

Trusted AI

Sicherer und ganzheitlicher KI-Einsatz
in Unternehmen



Sicherer und ganzheitlicher KI-Einsatz in Unternehmen

Unternehmen befinden sich ab 2026 in einem gefährlichen Zangengriff: Der wirtschaftliche und demografische Druck macht den KI-Einsatz alternativlos, während der EU AI Act die Hürden für die "License to Operate" drastisch erhöht. Auf der einen Seite diktiert die Marktrealität ein sofortiges Handeln – mit 20 Millionen Erwerbstätigen, die bis 2036 in Rente gehen, lässt sich die drohende Lücke nur durch die prognostizierten 45% Effizienzsteigerung und 10–20% Produktivitätsgewinne mittels KI schließen. Wer hier zögert, verliert den Anschluss, da bereits 56% der Wettbewerber durch KI-Nutzung signifikante Marktvorteile realisieren.

Die operative Realität ist alarmierend: Nur die wenigsten Unternehmen haben Security-Tools speziell für KI. Trotzdem müssen mehr als dreiviertel der Firmen neues Personal aufbauen, um die KI-Sicherheit überhaupt bewerten zu können. Das größte Risiko bleibt aber der Faktor Mensch: Nur 6% der Mitarbeitenden sind aktuell bereit oder befähigt, KI tiefgreifend in ihre Prozesse zu integrieren. Ohne Akzeptanz und Absicherung wird das KI-Investment zum unkalkulierbaren Risiko.

56%

der Unternehmen,
die KI nutzen, haben
Wettbewerbs-
vorteile.

89%

der Unternehmen
setzen KI ein, ohne
über spezialisierte
Security-Tools zu
verfügen.

20 Millionen

Erwerbstätige
gehen bis 2036 in
Rente.

Warum Ihre Firewall gefährdet ist

Klassische IT-Sicherheitsarchitekturen sind blind für die Risiken der algorithmischen Ära. Während Firewalls die Infrastruktur schützen, zielen moderne Angriffsvektoren (MITRE ATLAS) direkt auf die geschäftskritische Logik Ihrer Modelle. Das Risikoprofil verschiebt sich damit fundamental von der IT-Infrastruktur hin zu den Assets:

- Operative Manipulation: Prompt Injection und Jailbreaking hebeln interne Sicherheitsbarrieren aus und erzwingen fatale Fehlfunktionen.
- Strategische Sabotage: Data Poisoning korrumpiert schleichend die Entscheidungsgrundlage Ihres Unternehmens und entwertet Investitionen.
- Erosion von Intellectual Property: Durch Model Extraction wird Ihr wertvollstes Asset – das trainierte Modell – zur leichten Beute für direkte Replikation durch Wettbewerber.

Unser Ansatz: Trusted AI als Business Enabler. Wir schließen die Lücke zwischen technologischer Euphorie und regulatorischer Notwendigkeit. Wir transformieren KI-Sicherheit von einer technischen Pflichtübung in einen Werttreiber: Unsere Härting macht Ihre Systeme nicht nur resilient gegen Angriffe und rechtssicher im Sinne des AI Acts, sondern sichert die Profitabilität Ihrer Innovation nachhaltig ab. Damit scheitert Ihre KI-Strategie nicht an Hackern oder Gesetzen, sondern stiftet messbaren Wert.

Das Ende der Schonfrist

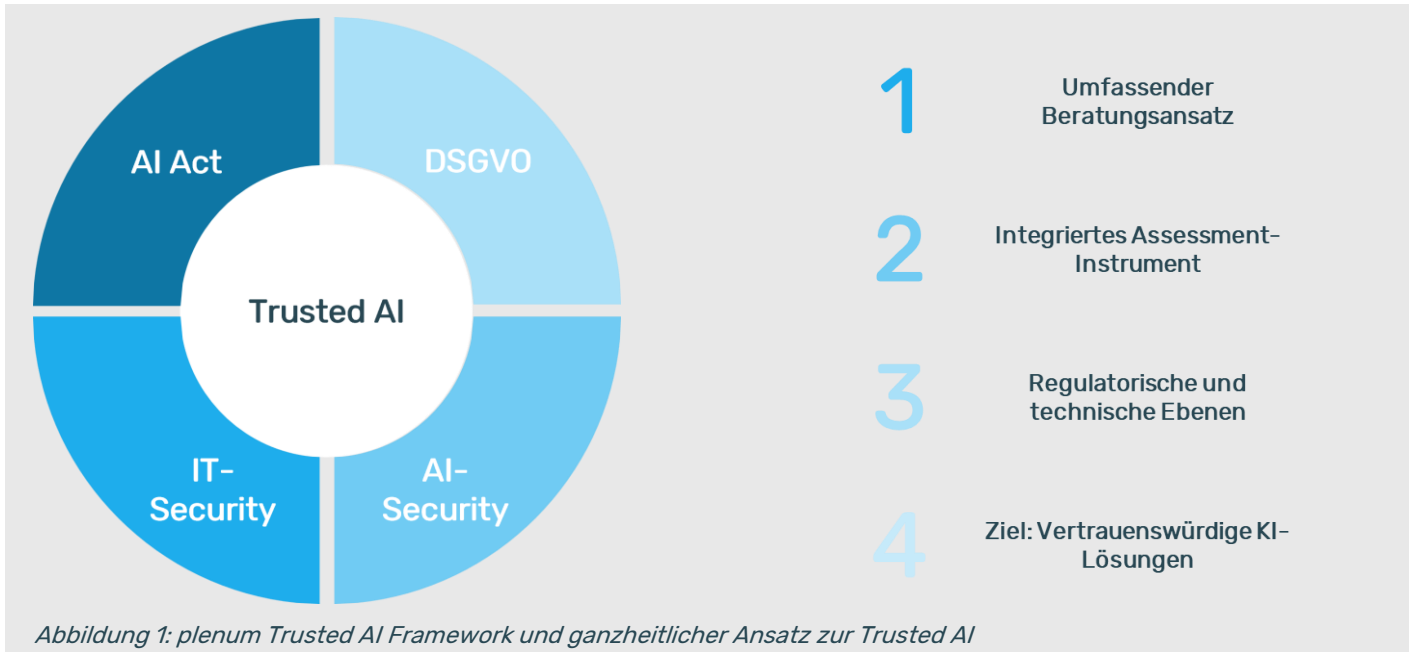
Das kumulierte Sanktionspotenzial ist beispiellos: Zur 7%-Umsatzstrafe des EU AI Act addiert sich weiterhin das 4%-Risiko der DSGVO. Viel kritischer als die monetäre Strafe ist jedoch die operative Bedrohung: Bei Non-Compliance droht der sofortige Marktausschluss oder die zwangsweise Abschaltung geschäftskritischer KI-Systeme ("Kill Switch"), was im laufenden Betrieb einem Herzstillstand gleichkommt.

Verschärft wird dieses Szenario durch eine interne Zerreißprobe. Während 70% der Belegschaft bereit sind, Aufgaben an KI zu delegieren, lähmt die Angst vor dem Arbeitsplatzverlust (49%) und eine alarmierend niedrige "High Readiness"-Quote von nur 6% die operative Umsetzung. Hinzu kommt das "Legacy-Cliff" im Jahr 2030: Bestandsinvestitionen ohne rechtzeitiges Retrofit werden rechtlich unverwendbar und zwingen Unternehmen in einen teuren Modernisierungstau.

Vom regulatorischen Haftungsrisiko zum strategischen Wettbewerbsvorteil durch lückenlose Absicherung der algorithmischen Wertschöpfung.



Trusted AI: Vier Säulen für vertrauenswürdige KI



Säule: EU AI Act - Rechtskonformität sicherstellen

Vier Risikokategorien mit unterschiedlichen Anforderungen an Qualitätsmanagement, Data Governance, technische Dokumentation und Cybersicherheit.

Kritische Unterscheidung: Anbieter haben wesentlich höhere Anforderungen als Betreiber. Achtung: Bei wesentlichen Modelländerungen wird der Betreiber zum Anbieter.

Unsere Leistung: Risikoklassifizierung, technische Dokumentation nach Anhang IV (Art. 11), Qualitäts- und Risikomanagementsysteme nach Art. 9 und 27, Grundrechte-Folgenabschätzung.

Säule IT-Security - Das notwendige Fundament

Standards: ISO 27001, BSI Grundschutz, MITRE ATT&CK für Frontend, API und Host-Infrastruktur.

Drei Hauptbedrohungen bei KI:

1. Manipulation & Steuerungsumlenkung (Prompt Injection, Jailbreaking)
2. Ungewollter Datenabfluss (sensible Daten in Training/Output)
3. Modellfehler & Vergiftung (Data Poisoning)

Unsere Leistung: Spezialisiertes Pentesting für KI-Interfaces, Systemhärtung, Segmentierung, MFA, robuste Zugriffskontrolle.

Säule: DSGVO - Datenschutz in KI-Kontexten

Kernpflichten: Transparenz, Betroffenenrechte (Auskunft, Löschung, Übertragbarkeit), Nachweispflicht (Accountability), DSFA (Datenschutz-Folgenabschätzung) nach Art. 35 für Hochrisiko-KI.

Besondere Herausforderung: Sensible Daten (Gesundheit, Ethnie) - AI Act erlaubt Nutzung zur Bias-Korrektur, DSGVO ist restriktiv. Sorgfältige Abwägung erforderlich.

Unsere Leistung: DSFA-Durchführung, Privacy by Design, Rechtsgrundlagen, Datenminimierung, Verschlüsselung, Berechtigungskonzepte.

Säule KI-Security - Modelle gezielt schützen

Frameworks: MITRE ATLAS (AI-Threat-Intelligence), ISO/IEC 22989 (Modellklassifizierung), ENISA AI (sichere Bereitstellung).

Sechs Kernbereiche: Modellklassifizierung & Risikoarchitektur, AI-Threat-Intelligence & TTPs, Datenherkunft & Input-Sicherheit, Modellrobustheit & Adversarial Resilience, sichere Bereitstellung & Laufzeitüberwachung, KI-Supply-Chain-Security.

Unsere Leistung: Adversarial Robustness Testing, Guardrails (Input/Output-Filter), Drift/Bias-Monitoring, Supply-Chain-Security, Validierung gegen Poisoning.



Brückenschlag zwischen Regulierung und Security

Der wirtschaftliche Erfolg von KI-Projekten entscheidet sich an der Schnittstelle von Recht und Technik. Unser Ansatz löst die typischen Reibungsverluste in zwei kritischen Dimensionen auf:

1. Compliance-Symbiose: EU AI Act trifft DSGVO Wir nutzen gezielte Synergien bei Transparenzpflichten und im Risikomanagement. Während der AI Act eine **Grundrechte-Folgenabschätzung** fordert, verlangt die DSGVO eine **DSFA**. Wir integrieren beide Prüfungen in einen hocheffizienten Prozess. Dabei lösen wir den inhärenten Zielkonflikt: Der AI Act erlaubt die Nutzung sensibler Daten zur Bias-Korrektur, während die DSGVO hier restriktiv agiert. Unsere Lösung: Priorisierte DSFA-Workflows und rechtssichere Zweckbindungen, die Innovation und Datenschutz harmonisieren.

2. Schutzschild-Erweiterung: IT-Security trifft KI-Security Klassische IT-Sicherheit schützt Ihre Infrastruktur, Netzwerke und APIs – doch sie ist blind für die Logik der Modelle. Firewalls und WAFs erkennen weder **Prompt Injection** noch **Data Poisoning**. Wir erweitern Ihre Basis-Sicherheit um eine spezifische KI-Schutzschicht für Modelle, Pipelines und Output-Logik. Durch klare Zuständigkeiten zwischen **CISO und AI Security Lead** sowie spezialisiertes **AI Pentesting** schließen wir die Lücke, bevor sie zum Einfallstor wird.

Das plenum Trusted AI-Assessment: Fünf Phasen, sechs Wirkungsfelder

Unser systematischer Fünf-Phasen-Ansatz transformiert regulatorische Anforderungen in ein messbares Zielbild entlang der sechs entscheidenden Wirkungsfelder der Trusted AI. Damit gewährleisten wir eine lückenlose Absicherung Ihrer Systeme – von der initialen Ist-Analyse über die technische Härtung bis hin zur dauerhaften Verankerung rechtskonformer Betriebsprozesse:



Sechs Wirkungsfelder im Überblick

- KI-Register & Rollen:** Inventarisierung aller KI-Systeme inklusive GPAL und Dritt-APIs. Etablierung einer RACI-Matrix für Anbieter, Betreiber und Händler gemäß AI Act Art. 3, 5 sowie 22-26.
- Risiko-Klassifizierung:** Systemisches Mapping auf AI-Act-Risikostufen inklusive Durchführung von Grundrechte-Folgenabschätzungen und Identifikation von DSFA-Triggern nach Art. 9, 27 AI Act sowie DSGVO Art. 35.
- Daten-Governance:** Sicherstellung von Herkunft, Qualität, Zweckbindung und Minimierung. Implementierung von pbD-Handling durch Pseudonymisierung und Anonymisierung auf Basis klarer Rechtsgrundlagen (AI Act Art. 10; DSGVO Art. 5-6).
- IT- & KI-Security:** Verzahnung von IT-Security (ISO 27001/BSI) mit KI-Security (MITRE ATLAS, ENISA) durch Pentesting. Einsatz von Guardrails sowie Robustheits- und Adversarial Tests gemäß Art. 15.
- Verantwortung & Rollen:** Umsetzung von Interaktionskennzeichnung, Erklärbarkeit und Eingriffsrechten. Erstellung von Betriebsanleitungen und Nutzerinformationen gemäß Art. 13-14 AI Act sowie DSGVO Art. 12-22.
- Betrieb & Monitoring:** Kontinuierliches Drift- und Bias-Monitoring. Etablierung eines Incident- und Schwachstellenmanagements inklusive der Meldepflicht für schwerwiegende Vorfälle nach Art. 73 AI Act.



Strategische Herausforderungen und Lösungsansätze

Die erfolgreiche Implementierung von KI-Systemen scheidert heute selten an der Technologie, sondern an der Beherrschung der regulatorischen und organisatorischen Komplexität. Unser Assessment adressiert die kritischen Schmerzpunkte moderner Unternehmen:

Aktuelle Ausgangssituation: Das Risiko der fragmentierten Verantwortung Die rapide Diffusion von KI-Systemen in alle Unternehmensbereiche hat ein gefährliches Verantwortlichkeits-Vakuum entstehen lassen. In der Praxis führt dies zu einem Rollen-Chaos, bei dem traditionelle Instanzen wie CISO, Datenschutzbeauftragte und die IT mit den neuartigen Anforderungen des EU AI Act und der KI-Governance kollidieren. Ohne klare Strukturen entstehen Reibungsverluste an den Schnittstellen zu den operativen Fachbereichen, was sowohl die Innovation als auch die Compliance gefährdet.

Erforderliche Zielsetzung: Trusted AI als operativer Standard Um Trusted AI nachhaltig zu gewährleisten, ist eine strikte Definition, Abgrenzung und Ordnung der Zuständigkeiten entlang der vier Säulen unseres Frameworks (EU AI Act, DSGVO, IT-Security und KI-Security) zwingend erforderlich. Unser Ziel ist die Etablierung einer lückenlosen Kontrolle: Nur durch die Verzahnung von strategischer Governance, risikobasiertem Management und tiefgreifender technischer Kontrolle stellen wir sicher, dass Ihre KI-Systeme resilient, rechtskonform und wertstiftend agieren.

Profitieren Sie von unseren Erfahrungen

Unser Anspruch ist es, nachhaltige und wirksame Lösungen für Sie zu schaffen.

- Wir sind über 150 motivierte Berater mit langjähriger Erfahrung und verstehen uns als Partner unserer Kunden
- Wir kennen die Anforderungen der modernen Versicherungen und wissen welche Maßnahmen notwendig sind – wir kennen aber auch die Gestaltungsspielräume
- Wir bieten exzellente Berater, die in allen Prüfungsfeldern die praktische Umsetzungserfahrung bereitstellen
- Wir schaffen gemeinsam mit unseren Kunden, die Risiken zu steuern bei gleichzeitiger Hebung der Geschäftschancen

Vom Paragraphen zur Praxis: Das operative Zielbild Ihrer Trusted AI

Die regulatorischen Anforderungen des EU AI Act und der DSGVO lassen sich nicht durch punktuelle Maßnahmen erfüllen – sie erfordern eine systematische Verankerung in der Unternehmensarchitektur. Unser Framework übersetzt komplexe Gesetzestexte in ein greifbares, operatives Zielbild. Durch die Verzahnung von verbindlichen Artefakten, technischen Metriken und klar definierten Rollen transformieren wir Compliance von einer Innovationsbremse in ein Qualitätsmerkmal.

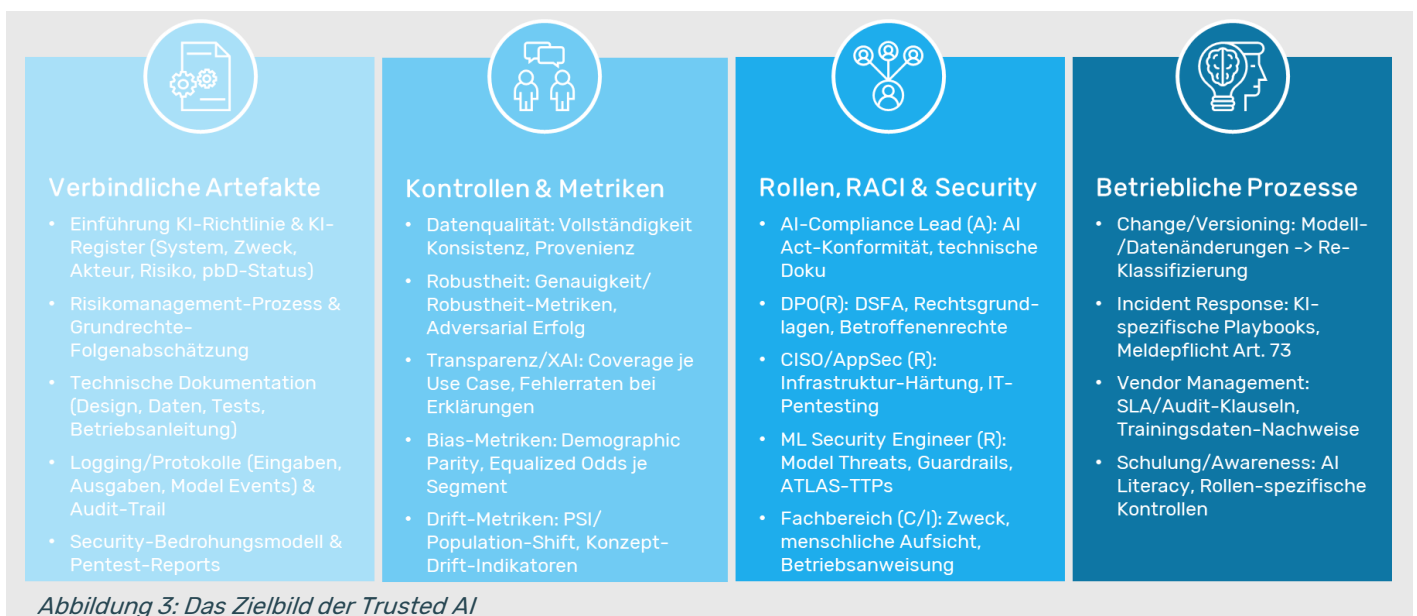


Abbildung 3: Das Zielbild der Trusted AI