novaCapta

Whitepaper

Agentic AI: vom Blindflug zur sicheren Skalierung

Vertrauen entsteht, wenn Governance, Identitäten, Daten und Security Operations gemeinsam gedacht werden und KI-Agenten sicher skaliert werden.



DE
novaCapta GmbH
Im Mediapark 5c
50670 Köln

T +49 (0)221 58919 343
M info@novacapta.com
W www.novacapta.com

CH
novaCapta Schweiz AG
Theaterstrasse 17
8400 Winterthur

T +41 (0)41 392 20 00
M info.schweiz@novacapta.com
W www.novacapta.ch

Agentic AI ist angekommen

Die neue Realität: KI handelt zunehmend selbstständig

Agenten analysieren, entscheiden und führen Aktionen über Systemgrenzen hinweg aus. KI handelt damit zunehmend selbstständig. Drei Kennzahlen verdeutlichen diese Entwicklung:

> 80 % der befragten Führungskräfte erwarten Agenten in ihren Organisationen innerhalb von 12 bis 18 Monaten.

1,3 Mrd. KI-Agenten prognostiziert IDC bis 2028.

29 % der Mitarbeitenden nutzen bereits nicht autorisierte KI-Agenten für Arbeitsaufgaben.

MANAGEMENT-IMPLIKATION

Die Einführung lässt sich nicht allein über Verbote steuern. Unternehmen benötigen Sichtbarkeit, Leitplanken und einen kontrollierten Weg vom Test bzw. PoC in den Betrieb.

Warum das bestehende Sicherheitsmodell nicht ausreicht

Autonome Geschwindigkeit und breite Zugriffsrechte verstärken bekannte Schwachstellen. Vier Risiken lassen sich unterscheiden:

1. Unbekannte Agenten
Es fehlt Transparenz darüber, wo Agenten existieren und was sie tun.

2. Überprivilegierte Zugriffe
Agenten können Berechtigungen ansammeln und sensible Daten erreichen.

3. Neue Angriffswege
Dazu gehören Prompt-Manipulation, Model-Tampering und agentenbasierte Angriffsketten.

4. Kaskadierende Risiken
Ein Vorgang kann E-Mail, Dokumente, SaaS, Endpunkte, Cloud und Legacy-Systeme verbinden.



Quelle: Forrester Total Economic Impact Studie – AI 10/2026

Das Zielbild: Sicherheit auf drei Ebenen

Eine wirksame Strategie verbindet Agenten-Governance, sichere Grundlagen und KI-gestützte Abwehr.

1

KI-Agenten absichern: vom Blindflug zur Steuerbarkeit

Ein gemeinsamer Kontrollrahmen schafft Transparenz, Verantwortlichkeit und Begrenzung. Er besteht aus vier Bausteinen, die jeweils eine Leitfrage beantworten:

1. Beobachten
Welche Agenten existieren, wo laufen sie und was tun sie?

2. Governance
Welche Regeln, Freigaben, Aufbewahrung und Audits gelten?

3. Identität & Zugriff
Welche Ressourcen darf ein Agent im jeweiligen Kontext nutzen?

4. Daten & Bedrohungen
Welche Daten sind erlaubt und wie werden KI-spezifische Angriffe erkannt?

2

Grundlagen absichern: Identitäten und Daten zuerst

Agenten erben die Stärken und Schwächen der digitalen Umgebung, in der sie arbeiten. Vier Microsoft-Lösungen bilden die Grundlage:

1. Microsoft Entra & Agent 365
Agentenidentitäten, Zugriffssteuerung und Begrenzung übermäßiger Berechtigungen.

2. Microsoft Purview
Datenschutz, Compliance und verantwortliche Verarbeitung sensibler Informationen.

3. Defender & Intune
Schutz von Anwendungen, Endpunkten und Cloud-Umgebungen.

4. Microsoft Sentinel
Zentrale Sichtbarkeit und Security Operations über die Umgebung hinweg.

LEITSATZ

Kein Agent darf mehr Zugriffe erhalten, als es die zugrunde liegenden Identitäten, Daten und Systeme rechtfertigen.

3

Security Operations mit Agenten skalieren

Die Verteidigung muss mit der Geschwindigkeit autonomer Angriffe Schritt halten. Dabei lassen sich drei Aufgaben unterscheiden:

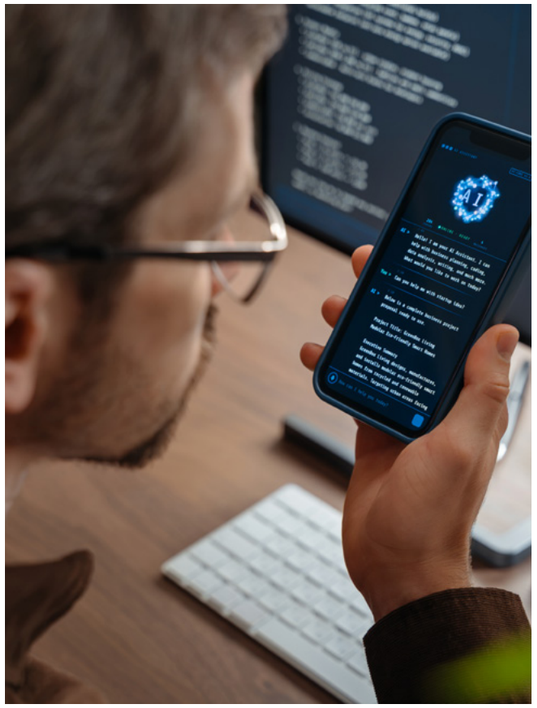
1. Erkennen
Signale zusammenführen, Bedrohungen priorisieren und KI-spezifische Angriffsmuster identifizieren.

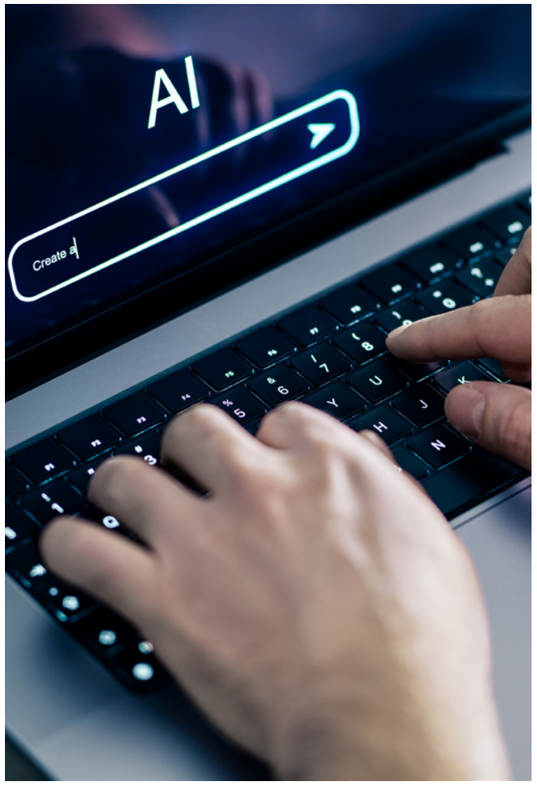
2. Analysieren
Triage und Kontextanreicherung mit Security Copilot und spezialisierten Security Agents beschleunigen.

3. Reagieren
Definierte Maßnahmen automatisieren, während Menschen Governance und Eskalation verantworten.

ZIEL

Menschen entscheiden über Leitplanken und Ausnahmen. Agenten übernehmen Geschwindigkeit, Volumen und wiederkehrende Analysen.







Der Business Case: Konsolidierung statt zusätzlicher Silos

Integrierte Sicherheit soll Risiko reduzieren und gleichzeitig den Betriebsaufwand senken. Vier Kennzahlen zeigen den erwarteten Nutzen.

124%
ROI über drei Jahre

30%
geringere Wahrscheinlichkeit eines Sicherheitsvorfalls

25%
geringere Kosten für die Behebung von Sicherheitsverletzungen

>20%
geringere Ausgaben für Sicherheitstechnologie

EINORDNUNG

Die Werte beziehen sich laut Microsoft auf eine zusammengesetzte, fiktive Organisation in einer Forrester-TEI-Studie und sind nicht automatisch auf jedes Unternehmen übertragbar.

Studie: Forrester Total Economic Impact Studie – AI 2026

Empfohlene Management-Roadmap

Jetzt starten, aber nur innerhalb klarer Verantwortlichkeiten und messbarer Leitplanken und so Wettbewerbsvorteile sichern. Die Roadmap umfasst drei Schritte:

1 Transparenz schaffen

Agenten inventarisieren, Verantwortliche benennen, Daten- und Systemzugriffe sichtbar machen.

2 Leitplanken definieren

Zulässige Use Cases, Identitäten, Berechtigungen, Datenklassen und Freigabeprozesse festlegen.

3 Sicher skalieren

Security Operations integrieren, Metriken etablieren und erfolgreiche Use Cases kontrolliert ausrollen.



Umsetzung: Beispiele für eine Umsetzungsmatrix

Die folgende Matrix zeigt Beispiele für Handlungsfelder mit Maßnahme, Ergebnis, Verantwortlichen, Priorität und messbarem Erfolg.

Handlungsfeld	Maßnahme	Ergebnis	Verantwortlich	Prio	Messbarer Erfolg
AI Governance	Agentenrichtlinie und Risikoklassen definieren	Verbindliche Policy, Freigabegates & Eskalationsweg	CIO/CISO/ Datenschutz	A	Policy verabschiedet; alle Piloten klassifiziert
Transparenz	Agenten und Agentenplattformen inventarisieren	Zentrales Register mit Owner, Zweck, Datenzugriff und Status	IT Governance	A	Anteil registrierter Agenten; ungeklärte Owner; Agent365
Anwendungsfall-Steuerung	Pilotportfolio priorisieren und bewerten	Bewertete Liste mit Nutzen, Risiko und Freigabestatus	Fachbereiche / KI-Team	A	Jeder Pilot besitzt Sponsor, Owner und Risikobewertung
Identität	Eindeutige Agentenidentitäten etablieren	Nachvollziehbare Identität je produktivem Agenten	IAM / Entra Team	A	Keine produktiven Shared Identities; vollständige Zuordnung
Berechtigungen	Least Privilege und regelmäßige Reviews umsetzen	Rollenmodell, Access Reviews und Rezertifizierung	IAM / Applikationsowner	A	Überprivilegierte Zugriffe reduziert; Reviews fristgerecht
Daten	Datenklassifizierung und zulässige Zugriffe definieren	Datenklassen, DLP-Regeln, dokumentierte Ausnahmen	Data Governance / Datenschutz	A	Kritische Datenquellen klassifiziert; Verstöße messbar
Bedrohungsschutz	KI-spezifische Angriffsszenarien ergänzen	Threat Model für Prompt Injection, Manipulation, Angriffsketten	Security Architecture	A	Threat Model je kritischem Use Case vorhanden
Security Operations	Monitoring, Triage und Reaktion integrieren	Anwendungsfälle, Alarmierung, Playbooks, Eskalationsregeln	SOC / CISO	A	Relevante Agentenereignisse im SOC sichtbar; Reaktionswege getestet
Lifecycle	Entwicklung, Test, Freigabe und Stilllegung regeln	Lifecycle-Prozess mit dokumentierten Kontrollpunkten	KI-Team / ITSM	B	Produktive Agenten durchlaufen definierte Gates
Compliance	Protokollierung, Aufbewahrung und Auditfähigkeit festlegen	Logging- & Retention-Konzept; prüfbare Nachweise	Compliance / Legal	B	Nachweise für kritische Agenten vollständig
Betriebsmodell	RACI und Service-Ownership festlegen	Klare Rollen für Business Owner, Technical Owner, Security und Betrieb	CIO / Fachbereiche	A	Keine produktiven Agenten ohne benannte Owner
Qualifizierung	Rollenbasierte Schulungen durchführen	Trainings für Fachbereiche, Entwickler, Betrieb und SOC	HR / KI-Team / Security	B	Relevante Rollen geschult; Wissensnachweise dokumentiert
Messung	KPI- und Risikodashboard etablieren	Management-Dashboard zu Bestand, Risiken, Vorfällen und Nutzen	IT Governance / Controlling	B	Regelmäßiges Reporting und definierte Schwellenwerte
Skalierung	Freigabe erfolgreicher Piloten nach Freigabeprüfung	Skalierungsentscheidung mit Restrisiko und Betriebsfreigabe	Steuerungsgremium	B	Alle Skalierungen formal entschieden und dokumentiert

Microsoft-Produkte und ein passendes Lizenzmodell

Eine Plattform für Governance, Identitäten, Daten, Endpunkte und Security Operations.

1 Agenten steuern:

Agenten erfassen, Owner zuordnen, Zugriffe steuern und den Lebenszyklus kontrollieren.

2 Grundlagen schützen

Identitäten absichern, Daten klassifizieren, Endpunkte schützen und Angriffe erkennen.

3 Security Operations skalieren

Signale bündeln, Vorfälle priorisieren und Reaktionen schneller ausführen.

Empfehlung: Microsoft 365 E7 als strategisches Zielmodell

Microsoft 365 E7 bietet eine breite Security- und Compliance-Basis, Agent 365 als zentrale Kontrollebene und Security Copilot für Security-Teams.

Wichtig – Kostenkontrolle ist notwendig bei autonomen Agenten oder Copilot Cowork etc.

Zusätzliche, nutzungsabhängige Bausteine

Microsoft Sentinel
Abrechnung nach aufgenommenen Daten. E5/E7 kann ein Datenkontingent für Microsoft-365-Daten erhalten.muster identifizieren.

Copilot Studio
Für eigene Agenten und Automatisierungen. Die Kosten sind abhängig von Nutzung und gewähltem Modell.

Optionale Zusatzkapazität
Security Copilot SCUs nur, wenn die enthaltene Leistung für den Betrieb nicht ausreicht.

MANAGEMENT-ENTSCHEIDUNG

E7 für Security-, Compliance- und Schlüsselrollen priorisieren; Sentinel und Agentenverbrauch getrennt steuern.

Sovereign Landing Zone für besonders schützenswerte Workloads

Grundlage ist die Azure Sovereign Landing Zone. Sie umfasst Management Groups, Azure Policy, Netzwerk, Logging sowie Rollen und Berechtigungen. Darauf setzen vier Komponenten auf:

1. Sovereign Control Panel
Vorgaben bereitstellen, überwachen und als wiederholbaren Standard betreiben.

2. Data Guardian – Betriebszugriffe überwachen
Remotezugriffe durch Microsoft werden regional überwacht und nachvollziehbar protokolliert (EU Data Boundary, Azure Touch Ledger).

3. Customer Lockbox – Zugriff freigeben
Zugriff auf Kundendaten wird im Supportfall vom Kunden geprüft und genehmigt oder abgelehnt (für M365 und Azure, Aktivierung notwendig).

4. Azure Key Vault – Schlüssel kontrollieren
Kundeneigene Schlüssel verwalten. Für höhere Anforderungen: Premium oder Managed HSM, eigener Vertrauensanker im eigenen RZ mit uTrust Anchor und ESKM.

Ergänzung für besonders kritische Daten

Azure Confidential Computing & novaEnqlaive
Schützt ausgewählte Daten auch während der Verarbeitung vor Fremdzugriff (Betreiberausschluss ist gewährleistet). Es besteht eine Abhängigkeit zur Azure Landing Zone und zu den verfügbaren Komponenten.

Lizenz- und Kostenmodell

Azure-Verbrauch nach Nutzung. Customer Lockbox benötigt mindestens einen Developer Support Plan oder M365 E5+. Key Vault und Managed HSM (BYOK, HYOK mit Utimaco) erzeugen zusätzliche Kosten.

MANAGEMENT-ENTSCHEIDUNG

Sovereign Landing Zone für sensible Workloads nutzen und Schlüssel-, Zugriffs- und Betriebsmodell gemeinsam festlegen.



Fazit: Zentrale Entscheidung

Agentic AI nicht bremsen. Vertrauen aufbauen.

Die Management-Prioritäten:

1. Ein unternehmensweites Governance-Modell für Agenten etablieren.
2. Identitäten, Berechtigungen und Datenzugriffe (evtl. souverän) technisch durchsetzen.
3. Security Operations auf autonome Geschwindigkeit ausrichten. Project Perception & MDASH.
4. Skalierung an Transparenz, Messbarkeit und klare Verantwortlichkeit koppeln.

ERSTER SCHRITT

Executive Sponsor, Security, IT, Datenschutz, Betriebsrat und Fachbereiche auf ein gemeinsames Zielbild und Pilotportfolio ausrichten.

„Die entscheidende Frage bei Agentic AI ist für mich nicht, ob wir KI-Agenten einsetzen, sondern wie wir ihre zunehmende Autonomie beherrschbar machen. Unternehmen brauchen Transparenz darüber, welche Agenten aktiv sind, worauf sie zugreifen und wer die Verantwortung trägt. Erst wenn Governance, Identitäten, Daten und Security zusammenspielen, entsteht das nötige Vertrauen. Genau dieses Vertrauen ist die Voraussetzung, um Agentic AI sicher zu skalieren."

Jürgen Dick
Director Business Development
novaCapta

Über den Autor



Jürgen Dick ist Cloud-Architekt und IT-Stratege mit umfassender Erfahrung in der Entwicklung und Umsetzung von Cloud- und Infrastrukturprojekten im Microsoft-Umfeld. Seine beruflichen Stationen, unter anderem bei Microsoft, und seine Expertise in Microsoft Azure, Security und Governance prägen seinen Blick auf die Frage, wie neue Technologien sicher und kontrolliert in bestehende IT-Landschaften integriert werden können.

Als **Director Business Development** bei novaCapta beschäftigt er sich im Kontext von Agentic AI insbesondere mit den Herausforderungen, die durch zunehmend autonom handelnde KI-Agenten für Transparenz, Zugriffe, Governance und Security entstehen.

novaCapta

Damit KI-Innovationen sicher im Unternehmen eingesetzt und skaliert werden können, müssen AI, Cloud, Daten, Identitäten und Security zusammengedacht werden. Genau hier verbindet novaCapta umfassende Microsoft-Technologiekompetenz mit der Erfahrung aus der Beratung und Umsetzung anspruchsvoller IT-Projekte und begleitet Unternehmen von der Strategie über die technische Umsetzung und das Change Management bis zum laufenden Betrieb.

Als Microsoft Solutions Partner „Microsoft Cloud" ist novaCapta in allen drei Microsoft Solution Areas **Cloud & AI Platforms, AI Business Solutions und Security** ausgezeichnet. Ergänzt wird diese Kompetenz durch Microsoft Specializations unter anderem für Agentic Business Solutions, AI Apps on Azure, Microsoft 365 Copilot, Cloud Security, Identity & Access Management und Threat Protection.

Ihr Microsoft Premium Partner

Kontaktieren Sie uns!

Bei Fragen zu unseren Themen sind wir gerne für Sie da und finden gemeinsam mit Ihnen die beste Auswahl aus den Microsoft Bausteinen.



DE
novaCapta GmbH
Im Mediapark 5c
50670 Köln

T +49 (0)221 58919 343
M info@novacapta.com
W www.novacapta.com

CH
novaCapta Schweiz AG
Theaterstrasse 17
8400 Winterthur

T +41 (0)41 392 20 00
M info.schweiz@novacapta.com
W www.novacapta.ch