WHITEPAPER

Künstliche Intelligenz als Schlüssel zur Cyberresilienz: Sichere Integration, Schutz vor Angriffen und intelligente Verteidigung

Mitwirkende Autor:innen/ Unternehmen

Ralf Benzmüller
Executive Speaker Security Labs G Data CyberDefense

Lisa Fröhlich
Unternehmenssprecherin Link11 GmbH

Christof Klaus
Director Global Network Defense Myra Security

Olaf Pursche
Leiter Kompetenzgruppe Sicherheit eco – Verband der Internetwirtschaft e. V.
Pursche Tactical Consulting (in Gründung)

Cornelia Schildt
Senior Projektmanagerin IT-Sicherheit eco – Verband der Internetwirtschaft e. V.

Dr. Or Sela
Channel Account Manager F5

Maurice Striek
Senior Consultant Cyber Strategy &Architecture NVISO Security

Marcel Rieger
Co-Founder Jamorie Consulting

Inhaltsverzeichnis

Vorwort

Künstliche Intelligenz ist längst geschäftliche Realität: Unternehmen stehen nicht mehr vor der Frage, ob sie KI einsetzen, sondern vor der Aufgabe, sie sicher und effektiv in ihre Strukturen zu integrieren. Doch die rasante Entwicklung der Technologie hat in den letzten Jahren nicht nur innovative Lösungen für Unternehmen hervorgebracht, sondern auch neue Angriffsvektoren für Cyberkriminelle geschaffen. Aktuelle Forschungsergebnisse belegen eindringlich diese Entwicklung. Gemäß aktuellen Untersuchungen hat sich die Qualität und Quantität KI-generierter Fehlinformationen dramatisch erhöht. Renommierte Sicherheitsexpert:innen, darunter das unabhängige Forschungsinstitut AV-TEST, dokumentieren einen besorgniserregenden Anstieg von täuschend echten Deepfakes und maßgeschneiderten Phishing-Kampagnen, die mit Hilfe fortschrittlicher KI-Modelle erstellt werden.

Zu beachten ist auch das aufkommende Phänomen des KI-Poisonings, bei dem Akteure versuchen, Trainingsdaten oder Modellparameter zu beeinflussen, um die Funktionsweise von KI-Systemen zu manipulieren. Obwohl diese Art von Angriffen auf Trainingsdaten derzeit noch weniger verbreitet ist und vorwiegend in spezifischen Kontexten auftritt, verdient sie dennoch Aufmerksamkeit als potenzielle zukünftige Herausforderung. KI-gestützte Täuschungsmanöver und Manipulationen können auf Geschäftsprozesse und Mitarbeitende abzielen und sowohl finanzielle Einbußen als auch Reputationsschäden nach sich ziehen.

Dieses Whitepaper verfolgt einen praxisorientierten Ansatz und richtet sich an Eintscheidungsträger:innen und IT-Sicherheitsverantwortliche, die vor der Herausforderung stehen, KI sowohl als Werkzeug für betriebliche Optimierung als auch als Verteidigungsmechanismus gegen immer raffiniertere Bedrohungen einzusetzen. Es bietet konkrete Handlungsempfehlungen, die auf Erfahrungen führender Expert:innen und erfolgreichen Implementierungen in verschiedenen Branchen basieren.

Von der sicheren Integration von KI-Lösungen in bestehende Unternehmensstrukturen bis hin zur aktiven Abwehr KI-gestützter Cyberangriffe – Ziel ist es, praxisnahes Wissen zu vermitteln, um Unternehmen zu helfen, das transformative Potenzial künstlicher Intelligenz zu nutzen und gleichzeitig die damit verbundenen Risiken zu minimieren. Besonderes Augenmerk liegt dabei auf regulatorischen Anforderungen wie dem EU AI Act und der Gewährleistung der Compliance in einem sich ständig weiterentwickelnden rechtlichen Umfeld.

Die hier zusammengetragenen Erkenntnisse unterstreichen, dass die erfolgreiche Nutzung von KI nicht nur technologische Expertise erfordert, sondern auch ein tiefgreifendes Verständnis für Sicherheitsaspekte und die strategische Einbettung in bestehende Geschäftsprozesse. Die präsentierten Fallstudien demonstrieren, wie Unternehmen KI nicht nur als Innovationstreiber, sondern auch als wirksamen Schutzschild gegen Cyberbedrohungen einsetzen können.

Während KI-gestützte Angriffe an Komplexität und Häufigkeit zunehmen, bietet dieses Whitepaper einen wertvollen Leitfaden für Unternehmen, die ihre digitale Resilienz stärken und gleichzeitig die Chancen der KI-Revolution nutzen möchten. Die folgenden Seiten dienen als Ressource für die KI-Transformation – mit dem Ziel, Innovationen zu fördern und gleichzeitig ein hohes Maß an Sicherheit zu gewährleisten.

Olaf Pursche
Leiter Kompetenzgruppe Sicherheit
eco – Verband der Internetwirtschaft e. V.
Pursche Tactical Consulting (in Gründung)

I. Einleitung

Künstliche Intelligenz (KI) ist heute fester Bestandteil vieler Geschäftsprozesse. Die zentrale Frage lautet nicht mehr, ob Unternehmen KI einsetzen, sondern wie sie sicher und verantwortungsvoll integriert werden kann. Parallel zur wachsenden Nutzung nehmen auch die Bedrohungen zu: Deepfakes, KI-gestütztes Phishing, automatisierte Schadsoftware und Manipulationen von Trainingsdaten sind nur einige Beispiele.

Dieses Whitepaper gibt einen Überblick über zentrale Herausforderungen und zeigt, wie Organisationen KI erfolgreich implementieren und gleichzeitig Schutzmechanismen gegen neue Angriffsformen etablieren können.

II. Sichere Implementierung von KI-Werkzeugen in die Unternehmensstruktur

Die Einführung von KI verändert Prozesse, Rollen und Entscheidungswege. Damit der Einsatz sicher und erfolgreich gelingt, braucht es ein strukturiertes Vorgehen.

Lernmodelle und Trends

KI-Systeme basieren auf verschiedenen Lernansätzen – vom überwachten und unüberwachten Lernen bis zu Reinforcement und Transfer Learning. Besonders im Fokus stehen derzeit generative Modelle, die Texte, Bilder oder Code erzeugen, sowie hochspezialisierte Mustererkennungs- und Expertensysteme.

Einsatzfelder

KI bietet Potenziale in vielen Bereichen: von Marketing und Kundenservice über medizinische Diagnostik bis hin zu Cybersecurity und autonomem Fahren.

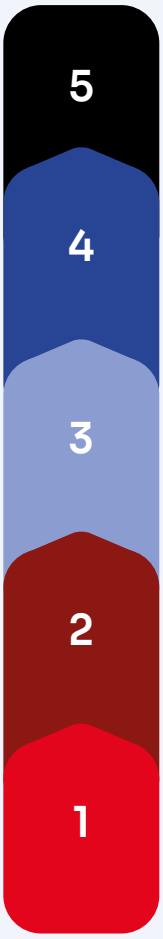
Herausforderungen

Unternehmen müssen Datenschutz, Zuverlässigkeit, Haftung und Akzeptanz berücksichtigen. Die Balance zwischen Innovationsgeschwindigkeit und Sicherheit ist entscheidend.

Vorgehensmodell

Das folgende Vorgehensmodell zeigt einen strukturierten und praxisnahen Ansatz zur Einführung von KI im Unternehmen. In fünf aufeinander aufbauenden Schritten wird dargestellt, wie Organisationen systematisch Kompetenzen aufbauen, klare Zielbilder entwickeln und geeignete Anwendungsfelder identifizieren. Ergänzt durch iterative Prototypenentwicklung und die frühzeitige Integration von Sicherheits- und Compliance-Aspekten bietet das Modell eine kompakte Orientierung für eine verantwortungsvolle und erfolgreiche KI-Implementierung.

Implementierung eines praxisnahen Ansatzes



Sicherheit integrieren

Berücksichtigen Sie rechtliche und technische Vorgaben von Anfang an, um die Einhaltung sicherzustellen.

Prototypen entwickeln

Testen und sammeln Sie Erfahrungen in kleinen Schritten, um die Lösung zu verfeinern.

Anwendungsfelder priorisieren

Bewerten Sie Nutzen, Risiko und Datenlage, um sich auf die wichtigsten Bereiche zu konzentrieren.

Zielbild definieren

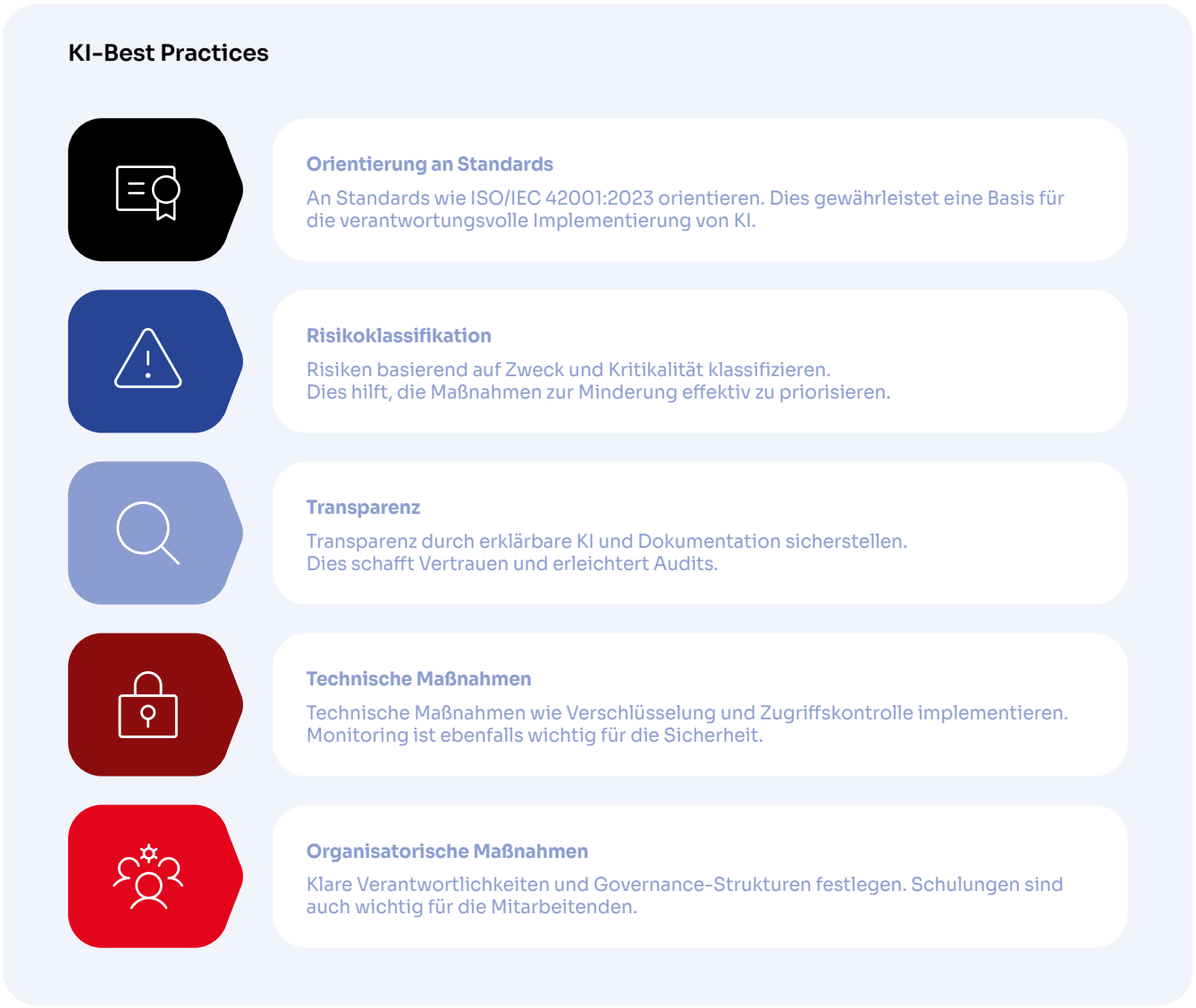
Legen Sie Governance, Rollen und klare Ziele fest, um die Richtung zu bestimmen.

Kompetenz aufbauen

Stärken Sie das interne Fachwissen, um eine solide Grundlage zu schaffen.

Best Practices

Der Einsatz von KI-Werkzeugen eröffnet Unternehmen vielfältige Chancen, bringt jedoch zugleich neue Anforderungen an Sicherheit, Transparenz und Governance mit sich. Um KI verantwortungsvoll und nachhaltig in bestehende Unternehmensstrukturen zu integrieren, sind klar definierte Best Practices erforderlich. Die folgende Grafik gibt einen kompakten Überblick über zentrale Handlungsfelder für eine sichere Implementierung von KI-Werkzeugen.



Blick in die Praxis

In datenintensiven Bereichen zeigt sich der Nutzen besonders deutlich: In der medizinischen Bildgebung unterstützt KI die Erkennung kleinster Anomalien in MRT-Daten und verbessert so die Diagnostikqualität. Auch personalisierte Inhalte in digitalen Plattformen sowie Betrugserkennung in großen Datenströmen profitieren von Mustererkennung und, wo sinnvoll, generativen Verfahren. Darüber hinaus reichen die Einsatzfelder von Sicherheitsanalysen über autonomes Fahren bis hin zu Lagerhaltung und Gaming, wo KI-Abläufe optimiert und komplexe Entscheidungen automatisiert werden. Diese Beispiele verdeutlichen, wie kombinierte Lernansätze (supervised/unsupervised/transfer) in realen Umgebungen Mehrwert stiften.

Rechtlicher Rahmen für den Einsatz künstlicher Intelligenz: AI Act – die europäische KI-Regulierung

Der EU AI Act, auch als Artificial Intelligence Act bezeichnet, ist eine EU-Verordnung, deren Ziel in der Regulierung von künstlicher Intelligenz (KI) besteht. Dies soll durch die Schaffung eines einheitlichen Rechtsrahmens für vertrauenswürdige und sichere KI-Systeme erreicht werden. Der AI Act beinhaltet eine Reihe von Kernpunkten, darunter die Definition verbotener KI-Systeme, die Einstufung von Risiken und die Festlegung von Umsetzungsregeln.

Definition KI gemäß AI Act

Ein KI-System ist ein maschinengestütztes System, das für den Betrieb mit unterschiedlichen Autonomiegraden ausgelegt ist, das nach seinem erstmaligen Einsatz Anpassungsfähigkeit zeigen kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie es Ergebnisse, wie Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erzeugt, die die physische oder virtuelle Umgebung beeinflussen können.

Transparenzverpflichtungen

Die Anbieter von KI-Systemen sind dazu angehalten, die Nutzer:innen über die Interaktion mit der KI zu informieren. Insbesondere die Interaktion mit KI-generierten Inhalten, wie Deep Fakes, ist deutlich zu kennzeichnen. Die EU entwickelt Verhaltenskodizes und Prüfverfahren, um die Kennzeichnung künstlich erzeugter Inhalte zu erleichtern.

Risikoklassifizierung von KI-Systemen

Die Risikoklassifizierung von KI-Systemen wird im AI Act in vier Risikokategorien unterteilt:

- Unannehmbares Risiko: Verbotene KI-Systeme (beispielsweise Social Scoring, biometrische Identifikation in Echtzeit mit wenigen Ausnahmen).
- Hohes Risiko: Es sind strenge Prüfungen erforderlich, beispielsweise für Systeme in der Strafverfolgung, kritischen Infrastruktur oder Medizin.
- Geringes Risiko: Transparenzpflichten für die Nutzer:innen, z. B. bei automatisierten Empfehlungen.
- Minimales Risiko: Keine spezifischen Vorschriften, z. B. KI-gestützte Videospiele.

Anforderungen an Hochrisiko-KI

- Einrichtung eines Risikomanagementsystems
- Qualitätsanforderungen an Trainingsdaten zur Vermeidung von Bias
- Technische Dokumentation zur Bewertung der Systemkonformität
- Transparente und nachvollziehbare Gebrauchsanweisung
- Menschliche Überwachung mit Notfall-Abschaltung

Konformitätsbewertung vor Marktzugang

- Interne Kontrolle: Hersteller prüfen die Einhaltung selbst
- Benannte Stellen: staatlich autorisierte Prüfstellen für biometrische Anwendungen und andere risikoreiche Systeme

Anwendungsbereiche

Die Verordnung gilt nur für Anwendungsbereiche des EU-Rechts (also nicht für Zuständigkeitsbereiche der Mitgliedsstaaten oder Bereiche der nationalen Sicherheit). Von der Verordnung ausgenommen sind KI-Systeme, die ausschließlich militärischen oder verteidigungspolitischen Zwecken dienen. Gleiches gilt für KI-Systeme, die ausschließlich für Forschung und Entwicklung entwickelt und in Betrieb genommen werden oder die für Forschungs-, Test- oder Entwicklungstätigkeiten eingesetzt werden, bevor sie in Verkehr gebracht oder in Betrieb genommen werden. Ebenfalls ausgenommen sind Personen, die KI-Systeme zu nicht gewerblichen Zwecken nutzen.

Im Falle eines Verstoßes gegen die Sanktionsbestimmungen kann eine Geldbuße verhängt werden, deren Höhe sich nach dem weltweiten Jahresumsatz des betreffenden Unternehmens richtet und einen Höchstbetrag von 7 % nicht überschreiten darf. Bei Unternehmen der Kategorie KMU sowie bei Start-ups ist eine gestaffelte Bemessungsgrundlage zulässig.

III. Abwehr von KI-gestützten Angriffen auf Unternehmen

Bedrohungsbild

Generative Modelle senken die Einstiegshürden für Angreifer:innen und beschleunigen bekannte Taktiken: hochgradig personalisiertes Phishing, täuschend echte Deepfakes in Bild und Ton, automatisiert erzeugte oder verschleierte Malware sowie schnell angepasste Exploits nach der Veröffentlichung neuer Schwachstellen. Dadurch steigen Tempo, Präzision und Skalierung von Kampagnen, während traditionelle Abwehrmechanismen unter Druck geraten. Dokumentierte Vorfälle mit Deepfake-Videokonferenzen und gefälschten Freigaben zeigen, wie leicht etablierte Kommunikations- und Freigabeprozesse ausgenutzt werden können.

Erkennung und Analyse

Technisch unterscheidet sich die Identifikation KI-gestützter Angriffe nur teilweise von klassischer Erkennung, erfordert jedoch feinere Telemetrie und eine stärkere Kontextauswertung. In menschenzentrierten Szenarien (Phishing, Deepfakes) sind Sensibilisierung und Tools zur Erkennung synthetischer Inhalte zentral. In Anwendungen und Web-Infrastrukturen braucht es kontinuierliches, granuläres Monitoring von Protokollen und Parametern, um Bot-Verhalten von legitimen Sessions zu trennen; bei verschlüsseltem Traffic helfen ergänzend Reputation, Geo-Signale und Heuristiken. Gegen neuartige Malware-Varianten stoßen Signaturen an Grenzen – signaturlose, verhaltensbasierte Verfahren und ML-gestützte Klassifikatoren gewinnen an Bedeutung.

Verteidigungsprinzipien

Eine rein signaturbasierte Abwehr reicht nicht mehr aus. Notwendig sind Baselines für normales System- und Nutzerverhalten, Anomalie-Erkennung (z. B. UBA/UEBA-Ansätze) und Telemetrie, die Abweichungen früh sichtbar macht. Ergänzend sollten Reaktionsprozesse KI-spezifische Vektoren berücksichtigen (Prompt-Manipulation, modellseitiger Missbrauch) und technische wie organisatorische Maßnahmen zusammenführen.

Schutzmaßnahmen und Gegenstrategien

Priorität haben starke Identitäten (MFA, wo sinnvoll auch passwortlose Verfahren), durchgängige Verschlüsselung mit kontrollierter Inspektion, strikte Zugriffskontrollen sowie Protokollierung und Analyse aller Interaktionen. KI-spezifische Kontrollen adressieren Prompt-Injection, Modell-Manipulation und riskante Ausgaben; Traffic-Management hilft, volumetrische Angriffe zu mitigieren, ohne legitimen Verkehr zu stören. Übungen (Tabletop), Penetrationstests inkl. KI-gestützter Werkzeuge und realistische Phishing-Simulationen härten Prozesse und Teams.

Rolle von Mensch und KI

KI-gestützte Erkennung (IDS/IPS mit ML) identifiziert Abweichungen in Echtzeit und entlastet Teams – dennoch bleiben erfahrene Analystinnen und Analysten unverzichtbar: LLMs und ML beschleunigen Routine, können aber halluzinieren oder Kontext falsch bewerten; kritische Entscheidungen brauchen menschliche Prüfung und klare Verantwortlichkeiten.

Lessons Learned aus Vorfällen

Deepfake-basierte Anweisungen und freigegebene Zahlungen zeigen: Technische Kontrollen müssen mit Prozess-Leitplanken zusammenwirken – z. B. verbindliche Rückruf-Kanäle, Vier-Augen-Prinzip, Limit-Policies für Transaktionen sowie Identitätsprüfung außerhalb der gerade genutzten Kommunikationsumgebung. So lassen sich täuschend echte, aber unplausible Anfragen zuverlässig ausfiltern.

Blick in die Praxis

Mehrere dokumentierte Vorfälle zeigen, wie Deepfakes in Video-Calls oder per Audio Führungskräfte täuschend echt imitieren, um Freigaben oder Zahlungen zu erschleichen. Solche Angriffe umgehen gewohnte Kommunikationswege und unterstreichen die Notwendigkeit kombinierter Maßnahmen: technische Erkennung synthetischer Inhalte, starke Identitätsprüfungen und prozessuale Leitplanken wie Rückruf über bekannte Kanäle und Vier-Augen-Prinzip. Parallel stoßen signaturbasierte Verfahren gegen neuartige, automatisiert veränderte Malware rasch an Grenzen; verhaltensbasierte, signaturlose Ansätze und ML-gestützte Klassifikation erhöhen hier die Erkennungsqualität.

IV. KI-gestützte Abwehr von Cyberangriffen

Rolle der KI in der Verteidigung

KI ist nicht nur ein Werkzeug für Angreifer, sondern auch ein zentraler Baustein moderner Cyberabwehr. Sie verarbeitet enorme Datenmengen, erkennt Muster, die menschlichen Analyst:innen verborgen bleiben würden, und kann Bedrohungen in Echtzeit identifizieren.

Erkennung und Vorhersage

KI-Modelle analysieren Log-Daten, Netzwerkflüsse und Nutzerverhalten, um Abweichungen von etablierten Normalwerten früh zu erkennen. Auf dieser Basis ermöglichen prädiktive Verfahren eine antizipative Absicherung, indem sie wahrscheinliche Angriffspfade und Schwachstellen aus historischen Mustern ableiten. Auch das Threat Hunting gewinnt an Tempo, weil Hinweise aus externen Quellen (z. B. Foren, Social Media oder Untergrund-Marktplätzen) automatisiert ausgewertet, in den Kontext gestellt und priorisiert werden.

Automatisierte Reaktionen

In SOAR-Umgebungen verknüpft KI Erkennungsereignisse mit vordefinierten Playbooks: verdächtige Aktivitäten werden automatisch gedrosselt oder blockiert, kompromittierte Systeme isoliert und Richtlinien dynamisch angepasst. Kontinuierliche Feedback-Schleifen mit Human-in-the-Loop sichern die Qualität, reduzieren Fehlalarme und verbessern die Modelle im laufenden Betrieb.

Integration in bestehende Sicherheitsarchitekturen

KI-basierte Funktionen ergänzen etablierte Kontrollen wie Firewalls, EDR/XDR, API-Sicherheit oder DDoS-Schutz. Voraussetzung sind klare Zuständigkeiten, standardisierte Telemetrieschnittstellen und lückenlose Protokollierung. In großskaligen Infrastrukturen stellen GPU-/DPU-gestützte Pipelines die nötige Performance für Analyse und Inferenz sicher, ohne Kernprozesse auszubremsen.

KI-Sicherheitsfunktionen reichen von reaktiver bis proaktiver Verteidigung

Blick in die Praxis

In produktiven Umgebungen analysieren KI-gestützte Systeme Netzwerkströme in Echtzeit, erkennen DDoS-Muster früh und leiten automatisierte Gegenmaßnahmen ein – etwa dynamisches Drosseln oder das Isolieren verdächtiger Quellen. Für API- und Applikationsschutz werden KI-fähige WAF/IDS/IPS-Funktionen mit User-/Entity-Behavior-Analysen kombiniert, um abweichendes Verhalten schneller zu identifizieren. Wo verschlüsselter Traffic die Sicht einschränkt, sorgen kontrollierte TLS-Inspektions- und Orchestrierungsmechanismen dafür, dass Anomalien trotz Ende-zu-Ende-Verschlüsselung sichtbar werden. Zur Leistungs- und Latenzsicherung setzen moderne Architekturen zudem auf DPU-/GPU-beschleunigte Pfade, die Kryptografie, Paketverarbeitung und Telemetrie entlasten und so Erkennungs- und Reaktionszeiten weiter verkürzen.

V. Schlussfolgerungen und Ausblick

KI ist zugleich Werttreiber und Angriffsfläche. Erfolgreich ist, wer Governance, Sicherheit, Transparenz und Compliance von Beginn an integriert und den Betrieb messbar macht (Metriken, SLOs, Drift-Kontrollen, Incident-Playbooks) – kurz gesagt: wer kleine, risikoarme Schritte konsequent dokumentiert und verbessert.

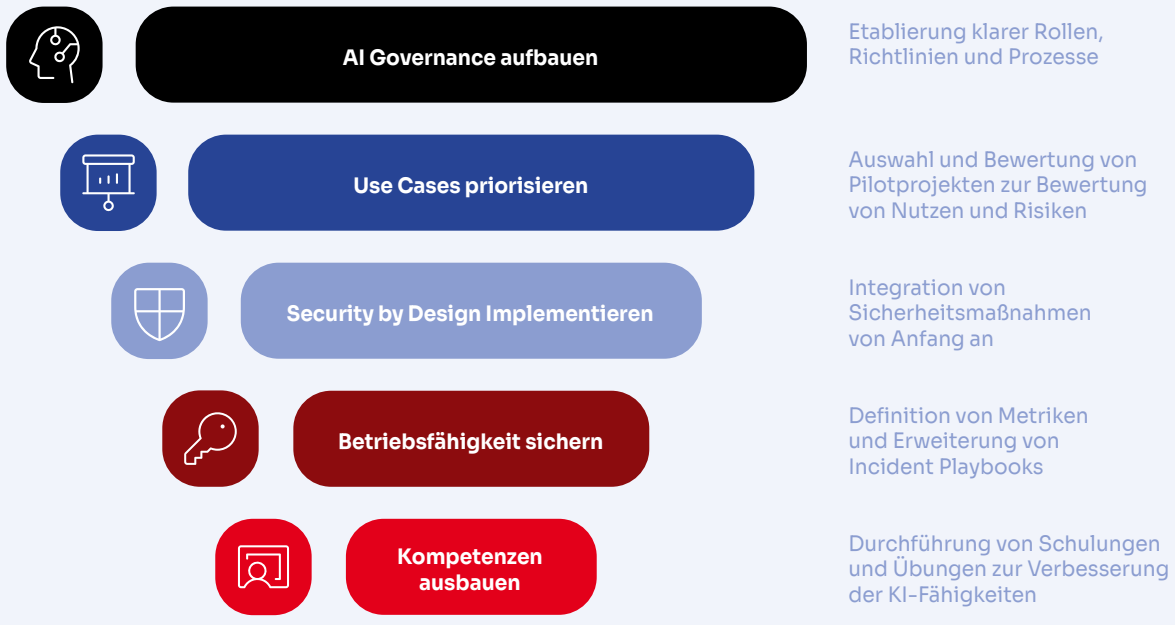
Nächste Schritte

Für eine sichere und nachhaltige Nutzung von KI sollten Unternehmen zunächst eine belastbare AI-Governance aufbauen. Dazu gehören klare Rollen, Richtlinien, Freigabeprozesse sowie eine fundierte Risiko- und Impact-Bewertung. Darauf aufbauend gilt es, die Use Cases zu priorisieren: Pilotprojekte helfen, Nutzen, Risiken und Datenlage transparent zu bewerten und erste Erfahrungen kontrolliert zu sammeln.

Von Beginn an sollte Security by Design verankert werden – mit durchgängigen Zugriffskontrollen, Verschlüsselung, lückenloser Protokollierung, klaren Output- und Prompt-Sicherheitsmechanismen sowie Drift-Monitoring für die Modelle. Ebenso wichtig ist es, die Betriebsfähigkeit zu sichern, indem relevante Metriken und Service Level Objectives definiert, Incident Playbooks um KI-spezifische Szenarien erweitert und Telemetriedaten konsistent erfasst werden.

Schließlich müssen die Kompetenzen ausgebaut werden: Dazu gehören Awareness-Maßnahmen für alle Mitarbeitenden, vertiefte Trainings für Entwicklung, Betrieb und Rechtsabteilungen sowie regelmäßige Übungen, die auch KI-basierte Angriffsszenarien einbeziehen.

Aufbau einer sicheren und nachhaltigen KI-Nutzung



Impressum

eco – Verband der Internetwirtschaft e. V.

Lichtstraße 43h
50825 Köln

Tel.: +49 221 70 00 48-0
E-Mail: info@eco.de

www.eco.de

Ansprechpartner:innen

Cornelia Schildt
Senior Projektmanagerin IT-Sicherheit

E-Mail: cornelia.schildt@eco.de

Olaf Pursche
Leiter der eco Kompetenzgruppe Sicherheit

E-Mail: sicherheit@eco.de

Autor:innen der Studie

Ralf Benzmüller, Lisa Fröhlich, Christof Klaus, Olaf Pursche, Cornelia Schildt, Dr. Or Sela, Maurice Striek, Marcel Rieger

Disclaimer

Diese Publikation dient ausschließlich der allgemeinen Information und stellt keine rechtliche oder sonstige fachliche Beratung dar. Sie kann und soll eine individuelle Beratung unter Berücksichtigung der besonderen Umstände des Einzelfalls nicht ersetzen.

Die Inhalte wurden mit der gebotenen Sorgfalt erstellt; eine Gewähr für die Richtigkeit, Vollständigkeit und Aktualität der Informationen wird jedoch nicht übernommen.

Eine Haftung für Schäden, die aus der Nutzung dieser Publikation entstehen, ist – gleich aus welchem Rechtsgrund – ausgeschlossen, soweit dies gesetzlich zulässig ist.