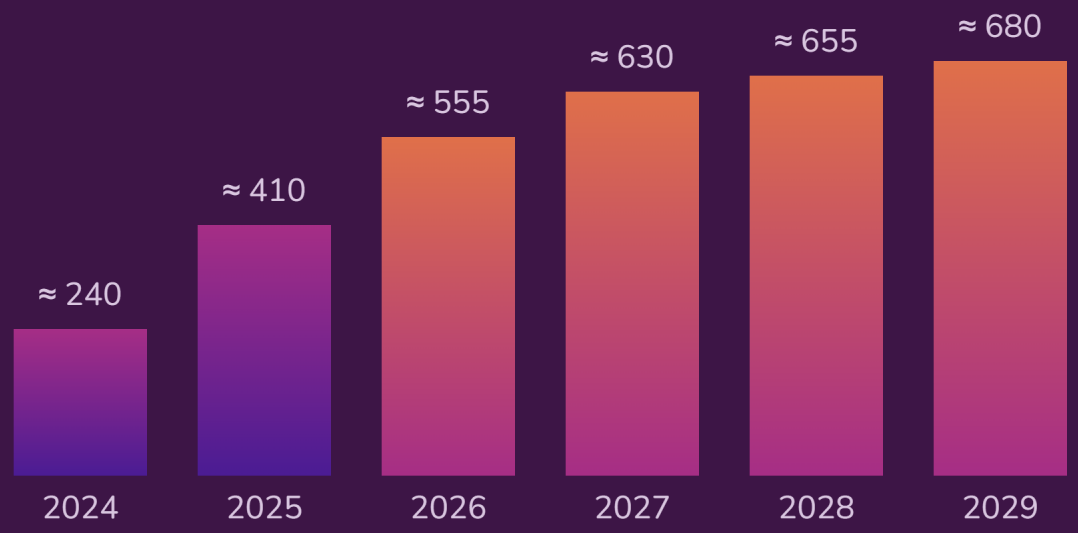
Sarah Günther · Geschäftsführerin protectifAI

KI IN UNTERNEHMEN · UNTENEHMERTAG FILDERSTADT

// Einordnung · Der Hype

Das Tempo ist atemberaubend

KI-Investitionen der US-Hyperscaler, Mrd. $



Microsoft, Amazon, Alphabet, Meta, Oracle · ab 2026 Prognose

KI-Modell Claude findet in 2 Wochen über 100 Firefox-Bugs, mehr als die Community in 2 Monaten

Heise, 6.3.2026

Rund 890 Mio. € für Yann LeCuns Ami Labs: größte Anschubfinanzierung Europas

Handelsblatt, 11.3.2026

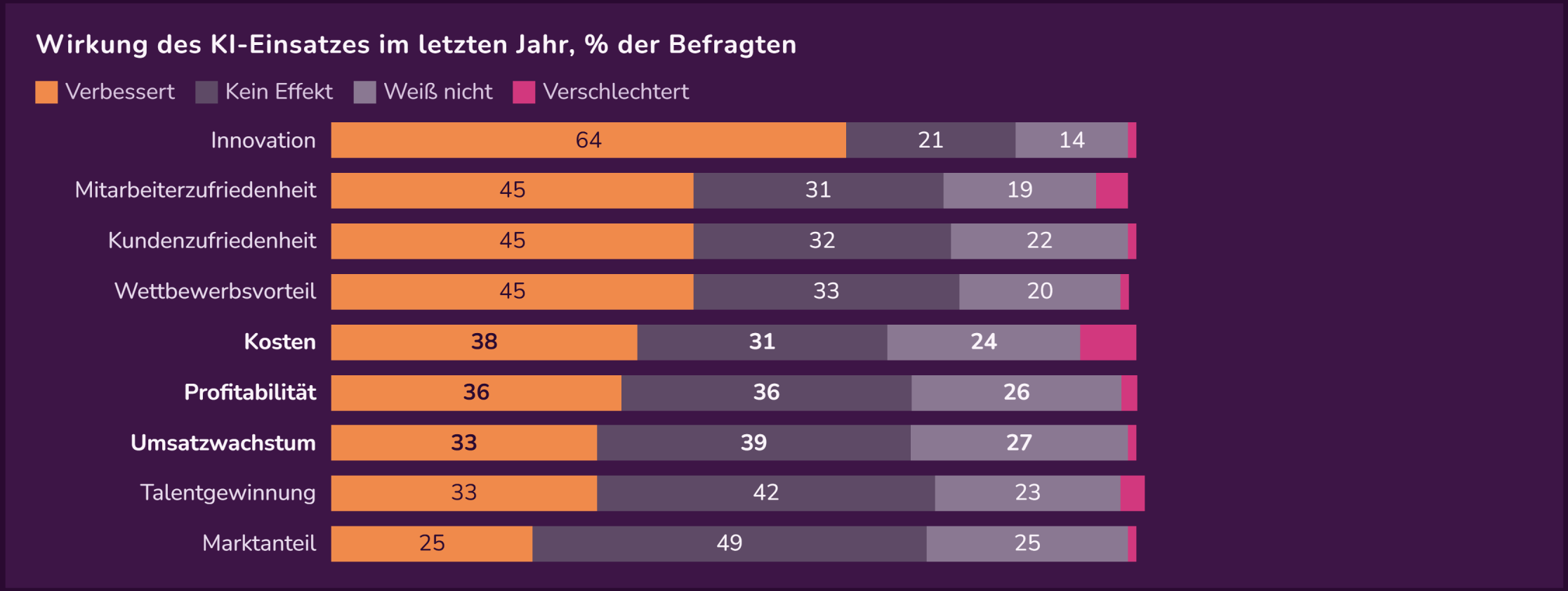
KI-Streit: Pentagon stuft Anthropic offiziell als Sicherheitsrisiko ein

Handelsblatt, 6.3.2026

Investitionen: Werte aus Grafik gerundet

// Einordnung · Die Realität

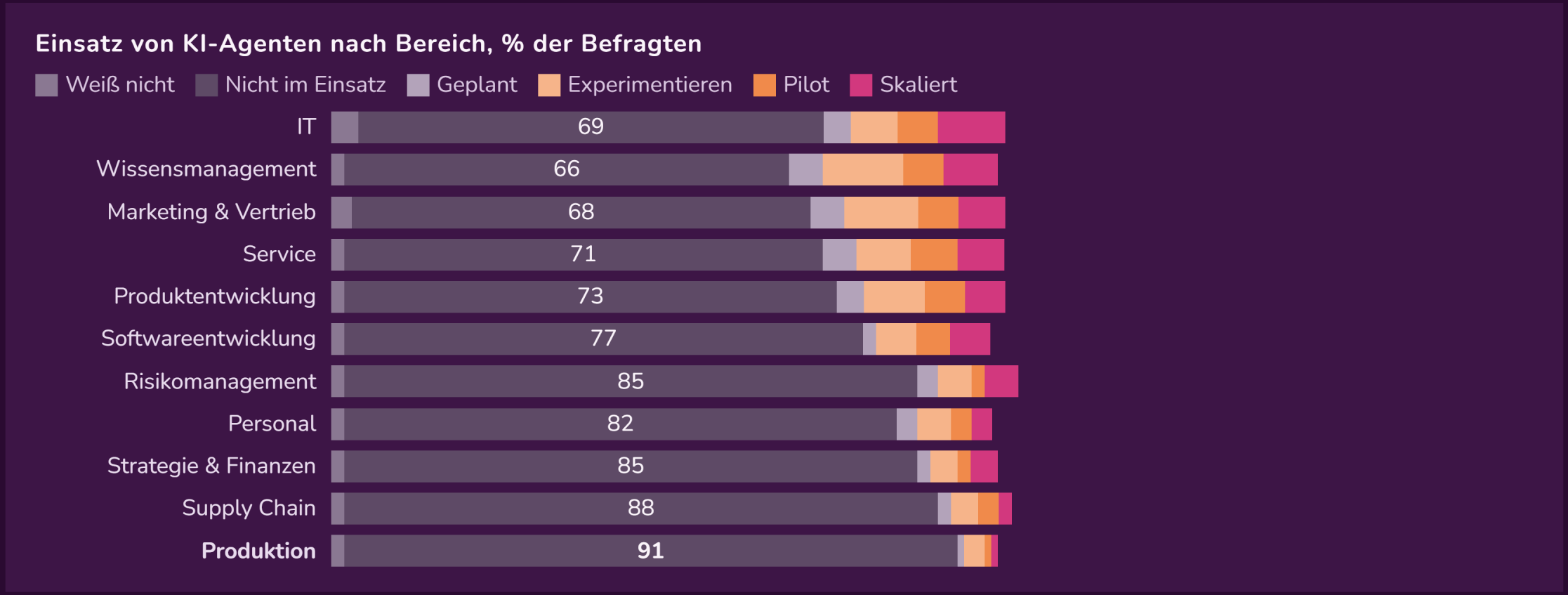
Reifegrad generativer KI im Unternehmen



Quelle: McKinsey Global Survey „The state of AI“, 2025 · rund 2.000 Befragte mit regelmäßiger KI-Nutzung

// Einordnung · Die Realität

Reifegrad agentischer KI in Unternehmen



Quelle: McKinsey Global Survey „The state of AI“, 2025, rund 2.000 Befragte

VIER ECHTE FÄLLE AUS DEM UNTERNEHMENSALLTAG

Kein einziger davon war ein Hackerangriff.

Es waren Mitarbeitende, die produktiv sein wollten.

// Fall 1 · Schatten-KI

Ihr Wissen landet in fremden KIs

Samsung, 2023

Ingenieure kopieren Quellcode und Besprechungsprotokolle in ChatGPT. Die Daten liegen beim Anbieter, der Konzern sperrt die Nutzung weltweit.

Geteilte ChatGPT-Chats, 2025

Per Link geteilte Chats tauchen massenhaft in der Google-Suche auf, darunter Geschäftsinterna, Bewerbungen und Kundendaten.

// Fall 1 · Schatten-KI

Ihr Wissen landet in fremden KIs

Samsung, 2023

Ingenieure kopieren Quellcode und Besprechungsprotokolle in ChatGPT. Die Daten liegen beim Anbieter, der Konzern sperrt die Nutzung weltweit.

Geteilte ChatGPT-Chats, 2025

Per Link geteilte Chats tauchen massenhaft in der Google-Suche auf, darunter Geschäftsinterna, Bewerbungen und Kundendaten.

Bei Ihnen

Daten werden von Mitarbeitern in private Accounts eingegeben, weil es kein offizielles Tool gibt. Daten landen ohne ihr Wissen beim Anbieter.

Ab morgen

Freigegebenes Business-Tool plus Daten-Ampel: Grün darf rein, Gelb nur ins freigegebene Tool, Rot nie.

// Fall 2 · Berechtigungen und Agenten

Die KI sieht alles und handelt selbst

Microsoft Copilot

Viele Unternehmen stoppen den Rollout: Über alte Freigaben findet die KI Gehaltslisten, Vorstandsunterlagen und Personalakten für jeden.

PocketOS, 2026

Ein KI-Agent löscht die Produktivdatenbank des eignen Unternehmens in 9 Sekunden, letztes Back-Up 3 Monate alt, mehrere Wochen Ausfall des Systems, Kunden handlungsunfähig.

// Fall 2 · Berechtigungen und Agenten

Die KI sieht alles und handelt selbst

Microsoft Copilot

Viele Unternehmen stoppen den Rollout: Über alte Freigaben findet die KI Gehaltslisten, Vorstandsunterlagen und Personalakten für jeden.

PocketOS, 2026

Ein KI-Agent löscht die Produktivdatenbank des eignen Unternehmens in 9 Sekunden, letztes Back-Up 3 Monate at, mehrere Wochen Ausfall des Systems, Kunden handlungsunfähig.

Bei Ihnen

Alte Freigaben wirken sich auf die KI-Rechte aus. Mitarbeiter kommen an Daten, die nicht für Sie vorgesehen waren, werden gelöscht oder gelangen nach außen.

Ab morgen

Data-Security und Hygiene. Die KI sieht nur, was der Fragende sehen darf. Agenten bekommen Praktikanten-Rechte (Harness): kein Zahlen, kein Löschen ohne Freigabe.

// Hintergrund · So funktioniert ein Sprachmodell

Die KI rät das wahrscheinlichste nächste Wort

1 · Text wird in Bausteine (Tokens) zerlegt

Die Kraft fahr zeug -H aft pflicht vers icher ung ist eine Pflicht v ers icherung , die jedes Auto braucht , um auf der Stra ße fahr en zu dürfen .

// Hintergrund · So funktioniert ein Sprachmodell

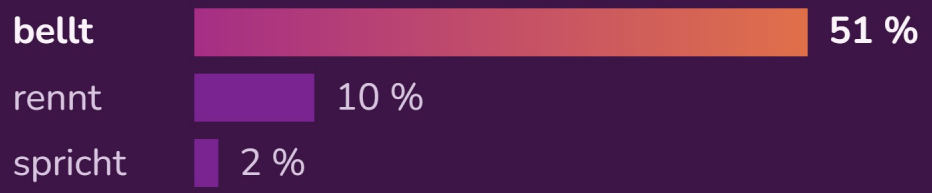
Die KI rät das wahrscheinlichste nächste Wort

1 · Text wird in Bausteine (Tokens) zerlegt

Die | Kraft | fahr | zeug | -H | aft | pflicht | vers | icher | ung | ist | eine | Pflicht | v | ers | icherung | , | die | jedes | Auto | braucht | , | um | auf | der | Stra | ße | fahr | en | zu | dürfen | .

2 · Token für Token wird das wahrscheinlichste nächste vorhergesagt

Der | kleine | Hund | im | Park | ?



Trainiert auf riesigen Textmengen, aber ohne Begriff von Wahrheit: Statistik, keine Magie.

// Hintergrund · Halluzinationen

Klingt plausibel, ist erfunden

Prompt

Wie lange haften wir als Zulieferer für Mängel an gelieferten Teilen?

Antwort der KI

Nach § 438 BGB beträgt die Gewährleistung im B2B-Geschäft 6 Monate. Eine Begrenzung der Haftung auf den Warenwert ist gesetzlich vorgesehen, Sie müssen dafür nichts vereinbaren.

Orange = falsch. Richtig: in der Regel 2 Jahre (§ 438 Abs. 1 Nr. 3 BGB), eine Haftungsgrenze gibt es nur per Vertrag.

// Hintergrund · Halluzinationen

Klingt plausibel, ist erfunden

Prompt

Wie lange haften wir als Zulieferer für Mängel an gelieferten Teilen?

Antwort der KI

Nach § 438 BGB beträgt die Gewährleistung im B2B-Geschäft 6 Monate. Eine Begrenzung der Haftung auf den Warenwert ist gesetzlich vorgesehen, Sie müssen dafür nichts vereinbaren.

Orange = falsch. Richtig: in der Regel 2 Jahre (§ 438 Abs. 1 Nr. 3 BGB), eine Haftungsgrenze gibt es nur per Vertrag.

Warum?

Das Modell antwortet immer. Es weiß nicht, ob seine Aussage stimmt, es liefert nur eine wahrscheinliche Wortfolge.

Gegenmittel

KI an geprüfte Firmendaten anbinden (Grounding, RAG), Quellprüfung / Ergebnisse von Menschen prüfen lassen.

// Fall 3 · Fehler und Haftung

Die KI irrt, Sie haften

Air Canada, 2024

Der Chatbot erfindet eine Erstattungsregel. Das Gericht: Das Unternehmen haftet für alles, was sein Bot sagt.

Deloitte Australien, 2025

Ein Regierungsgutachten enthält von KI erfundene Quellen. Teilrückzahlung des Honorars, Schlagzeilen weltweit.

Amtsgericht Köln, 2025

Ein Anwalt reicht einen Schriftsatz mit von KI erfundenen Urteilen ein. Das Gericht rügt das öffentlich.

// Fall 3 · Fehler und Haftung

Die KI irrt, Sie haften

Air Canada, 2024

Der Chatbot erfindet eine Erstattungsregel. Das Gericht: Das Unternehmen haftet für alles, was sein Bot sagt.

Deloitte Australien, 2025

Ein Regierungsgutachten enthält von KI erfundene Quellen. Teilrückzahlung des Honorars, Schlagzeilen weltweit.

Amtsgericht Köln, 2025

Ein Anwalt reicht einen Schriftsatz mit von KI erfundenen Urteilen ein. Das Gericht rügt das öffentlich.

Bei Ihnen

Der KI-Anbieter schließt seine Haftung per AGB aus. Sie als Betreiber können das nicht.

Ab morgen

Keine KI-Ausgabe geht ungeprüft nach außen. Ein Mensch zeichnet ab / Human in the loop.

// Fall 4 · Personalauswahl

Die KI entscheidet, wer eingestellt wird

Amazon

Ein KI-Recruiting-Tool sortiert Bewerbungen von Frauen systematisch aus, weil es aus alten Mustern gelernt hat. Das Tool wird eingestellt.

Workday, USA

Gegen den HR-Softwareanbieter läuft eine Sammelklage wegen Altersdiskriminierung durch die KI-Vorauswahl.

// Fall 4 · Personalauswahl

Die KI entscheidet, wer eingestellt wird

Amazon

Ein KI-Recruiting-Tool sortiert Bewerbungen von Frauen systematisch aus, weil es aus alten Mustern gelernt hat. Das Tool wird eingestellt.

Workday, USA

Gegen den HR-Softwareanbieter läuft eine Sammelklage wegen Altersdiskriminierung durch die KI-Vorauswahl.

Bei Ihnen

In Deutschland ein AGG-Fall mit Entschädigungsansprüchen. Der AI Act stuft KI in der Personalauswahl als Hochrisiko ein.

Ab morgen

KI darf vorsortieren. Entscheiden muss ein Mensch, und das wird dokumentiert.

// Und die Gegenseite?

Angreifer nutzen dieselbe KI

Angriff

- Stimmklon des Chefs aus 10 Sekunden Audio
- Phishing in perfektem Deutsch / Dialekt
- Automatisierte Schwachstellensuche rund um die Uhr

Abwehr

- KI-gestützte Erkennung, als Managed Service auch für kleine Firmen
- Kontrolle von Schatten-KI und Agenten-Rechten
- Deepfake-Erkennung

// Und die Gegenseite?

Angreifer nutzen dieselbe KI

Angriff

- Stimmklon des Chefs aus 10 Sekunden Audio
- Phishing in perfektem Deutsch / Dialekt
- Automatisierte Schwachstellensuche rund um die Uhr

Abwehr

- KI-gestützte Erkennung, als Managed Service auch für kleine Firmen
- Kontrolle von Schatten-KI und Agenten-Rechten
- Deepfake-Erkennung

Ab morgen · kostet null Euro

Rückruf auf bekannte Nummer, Codewort für Zahlungen, Vier-Augen-Prinzip ab festem Betrag.

// Regulierung

Hersteller oder Dienstleister: Eins trifft Sie

Regelwerk	Wen es trifft	Was gilt	Seit / ab
EU AI Act	Alle, die KI nutzen	Maßnahmen zur KI-Kompetenz der Mitarbeitenden (Art. 4), Hochrisiko-Pflichten ab Dez. 2027 bzw. Aug. 2028	in verschiedenen Stufen
Cyber Resilience Act	Hersteller von Produkten mit Software	Meldepflicht für ausgenutzte Schwachstellen, Security by Design, sonst kein CE	Meldepflicht seit Sept. 2026, voll ab Dez. 2027
Neue Produkthaftung	Hersteller, auch von Software und KI	Software gilt als Produkt, Haftung ohne Verschulden	Umsetzung bis Dez. 2026
NIS2	Betroffene Sektoren und ihre Lieferkette	Risikomanagement, Meldepflichten, Nachweise gegenüber Kunden	in Deutschland seit 6. Dez. 2025, Lieferkette folgt noch

Bußgelder bis: AI Act 35 Mio. € / 7 % (verbotene KI-Praktiken), sonst 15 Mio. € / 3 % · CRA 15 Mio. € / 2,5 % · NIS2 10 Mio. € / 2 % (wichtige Einrichtungen 7 Mio. € / 1,4 %) · jeweils vom weltweiten Jahresumsatz

// NIS2 · Geschäftsleitung

Chefsache. Mit Ihrem Namen dran.

Die Geschäftsleitung billigt und überwacht die Maßnahmen, haftet persönlich und muss sich schulen lassen. Delegieren geht nicht.

Und wer selbst nicht darunter fällt, bekommt den Fragebogen vom Kunden. Wer nicht liefert, fliegt aus der Lieferkette.

// Absicherung

Restrisiko bleibt. Ist es versichert?

Stillschweigend

Viele Cyber-, Haftpflicht- und D&O-Policen sagen nichts zu KI. Das ist keine verlässliche Deckung.

Deepfake-Überweisung

Oft nicht über die Cyberpolice gedeckt, sondern nur über eine Vertrauensschadenversicherung, wenn überhaupt.

KI-Fragen im Antrag

Versicherer fragen die KI-Nutzung ab. Falsche Angaben gefährden den Versicherungsschutz.

Ein Risiko kann immer nur so gut abgesichert werden, wie es kommuniziert wird. Erst dann lässt sich klären, ob der Versicherungsschutz es deckt.

// Ihr 5-Fragen-Check

Beantworten Sie für sich: Ja oder Nein?

1. Weiß ich, welche KI-Tools in meinem Betrieb genutzt werden?
2. Gibt es eine schriftliche KI-Regel und eine Schulung?
3. Würde meine Buchhaltung einen Anruf mit meiner Stimme hinterfragen?
4. Weiß ich, ob AI Act, CRA oder NIS2 mich betreffen?
5. Ist ein KI-Schaden in meiner Versicherung gedeckt?

Gezögert? Sprechen Sie mich gleich beim Imbiss an.

protectifAI

Wir machen KI im Mittelstand sicher und versicherbar

Neue Risiken

Wir entwickeln gemeinsam mit Mandanten und Risikoträgern neue und innovative Versicherungslösungen (RaaS, Cloud-Ausfall, KI-Garantie etc.).

Risikoanalyse & Transfer

Vollständige Unternehmensanalyse und entsprechender Transfer an Risikoträger.

Risikomanagement

Umsetzung und Begleitung von Risikomanagement in KMU hier in der Region - nach Ihren Bedürfnissen.

KI-Versicherungs-Check

Wir prüfen Ihre Policen auf Vollständigkeit sowie KI- und Betrugs-Lücken und schließen sie.

Vielen Dank.

> *„Der größte Produktivitätssprung unserer Zeit ist auch ein großes Risiko. Die Frage ist, ob Sie darauf vorbereitet sind.“*

Scannen & vernetzen

protectifAI

Sarah Günther · Geschäftsführerin

0170-7956221 · sarah.guenther@protectifai.de · www.protectif.ai