Marcel Küppers

Cyber-sicherheit & KI

Strategien und Praxis für IT- und Security-Verantwortliche

mitp

Inhaltsverzeichnis

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Kapitel 1

Das neue Paradigma: KI trifft Cybersicherheit

Cybersicherheit war schon immer ein Wettlauf – aber lange Zeit ein Wettlauf mit vertrauten Regeln: Angreifer entwickeln neue Methoden, Verteidiger bauen Detektion und Kontrollen nach, und die Effizienzgewinne entstehen hauptsächlich durch bessere Tools, mehr Automatisierung und mehr Daten. Mit moderner KI – insbesondere großen Sprachmodellen (LLMs) und generativen Modellen – verschieben sich diese Regeln gerade spürbar. Nicht, weil KI »noch ein Tool« ist, sondern weil sie eine neue Klasse von Fähigkeiten in die Breite bringt: Verständnis, Synthese und Handlungsfähigkeit in natürlicher Sprache und Code.

Zum ersten Mal verfügen Angreifer und Verteidiger über Werkzeuge, die nicht nur Muster erkennen, sondern Bedeutung ableiten können: aus Logdaten, Netzwerktelemetrie, Identitäts- und Verhaltenssignalen. Sie können natürliche Sprache verarbeiten – also genau das Medium, in dem ein großer Teil moderner Arbeit stattfindet: E-Mails, Support-Tickets, Runbooks, Dokumentation, Chat-Verläufe, Policies. Und sie können Code nicht nur lesen, sondern generieren, variieren und testen – von Scripts über Infrastructure-as-Code bis hin zu Exploit-ähnlichen Proofs of Concept. Was früher Spezialwissen, Zeit und mehrere Rollen erforderte, kann heute teilweise in Minuten orchestriert werden.

Damit ist KI kein inkrementelles Upgrade der bisherigen Security-Landschaft. Es ist ein Paradigmenwechsel: Wir gehen weg von überwiegend deterministischen, regelbasierten Systemen (»wenn X, dann Y«) hin zu lernenden, probabilistischen Systemen, die Wahrscheinlichkeiten und Kontext bewerten. Diese Systeme lassen sich nicht vollständig »durchkonfigurieren« wie klassische Security-Controls. Stattdessen müssen sie gesteuert werden: durch Datenqualität, klare Aufgabenabgrenzung, überprüfbare Outputs, robuste Grenzen (Guardrails) und kontrollierte Integrationen. Kurz: Wir konfigurieren nicht nur Tools – wir managen ein neues Risikoprofil.

Dieser Wandel wirkt in mehrere Richtungen:

Erstens: Verteidiger können schneller werden. KI kann operative Last reduzieren, Analysen beschleunigen, triagieren, korrelieren, zusammenfassen, Prioritäten vorschlagen und sogar Gegenmaßnahmen vorbereiten. Das betrifft SOC-Prozesse genauso wie Secure SDLC, Threat Modeling, Asset Discovery oder Third-Party-

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Risk. Die Chance ist real: mehr Wirkung pro Kopf, bessere Reaktionszeiten, konsistentere Entscheidungen.

Zweitens: Angreifer werden skalierbarer und glaubwürdiger. Social Engineering wird qualitativ besser und quantitativ massentauglich, Malware- und Script-Varianten entstehen schneller, Reconnaissance und Targeting werden effizienter, und Angreifer können Verteidigungsmaßnahmen leichter umgehen, indem sie Sprache, Verhalten und Infrastruktur dynamisch anpassen. KI senkt nicht nur die Kosten – sie senkt vor allem die Reibung. Und das ist im Angreifer-Ökosystem ein enormer Hebel.

Drittens (und oft unterschätzt): Der Einsatz von KI erzeugt neue Angriffsflächen und neue Compliance-Risiken. Prompt Injection, Datenabfluss über Modelle, Schatten-KI im Unternehmen, unkontrollierte Tool-Integrationen, fehlerhafte oder halluzinierte Outputs, Lizenz- und Urheberrechtsfragen, Datenschutz, sowie regulatorische Anforderungen – all das kann aus einem »Produktivitätsgewinn« schnell ein strukturelles Risiko machen, wenn Governance, Architektur und Betrieb nicht mitwachsen.

Für IT- und Security-Verantwortliche ergeben sich daraus drei zentrale Leitfragen, die über die nächsten Jahre entscheiden werden:

1. Wie nutzen wir KI so, dass unsere Verteidigung messbar stärker wird?

   Nicht als Spielerei, sondern als Capability: schnellere Detection, bessere Priorisierung, geringere MTTR, robustere Engineering-Prozesse, weniger manuelle Fehler.

2. Wie schützen wir uns gegen Angreifer, die KI als Beschleuniger einsetzen?

   Das umfasst neue Phishing-Qualität, automatisierte Recon, variantenreiche Payloads, aber auch KI-gestützte Umgehung von Kontrollen – also eine Verschiebung im Threat Model.

3. Wie verhindern wir, dass KI selbst zum Risiko wird – technisch, rechtlich und organisatorisch?

   Governance, Datenklassifizierung, Zugriffskontrollen, Logging, Lieferkette, Modell- und Tool-Risiken sowie klare Verantwortlichkeiten müssen Teil des Security-Programms werden – nicht ein Add-on.

Dieses Kapitel legt das Fundament für genau diese Neubewertung. Es klärt zentrale Begriffe, grenzt KI-basierte Ansätze von klassischen Security-Mechanismen ab und zeigt, warum sich Sicherheitsstrategien in den kommenden Jahren verändern müssen: weniger Fokus auf starre Regeln als alleinige Wahrheit – mehr Fokus auf Resilienz, Überprüfbarkeit, kontrollierte Automatisierung und ein Threat Model, in dem sowohl Menschen als auch Maschinen auf beiden Seiten des Spielfelds agieren.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

1.1 Was wir mit KI in der Cybersicherheit wirklich meinen

Bevor wir über Architektur, Risiken und konkrete Use Cases sprechen, müssen wir ein gemeinsames Begriffsverständnis schaffen. »KI« ist in der Sicherheitsdebatte inzwischen ein Sammelbegriff für sehr unterschiedliche Technologien – und genau das führt regelmäßig zu Missverständnissen: Manche meinen klassische Anomalieerkennung in einem SIEM, andere denken an Chatbots, wieder andere an autonom handelnde Agents, die Tickets schließen und Changes ausrollen. Wenn wir nicht sauber trennen, reden wir schnell aneinander vorbei – und treffen am Ende falsche Entscheidungen bei Tooling, Governance und Risikobewertung.

Wichtig ist auch die Abgrenzung: Nicht alles, was »KI« genannt wird, ist neu. Viele Security-Produkte nutzen seit Jahren statistische Verfahren, Heuristiken oder ML-basierte Klassifikatoren. Neu ist weniger die Existenz von ML – neu ist, dass generative Modelle und insbesondere LLMs Fähigkeiten in die Breite bringen, die bisher stark menschlich geprägt waren: Sprache verstehen, Wissen synthetisieren, Code erzeugen, kontextbezogen argumentieren und Aufgabenketten orchestrieren. Genau diese Fähigkeiten verschieben die operative Realität im Security-Alltag.

1.1.1 Künstliche Intelligenz als Oberbegriff

Künstliche Intelligenz (KI) beschreibt grundsätzlich Systeme, die Aufgaben übernehmen, für die bislang menschliche Intelligenz erforderlich war. In der Cybersicherheit betrifft das vor allem vier Klassen von Fähigkeiten:

- Muster erkennen und klassifizieren:

  KI kann aus großen Datenmengen wiederkehrende Strukturen ableiten – etwa typische Merkmale von Malware, verdächtige Kommunikationsprofile oder auffällige Login-Muster. Der Kern ist: Aus »viel Telemetrie« wird »eine bewertbare Hypothese«.

- Natürliche Sprache verstehen und verarbeiten:

  Ein erheblicher Teil von Security passiert in Textform: Phishing-Mails, Ticket-Kommentare, Incident-Reports, Policies, Chat-Verläufe, Post-Mortems. KI kann diese Inhalte zusammenfassen, einordnen, priorisieren und in Maßnahmen übersetzen – also genau das, was bisher viel menschliche Arbeitszeit gebunden hat.

- Entscheidungen unter Unsicherheit unterstützen:

  Security ist selten schwarz-weiß. Meist geht es um Wahrscheinlichkeiten, unvollständige Informationen, widersprüchliche Signale. KI kann hier helfen, Optionen zu bewerten, Risiken zu gewichten und sinnvolle nächste Schritte vorzuschlagen – nicht als Ersatz für Verantwortung, aber als Beschleuniger für Analyse und Triage.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Neue Inhalte generieren:

  Dazu zählen Texte, Playbooks, Detection-Queries, Code, Tests, IaC-Snippets, aber auch verständliche Erklärungen für unterschiedliche Zielgruppen (Engineering vs. Management). Das ist einer der größten Produktivitätshebel – und gleichzeitig eine der größten Risikoquellen, wenn Outputs ungeprüft übernommen werden.

Im Alltag der Cybersicherheit sind für uns vor allem vier Teilbereiche relevant, die sich teilweise überlappen, aber unterschiedliche Stärken und Risiken haben:

- Machine Learning (ML):

  ML-Systeme lernen Muster aus Daten, statt Regeln ausschließlich manuell vorzugeben. Typische Security-Anwendungen sind Klassifikation (»bösartig vs. gutartig«), Scoring (Risiko- oder Prioritätswerte), Clustering (Gruppierung ähnlicher Ereignisse) und Anomalieerkennung (Abweichungen vom Normalverhalten). ML ist oft dort stark, wo viele strukturierte Daten vorliegen und ein klarer Feedback-Loop möglich ist.

- Deep Learning (DL):

  Deep Learning ist eine Unterkategorie von ML und nutzt komplexe neuronale Netze, um sehr hochdimensionale Daten zu verarbeiten – z. B. Sprache, Text, Sequenzen oder Binärdaten. In Security kann DL etwa bei der Analyse von Malware-Familien, bei der Auswertung großer Textmengen oder bei Verhaltensmustern helfen. Der Preis ist häufig: weniger Transparenz, höhere Anforderungen an Daten und Rechenleistung und schwierigere Erklärbarkeit.

- Large Language Models (LLMs):

  LLMs sind Modelle, die natürliche Sprache verstehen und generieren können. Praktisch werden sie oft als »Assistenten« genutzt: sie lesen und schreiben Text, erklären Zusammenhänge, erstellen Zusammenfassungen, formulieren Tickets, generieren Code, helfen bei Root-Cause-Analysen und können zunehmend auch Werkzeuge ansteuern (z. B. Abfragen gegen Logsysteme), wenn man sie entsprechend integriert. Ihr großer Vorteil ist ihre Schnittstelle: Sprache – und damit die Fähigkeit, Security-Wissen und Security-Arbeit in natürlicher Form zu skalieren.

- Generative KI (GenAI):

  Generative KI ist der Oberbegriff für Modelle, die neue Inhalte erzeugen können – Text, Code, Bilder, Audio und mehr. LLMs sind ein Teil davon (GenAI für Text/Code). In Security ist GenAI besonders relevant für Automation, Dokumentation, Detection Engineering, Secure Coding Support und Training. Gleichzeitig erhöht GenAI das Risiko für überzeugende Täuschung (Phishing, Impersonation) und für unkontrollierte Content-Erzeugung (z. B. unsichere Code-Snippets oder falsche Handlungsempfehlungen).

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Die zentrale Konsequenz aus dieser Begriffsarbeit ist simpel, aber entscheidend: Wenn wir »KI in der Cybersicherheit« sagen, meinen wir nicht ein einzelnes Produkt und nicht eine einzelne Methode. Wir meinen ein Spektrum von Fähigkeiten – von statistischer Klassifikation bis hin zu sprach- und handlungsfähigen Systemen. Und je nachdem, welche Kategorie wir einsetzen, ändern sich Nutzen, Grenzen, Angriffsfläche, Governance-Anforderungen und die Art, wie wir Sicherheit »bauen und betreiben«.

1.1.2 Machine Learning vs. Regeln – der fundamentale Unterschied

Viele der klassischen Security-Kontrollen, mit denen wir groß geworden sind, funktionieren nach einem klaren Prinzip: Regeln und bekannte Muster. Signaturbasierter Virenschutz, IDS/IPS-Regeln, YARA-Signaturen, SIEM-Korrelationen oder starre DLP-Policies sind im Kern deterministisch aufgebaut. Sie folgen Logiken wie: »Wenn Bedingung A und B erfüllt sind, dann ist das Ereignis bösartig« – oder zumindest alarmwürdig.

Das hat Vorteile: Solche Systeme sind oft transparent, reproduzierbar und gut zu auditieren. Wenn ein Alarm auslöst, kann man häufig nachvollziehen, welche Regel gegriffen hat. Der Nachteil ist ebenso offensichtlich: Diese Ansätze funktionieren am besten dort, wo das Problem bereits bekannt ist. Sie sind stark gegen »das, was wir schon gesehen haben« – und schwächer gegen Varianten, neue Taktiken oder bewusstes Evasion-Verhalten.

Machine Learning (ML) dreht diese Logik teilweise um. Statt dass Menschen jede Erkennungslogik explizit formulieren, lernt ein Modell Muster aus Daten. Drei Punkte sind dabei entscheidend:

- Daten als Grundlage:

  ML-Systeme brauchen Trainingsdaten (und oft auch Validierungsdaten), die das »Normal« und das »Verdächtig« abbilden. Die Qualität der Ergebnisse hängt direkt davon ab, wie gut diese Daten sind: vollständig, repräsentativ, sauber gelabelt und aktuell.

- Wahrscheinlichkeiten statt Wahrheiten:

  ML entscheidet selten binär. Statt »gut/böse« liefert es häufig ein Score, eine Wahrscheinlichkeit oder eine Rangfolge: »mit 93 % verdächtig« oder »Top 5 % auffällig im Vergleich zur Baseline«. Das ist in der Praxis extrem nützlich – aber es zwingt uns, mit Unsicherheit professionell umzugehen.

- Verallgemeinerung auf Unbekanntes:

  Der große Hebel von ML ist die Fähigkeit, Muster zu erkennen, die nicht exakt im Regelwerk stehen. Ein Modell kann Varianten auffangen, die keiner Signatur entsprechen, und Abweichungen markieren, die noch nie exakt so vorkamen – gerade bei Verhalten, Sequenzen oder Kombinationsmustern.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Für Sie als IT- und Security-Verantwortlichen hat das eine konkrete Konsequenz: Sie »programmieren« die Erkennung nicht mehr primär über einzelne Regeln, sondern über Rahmenbedingungen, in denen das System lernt und entscheidet. Das bedeutet in der Praxis:

- Sie entscheiden, welche Datenquellen überhaupt in das Modell fließen (Identitätsdaten, Endpoint-Telemetrie, Cloud-Logs, Netzwerkdaten, Tickettexte etc.).
- Sie definieren, wie Erfolg gemessen wird: Was ist ein False Positive? Was kostet ein False Negative? Was ist die akzeptable Alarmrate?
- Sie bestimmen, welche Autonomie das System erhält: Nur Vorschläge machen? Automatisch Tickets erstellen? Quarantäne auslösen? Accounts sperren?
- Sie bauen Kontrollen um das Modell herum: Monitoring von Drift, Qualitätschecks, human-in-the-loop, Rollback-Mechanismen.

Kurz gesagt: Bei regelbasierten Systemen ist die Hauptarbeit »Regeln schreiben und pflegen«. Bei ML-basierten Systemen ist die Hauptarbeit »Daten, Bewertung und Leitplanken managen«. Beides ist Security Engineering – aber es ist ein anderer Hebel und ein anderes Betriebsmodell.

1.1.3 LLMs und GenAI – warum sie Security besonders verändern

LLMs und generative KI verändern Security nicht nur, weil sie »bessere ML-Modelle« sind, sondern weil sie eine neue Schnittstelle in den Betrieb bringen: Sprache und Code als universelle Arbeitsoberfläche. Und genau darin liegt die Sprengkraft – positiv wie negativ.

LLMs sind für Security deshalb so relevant, weil sie drei Fähigkeiten kombinieren, die bisher selten in einem System zusammenkamen:

- Natürliche Sprache verstehen und erzeugen:

  Ein großer Teil der Security-Arbeit steckt in unstrukturierten Artefakten: Phishing-Mails, Incident-Notes, Chat-Verläufe, Post-Mortems, Policies, Risikoausnahmen, Audit-Fragen, Vendor-Fragebögen. LLMs können diese Inhalte lesen, zusammenfassen, strukturieren, vergleichen und in Handlungsvorschläge übersetzen.

- Code lesen, schreiben und erklären:

  Security lebt von Code: Detection-Queries, Regex/YARA, Sigma-Regeln, SIEM-Korrelationen, Scripts, IaC, CI/CD-Konfigurationen. LLMs können helfen, Code zu reviewen, Schwachstellen zu erklären, Fixes vorzuschlagen, Tests zu erzeugen oder Regeln schneller zu formulieren. Das beschleunigt besonders DevSecOps- und Detection-Engineering-Workflows.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Interaktion und Orchestrierung über Prozesse hinweg:

  LLMs sind nicht nur »Textgeneratoren«, sondern werden in der Praxis als Interface genutzt: als Chatbot im SOC, als Assistenz im Change-Prozess, als Unterstützung im Incident-Management. In Kombination mit Tool-Integrationen können sie Informationen abfragen, Tickets anlegen, Reports erstellen oder standardisierte Workflows anstoßen.

In der Security-Praxis sehen wir LLMs und GenAI bereits in sehr konkreten Rollen, zum Beispiel:

- automatische SOC-Lageberichte aus Events, Alerts und Tickets,
- Incident-Zusammenfassungen inkl. Timeline, Impact und Next Steps,
- Assistenz bei der Log-Analyse (»Was ist hier auffällig? Welche Hypothesen passen?«),
- Policy-Entwurf und Review (z. B. Abgleich mit Standards/Controls),
- Unterstützung im Secure SDLC (Code Review, Threat Modeling, Abuse Cases, Security Tests).

Der entscheidende Punkt ist jedoch: Die gleichen Fähigkeiten stehen auch Angreifern zur Verfügung – und dort wirken sie als Skalierungsfaktor. Besonders relevant sind:

- Hochqualitative, überzeugende Phishing- und Impersonation-Mails (stilistisch sauber, kontextbezogen, sprachlich passend),
- Automatisierte Reconnaissance (Auswertung öffentlicher Informationen, Ableitung von Rollen/Prozessen, Identifikation von Angriffswegen),
- Schnelle Variation von Payloads und Exploit-ähnlichem Code (nicht zwingend »Hollywood-Malware«, aber genug, um Defenses zu testen, Evasion zu betreiben oder Initial Access zu erleichtern).

Wichtig ist dabei die richtige Einordnung: Viele Modelle sind durch Policies und technische Schutzmechanismen begrenzt und nicht jede direkte »Malware-Generierung« ist trivial. Aber selbst ohne »vollautomatische Exploit-Fabrik« reicht die Kombination aus Sprachqualität, Automatisierung und Codeassistenz aus, um die Effektivität vieler Angriffe deutlich zu erhöhen – vor allem bei Social Engineering und bei der Geschwindigkeit, mit der Angreifer iterieren können.

Die Konsequenz für Security-Programme ist klar: LLMs/GenAI sind zugleich Verteidigungsbeschleuniger und Angriffsverstärker. Wer sie nur als Produktivitäts-Tool betrachtet, unterschätzt das Risiko. Wer sie nur als Risiko betrachtet, verschenkt einen realen Hebel. Die richtige Antwort ist: gezielter Einsatz mit klaren Grenzen, messbaren Ergebnissen und einem aktualisierten Threat Model.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

1.2 Von regelbasierten zu lernenden Security-Systemen

Um zu verstehen, warum KI in der Cybersicherheit mehr ist als ein weiteres Feature-Upgrade, lohnt sich ein kurzer Blick zurück. Security-Mechanismen haben sich historisch entlang einer klaren Linie entwickelt: erst Signaturen und harte Regeln, dann datengetriebene Modelle, und jetzt – mit GenAI und LLMs – Systeme, die nicht nur erkennen, sondern kommunizieren, erklären und (teilweise) handeln können. Jede Stufe hat unsere Verteidigung leistungsfähiger gemacht, aber auch neue Betriebs- und Risikofragen geschaffen.

1.2.1 Die Ära der Signaturen und starren Regeln

Die lange dominierende Phase der IT-Sicherheit war geprägt von deterministischen Kontrollen: Ein Ereignis wird gegen ein vordefiniertes Muster geprüft, und daraus folgt eine eindeutige Entscheidung. Typische Vertreter dieser Ära sind:

- Virenscanner mit Signaturen: Hashes, Bytefolgen, bekannte Malware-Familien.
- Firewalls mit festen Regeln: »Port X nach Y blockieren«, »Subnetz A darf nicht mit Subnetz B sprechen«.
- IDS/IPS mit statischen Mustern: bekannte Angriffssignaturen, Protokollanomalien, feste Regeln.
- SIEM-Korrelationen: händisch gepflegte Logik wie »wenn ungewöhnlicher Login + Admin-Änderung + Datenexfil-Indikator, dann Alarm«.

Das war über viele Jahre der Standard – aus guten Gründen. Diese Kontrollen haben drei klassische Stärken:

1. Nachvollziehbarkeit: Man kann meist erklären, warum etwas blockiert oder alarmiert wurde.
2. Determinismus: Gleiche Inputs führen zu gleichen Outputs; das ist operativ stabil.
3. Compliance-Freundlichkeit: Kontrollen lassen sich gut dokumentieren (»Regel X mitigiert Risiko Y«), auditieren und reproduzieren.

Mit der Zeit wurden jedoch die Grenzen immer deutlicher – besonders in dynamischen, hochvernetzten Umgebungen:

- Reaktivität: Regel- und signaturbasierte Systeme erkennen primär das, was bereits bekannt ist. Neue Varianten, neue Taktiken und Evasion laufen leichter durch.
- Pflegeaufwand: Signaturen, Regeln, Ausnahmen, Whitelists – alles muss dauerhaft gepflegt werden. Das skaliert nur begrenzt, vor allem bei hoher Change-Rate.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Schwache Skalierung bei Datenvolumen und Komplexität: Mit Cloud, SaaS, Container-Plattformen und verteilten Identitäten explodieren Events und Kontextsignale. »Mehr Regeln« ist dann selten die Lösung – oft ist es der direkte Weg in Alarmmüdigkeit.

Die Kernerkenntnis: Regelwerke funktionieren gut für bekannte, klar abgrenzbare Muster. Sie versagen dort, wo Verhalten, Kontext und Variabilität dominieren.

1.2.2 Übergang zu ML-basierten Detektionen

Mit zunehmender Komplexität moderner IT-Landschaften – Cloud, Microservices, Remote Work, BYOD, Identity-first-Architekturen – wurde ein Problem offensichtlich: Wir können nicht mehr jede relevante Event-Kombination sinnvoll per Hand modellieren. Nicht nur, weil es zu viele sind, sondern weil sich »Normalverhalten« ständig verändert: Releases, neue Tools, neue Teams, saisonale Schwankungen, neue Workloads.

An dieser Stelle wurden ML-Ansätze in vielen Security-Bereichen populär, unter anderem:

- Anomalieerkennung im Netzwerkverkehr: Abweichungen in Flows, Protokollen, Datenmengen oder Kommunikationsbeziehungen.
- UEBA (User and Entity Behavior Analytics): Verhalten von Benutzerkonten, Service Accounts, Endpoints oder Workloads wird modelliert, um »ungewöhnliche« Aktivitäten zu identifizieren.
- ML-basierte Malware-Erkennung: Statt nur auf Signaturen zu reagieren, werden Features (z. B. Verhalten, Struktur, Aufrufmuster, Metadaten) genutzt, um neue oder leicht veränderte Samples einzuordnen.

Hier beginnt der eigentliche Paradigmenwechsel: Security wird daten- und modellgetrieben, nicht mehr ausschließlich regelgetrieben. Das verändert die Aufgaben des Security-Teams fundamental.

Wo früher die Kernarbeit war: »Regeln schreiben und Exceptions pflegen«, verschiebt sich der Schwerpunkt nun zu:

- Daten verstehen und kuratieren: Welche Telemetrie ist vollständig? Welche Felder fehlen? Welche Daten sind zuverlässig? Wo gibt es Blind Spots?
- Modelle evaluieren: Wie gut sind Precision/Recall? Wie stabil ist das Modell bei Änderungen? Wie reagieren wir auf Drift?
- Outputs interpretieren und in Maßnahmen übersetzen: Ein Score allein löst kein Incident. Das Team muss Kontext herstellen, Prioritäten setzen, Maßnahmen ableiten und das Ganze operationalisieren.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Das ist eine andere Art von Security-Betrieb: weniger Handwerk am Regelwerk, mehr Engineering an Datenpipelines, Qualitätsmetriken, Feedback-Loops und operativen Leitplanken.

1.2.3 Der Sprung mit GenAI und LLMs

Der nächste Entwicklungsschritt ist nicht nur »mehr ML«, sondern qualitativ anders: GenAI und LLMs bringen Fähigkeiten in Security-Systeme, die bisher überwiegend menschlich waren – nämlich Sprache, Synthese und Wissensarbeit.

Plötzlich können Security-Systeme mit Ihnen interagieren: im SOC, in GRC-Prozessen, in der Softwareentwicklung, im Change Management. Und sie können dabei nicht nur strukturierte Events verarbeiten, sondern auch unstrukturierte Informationen, die in klassischen Pipelines oft »außen vor« waren, zum Beispiel:

- E-Mails und Phishing-Threads
- Chat-Logs aus Slack/Teams
- Runbooks und Confluence-Seiten
- PDF-Reports, Audit-Dokumente, Lieferantenunterlagen
- Ticket-Historien und Incident-Kommunikation

Zusätzlich können sie neue Artefakte erzeugen – also nicht nur bewerten, sondern produzieren:

- Reports (Lagebilder, Executive Summaries, Audit-Readouts)
- Playbooks (Incident-Workflows, Standard Responses, Runbooks)
- Detections (Queries, Regeln, Sigma/YARA-ähnliche Patterns)
- Code und Konfigurationen (Scripts, IaC-Änderungen, CI/CD-Security-Checks)

Damit entsteht ein neues Rollenmodell im Sicherheitsbetrieb: KI ist nicht mehr nur Sensor (erkennen) oder Filter (blockieren), sondern ein aktiver Assistent – und in manchen Architekturen sogar ein Akteur, der Entscheidungen vorbereitet oder unter klaren Bedingungen selbst ausführt.

Und genau hier liegt die strategische Relevanz: Sobald KI nicht nur »Signal« liefert, sondern Arbeitsergebnisse erzeugt und Prozesse antreibt, müssen wir Sicherheit neu denken – inklusive Governance, Verantwortlichkeiten, Qualitätssicherung, Auditierbarkeit und Kontrollmechanismen. Der Paradigmenwechsel ist nicht nur technisch. Er ist organisatorisch und operativ.

Die Entwicklung der Cybersicherheit lässt sich grob in drei Phasen einteilen: von klassisch signatur- und regelbasierten Systemen über erste ML-gestützte Anomalieerkennung hin zu generativer KI und LLMs, die heute ganze Analyse- und Kommunikationsaufgaben übernehmen können. Abbildung 1.1 stellt diese Evolution und den damit verbundenen Paradigmenwechsel übersichtlich dar.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

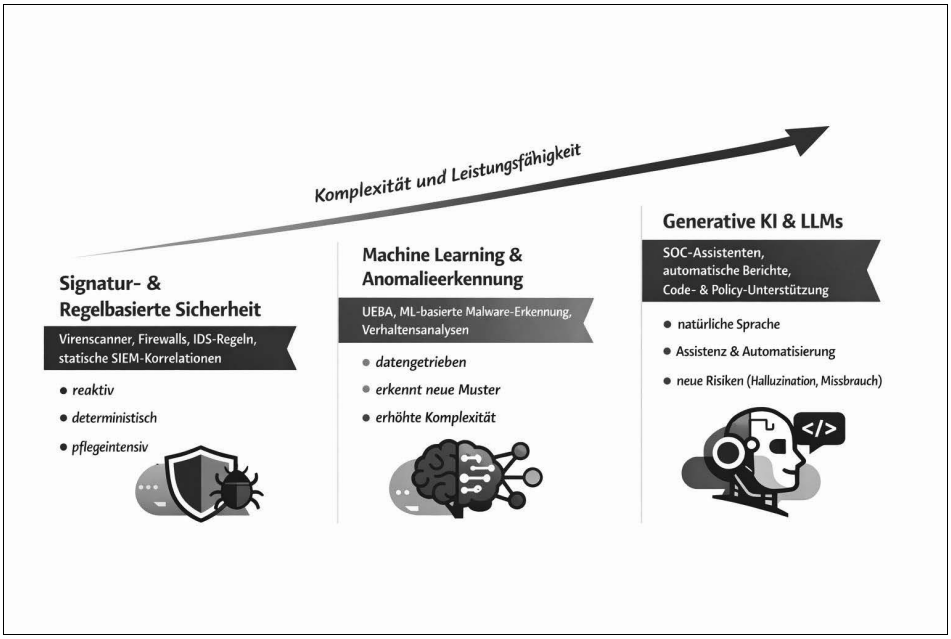

Abb. 1.1: Evolution der Cybersicherheit

1.3 Drei Rollen von KI in der Cybersicherheit

Wenn Sie eine KI-Strategie für Security entwickeln wollen, hilft es enorm, KI nicht als »eine Technologie« zu betrachten, sondern als Akteur in drei Rollen. Denn je nachdem, in welcher Rolle KI auftritt, ändern sich Zielbild, Architektur, Controls und Risikoprofil komplett. In der Praxis begegnet Ihnen KI gleichzeitig als Verstärker Ihrer Teams, als Beschleuniger auf Angreiferseite und als eigenständige Risikoquelle, die Sie wie ein neues System mit eigenen Failure-Modes behandeln müssen.

1.3.1 KI als Verstärker der Verteidiger

In der ersten und naheliegendsten Rolle fungiert KI als Produktivitäts- und Qualitätshebel für Verteidiger. Sie ersetzt nicht die Verantwortung des Security-Teams, aber sie kann große Teile der repetitiven Arbeit beschleunigen und die »kognitive Last« reduzieren – vor allem dort, wo Menschen bisher viel Zeit mit Suchen, Zusammenfassen und Übersetzen von Informationen verbringen.

Typische Einsatzfelder sind:

- SOC-Augmentation (Triage & Analyse):

  KI kann Alerts zusammenfassen, Kontext aus mehreren Systemen anreichern (z. B. Asset-Info, Identity-Kontext, bekannte Kampagnenmuster), ähnliche Fälle

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

clustern und Playbook-Schritte vorschlagen. Der operative Effekt: weniger Zeit, um »den Fall zu verstehen«, mehr Zeit, um Entscheidungen zu treffen.

- Threat Hunting & Detection Engineering:

  Viele Hunting-Iterationen scheitern nicht an fehlenden Ideen, sondern an Reibung: Query-Formulierung, Datenquellen verstehen, Ergebnisinterpretation. KI kann beim Erstellen komplexer Queries helfen, Hypothesen in Suchlogik übersetzen und Ergebnisse erklären – inklusive Vorschläge für »nächste Pivot-Schritte«.

- Incident Response & Kommunikation:

  In Incidents ist Zeit der kritische Faktor – und gleichzeitig steigt der Kommunikationsdruck (Stakeholder, Management, Kunden, ggf. Behörden). KI kann Drafts für Tickets, Status-Updates, Benachrichtigungen, Executive Summaries und Post-Incident-Reports erstellen, basierend auf den vorhandenen Fakten aus Timeline und Artefakten. Das entlastet Teams, ohne dass die inhaltliche Verantwortung abgegeben wird.

- Knowledge Management & »Security-Wissen on demand«:

  Viele Organisationen haben Security-Wissen – aber es ist verteilt: Policies hier, Architekturdiagramme dort, Lessons Learned in Post-Mortems, Ausnahmen in Tickets. Über RAG-Ansätze (Retrieval-Augmented Generation) kann KI dieses Wissen kontextsensitiv verfügbar machen: »Welche Policy gilt für Datenklasse X?« oder »Wie war das Runbook für Incident-Typ Y?«. Das reduziert Suchzeiten und erhöht Konsistenz.

Das Ziel dieser Rolle ist nicht »KI macht Security«, sondern: KI reduziert Low-Value-Arbeit und erhöht die Schlagkraft der Experten. In der Praxis heißt das:

- weniger »Copy & Paste« und Formatierung,
- weniger Zeit mit »Wo finde ich diese Information?«,
- mehr Zeit mit »Was bedeutet das – und was tun wir jetzt?«.

Wenn KI in dieser Rolle richtig eingesetzt wird, verbessert sie Geschwindigkeit, Qualität und Konsistenz – und verschiebt Kapazität zurück in Analyse, Engineering und Risikosteuerung.

1.3.2 KI als Werkzeug der Angreifer

Die zweite Rolle ist unangenehm, aber zentral für ein realistisches Threat Model: Dieselben Fähigkeiten, die Verteidiger produktiver machen, senken auch auf Angreiferseite die Kosten und erhöhen die Erfolgswahrscheinlichkeit – vor allem durch Skalierung und Qualität.

Typische Angreifer-Use-Cases sind:

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Phishing auf Industriestandard:

  Angriffe werden sprachlich fehlerfrei, stilistisch passend, in Landessprache, mit überzeugender Tonalität und hoher Kontextnähe formuliert. Das erhöht nicht nur die Klickrate, sondern auch die Glaubwürdigkeit bei Rückfragen (»Conversation Hijacking«, längere Mail-Threads, präzise Referenzen).

- Social Engineering & Identitätsbetrug:

  Generierte Personas, Lebensläufe, Profile und Kommunikationsmuster machen es leichter, Vertrauen aufzubauen – bis hin zu Deepfake-gestützter Audio-/Video-Interaktion in ausgewählten Szenarien. Der Kern ist nicht »Hollywood«, sondern: weniger Reibung bei der Täuschung.

- Malware- und Exploit-nahe Entwicklung:

  KI kann Codevarianten erzeugen, Obfuskation unterstützen, Evasion-Ideen liefern oder Komponenten schneller zusammensetzen (z. B. Loader, Dropper, C2-Kommunikation). Nicht jede Modellinstanz liefert »fertige Malware«, aber der Output reicht oft, um Iterationszyklen zu beschleunigen und Defenses gezielt zu testen.

- Reconnaissance als Automationspipeline:

  Angreifer können Informationen aus GitHub, LinkedIn, Stellenanzeigen, Dokumentation, öffentlichen Repos, Leak-Datenbanken und Metadaten automatisch auswerten, verdichten und in konkrete Angriffspläne übersetzen: Wer nutzt welche Tools? Welche Rollen gibt es? Wo liegen wahrscheinlich Secrets? Welche Third Parties sind im Einsatz? Welche Prozesse sind anfällig?

Das Ergebnis ist eindeutig: Angriffe werden massiv skaliert und professionalisiert – und zwar nicht nur durch Top-Tier-Gruppen, sondern auch durch weniger technisch versierte Täter, die mit KI »Kompetenz einkaufen«. Das verändert die Baseline: Mehr Angriffe, bessere Qualität, schnellere Iteration. Für Verteidiger bedeutet das: Der »Durchschnittsgegner« wird stärker.

1.3.3 KI als »unberechenbarer Dritter«

Die dritte Rolle wird in vielen Strategien unterschätzt: KI ist nicht nur Tool der Guten oder der Bösen – KI ist auch ein System mit eigenen Fehlermodi. Moderne Modelle sind häufig:

- probabilistisch (sie liefern Wahrscheinlichkeiten, keine Gewissheiten),
- intransparent (die Entscheidungslogik ist nicht vollständig erklärbar),
- nicht deterministisch (gleiche Eingabe kann variierende Ausgaben erzeugen, je nach Kontext/State).

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Daraus entstehen spezifische Risiken, die Sie in Security-Architektur und Governance einplanen müssen:

- Halluzinationen:

  KI kann falsche Aussagen erzeugen, die plausibel klingen – besonders gefährlich, wenn sie in operative Entscheidungen einfließen (z. B. falsche Root-Cause, falsche Remediation, falsche Policy-Auslegung).

- Fehleinschätzungen und Bias:

  Ein Modell kann eine Situation als »harmlos« bewerten, obwohl sie kritisch ist – oder umgekehrt Alarmismus erzeugen. Wenn solche Fehler systematisch auftreten (z. B. aufgrund unausgewogener Trainings-/Kontextdaten), werden sie schwer erkennbar und teuer im Betrieb.

- Mangelnde Auditierbarkeit und Erklärbarkeit:

  In regulierten Umfeldern – oder generell, wenn Sie Entscheidungen später rechtfertigen müssen – ist »das Modell hat es so gesagt« keine ausreichende Begründung. Ohne saubere Protokollierung, Quellenbezug und Kontrollmechanismen wird KI operativ riskant.

Für Security-Teams folgt daraus eine sehr praktische Leitlinie: Behandeln Sie KI wie ein extrem leistungsfähiges, aber fehleranfälliges Teammitglied. Das heißt:

- klare Verantwortlichkeiten (wer entscheidet wirklich?),
- klare Grenzen (welche Aktionen darf KI nie/immer/unter Bedingungen ausführen?),
- klare Kontrollmechanismen (Review, Logging, Validierung, Rollback),
- und ein Betriebsmodell, das mit Unsicherheit umgehen kann.

So eingesetzt wird KI zu einem Multiplikator – nicht zu einem unkontrollierten Risiko.

1.4 Chancen für Verteidiger: Wo KI wirklich hilft

Für IT- und Security-Verantwortliche ist die entscheidende Frage nicht, ob KI »spannend« ist, sondern ob sie konkret messbaren Mehrwert liefert: weniger Risiko, schnellere Reaktion, bessere Qualität – oder schlicht mehr Output pro Team. Der größte Hebel liegt dabei nicht in »vollautonomen« Systemen, sondern in gut designten Assistenz- und Automationsbausteinen, die Engpässe im Betrieb entlasten und vorhandene Daten endlich nutzbar machen.

1.4.1 Entlastung im SOC

Das Security Operations Center (oder die SOC-Funktion) ist oft der Ort, an dem Komplexität und Volumen am brutalsten aufschlagen: zu viele Alerts, zu wenig

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Kontext, zu viele Systeme, zu wenig Zeit. KI kann hier besonders effektiv helfen, weil ein großer Teil der Arbeit aus Triagieren, Zusammenfassen, Priorisieren und Kommunizieren besteht.

Typische KI-gestützte Einsatzfelder sind:

- Automatisierte Lageberichte (täglich / wöchentlich):

  Statt dass Analysten mühsam Tickets, Alerts und Trends manuell konsolidieren, kann KI daraus ein konsistentes Lagebild erstellen: Top Incidents, relevante Veränderungen, wiederkehrende Muster, offene Risiken, SLA-Status.
- Clusterung von Alerts nach Kampagnen oder Mustern:

  Häufig sind 50 »einzelne Alerts« in Wahrheit 1–2 zusammenhängende Ereignisse. KI kann Signale gruppieren (z. B. nach Host, User, TTP-Ähnlichkeit, zeitlicher Nähe), damit Teams nicht in Ticket-Splitting untergehen, sondern kampagnenorientiert arbeiten.
- Priorisierung von Incidents:

  Nicht jeder Alarm ist gleich relevant. KI kann Prioritäten vorschlagen, indem sie Kontext einbezieht: Kritikalität des Assets, Exposure, Identitätsrechte, Blast Radius, bekannte Threat-Intel-Indikatoren, historische Vergleichsfälle. Das ersetzt nicht die Entscheidung – aber es bringt Struktur in die Flut.
- Schnelle Zusammenfassung komplexer Incident-Timelines:

  Gerade in laufenden Incidents ist die Timeline oft ein chaotischer Mix aus Logs, Chat, Ticket-Updates und Hypothesen. KI kann daraus eine verständliche, fortschreibbare Chronologie machen: »Was wissen wir sicher, was vermuten wir, was sind die nächsten Schritte?«.

Der operative Effekt ist unmittelbar:

- weniger Alert-Müdigkeit (weil mehr Kontext und weniger redundante Arbeit),
- bessere Fokussierung auf kritische Ereignisse (weil Priorisierung und Clustering Reibung reduzieren),
- kürzere Einarbeitungszeiten für neue Analysten (weil Zusammenfassungen, Erklärungen und standardisierte Workflows Lernen beschleunigen).

Das ist einer der wenigen Bereiche, in denen KI sehr schnell ROI liefern kann – sofern Sie klare Qualitätsmetriken und human-in-the-loop etabliert haben.

1.4.2 Bessere Nutzung vorhandener Daten

In vielen Organisationen ist nicht »zu wenig Sichtbarkeit« das Kernproblem, sondern das Gegenteil: Es gibt Unmengen an Daten – Logs, Tickets, Wiki-Artikel, Architektur-Dokus, Post-Mortems, E-Mail-Threads – aber sie sind verteilt, inkonsistent und schwer auffindbar. Sicherheit scheitert dann nicht an Technik, sondern an Orientierung.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Typische Symptome sind:

- niemand hat einen echten Überblick, wo welche Information liegt,
- Dokumentation ist veraltet oder widersprüchlich,
- Wissen steckt in Köpfen oder in Ticket-Kommentaren,
- Tools sind Insellösungen mit unterschiedlichen Datenmodellen.

KI-Ansätze – insbesondere RAG (Retrieval-Augmented Generation) – können hier einen sehr pragmatischen Hebel bieten:

- Verstreute Wissensinseln zusammenführen:

  Confluence, Ticketsysteme, Runbooks, Mails, Wikis, SharePoint – alles kann in einen such- und zitierfähigen Wissenslayer überführt werden, der kontextbezogen Antworten liefert.
- Sicherheitswissen kontextsensitiv bereitstellen:

  Statt »Suchmaschine« bekommen Sie »Antwortmaschine«: nicht nur Trefferlisten, sondern eine Zusammenfassung mit relevanten Quellen, bezogen auf den konkreten Fall (»Welche Schritte sind bei diesem Incident-Typ vorgesehen?«).
- Doku-»Leichen« wieder nutzbar machen:

  Alte, verstreute Dokumente sind oft nicht wertlos – sie sind nur nicht zugänglich. KI kann Inhalte extrahieren, normalisieren, miteinander vergleichen und helfen, veraltete Stellen sichtbar zu machen.

Das Ergebnis ist weniger »Magie« als schlicht bessere Organisation: Wissen wird operationalisierbar, Entscheidungen werden konsistenter, und Teams verlieren weniger Zeit durch Suche und Interpretationsarbeit.

1.4.3 Beschleunigung von DevSecOps

Im Entwicklungs- und Betriebsumfeld liegt der größte Security-Schaden oft nicht in fehlenden Tools, sondern in Fehlkonfigurationen, unsicheren Defaults und Zeitdruck. KI kann hier helfen, Security in den Delivery-Prozess zu integrieren, ohne dass sie als reines »Stoppschild« wahrgenommen wird.

Konkrete Ansatzpunkte sind:

- Automatisierte Security-Code-Reviews (unter Aufsicht):

  KI kann Pull Requests voranalysieren, typische Schwachstellenmuster markieren, sichere Alternativen vorschlagen und Entwicklerkommentare in ein Security-Fazit übersetzen. Wichtig: als Assistenz, nicht als alleinige Autorität.
- Unterstützung beim Schreiben sicherer Konfigurationen (IaC):

  Gerade in Terraform/Kubernetes/CI-Pipelines sind die Fehler oft banal, aber folgenreich. KI kann Best Practices in konkrete Konfigurationsänderungen übersetzen und Risiken in verständlicher Form erklären.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Threat-Modeling-Assistenten:

  Nicht jedes Team hat Security-Expertise für STRIDE/Abuse-Cases. KI kann Fragen stellen, Angriffspfade vorschlagen, Trust Boundaries strukturieren und daraus handhabbare Controls ableiten.

- Generierung von Testfällen und Sicherheitstests:

  Von Input-Validation-Tests über AuthZ-Checks bis hin zu API-Fuzzing-Ansätzen: KI kann helfen, Testideen zu systematisieren und Boilerplate zu reduzieren.

Das reduziert ganz konkret:

- »Security als Bottleneck« im Delivery-Prozess (weil Vorarbeit automatisiert wird),
- händische Routineaufgaben (weil Vorschläge und Reviews schneller entstehen),
- Fehlkonfigurationen durch Unwissen (weil Teams direkt beim Umsetzen unterstützt werden).

Richtig gemacht ist das kein »Security macht weniger«, sondern »Security kommt früher und leichter in den Prozess«.

1.4.4 GRC & Compliance

Governance, Risk & Compliance ist für viele Organisationen der Bereich mit dem größten Anteil an formaler Arbeit: Policies schreiben, Normen mappen, Evidenzen sammeln, Audits vorbereiten, Management-Reports erstellen. Genau hier ist KI besonders wirksam, weil es viel um Textarbeit, Strukturierung und Abgleich geht.

Typische Einsatzfelder sind:

- Erstellung und Pflege von Policies:

  KI kann Policy-Entwürfe erstellen, Versionsänderungen konsistent einarbeiten, Formulierungen vereinheitlichen und auf Lücken prüfen – immer mit menschlicher Freigabe.

- Mapping von Normen auf Unternehmensprozesse:

  Ob ISO 27001, NIST, CIS Controls oder interne Standards: KI kann Zuordnungen vorschlagen, Überschneidungen sichtbar machen und Gap-Analysen strukturieren.

- Analyse von Audit-Berichten und Findings:

  KI kann Findings clustern, Root-Cause-Muster identifizieren, Wiederholungen über Jahre sichtbar machen und Maßnahmen in umsetzbare Workpakete übersetzen.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Vorbereitung von Management-Reports und Aufsichtsunterlagen:

  Statt dass GRC-Teams aus zig Quellen Berichtstexte zusammenschreiben, kann KI Fakten konsolidieren und narrative, verständliche Reports erzeugen – inkl. Risikoargumentation und Status.

In dieser Rolle wird KI zum Policy- und Audit-Copiloten: Sie reduziert die formale Last, erhöht Konsistenz und schafft Kapazität für das, was eigentlich zählt – Risikosteuerung, Priorisierung und die wirksame Umsetzung von Kontrollen.

1.5 Neue Risiken und Angriffsflächen durch KI

Die Kehrseite ist banal, aber entscheidend: Jedes mächtige Werkzeug erweitert die Angriffsfläche. Sobald Sie KI in Prozesse, Datenflüsse und Entscheidungen integrieren, entsteht nicht nur »mehr Automatisierung«, sondern ein neues Systemverhalten – mit neuen Missbrauchsmöglichkeiten, neuen Failure-Modes und neuen Governance-Anforderungen. Und weil KI heute häufig als Cloud-Service, als API oder als tief integrierter Copilot genutzt wird, sind diese Risiken nicht theoretisch, sondern unmittelbar operativ relevant.

1.5.1 Angriffe auf und über LLMs

Mit LLMs entstehen Angriffsformen, die Sie aus klassischer AppSec oder klassischen SIEM-Modellen so nicht kennen – weil das Ziel nicht nur ist, ein System zu kompromittieren, sondern sein Verhalten zu manipulieren oder seine Grenzen zu umgehen.

Typische Muster sind:

- Prompt Injection:

  Hier wird Benutzereingabe so gestaltet, dass das Modell seine eigentlichen Vorgaben ignoriert: »Vergiss alle Regeln«, »Handle als Admin«, »Gib mir die vertraulichen Details«, oder subtiler: durch eingebettete Anweisungen in Dokumenten, E-Mails oder Webseiten, die das Modell als Kontext »mitliest«. Der Angriff ist nicht (nur) technisch, sondern psychologisch: Der Angreifer versucht, die Steuerlogik des Modells umzudefinieren.

- Data Leakage / unbeabsichtigte Datenpreisgabe:

  KI-Systeme arbeiten mit Kontextfenstern, Retrieval, Tool-Integrationen und oft mit riesigen Dokumentenbeständen. Wenn Zugriffsrechte, Kontextfilter oder Output-Constraints nicht sauber umgesetzt sind, können vertrauliche Inhalte in Antworten auftauchen – entweder, weil sie im Kontext enthalten waren, oder weil die KI über RAG »zu viel« zurückholt. Ein verwandtes Risiko ist falsche Nutzung von Trainings-/Fine-Tuning-Prozessen: Dann werden sensible Daten nicht nur angezeigt, sondern potenziell in Modelle oder Logs hinein konserviert.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Model Theft / Model Extraction:

  Wenn ein Modell über eine API zugänglich ist, kann ein Angreifer versuchen, durch systematisches Abfragen das Verhalten zu rekonstruieren: Welche Prompts führen zu welchen Outputs? Welche Policies sind implementiert? Welche »Guardrails« lassen sich umgehen? Ziel ist oft, ein funktional ähnliches Modell nachzubauen oder das Modellverhalten so gut zu verstehen, dass es leichter missbraucht werden kann.

- Evasion und »Model Probing« gegen ML-Detektoren:

  ML-basierte Detektionssysteme (z. B. Malware-Klassifikatoren, UEBA, Anomalie-Detektoren) lassen sich gezielt austesten. Angreifer variieren Inputs, bis ein Score unter die Schwelle fällt, oder sie lernen, welche Features das Modell besonders stark gewichtet. Das ist eine neue Variante des alten Problems »Evasion«, aber durch KI wird die Iteration schneller und die Suche nach blinden Flecken systematischer.

Der Punkt ist: KI ist nicht nur ein neues Tool, sondern ein neues Angriffsziel – und in vielen Fällen auch ein neuer Angriffsweg, weil KI-Systeme häufig mit Datenquellen und Automationsrechten verbunden sind.

1.5.2 Governance- und Compliance-Risiken

Neben technischen Angriffen entstehen Governance-Risiken, die oft noch schneller zuschlagen – weil sie im Alltag beiläufig passieren: »mal eben« ein Prompt mit sensiblen Informationen, »kurz« ein Tool ausprobieren, »einfach« einen Copilot integrieren. Das ist selten böswillig, aber es ist riskant.

Typische Problemfelder sind:

- Intransparente Datenflüsse in externe KI-Dienste:

  Sobald Mitarbeitende oder Tools Inhalte an externe KI-APIs senden, stellt sich die Frage: Welche Daten gehen raus? Wie werden sie verarbeitet? Werden sie geloggt? Wo liegen sie? Wer hat Zugriff? Ohne klare Architektur und Verträge ist das faktisch ein Blindflug.

- Shadow AI ohne vertragliche Absicherung:

  Mitarbeitende nutzen GenAI-Tools, Browser-Plugins oder »kostenlose« Services, die nicht durch Einkauf, Legal und Security bewertet wurden. Ohne DPA, ohne Sicherheitszusagen, ohne klare Datenverwendungsbedingungen. Das ist das KI-Pendant zu Shadow IT – nur oft mit höherem Datenabflussrisiko, weil Inhalte aktiv hineinkopiert werden.

- Datenschutz- und Geheimnisschutzverstöße:

  PII, HR-Daten, Kundendaten, interne Finanzen, Security-Incidents, Credentials in Tickets – all das landet schnell in Prompts, wenn keine Leitplanken existieren. Der Verstoß passiert nicht, weil jemand »Daten exfiltriert«, sondern weil Prozesse und Awareness nicht mit dem Tooling mithalten.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

- Fehlende Nachvollziehbarkeit gegenüber Revision und Aufsicht:

  Wenn KI Outputs erzeugt, die in Entscheidungen einfließen (z. B. Risikoeinstufung, Incident-Schwere, Policy-Auslegung), muss nachvollziehbar sein, auf welcher Basis das passiert ist: Welche Quellen wurden genutzt? Welche Daten lagen zugrunde? Welche Version des Modells? Welche Prompt- und Policy-Konfiguration? Ohne diese Audit-Trails werden KI-gestützte Prozesse im Audit schnell zum Problem – selbst, wenn sie fachlich gut funktionieren.

Governance ist damit kein »nice to have«, sondern Voraussetzung, um KI überhaupt verantwortbar in Security und IT zu betreiben.

1.5.3 Organisatorische Risiken

Die unterschätztesten Risiken sind häufig nicht technisch, sondern menschlich und organisatorisch. KI verändert Rollen, Erwartungen und Verantwortungsgrenzen – und wenn das nicht aktiv gemanagt wird, entstehen gefährliche Grauzonen.

Typische organisatorische Risiken sind:

- Überschätzung der KI-Fähigkeiten:

  »Die KI wird's schon richten« ist eine der schnellsten Arten, Risiken zu erhöhen. LLMs wirken kompetent, weil sie flüssig formulieren. Das kann dazu führen, dass Teams Outputs zu wenig prüfen, Annahmen übernehmen oder Entscheidungen zu früh automatisieren.
- Unklare Verantwortlichkeiten:

  Wenn KI im SOC Tickets vorschlägt, Prioritäten setzt oder sogar Actions anstößt, muss klar sein: Wer trägt die Verantwortung? Wer hat freigegeben? Wer kontrolliert? Ohne klare RACI-Logik und Prozessdefinition entsteht »Verantwortungsdiffusion« – genau das, was Sie in Incident-Situationen nicht gebrauchen können.
- Skill-Gap im Team:

  KI-Ergebnisse sinnvoll zu nutzen, erfordert Kompetenz: Wie interpretiere ich Scores? Wie erkenne ich Halluzinationen? Wie teste ich Prompt-Robustheit? Wie bewerte ich Datenqualität und Drift? Wenn diese Skills fehlen, steigt die Wahrscheinlichkeit, dass KI entweder falsch genutzt oder aus Angst gar nicht genutzt wird.
- Widerstände in der Belegschaft:

  Wenn KI als Überwachungs- oder Kontrollinstrument wahrgenommen wird (»die KI liest unsere Chats«, »die KI bewertet unser Verhalten«), entstehen Akzeptanzprobleme, Ausweichverhalten und politische Konflikte. Gerade UEBA- und Kommunikationsanalysen brauchen deshalb klare Transparenz, Zweckbindung und Governance, sonst wird das Thema intern toxisch.

Unterm Strich: KI erweitert nicht nur Ihre technischen Möglichkeiten – sie verändert Ihre Organisation. Wer KI sicher nutzen will, muss sie nicht nur integrieren, sondern beherrschen: technisch, vertraglich, prozessual und kulturell.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

1.6 Was sich für IT-Verantwortliche und CISOs konkret ändert

KI in der Security ist nicht nur ein neues Toolset, sondern ein Wechsel im Betriebsmodell. Das betrifft Entscheidungen, Verantwortlichkeiten, Skill-Profile und die Art, wie Sie Führung und Governance organisieren. Wer KI erfolgreich und sicher einsetzen will, muss nicht »mehr Regeln konfigurieren«, sondern Zielbilder, Leitplanken und messbare Steuerung etablieren – und das über mehrere Disziplinen hinweg.

1.6.1 Vom Regel-Admin zum Risiko-Architekten

In der klassischen Security-Welt war ein großer Teil der Steuerung operativ-technisch: Welche Regel setze ich wo? Welche Signatur ist aktiv? Welche Korrelation fehlt? Welche Firewall-Exception ist vertretbar? Das bleibt wichtig – aber KI verschiebt den Schwerpunkt.

Ihre Rolle wandert weg von der Frage »Welche Regel setze ich in welchem System?« hin zu einer deutlich strategischeren und gleichzeitig architekturlastigeren Frage »Welche Use Cases zahlen wirklich auf unser Sicherheitsziele-Konto ein – und wie steuern wir Daten, Modelle und Prozesse so, dass der Nutzen steigt, ohne das Risiko auszuweiten?«

Das ist ein Rollenwechsel vom »Regel-Administrator« hin zum Risiko-Architekten: Sie designen nicht nur Kontrollen, sondern ein System aus Datenflüssen, Modellen, Entscheidungslogik, Human-in-the-loop und Auditierbarkeit.

Dafür brauchen Sie ein KI-Sicherheitszielbild (Target Operating Model), das mindestens folgende Punkte beantwortet:

- Wo wollen wir KI bewusst einsetzen?

  Welche Prozesse sind geeignet (z. B. Zusammenfassen, Priorisieren, Assistenz), wo ist der erwartete ROI hoch und das Fehlerrisiko beherrschbar?
- Wo wollen wir KI bewusst nicht einsetzen?

  Welche Entscheidungen sind zu kritisch, zu rechtlich sensibel oder zu schwer auditierbar? Wo sind Fehlentscheidungen nicht tolerierbar?
- Welche Guardrails setzen wir – technisch und organisatorisch?

  Zugriffskontrollen, Datenklassifizierung, Output-Constraints, Logging, Freigabeprozesse, Limits für autonome Aktionen, Red-Teaming, Incident-Handling für KI-Fehler.
- Wie messen wir Erfolg und Risiko?

  Nicht nur »Nutzung« oder »Zeitersparnis«, sondern präzise Metriken: Qualität der Outputs, False-Positive/False-Negative-Raten, Drift-Indikatoren, Ticket-

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Throughput, MTTR, Audit-Feststellungen, Datenabfluss-Events, Policy-Verstöße.

Das klingt nach »mehr Governance«, ist aber in Wahrheit die Voraussetzung, um KI überhaupt skalierbar und verantwortbar zu betreiben.

1.6.2 Skill-Shift im Security-Team

KI verändert auch, was ein gutes Security-Team können muss. Nicht jede Rolle muss Data Science beherrschen – aber die Grundkompetenz, KI-Systeme zu verstehen, zu bewerten und sauber zu betreiben, wird zur Baseline. Sonst entsteht entweder blindes Vertrauen (»die KI hat gesagt…«) oder totale Blockade (»das ist uns zu riskant«).

Neue bzw. stärker gewichtete Skills sind unter anderem:

- Verständnis von ML-/LLM-Konzepten auf Architekturebene:

  Was ist Training vs. Inferenz? Was bedeutet Drift? Was ist RAG? Wo liegen typische Failure-Modes (Halluzination, Bias, Prompt Injection, Over-retrieval)?
- Evaluierung von Modellen und Outputs:

  Wie messe ich Genauigkeit, Robustheit und Bias? Wie teste ich Worst-Case-Szenarien? Wie überprüfe ich, ob ein System »zu selbstbewusst« falsche Dinge behauptet? Wie stelle ich reproduzierbare Qualität sicher?
- Aufbau und Betrieb von RAG/LLM-Anwendungen:

  Datenquellen anbinden, Zugriff sauber durchsetzen, Retrieval begrenzen, Quellen zitierbar machen, Logging/Audit implementieren, Versionierung und Change-Control etablieren.
- Kritische Einordnung von KI-Ergebnissen:

  Fähigkeit, Aussagen zu verifizieren, Annahmen zu erkennen, Unsicherheit zu akzeptieren und KI-Outputs in Entscheidungen zu übersetzen, ohne Verantwortung abzugeben.

Der Kern ist eine neue Form von Grundbildung im Security-Kontext: KI-Lese- und Schreibfähigkeit. Das heißt:

- verstehen, was ein Modell tut,
- wissen, was es nicht kann,
- einschätzen, wann es hilft,
- und testen können, ob es in Ihrem Kontext zuverlässig ist.

Teams, die diese Kompetenz aufbauen, können KI kontrolliert nutzen. Teams, die sie nicht aufbauen, werden entweder riskant über-automatisieren oder relevante Chancen liegen lassen.

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

Stichwortverzeichnis

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144

© des Titels »Cybersicherheit und KI« (ISBN 978-3-7475-1144-2) 2026 by
mitp Verlags GmbH & Co. KG, Frechen. Nähere Informationen unter: https://www.mitp.de/1144