

Trusted AI

Sicherer und ganzheitlicher KI-Einsatz
in Unternehmen



Warum Trusted AI jetzt zur Pflicht wird

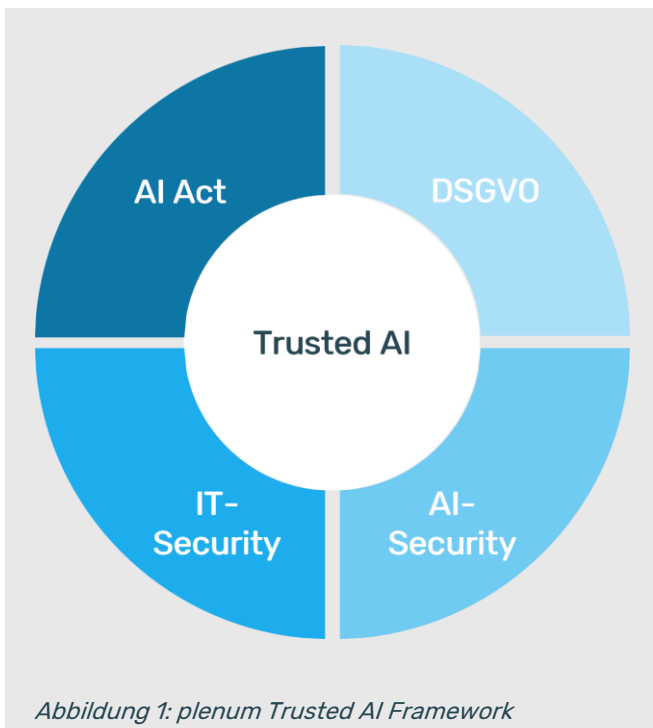
Künstliche Intelligenz ist kein Hype mehr, sondern ein messbarer Wirtschaftsfaktor: Im Dienstleistungssektor sind Produktivitätssteigerungen von 10–20% und Effizienzgewinne von bis zu 45% möglich. Kein Wunder, dass bereits 56% der KI-Nutzer signifikante Wettbewerbsvorteile gegenüber der Konkurrenz vermelden.

Trotz des Booms herrscht eine gefährliche Sorglosigkeit: 89% der Unternehmen nutzen bisher keine spezialisierten KI-Security-Tools, obwohl 76% bereits jetzt zusätzliches Personal für KI-Sicherheit einstellen müssten. Hinzu kommt ein kulturelles Hindernis: Bisher zeigen nur 6% der Mitarbeitenden eine hohe KI-Nutzungsbereitschaft.

Die Schonfrist endet. Mit dem EU AI Act drohen Bußgelder von bis zu 35 Mio. EUR oder 7% des weltweiten Umsatzes. Parallel sanktioniert die DSGVO Verstöße mit bis zu 20 Mio. EUR oder 4%. Kritisch: Bei Non-Compliance können Systeme unmittelbar vom Markt genommen werden. Bestehende Alt-Systeme genießen nur bis 2030 Bestandsschutz und das auch nur bei unveränderter Nutzung.

Traditionelle IT-Security wie Firewalls versagt bei KI-spezifischen Angriffen. Das Framework MITRE ATLAS dokumentiert eine völlig neue Dimension von Risiken: Prompt Injection, Jailbreaking, Data Poisoning, Adversarial Attacks sowie Model Extraction und Model Theft gefährden IP und Integrität.

Wir schaffen Resilienz. Durch die Verzahnung von technischer Sicherheit, rechtlicher Compliance, Einhaltung der Regulatorik und organisatorischer Verankerung machen wir Ihre KI-Infrastruktur integer, fair und jederzeit erklärbar.



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.

Vier Säulen für vertrauenswürdige KI

EU AI Act - Compliance sichern

Risikoklassifizierung (verboten, Hochrisiko, begrenzt, minimal), technische Dokumentation nach Anhang IV, Qualitäts- und Risikomanagementsysteme. Unterscheidung Anbieter vs. Betreiber - bei Modell-änderungen kann es einen Rollentausch geben.

DSGVO - Datenschutz gewährleisten

DSFA (Datenschutz-Folgenabschätzung) für KI-Systeme, Privacy by Design, Transparenz und Betroffenenrechte. Nachweispflicht und Meldepflichten. Sanktionen bis 20 Mio. EUR oder 4% des Jahresumsatzes.

IT-Security - Infrastruktur schützen

ISO 27001, BSI Grundschatz. Spezialisiertes Pen-testing gegen Hauptbedrohungen: Manipulation & Steuerungsumlenkung, ungewollter Datenabfluss, Modellfehler & Vergiftung.

KI-Security - Modelle absichern

Basierend auf MITRE ATLAS, ISO/IEC 22989 und ENISA AI. Adversarial Robustness Testing, Guardrails, Drift/Bias-Monitoring, Supply-Chain-Security. Schutz vor neuartigen KI-spezifischen Angriffen.

Wir garantieren eine lückenlose Absicherung durch Transformation regulatorischer Anforderungen in ein messbares Zielbild entlang von fünf Phasen und sechs Wirkungsfeldern.



Sicherheit, Recht und Technik im Einklang

Ein rechtssicherer KI-Einsatz erfordert das nahtlose Ineinandergreifen unterschiedlicher Disziplinen. Unser Framework deckt alle kritischen Dimensionen ab:

Strategische Inventarisierung: Wir schaffen Transparenz durch ein KI-Register und klare RACI-Matrizen, um Verantwortlichkeiten zwischen Anbietern und Betreibern zweifelsfrei zu klären.

Risikobasierte Steuerung: Durch präzises Mapping auf AI-Act-Risikoklassen, flankiert von Grundrechte-Folgenabschätzungen (FRIA) und DSFA-Triggern, machen wir Risiken beherrschbar.

Data Governance & Security: Wir schützen Ihr wertvollstes Gut. Durch "Privacy by Design" (Art. 10 AI Act) und modernste Security-Validierungen gegen Bedrohungsszenarien, siehe **MITRE ATLAS**, sichern wir Ihre Modelle gegen Angriffe und Bias ab.

Transparenz & Monitoring: Wir implementieren die notwendigen Mechanismen für Erklärbarkeit und Incident-Management, damit Ihre KI-Systeme nicht nur performant, sondern auch jederzeit prüffähig bleiben.

Die fünf Phasen des plenum Trusted AI-Assessments

Unser systematisches Assessment verzahnt die regulatorisch-strategische mit der technisch-operativen Ebene:

Phase 1 - Ist-Analyse: Inventarisierung aller KI-Systeme (inkl. General-Purpose-Modelle, Dritt-APIs), Analyse der IT-Landschaft und Governance-Strukturen. Verständnis über Rollen (Anbieter/Betreiber/Händler).

Phase 2 - Gap-Analyse: Identifikation regulatorischer Lücken (AI Act, DSGVO, ISO-Normen). Betrachtung des Risikomanagement-Rahmens sowie der IT- und KI-Security-Strategie.

Phase 3 - Anforderungsübersetzung: Ermittlung des Bedarfs von Richtlinien und Dokumentationen. Festlegung notwendiger Schulungen, behördlicher Meldeprozesse und Vorbereitung auf Prüfungen.

Phase 4 - Implementierung: Erstellung des KI-Registers, notwendiger Richtlinien und technischer Dokumentationen. Einbettung in Risikomanagement und KI-/IT-Sicherheitsstrategie.

Phase 5 - Verankerung: Vereinheitlichung der Prozesse, Behebung des regulatorischen Deltas. Implementierung von Kontrollen, fortlaufender Überprüfung und Verankerung über alle Stakeholder.

Die sechs Wirkungsfelder der Trusted AI

Wir transformieren regulatorische Anforderungen in ein messbares Zielbild. Damit gewährleisten wir eine lückenlose Absicherung Ihrer Systeme – von der initialen Ist-Analyse über die technische Härtung bis hin zur dauerhaften Verankerung rechtskonformer Betriebsprozesse.

