Whitepaper

Schutz von KI-Agenten von der Entwicklung bis zum unternehmensweiten Einsatz

Inhalt

Executive Summary

KI-Agenten verändern nicht nur die Arbeitswelt – sie definieren das Verständnis von Identitäten selbst neu.

Als autonome Akteure sind sie unabhängig, zielorientiert und handeln immer häufiger ohne menschliche Aufsicht. Sie haben einen unersättlichen Datenhunger, analysieren permanent Informationen, schreiben Code, senden E-Mails und treffen systemübergreifend Entscheidungen. Ebenso wie ruhelose leistungsorientierte Menschen, die ihre Ziele schnell erreichen möchten, dehnen Agenten die Grenzen aus, um an neue Daten zu gelangen. Ohne zuverlässige Richtlinien können sie unbeabsichtigt außer Kontrolle geraten und Schaden und Chaos verursachen. Doch nur wenige Unternehmen haben Antworten auf einfache Fragen wie: Wo befinden sich die Agenten in meinem Ökosystem? Auf welche Daten und Systeme können sie zugreifen? Wer ist verantwortlich, wenn sie außer Kontrolle geraten? Laut dem Okta-Bericht AI at Work 2025 **nutzen 91 % der Unternehmen KI-Agenten**, wobei **44 % keine Governance-Maßnahmen implementiert haben**. Das führt zu einer neuen Sicherheitsherausforderung: eine explosionsartig wachsende Zahl autonomer nicht-menschlicher Identitäten, die ein einheitliches Framework für Authentifizierung, Autorisierung oder Transparenz benötigen.

Das Aufkommen von KI-Agenten stellt das klassische Identity and Access Management (IAM) infrage, dessen Kontrollen für Menschen entwickelt wurden. Sie können daher nicht mit Agenten Schritt halten, die ohne menschliche Aufsicht im großen Umfang komplexe Workflows und API-Ketten starten können. Damit Sie das **Vertrauen Ihrer Kunden aufrecht erhalten** können, benötigen Sie eine neue Generation von Identity-Sicherheitslösungen, die **mit der Geschwindigkeit, dem Umfang und der Intelligenz von KI-Agenten Schritt hält**.

In diesem Whitepaper zeigen wir, wie Sie **alle Agenten absichern** und Sicherheitsmaßnahmen von der ersten Codezeile bis zu einer unternehmensweiten Kontrollebene einbetten – denn im Zeitalter autonomer KI **geht es beim Identity-Management nicht nur um die Verifizierung der Identität, sondern auch darum, die Kontrolle zu behalten.**

Wichtige Themen dieses Whitepapers

Sie erhalten ein umfassendes Framework, mit dem Sie die zweifache Herausforderung bei der Absicherung von KI-Agenten bewältigen:

1. Für Entwickler – Standardmäßige Absicherung aller Agenten: Sie lernen die wichtigsten Sicherheitsmaßnahmen kennen, die Entwickler bei der Erstellung integrieren müssen, einschließlich robuster **Authentifizierung** für alle Benutzer, sicherem **Token Vaulting** für API-Zugriffe, feingranulare **Datenautorisierung** für RAG-Systeme und Kontrollen mit **Involvierung eines Menschen** für kritische Aktionen.

2. Für IT- und Security-Teams – Absicherung aller Agenten mit einer gemeinsamen Kontrollebene: Sie erfahren, welche unternehmensgerechten Funktionen erforderlich sind, um Agenten im großen Umfang zu verwalten. Dazu gehören **Agentenerkennung** (zum Finden von „Schatten-KI“), eine **Agentenregistrierung** (zum Festlegen von Identität und Zuständigkeit), **vollständige Zugriffskontrolle mit domainübergreifendem Vertrauen** (damit Agenten sicher über organisatorische Grenzen hinweg auf Ressourcen zugreifen können und der Benutzerkontext dennoch gewahrt bleibt), **vollständige Lebenszyklusverwaltung** sowie **Bedrohungserkennung** mit Universal Logout-Funktionen.

3.

Sie müssen alle Agenten **einzeln** und **als** Gesamtheit schützen.

Wichtige Erkenntnisse

- **Das größte Risiko ist die „Governance-Lücke“:** Das grundlegende Problem ist tatsächlich nicht KI als solche, sondern die Agenten-bereitstellung, die bei Weitem über die Möglichkeiten der Governance-Maßnahmen hinausgeht, die zur Kontrolle und Aufsicht zur Verfügung stehen. Dabei umfasst Governance sowohl präventive Kontrollen (Zugriffsrichtlinien, Durchsetzung des Least-Privilege-Prinzips, Autorisierungsregeln) als auch reaktive Kontrollen (Zertifizierungen, Zugriffsprüfungen, Verhaltensüberwachung). Unternehmen benötigen beides, um KI-Agenten effektiv verwalten zu können.
- **Sicherheit ist eine zweifache Herausforderung:** Eine vollständige Strategie muss Sicherheit bei der Entwicklung (die korrekte Erstellung von Agenten) sowie unternehmensweite Governance (Verwaltung im großen Umfang) umfassen.
- **Eine einheitliche Plattform ist unverzichtbar:** Sie benötigen eine einheitliche Identity-Plattform, die Agenten als vollwertige Identitäten behandelt und von der Erkennung und Registrierung bis zum Ende ihres Lebenszyklus verwaltet. Dadurch wird die Governance-Lücke geschlossen und das Datenschutzrisiko minimiert, sodass Unternehmen KI problemlos im großen Maßstab einsetzen können.

Standardmäßige Absicherung aller Agenten

Bei der Erstellung von KI-Agenten müssen Sie Sicherheitsanforderungen erfüllen, die sich grundsätzlich von der klassischen Anwendungsentwicklung unterscheiden. Sicherheit lässt sich nicht nachträglich implementieren: Sie muss ab der ersten Codezeile in die Agentenarchitektur integriert werden.

Dieser Abschnitt richtet sich an drei unterschiedliche Entwicklerzielgruppen:

- **B2C-SaaS-Entwickler**, die verbraucherorientierte KI-Agenten erstellen (Chatbots, persönliche Assistenten, Empfehlungsmodule)
- **B2B-SaaS-Entwickler**, die KI-Agenten für Geschäftskunden erstellen (Workflow-Automatisierung, Analysen, Enterprise-Tools)
- **Unternehmensentwickler**, die interne KI-Agenten für die spezifischen Workflows und Prozesse ihres Unternehmens erstellen

Auch wenn sich die Implementierungsdetails dieser Szenarien leicht unterscheiden können, gelten die grundlegenden Sicherheitsmuster für Authentifizierung, Token-Management, Autorisierung und Überwachung durch Menschen universell. Diese Lösung konzentriert sich auf die Bereitstellung der wichtigsten Funktionen, damit Entwickler sichere Agenten erstellen können, ohne bei Geschwindigkeit oder Innovationen Abstriche in Kauf nehmen müssen.

1. Authentifizierung: Festlegung der Benutzer-Identity

KI-Agenten müssen Benutzer zuverlässig identifizieren können, um personalisierte Experiences bereitstellen zu können und gleichzeitig Sicherheitseinschränkungen einzuhalten. Wichtig ist dabei, dass wir nicht den Agenten, sondern den Benutzer authentifizieren. Der Agent agiert im Namen dieses authentifizierten Benutzers. Unabhängig davon, ob Sie interaktive Chatbots oder Hintergrund-Worker nutzen, benötigen Agenten eine zuverlässige Authentifizierung, die sich nahtlos mit modernen Identity-Anbietern integriert.

Unternehmen benötigen folgende Funktionen:

- **Universelle Authentifizierung**, die mehrere Identity-Anbieter abdeckt und klassische Anmeldedaten sowie Social-Login-Optionen unterstützt
- **Standardbasierte Authentifizierung** auf Basis von OpenID Connect und OAuth 2.0, die Interoperabilität und Sicherheit gewährleistet
- **Übermittlung der Benutzer-Identity mit sicheren Token**, sodass Agenten verstehen, in wessen Namen sie agieren
- **Zuverlässiges Session-Management** mit angemessenen Timeouts und Sicherheitskontrollen
- **Multi-Faktor-Authentifizierung** für Szenarien mit erhöhten Sicherheitsanforderungen

Die **Entwickler-Experience** sollte die Integration mit wenigen Zeilen Code ermöglichen, sodass Entwickler reibungslos mit gängigen Frameworks arbeiten und automatisch komplexe Funktionen für Callback-URLs, Session-Management und Token-Validierung implementiert werden können.

Beispiel: Ein Kundensupport-Chatbot authentifiziert Benutzer per Google-SSO. Als Sarah sich anmeldet, erhält der Agent ihre Identity-Informationen und kann personalisierte Antworten geben, bei denen er Sicherheitseinschränkungen einhält.

2. Token-Austausch: Eine Brücke zwischen Vertrauensdomains

Da KI-Agenten über mehrere Systeme und Sicherheitsdomains hinweg agieren, benötigen sie oft Zugriff auf Ressourcen in unterschiedlichen Vertrauensbereichen. Durch Token-Austausch können Agenten Access Token mit dem benötigten Rechteumfang für Ressourcen außerhalb ihrer unmittelbaren Domain erhalten, wobei Benutzerkontext und Autorisierungsketten bewahrt werden.

Unternehmen benötigen folgende Funktionen:

- **Standard-Token-Austausch** für Szenarien mit einer einzigen Vertrauensdomain, sodass Agenten unterschiedliche Token-Typen oder Berechtigungsbereiche vom selben Autorisierungsserver anfordern können
- **Domainübergreifendes Vertrauen** für Szenarien, bei denen Zugriff auf mehrere Vertrauensbereiche erforderlich ist
- **Mechanismen zur Bewahrung der Benutzer-Identity** und des Authentifizierungskontexts über mehrere Vertrauensbereiche hinweg
- **Validierung von Vertrauensbeziehungen** zwischen unterschiedlichen Identity-Anbietern
- **Übertragung des Berechtigungsumfangs**, sodass Berechtigungen richtig zwischen verschiedenen Domains übertragen werden
- **Sichere Konvertierung von Anmeldedaten**, damit vertrauliche Token während der Übertragung niemals offengelegt werden

Bei Agenten, die in einer einzigen Autorisierungsserver-Umgebung aktiv sind, ermöglicht der Austausch von Standard-OAuth 2.0-Token effizientes Anmeldedaten-Management. Wenn Agenten jedoch auch außerhalb der organisatorischen Grenzen agieren müssen, erweitert domainübergreifendes Vertrauen diese Funktion für Vertrauensdomains.

Standardmäßige OAuth-basierte Zustimmungsverwaltung oder domainübergreifendes Vertrauen

Bei der Wahl zwischen zustimmungsbasierten Workflows und domainübergreifendem Vertrauen ist das Bereitstellungsmodell ein entscheidender Faktor:

B2C-Szenarien: Standardmäßige OAuth-basierte Zustimmungsverwaltung

- Verbraucherorientierte Anwendungen mit Endbenutzern, die Besitzer ihrer Daten sind
- Benutzer gewähren einer Anwendung die Berechtigung, auf eine andere Anwendung zuzugreifen (z. B. „erlaube TravelBot den Zugriff auf deinen Google-Kalender“)
- Die Abfrage der Zustimmung ist erforderlich, weil Benutzer individuelle Entscheidungen über ihre eigenen Daten treffen
- **Beispiel:** Eine App zur Mahlzeitengestaltung fordert Zugriff auf die Fitnesstracker-Daten eines Benutzers an.

B2B- und Belegschaft-Szenarien: Domainübergreifendes Vertrauen

- Unternehmensumgebungen, in denen IT-Administratoren die Richtlinien zentral verwalten
- B2B2E-Szenarien (Business-to-Business-to-Employee), bei denen Benutzer aus der Belegschaft unternehmensweiten Richtlinien unterliegen
- Zustimmungsabfragen einzelner Benutzer sind nicht erforderlich, da der Zugriff durch Unternehmensrichtlinien anstatt durch individuelle Benutzerentscheidungen geregelt ist
- Der unternehmenseigene Identity-Anbieter agiert anwendungsübergreifend als Vertrauensvermittler
- **Beispiel:** Der Vertriebsmitarbeiter eines Unternehmens greift auf das Salesforce-CRM sowie auf eine interne Datenbank für die Preisgestaltung zu. Dieser Zugriff ist nicht von der Zustimmung der Mitarbeiter, sondern von IT-Richtlinien abhängig.

Warum ist das wichtig?
Im Kontext von Belegschaften und B2B beseitigt domainübergreifendes Vertrauen unnötige Zustimmungsabfragen und hält zentrale IT-Governance-Vorschriften ein. Unternehmen legen im Vorfeld Vertrauensbeziehungen zwischen Anwendungen fest und der Identity-Anbieter setzt Unternehmensrichtlinien durch, damit einzelne Benutzer nicht für jede anwendungsübergreifende Interaktion eine Autorisierungsentscheidung treffen müssen.

3. Token Vaulting: Sicheres API-Zugriffsmanagement

KI-Agenten benötigen häufig Zugriff auf Drittanbieter-APIs (z. B. Salesforce, Slack, Google Workspace), um im Namen der Benutzer Aufgaben durchzuführen. Token Vaulting übernimmt die sichere Speicherung und Verwaltung dieser OAuth Access Token. Diese sind die bevorzugte Authentifizierungsmethode für die meisten APIs und vermeiden das Risiko einer Token-Kompromittierung in Code, Protokollen oder Konfigurationsdateien. Obwohl der Vault auch andere Arten von Anmeldedaten schützen kann (z. B. persönliche Access Token oder API-Schlüssel für Legacy-Systeme), sollten Sie standardmäßig OAuth-Token nutzen, da sie automatische Aktualisierung, die granulare Definition von Berechtigungsbereichen (Scoping) sowie einen sicheren Widerruf ermöglichen.

Unternehmen benötigen folgende Funktionen:

- **Sichere Vault-Speicherung** von OAuth-Token mit starker Verschlüsselung
- **Automatische Token-Lebenszyklusverwaltung** einschließlich proaktiver Aktualisierung vor dem Ablaufdatum, um Dienstunterbrechungen zu vermeiden
- **On-Demand-Token-Abruf**, wobei Anwendungscode niemals mit Anmeldedaten in Berührung kommt
- **Unterstützung verschiedener Token-Typen** für unterschiedliche Authentifizierungsverfahren
- **Zugriff basierend auf dem Berechtigungsbereich**, sodass gewährleistet ist, dass Token nur notwendige Berechtigungen besitzen
- **Integrationsmuster**, die mit modernen KI-Entwicklungs-Frameworks kompatibel sind

Sicherheitsprinzip
Token dürfen niemals in Code, Protokollen oder Konfigurationsdateien von Agenten enthalten sein. Der Vault verwaltet transparent den gesamten Token-Lebenszyklus und verringert dank eines zentralen und prüffähigen Systems das Risiko von Anmeldedatendiebstahl oder -missbrauch.

Schutz von Anmeldedaten bei Agentenausgaben
Unternehmen müssen nicht nur verhindern, dass Anmeldedaten in Code und Protokollen enthalten sind, sondern gleichzeitig sicherstellen, dass Token und Secrets nicht in Antworten oder Ausgaben von Agenten auftauchen. Wenn Agenten die Anmeldedaten für den API-Zugriff aus dem Vault abrufen, können die zurückgegebenen Daten vertrauliche Informationen enthalten. Die Anmeldedaten selbst dürfen jedoch niemals in Antworten, generierten Dokumenten oder für Endbenutzer sichtbaren Ausgaben des Agenten enthalten sein. Implementieren Sie Ausgabefilter und Validierung zur Erfassung aller versehentlich kompromittierten Anmeldedaten, bevor sie die Benutzer erreichen. Dies ist insbesondere bei Agenten wichtig, die Code-Snippets, Konfigurationsbeispiele oder Leitfäden zur Problembehebung generieren, die versehentlich Anmeldedaten enthalten können.

Beispiel: Ein KI-Vertriebsagent benötigt Zugriff auf den Salesforce-Account eines Kunden. Statt Salesforce-Anmeldedaten fordert der Agent ein Token vom Vault an. Der Vault liefert ein aktuelles Token mit dem benötigten Rechteumfang und aktualisiert es bei Bedarf. Der Agent führt seine Aufgabe aus, wobei der Agentencode niemals mit den Anmeldedaten in Berührung kommt.

4. Datensicherheit: Feingranulare Autorisierung für RAG

Wenn KI-Agenten mithilfe von Retrieval Augmented Generation (RAG) Fragen beantworten, dürfen sie nur auf Daten zugreifen, für die der Benutzer Berechtigungen hat. Ohne ausreichende Autorisierung können RAG-Systeme unbeabsichtigt vertrauliche Informationen offenlegen und damit eine schwerwiegende Sicherheitslücke schaffen.

Unternehmen benötigen folgende Funktionen:

- **Beziehungsbasierte Zugriffskontrollen**, die Benutzer-Dokument-Berechtigungen definieren
- **Durchsetzung der Autorisierung** am Punkt des Dokumentenabrufs, bevor Daten in den Agentenkontext gelangen
- **Integrationsmuster für Vektor-Datenbanken**, die in RAG-Architekturen verwendet werden
- **Feingranulare Berechtigungen**, die auf Dokumenten- oder Abschnittebene wirksam sind
- **Bewertung der Echtzeit-Autorisierung** während der Abfrageverarbeitung
- **Unterstützung komplexer Berechtigungsmodelle** einschließlich hierarchischer und attributbasierter Richtlinien

Funktionsweise des Musters
Dokumente werden mit eingebetteten Inhalten in einer Vektordatenbank gespeichert. Ein Autorisierungssystem hält die Beziehungen zwischen Benutzern und Dokumenten aufrecht. Wenn ein Agent Kontext abruft, validieren Autorisierungsfilter die Berechtigungen, bevor auch nur ein Dokument in den Agentenkontext gelangt. Das LLM generiert Antworten nur mit den Daten, für die der Benutzer autorisiert ist.

Beispiel: Ein Finanz-KI-Agent unterstützt Mitarbeiter beim Analysieren von Berichten. Als Alice die Ergebnisse des 3. Quartals anfordert, ruft die Vektordatenbank die relevanten Finanzdokumente ab. Bevor sie an das LLM übergeben werden, überprüfen die Autorisierungsfilter die Zugriffsrechte von Alice. Ihr werden nur die Berichte ihrer eigenen Abteilung angezeigt, nicht jedoch die Finanzdaten des gesamten Unternehmens. Dadurch wird eine nicht autorisierte Datenweitergabe verhindert.

5. Abgesicherte Tool-Aufrufe: Autorisierung mit Involvierung eines Menschen

KI-Agenten sind oft autonom im Hintergrund aktiv und benötigen zur Erfüllung ihrer Aufgaben Minuten, Stunden oder sogar Tage. Bei wichtigen Aktivitäten wie dem Genehmigen von Käufen, Übermitteln von Verträgen oder Gewähren von Zugriffen ist eine Bestätigung durch Menschen erforderlich (Human-in-the-Loop), die jedoch die Autonomie der Agenten bei Routineprozessen nicht beeinträchtigen darf.

Unternehmen benötigen folgende Funktionen:

- **Asynchrone Autorisierungsmuster**, die mit langfristigen Agenten-Workflows funktionieren
- **Umfangreiche Benachrichtigungsmechanismen**, die Prüfern den vollständigen Transaktionskontext bereitstellen
- **Bindungsnachrichten** mit wichtigen Informationen wie Mengen, Empfängern und beabsichtigten Aktionen
- **Genehmigungs-Workflows**, die sich per Mobilgerät und E-Mail abrufen lassen, ohne dass ein Desktop-Computer erforderlich ist
- **Zeitlich befristete Autorisierungsanfragen**, die automatisch ablaufen, wenn sie nicht explizit verlängert werden
- **Umfangreiche Audit-Trails**, die alle Genehmigungsentscheidungen und ihren Kontext dokumentieren

Funktionsweise des Musters
Entwickler legen fest, welche Agentenaktionen eine Benutzerbestätigung erfordern. Wenn ein Agent versucht, eine geschützte Aktion durchzuführen, erhält die betreffende Person eine entsprechende Autorisierungsanfrage mit vollständigem Kontext. Anschließend gewährt oder verweigert der Prüfer die Berechtigung. Im ersteren Fall erhält der Agent die Autorisierung und kann fortfahren, während er bei einer Verweigerung eine Fehlermeldung erhält.

Beispiel: Ein KI-Agent der Beschaffungsabteilung stellt fest, dass Software-Lizenzen fehlen, und will sie erwerben. Vor der Überweisung von 5.000 EUR sendet er eine Benachrichtigung an den Einkaufsmanager und nennt den Anbieter, die Menge und eine Begründung. Der Manager überprüft die Anfrage beim Mittagessen, sendet per Mobilgerät oder E-Mail eine Bestätigung und der Agent führt den Kauf durch. Dadurch sind Automatisierung und Kontrolle jederzeit gewährleistet.

Absicherung aller Agenten über eine gemeinsame Kontrollebene

Auch wenn das Integrieren von Sicherheitsmaßnahmen in individuelle Agenten während der Entwicklung wichtig ist, benötigen Unternehmen zusätzlich zentrale Transparenz, Kontrolle und Governance für ihren gesamten KI-Agentenbestand. Diese Lösung ist die Antwort auf die Herausforderung von Unternehmen, die für ihre Abteilungen, Anwendungsfälle und Systeme Hunderte oder Tausende KI-Agenten nutzen.

1. Agentenregistrierung: Festlegung vollwertiger Identities

Alle KI-Agenten müssen als vollwertige Identitäten mit klarer Zuständigkeit und Verantwortlichkeit registriert werden, da Unternehmen andernfalls keine Übersicht besitzen und grundlegende Fragen nicht beantworten können, zum Beispiel: Wer ist für diesen Agenten zuständig? Wofür ist er autorisiert? Wer ist bei einem Fehler verantwortlich?

Unternehmen benötigen folgende Funktionen:

- **Identity-Profile** für alle Agenten mit dauerhaften und eindeutigen Identifikatoren
- **Zuordnung der Zuständigkeit**, die Agenten mit verantwortlichen Teams oder Personen verbindet (Besitzer)
- **Metadaten-Systeme**, die Zweck, Anwendungsfall und Lebenszyklusphase des Agenten dokumentieren
- **Integration** mit Unternehmensstrukturen wie HR-Systemen und Berichtshierarchien
- **Verwaltung von Änderungen**, die Veränderungen an Agentenkonfigurationen und Berechtigungen erfasst
- **Zuordnung von Abhängigkeiten**, die Beziehungen zwischen Agenten und den Systemen zeigt, auf die sie Zugriff haben

Warum ist das wichtig?
Laut der Umfrage zum Bericht „AI at Work 2025“, besitzen nur 10 % aller Unternehmen eine gut ausgearbeitete Strategie zur Verwaltung nicht-menschlicher Identitäten. Die Registrierung bildet eine grundlegende Identity-Ebene, auf der alle anderen Governance- und Sicherheitskontrollen aufbauen. Ohne diese Schicht bleiben Agenten für Security-Teams unsichtbar, sodass nicht verwaltete „Schatten-KI“ entsteht.

Beispiel: Das Security-Team entdeckt einen KI-Agenten, der über einen Service-Account auf Salesforce zugreift. Der Agent wird in der zentralen Agentenregistrierung erfasst, der Vertriebsabteilung zugeordnet und sein Zweck dokumentiert: „Automatisierte Angebotsgenerierung für Unternehmens-Accounts“. Falls es zu unerwartetem Verhalten kommt, ist die Verantwortlichkeit geklärt und ein klarer Ansprechpartner angegeben, der kontaktiert werden kann und die Behebung übernimmt.

2. Zugriffskontrolle: Richtlinienbasierte Autorisierung

KI-Agenten benötigen feingranulare Autorisierung, die Kontext und Risiko in Echtzeit berücksichtigt. Mithilfe von Zugriffskontrollrichtlinien wird festgelegt, was ein Agent wann und unter welchen Umständen tun kann. Damit wird das Least-Privilege-Prinzip umgesetzt, ohne die Funktionsfähigkeit des Agenten einzuschränken.

Unternehmen benötigen folgende Funktionen:

- **Richtlinien-Engines**, die Berechtigungen anhand von Agenten-Identity, operativem Kontext und Risikoindikatoren definieren
- **Auf Standards basierende Authentifizierungsabläufe**, die moderne Protokolle unterstützen
- **API-Zugriffsmanagement** kontrolliert, wie Agenten mit geschützten Services interagieren
- **Funktionen für domainübergreifendes Vertrauen**, damit Agenten auf sichere Weise über Unternehmens- und Vertrauensdomain-Grenzen hinweg auf Ressourcen zugreifen können, wobei der Benutzerkontext bewahrt bleibt
- **Dynamische Richtlinienevaluierung**, die mehrere Faktoren wie Zeit, Standort und Verhaltensmuster berücksichtigt
- **Integrationsmuster** zum Verbinden mit der bestehenden IAM-Infrastruktur
- **Unterstützung für komplexe anbieterübergreifende Architekturen**, die unterschiedliche Sicherheitsdomains abdecken

Erweitertes Muster – verwaltete Verbindungen
Unternehmen müssen festlegen können, auf welche Autorisierungsserver ein Agent zugreifen und welche Berechtigungen er abrufen kann. Dazu gehören Richtlinienrahmen, die festlegen, welche Berechtigungen automatisch gewährt werden, welche zusätzliche Genehmigung erfordern und welche niemals gewährt werden. Dadurch ist gewährleistet, dass Agenten stets innerhalb klar definierter Grenzen agieren. Bei Agenten, die auf Ressourcen in anderen Vertrauensdomains zugreifen müssen, erweitert domainübergreifendes Vertrauen diese verwalteten Verbindungen und ermöglicht sichere unternehmensübergreifende Autorisierung. Dabei wird die zentrale Richtlinienkontrolle stets sichergestellt.

Beispiel: Wenn ein Agent Zugriff anfordert, wird diese Anfrage anhand festgelegter Richtlinien vom Autorisierungssystem überprüft. Routinemäßige Berechtigungen können automatisch lokal gewährt werden, während sensible Prozesse eine Rechtfertigung oder Genehmigung erfordern und gefährliche Aktionen direkt verweigert werden. All das wird programmgesteuert ohne manuelle Eingriffe durchgesetzt. Wenn ein Agent auf die API eines Partnerunternehmens zugreifen oder von der Cloud zu On-Premise-Systemen gelangen muss, ermöglicht XAA diesen Domain-übergreifenden Zugriff und gewährleistet die Transparenz und Richtliniendurchsetzung für die zentrale Kontrollebene.

3. Lebenszyklusverwaltung: Vollständige Agenten-Governance

Ebenso wie menschliche Mitarbeiter haben auch KI-Agenten einen Lebenszyklus, der in diesem Fall Onboarding, aktiven Betrieb, Rollenwechsel und letztendlich die Stilllegung umfasst. Die Lebenszyklusverwaltung automatisiert den Wechsel zwischen diesen Phasen und gewährleistet gleichzeitig angemessene Sicherheitskontrollen.

Unternehmen benötigen folgende Funktionen:

- **Automatisierte Provisionierungs-Workflows**, die Agenten-Identities mit den richtigen initialen Berechtigungen erstellen
- **Rollenbasierte Templates**, die Zugriffsmuster für gängige Agententypen standardisieren
- **Just-in-Time-Zugriffsmöglichkeiten** für zeitlich begrenzt erhöhte Berechtigungen mit automatischem Ablaufzeitpunkt
- **Geplante Prüfprozesse**, die gewährleisten, dass es keine unzulässigen Änderungen an Berechtigungen gibt
- **Deprovisionierungs-Workflows**, die systematisch den gesamten Zugriff sperren, wenn Agenten stillgelegt werden
- **Change-Management-Systeme**, die Änderungen und Genehmigungen von Berechtigungen nachverfolgen

Der vollständige Lebenszyklus
Agenten werden entsprechend ihrem Einsatzzweck mit initialen Berechtigungen bereitgestellt. Während ihrer Nutzung führen sie Aufgaben aus, wobei ihre Zugriffsrechte kontinuierlich überprüft werden. Wenn sich die Anforderungen ändern, werden die Berechtigungen mithilfe von Genehmigungs-Workflows entsprechend angepasst und regelmäßig kontrolliert, ob der Zugriff noch angemessen ist. Wenn ein Agent stillgelegt wird, werden alle seine Berechtigungen zurückgezogen, wobei die Audit-Trails zu Compliance-Zwecken gespeichert werden.

Beispiel: Ein KI-Agent der Marketingabteilung wird mit Zugriff auf die E-Mail-Plattform und die Kundendatenbank provisioniert. Nach sechs Monaten wird bei einer vierteljährlichen Prüfung festgestellt, dass der Agent (wegen eines geänderten Anwendungsfalls) keinen Datenbankzugriff mehr benötigt. Daher wird sein Zugriff automatisch gesperrt. Nach dem Ende der Kampagne wird der Agent deprovisioniert und alle Berechtigungen werden entfernt, wobei Audit-Protokolle zu Compliance-Zwecken gespeichert bleiben.

4. Privilegierte Anmeldedaten: Sicheres Secrets-Management

KI-Agenten benötigen für den Zugriff auf Systeme häufig privilegierte Anmeldedaten wie API-Schlüssel, Datenbankpasswörter, Service-Account-Anmeldedaten sowie Zertifikate. Unzureichendes Anmeldedaten-Management, festkodierte Schlüssel sowie niemals rotierte Secrets schaffen enorme Sicherheitsrisiken, die von Angreifern aktiv ausgenutzt werden.

Unternehmen benötigen folgende Funktionen:

- **Sicherer Vault-Speicher** mit starker Verschlüsselung, der gespeicherte Anmeldedaten schützt
- **Zeitpläne für automatisierte Rotationen**, damit Anmeldedaten regelmäßig aktualisiert werden
- **Just-in-Time-Provisionierungsmuster**, die das Gefährdungsfenster für Anmeldedaten minimieren
- **Unterstützung unterschiedlicher Authentifizierungsmethoden** einschließlich schlüssel- und zertifikatbasierter Verfahren
- **Automatisierte Zertifikat-Lebenszyklusverwaltung** einschließlich Erneuerung und Verteilung
- **Konsequente Isolierung**, sodass Secrets niemals in Code, Protokollen oder Konfigurationsdateien enthalten sind
- **Integrationsmuster** für externe Secret-Management-Systeme

Auswirkung auf die Sicherheit
Da Angreifer verstärkt auf gestohlene Anmeldedaten setzen, kann die Beseitigung langlebiger Secrets durch automatisierte Anmeldedatenrotation die Angriffsfläche erheblich verkleinern.

Beispiel: Ein Daten-Pipeline-Agent nutzt Datenbank-Anmeldedaten, die aller 30 Tage rotiert werden. Wenn die Rotation durchgeführt wird, ruft der Agent automatisch neue Anmeldedaten aus dem Vault ab, ohne dass dazu Eingriffe durch Menschen oder Dienstunterbrechungen erforderlich sind. Der größte Vorteil für die Sicherheit besteht darin, dass Anmeldedaten zu keinem Zeitpunkt in Protokollen oder Code enthalten sind (vollständige Isolierung). Falls Anmeldedaten auf irgendeine Weise offengelegt werden, verringert die automatisierte Rotation das Gefährdungsfenster auf maximal 30 Tage anstatt auf einen unbegrenzten Zeitraum, was die Angriffsfläche erheblich verkleinert.

5. Agentenerkennung: Identifizierung von Schatten-KI

Unternehmen können nur absichern, was sie sehen. Die Agentenerkennung liefert einen Überblick über alle KI-Agenten in der Umgebung, einschließlich Schatten-KI, die möglicherweise von Geschäftsabteilungen ohne Genehmigung oder eine Sicherheitsüberprüfung durch die IT-Abteilung eingesetzt wird.

Unternehmen benötigen folgende Funktionen:

- **Automatisierte Erkennungsmechanismen**, die nicht-menschliche Accounts in Cloud- und SaaS-Plattformen identifizieren können
- **Schatten-KI-Erkennung**, die Agenten findet, die außerhalb formeller Governance-Prozesse bereitgestellt wurden
- **Umfassendes Anmeldedateninventar**, das zeigt, welche Agenten auf welche Systeme Zugriff haben
- **Risikobewertungsverfahren**, die Berechtigungen, Aktivitätsmuster und Gefährdungsniveau berücksichtigen
- **Konfigurationsanalyse**, die Konfigurationsfehler und übermäßig privilegierte Accounts identifiziert
- **Integrationsmuster** für die Verbindung mit Cloud-Plattformen, Identity-Anbietern und Sicherheitstools

Ansätze für die Erkennung
Für effektive Agentenerkennung müssen mehrere Techniken kombiniert werden: Analyse der API-Nutzungsmuster zum Identifizieren von nicht-menschlichem Verhalten, Korrelation von Authentifizierungsprotokollen zum Erkennen von Service-Account-Aktivitäten, Scans von Cloud-Ressourcen zum Finden bereitgestellter Agenten, Überwachung des Netzwerkverkehrs zum Identifizieren von Agent-zu-Service-Kommunikation sowie Durchführung von Verhaltensanalysen zum Unterscheiden zwischen Agenten und menschlichen Benutzern.

Warum ist das wichtig?
Die „AI at Work 2025"-Umfrage hat gezeigt, dass 91 % der Unternehmen bereits KI-Agenten nutzen, aber nur 10 % über gut ausgearbeitete Governance-Strategien verfügen. Die Lücke zwischen Bereitstellung und Governance kann dazu führen, dass Schatten-KI-Agenten ohne Sicherheitskontrollen agieren – was Risiken schafft, von deren Existenz Security-Teams keine Kenntnis haben.

6. Universal Logout für Agenten: Schnelle Threat-Response

Wenn Bedrohungen wie kompromittierte Anmeldedaten, ungewöhnliches Verhalten oder ein Richtlinienverstoß entdeckt werden, müssen Unternehmen über Möglichkeiten verfügen, den Zugriff aller Agenten auf allen Systemen sofort zu sperren. Universal Logout für Agenten bietet einen solchen „Notausschalter" und gewährleistet gleichzeitig einen detaillierten Audit-Trail.

Unternehmen benötigen folgende Funktionen:

- **Mechanismen für den sofortigen Widerruf** aller aktiven Agenten-Sessions und Token
- **Systemübergreifende Reichweite**, damit die Abmeldung alle integrierten Anwendungen erreicht
- **Notfallrotation der Anmeldedaten**, die potenziell kompromittierte Secrets ersetzt
- **Workflows zur Bedrohungseindämmung**, die weitere nicht autorisierte Aktivitäten verhindern
- **Speicherung forensischer Daten**, um vollständige Audit-Protokolle für spätere Untersuchungen verfügbar zu haben
- **Integration** mit SOCs (Security Operations Centers) und Incident-Response-Systemen

Bedrohungserkennung durch Verhaltensanalysen
Für eine effektive universelle Abmeldung (Universal Logout) ist zwingend eine zuverlässige Bedrohungserkennung erforderlich. Das erfordert die Festlegung von Basiswerten für normales Agentenverhalten, die Erkennung von Anomalien wie ungewöhnliche Häufigkeiten oder Zeitpunkte für Anfragen, die Berechnung von Risiko-Scores anhand von Verhaltensmustern, die Auslösung automatisierter Reaktionen, wenn Risiken festgelegte Schwellenwerte überschreiten, sowie die Echtzeit-Benachrichtigung von Security-Teams bei verdächtigen Aktivitäten.

Beispiel: Ein Kundenservice-KI-Agent greift pro Tag normalerweise auf 10–15 Kundendatensätze zu. Plötzlich ruft er jedoch 500 Datensätze innerhalb von 10 Minuten ab, was eindeutig ungewöhnlich ist. Verhaltensanalysen entdecken die Abweichung, lösen automatisch eine universelle Abmeldung mit Sperrung aller Agentenzugriffe aus und benachrichtigen das Security-Team. Bei der Untersuchung wird festgestellt, dass API-Anmeldedaten gestohlen wurden. Der Angriff wird innerhalb weniger Minuten eingedämmt und eine vollständige Exfiltration der Datenbank verhindert.

Zusammenarbeit aller Komponenten

Diese beiden Lösungen sichern standardmäßig alle Agenten ab und stellen eine gemeinsame Kontrollebene bereit. Es sind keine separaten Lösungen, die eine Integration erfordern, sondern sich ergänzende Funktionen innerhalb einheitlicher Identity-Plattformen, die speziell für KI-Agenten entwickelt wurden.

Während der Entwicklung werden von der ersten Codezeile an Sicherheitsmaßnahmen in jeden Agenten integriert. Per Authentifizierung wird die Benutzer-Identity festgelegt, sodass gewährleistet ist, dass Agenten wissen, in wessen Namen sie aktiv sind, und angemessene Sicherheitseinschränkungen einhalten. Token Vaulting verwaltet den API-Zugriff, ohne dass Anmeldedaten im Anwendungscode offengelegt werden, und übernimmt automatisch die Token-Aktualisierung und Lebenszyklusverwaltung. Autorisierungskontrollen verhindern nicht autorisierte Datenzugriffe, indem sie feingranulare Berechtigungen implementieren, die den Zugriffsrechten der Benutzer folgen. Human-in-the-Loop-Genehmigungen kontrollieren kritische Aktionen und ermöglichen autonome Agentenaktivitäten, gewährleisten dabei jedoch die Kontrolle über sensible Prozesse.

In der Praxis sieht das wie folgt aus:

- **Universal Login für universelle Authentifizierung:** Ermöglicht nahtlose Benutzerauthentifizierung für Social-Media-Provider, passwortlose Flows und MFA für KI-Agenten.
- **Token Vault für sichere API-Zugriffe:** Speichert und verwaltet OAuth-Token für externe APIs, wobei Anmeldedaten automatisch aktualisiert werden, ohne dass Secrets mit dem Agentencode in Berührung kommen.
- **Auth0 FGA für feingranulare Autorisierung:** Implementiert beziehungsbasierte Zugriffskontrolle für RAG-Systeme und gewährleistet, dass Agenten nur Dokumente abrufen können, für die Benutzer über die entsprechenden Berechtigungen verfügen.
- **Asynchrone Autorisierung für Human-in-the-Loop-Autorisierungen:** Fordert für kritische Agentenaktionen Genehmigungen per Mobilgerät oder E-Mail mit CIBA (Client-Initiated Backchannel Authentication) und Rich Authorization Request (RAR) an.

Für IT- und Security-Teams in Produktionsumgebungen liefert die Kontrollebene einen kontinuierlichen Überblick über den gesamten Agentenbestand. Die Erkennungsfunktionen finden sämtliche in der Umgebung aktiven Agenten (einschließlich Schatten-KI), sodass Security-Teams einen vollständigen Überblick über ihre Agentenumgebung haben. Die Registrierung gewährleistet vollwertige Agenten-Identities mit klarer Zuständigkeit, sodass für jedem Agenten ein verantwortliches Team oder eine verantwortliche Person zugeordnet ist.

Governance für den Zugriff setzt dynamisch Least-Privilege-Richtlinien durch und passt Berechtigungen in Echtzeit basierend auf Kontext und Risiko an. Die Bedrohungserkennung erkennt Anomalien mithilfe von Verhaltensanalysen und reagiert mit automatisierter Eindämmung, bevor ein Angriff Erfolg hat.

In der Praxis sieht das wie folgt aus:

- **Universal Directory zur Agentenregistrierung:** Registriert jeden Agenten als vollwertige Identität mit Zuständigkeit, Verantwortlichkeit und umfangreichen Metadaten.
- **Identity Security Posture Management zur Agentenerkennung:** Erkennt verwaltete und Schatten-KI-Agenten in verschiedenen Umgebungen und beseitigt dadurch blinde Flecken.
- **Okta Identity Governance für Zugriffskontrolle und Lebenszyklusverwaltung:** Definiert, überprüft und zertifiziert Agentenberechtigungen mithilfe richtlinienbasierter Lebenszyklus-Governance.
- **Okta Privileged Access zum Vaulting privilegierter Anmeldedaten:** Sichert und rotiert Anmeldedaten für Agenten, die erhöhte Berechtigungen benötigen, und verringert dadurch die Angriffsfläche.
- **Universal Logout für Agenten:** Widerruft systemübergreifend sofort Agenten-Sessions, Token und Anmeldedaten, sobald ein Risiko oder eine Kompromittierung entdeckt wird.

Durch die Verbindung zwischen diesen Lösungen entsteht ein umfassendes Sicherheits-Framework. Agenten, die nach grundsätzlich sicheren Verfahren entwickelt wurden, können von der Kontrollebene automatisch erkannt werden. Dadurch ist ein vollständiger Überblick gewährleistet, ohne dass zusätzlicher Integrationsaufwand entsteht. Die Kontrollebene legt anhand von Richtlinien fest, wie sich Agenten authentifizieren und auf Ressourcen zugreifen. Dadurch werden Sicherheitskontrollen aus dem Entwicklungsprozess zu Governance-Maßnahmen zur Laufzeit. Die Aktivitäten der Agenten werden mit Verhaltensanalysen überwacht, was durch sichere Entwicklungsprozesse ermöglicht wird. Dadurch lässt sich erkennen, ob Agenten von erwarteten Verhaltensmustern abweichen. Governance-Workflows decken sowohl Agenten-Identities als auch die Ressourcen ab, auf die sie zugreifen, und ermöglichen die einheitliche Richtliniendurchsetzung im gesamten Ökosystem.

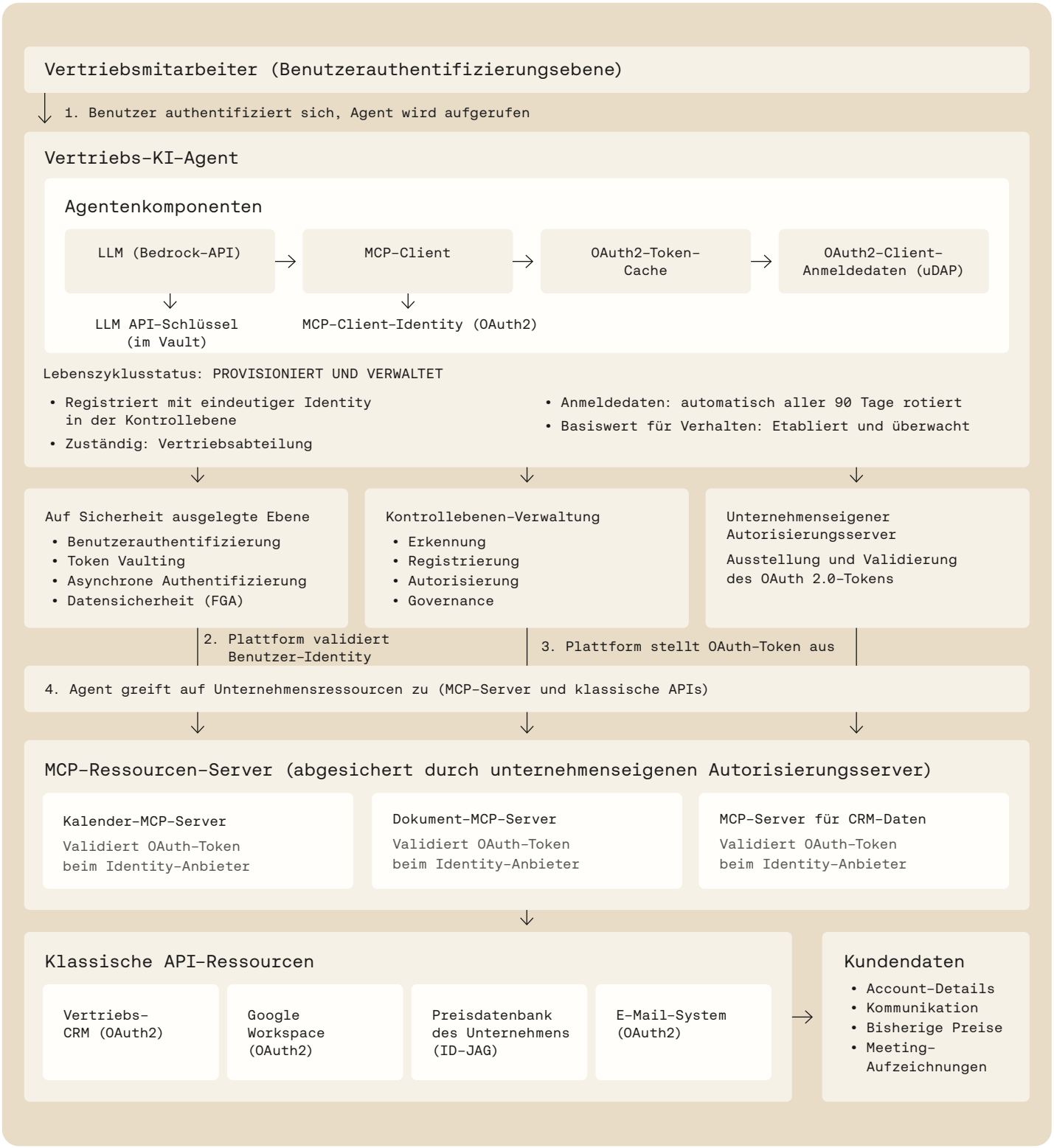
Unternehmen müssen sich nicht zwischen sicherer Entwicklung und Lebenszyklusverwaltung entscheiden – im Rahmen einer konsistenten Strategie sollte beides implementiert werden. Die folgende Referenzarchitektur demonstriert diesen einheitlichen Ansatz in einem realistischen Unternehmensszenario.

Referenz-architektur: Einheitliche Plattform in Aktion

Als Beispiel dafür, wie KI-Agenten mit einer einheitlichen Plattform während der Entwicklung sowie der Lebenszyklus-verwaltung abgesichert werden, dient ein unternehmenseigener Vertriebs-KI-Agent, der Vertriebsmitarbeiter durch automatisierte Kundenrecherche, die Generierung von Angeboten, den Abruf von CRM-Daten sowie die Planung von Folge-Meetings unterstützt. Dieser Agent demonstriert, wie beide Lösungen zusammenarbeiten: von Grund auf sichere Entwicklung in Kombination mit zentraler Lebenszykluskontrolle.

Der Agent ruft mithilfe des Model Context Protocol (MCP) auf sichere Weise Kontext aus mehreren Quellen ab und berücksichtigt dabei während des gesamten Lebenszyklus Benutzerberechtigungen, Token-Austausch (ID-JAG) für domainübergreifendes Vertrauen sowie umfassende Governance.

Übersicht über die Architektur



Die Plattform bietet Unterstützung bei der Absicherung von MCP-Implementierungen (Model Context Protocol). MCP-Server dienen als OAuth 2.0-Ressourcen-Server und delegieren Authentifizierung sowie Autorisierung an den unternehmenseigener Autorisierungsserver.

MCP-Sicherheitsmodell

Der Agent (MCP-Client) wird als OAuth 2.0-Client beim unternehmenseigenen Autorisierungsserver registriert und erhält Client-Anmeldedaten (client_id und client_secret oder zertifikatbasierte Anmeldedaten). Vor dem Zugriff auf eine beliebige MCP-Ressource ruft der Agent ein Access Token mit angemessenen Berechtigungsbereichen ab (z. B. mcp:crm:read, mcp:docs:read, mcp:calendar:read). Wenn der Agent eine Ressource wie crm://contacts/acme-corp anfordert, validiert der MCP-Server das Access Token beim Autorisierungsserver und überprüft Gültigkeit, Ablaufdatum, Zielgruppe sowie erforderliche Berechtigungen der Signatur, bevor er die Ressource übermittelt.

Dadurch müssen MCP-Server-Entwickler keine individuelle Authentifizierungslogik erstellen, sondern validieren stattdessen die OAuth-Token, die von der Plattform mit standardmäßiger OAuth 2.0-Token-Validierung ausgestellt wurden. Die Plattform übernimmt die Benutzerauthentifizierung, die Verwaltung des Token-Lebenszyklus, die Verwaltung der Berechtigungsbereiche sowie FGA-Prüfungen und gewährleistet dadurch die einheitliche Durchsetzung von Sicherheitsrichtlinien für alle MCP-Server sowie vollständige Audit-Trails im Systemprotokoll.

Die Referenzarchitektur demonstriert, wie die Plattform den MCP-basierten Kontextabruf absichert und umfassende Lebenszyklus-Governance für die KI-Agenten bereitstellt.

Detaillierter Ablauf: Agentenunterstützte Angebotsgenerierung mit MCP

Phase 1

Erkennung und Registrierung (Kontrollebene – Erkennung/Provisionierung)

Schritt 1.1: Erkennung von Schatten-KI

- Das Vertriebsteam stellt einen Prototyp-Agenten ohne Wissen der IT-Abteilung bereit.
- Okta erkennt den Agenten, der mit Service-Account-Anmeldedaten auf Salesforce zugreift.
- Das Security-Team erhält eine Warnung über einen nicht verwalteten KI-Agenten.
- Risiko-Score: HOCH (privilegierter Zugriff, keine Zuständigkeit, keine Governance)

Schritt 1.2: Registrierung des Agenten

- Das Security-Team registriert den Agenten in Okta als vollwertige Identität.
- Ein Agentenprofil mit einer eindeutigen Kennung wird erstellt: `sales-agent-prod-001`
- Die Zuständigkeit wird festgelegt: Vertriebsabteilung (John Smith, Vice President für Vertrieb)
- Der Zweck wird dokumentiert: „Automatisierte Angebotsgenerierung und Kundenrecherche“
- Die Anmeldedaten werden in einen sicheren Vault mit 90-Tage-Rotationsrichtlinie migriert.

Ergebnis: Der Agent wird von Schatten-KI zu einer verwalteten Identität mit klarer Zuständigkeit.

Phase 2

Benutzerauthentifizierung (Auf Sicherheit ausgelegte Ebene)
Schritt 2.1: Authentifizierung der Vertriebsmitarbeiterin

- Sarah (Vertriebsmitarbeiterin) ruft um 9:00 Uhr das Vertriebsportal auf.
- Auth0 Universal Login bietet verschiedene Authentifizierungsoptionen an.
- Sarah authentifiziert sich per Google SSO (Social Login).
- Auth0 validiert die Anmeldedaten beim Identity-Anbieter Google.
- Ein ID-Token mit Sarahs Profil und Authentifizierungs-Claims wird ausgestellt.

Schritt 2.2: Bindung des Agentenkontexts

- Ein Agent erhält den authentifizierten Benutzerkontext von Auth0.
- Das ID-Token enthält standardmäßige OIDC-Claims:

```
{
  "iss": "https://acmecorp.auth0.com/",
  "sub": "google-oauth2|108204567890123456789",
  "aud": "sales-agent-client-id",
  "exp": 1730480000,
  "iat": 1730477400,
  "name": "Sarah Johnson",
  "email": "sarah.johnson@acmecorp.com",
  "email_verified": true
}
```

Hinweis: Zusätzliche benutzerdefinierte Claims wie `role` oder `territory` lassen sich mit Auth0 Actions hinzufügen.

- Der Agent weiß jetzt, WER der Benutzer ist, und kann in dessen Namen agieren.

Phase 3

Kontextabruf per MCP (Datensicherheit und Autorisierung)

Schritt 3.1: Benutzeranfrage

Sarah fragt: „Erstelle ein Angebot für Acme Corp auf Grundlage vorheriger Käufe und aktueller Anforderungen.“

Schritt 3.2: MCP-Kontexterkennung

Der Agent nutzt den MCP-Client zur Erkennung verfügbarer Kontextquellen:
MCP-Server stellen strukturierte Ressourcen über Ressourcen-URIs bereit:

```
crm://contacts/acme-corp
docs://proposals/templates
calendar://availability/sales-team
pricing://enterprise-tier
```

Hierbei handelt es sich um einen **strukturierten Kontextabruf per MCP**, nicht um eine semantische Suche über eingebettete Inhalte (RAG).
MCP bietet direkten Zugriff auf spezifische Ressourcen basierend auf definierten Schemas und Ressourcen-URIs. Der Agent fordert konkrete Ressourcen anhand von Name/Pfad an, und MCP-Server geben strukturierte Daten zurück.

Schritt 3.3: Autorisierungsprüfung per Auth0 FGA

Der Agent fragt den FGA-Server (Fine-Grained Authorization) ab, um festzustellen, auf welche Ressourcen Sarah Zugriff hat:

- FGA evaluiert Beziehungs-Tupel für jede MCP-Ressource:
  - ✓ `user:sarah` hat `read`-Zugriff (Lesezugriff) auf `crm://contacts/acme-corp`
  - ✓ `user:sarah` hat `read`-Zugriff (Lesezugriff) auf `docs://proposals/templates`
  - ✗ `user:sarah` hat KEINEN Zugriff auf `pricing://executive-discounts`
- Der Kontextabruf ist nur bei autorisierten Ressourcen möglich.
- Dieser Zugriff nach dem Least-Privilege-Prinzip verhindert Datenlecks.
- Vor dem Abruf wird jede MCP-Ressourcenanfrage anhand von Sarahs Berechtigungen validiert.

In der Übersicht über die Architektur entspricht dies dem Eintrag „Datensicherheit (FGA)“ in der „Auf Sicherheit ausgelegten Ebene“.

Schritt 3.4: Token-Abruf aus Auth0 Token Vault

Für den Zugriff auf externe Systeme ruft der Agent OAuth2-Token aus dem Token Vault ab:

- Agent: „Ich benötige Zugriff auf Salesforce CRM für Acme Corp-Account-Daten".
- Token Vault validiert die Anforderung auf Basis der zulässigen Integrationen des Agenten.
- Token Vault gibt ein gültiges Access Token für Salesforce mit einem festgelegten Berechtigungsbereich zurück.
- Die Token-Berechtigungen sind auf schreibgeschützten Zugriff auf Kundendaten beschränkt.
- Der Agent nutzt das Token zum Abrufen von CRM-Daten per Salesforce-API.

Funktionen des Token Vaults:

- Sichere Speicherung der Anmeldedaten (keine festkodierten Token)
- Automatische Token-Aktualisierung beim Ablauf von Token
- Audit-Protokollierung aller Token-Zugriffe
- Ausgabe von Token mit beschränktem Berechtigungsbereich (Least-Privilege-Prinzip)

Schritt 3.5: Kontextzusammenstellung

Der Agent stellt den vollständigen Kontext aus autorisierten Quellen zusammen:

Aus CRM-System (per Token Vault → Salesforce):

- Acme Corp-Kontakt: CTO Jennifer Martinez
- Vorherige Käufe: 280.000 USD für KI-Schulungsservices (2024)
- Aktueller Vertrag: Support-Vertrag läuft im März 2026 aus

Aus Dokumentenbibliothek (per MCP):

- Unternehmenseigene Angebotsvorlage (genehmigte Version)
- Produktkatalog mit aktuellen Preisstufen
- Standard-Geschäftsbedingungen

Aus Kalender (per MCP):

- Sarahs Verfügbarkeit für Follow-up-Anrufe
- Vertriebsteam-Kapazität für Implementierungs-Support

NICHT enthalten (Autorisierung verweigert):

- Executive-Rabatte (Sarah hat keinen Zugriff)
- Vertrauliche Informationen aus Verhandlungen zu anderen Deals
- Interne Daten zur Kostenstruktur

Die Agent verfügt nun über umfassenden, autorisierten Kontext zum Generieren des Angebots.

Phase 4

Domainübergreifende Autorisierung mit ID-JAG (Token-Austausch)

Schritt 4.1: Zugriff auf Preisdatenbank des Unternehmens

- Der Agent benötigt Zugriff auf die interne Preisdatenbank: `pricing.acmecorp.internal`
- Dies ist eine separate Autorisierungsdomain von der Haupt-Identity-Plattform.
- Das Preissystem hat einen eigenen Autorisierungsserver, der ID-JAG-Token erfordert.

Schritt 4.2: Token-Austausch per ID-JAG

Der Agent sendet das ID-Token an den Autorisierungsserver für den Token-Austausch:

Anfrage:

```
POST /oauth2/token
Host: acmecorp.okta.com

grant_type=urn:ietf:params:oauth:grant-type:token-exchange
&subject_token=<Sarahs ID-Token>
&subject_token_type=urn:ietf:params:oauth:token-type:id_token
&requested_token_type=urn:ietf:params:oauth:token-type:id-jag
&audience=https://pricing.acmecorp.internal
&scope=pricing:read
```

Diese Schritte werden ausgeführt:

- Der Agent tauscht Sarahs ID-Token gegen ein ID-JAG-Token aus.
- ID-JAG (Identity Assertion JWT Authorization Grant) ist ein kryptografisch signiertes Token.
- Das ID-JAG ist an den Autorisierungsserver der Preisdatenbank adressiert.
- Das ermöglicht domainübergreifende Autorisierung und bewahrt den Benutzerkontext.

Schritt 4.3: Validierung des Autorisierungsservers

Der Autorisierungsserver führt mehrere Validierungsprüfungen durch:

- ID-Token-Validierung: Verifiziert die ID-Token-Signatur und Claims (Vertrauensbeziehung vorab etabliert).
- Überprüfung der verwalteten Verbindung: Validiert die verwaltete Verbindung des Agenten zum Autorisierungsserver der Preisdatenbank.
- Verwaltete Verbindung definiert zulässige Berechtigungsbereiche:
  - ✓ Gewährte Berechtigungen: `pricing:read`
  - ✗ **Verweigerte Berechtigungen:** `pricing:write`, `pricing:admin`

ID-JAG-Token-Ausstellung: Der Autorisierungsserver stellt ein ID-JAG-Token aus, das Sarahs Benutzerkontext bewahrt.

ID-JAG-Token-Claims:


```
{
 "iss": "https://acmecorp.authorization-server.com",
 "sub": "sarah.employee@acmecorp.com",
 "aud": "https://pricing.acmecorp.internal",
 "client_id": "sales-ai-agent",
 "jti": "9e43f81b64a33f20116179",
 "scope": "pricing:read",
 "exp": 1698583800,
 "iat": 1698580200,
 "auth_time": 1698580200,
 "amr": ["pwd", "mfa"]
}
```

Schritt 4.4: Ressourcenzugriff

Der Agent übergibt das ID-JAG Token an den Autorisierungsserver der Preisdatenbank:

- Der Autorisierungsserver der Preisdatenbank validiert die ID-JAG-Signatur mit den veröffentlichten öffentlichen Schlüsseln der Identity-Plattform (JWKS).
- Der Autorisierungsserver der Preisdatenbank überprüft folgende Claims:
  - `aud` **(audience, Zielgruppe): Stimmt der Claim mit der eigenen Aussteller-URL überein?**
  - `exp` **(expiration, Ablaufzeitdatum): Liegt das Ablaufdatum nicht in der Vergangenheit?**
  - `scope` **(Bereich): Ist der Berechtigungsbereich zulässig?**
  - `iss` (issuer, Aussteller): Ist der Aussteller ein vertrauenswürdiger Identity-Anbieter?
    - **Zugriff wird gewährt:** Der Agent ruft Preisdaten des Unternehmens schreibgeschützt ab.
    - Der Agent verfügt nun über autorisierten Zugriff auf die Preisinformationen für die Angebotsgenerierung

Plattform-Funktionen für Token-Austausch

Für domainübergreifende Autorisierungsszenarien unterstützt die Identity-Plattform den Token-Austausch per RFC 8693. Dadurch können KI-Agenten auf Ressourcen unterschiedlicher Autorisierungsserver zugreifen und dabei mithilfe kryptografisch signierter ID-JAG-Token den Benutzerkontext bewahren. Diese Plattformfunktion ist für entwickler- sowie für unternehmensorientierte Bereitstellungen verfügbar.

Phase 5

Asynchrone Autorisierung (Involvierung eines Menschen)

Schritt 5.1: Agent holt bei Bedarf Genehmigung ein

- Der Agent erkennt, dass für die Übermittlung eines Angebots in Höhe von 450.000 USD eine explizite Genehmigung durch einen Benutzer erforderlich ist.
- Dies löst den Workflow für die asynchrone Autorisierung aus.
- Der Agent initiiert eine Anfrage zur asynchronen Autorisierung und pausiert die Ausführung bis zum Erhalt der Genehmigung.

Warum ist asynchrone Autorisierung erforderlich?

- Hochpreisige Angebote übersteigen den Handlungsspielraum des Agenten.
- Die Unternehmensrichtlinien schreiben bei Angeboten ab 100.000 USD eine Benutzerbestätigung vor.
- Gewährleistet Verantwortlichkeit bei geschäftskritischen Entscheidungen.
- Verhindert nicht autorisierter Agentenaktionen.

Schritt 5.2: CIBA-Autorisierungsanfrage

Der Agent startet eine CIBA-Autorisierungsanfrage (Client-Initiated Backchannel Authentication):

Die Anforderung umfasst:

- **Benutzer-ID:** Sarahs Mitarbeiter-ID
- **Erforderliche Berechtigungen:** `email:send`, `drive:write`, `crm:update`
- **Kontext zur Aktion:** Angebotsdetails für die Prüfung durch Sarah
- **Callback-Endpoint:** Legt fest, wohin das Token nach der Genehmigung übermittelt werden soll

```
POST /bc-authorize
Host: acmecorp.authorization-server.com
scope=email:send drive:write crm:update
&login_hint=sarah.employee@acmecorp.com
&binding_message=Proposal Approval:
Acme Corp - $450.000
&client_notification_token=
8d67dc78-7faa-4d41-aabd-67707b374255
```

CIBA-Funktionen:

- Asynchrone Genehmigungs-Workflows (Agent blockiert keine anderen Prozesse)
- Out-of-Band-Benutzerauthentifizierung (Push-Benachrichtigung an Mobilgerät)
- Umfangreicher Kontext für Genehmigungsanforderungen (detaillierte Angebotsinformationen)
- Sicherer Callback-Mechanismus (Token wird nach Freigabe übermittelt)

Schritt 5.3: Push-Benachrichtigung

Der Autorisierungsserver sendet eine Push-Benachrichtigung an Sarahs Guardian-Mobilgeräte-App.

Sie enthält den folgenden Text:

```
Genehmigung des Angebots erforderlich
Kunde: Acme Corp
Betrag: 450.000 USD
Produkte: Enterprise AI Suite und Support
Empfänger: cto@acmecorp.com, cfo@acmecorp.com
Aktion: Angebot per E-Mail senden und in Drive speichern
[Genehmigen] [Verweigern]
```

Umfangreiche Benachrichtigungsfunktionen

- Detaillierter Kontext zu der Aktion, die eine Genehmigung erfordert
- Kundenname, Geldbetrag und inbegriffene Produkte
- Empfängerliste (aus Transparenzgründen)
- Klare Beschreibung der erforderlichen Schritte
- Klar verständliche Genehmigen/Verweigern-Abfrage

Hinweis: Dies ist die umfangreiche Benachrichtigungsfunktion von Guardian und NICHT die OAuth-Spezifikation Rich Authorization Requests (RAR - RFC 9396). Die Benachrichtigung bietet detaillierte Kontextinformationen, anhand derer Sarah eine informierte Genehmigungsentscheidung treffen kann.

Schritt 5.4: Genehmigung durch einen Benutzer

Sarah überprüft die Anfrage und genehmigt sie:

- Sarah überprüft die Details auf ihrem Mobilgerät.
- Sie überprüft folgende Informationen:
  - Kunde ist korrekt (Acme Corp)
  - Betrag stimmt (450.000 USD)
  - Empfänger sind richtig (CTO und CFO)
  - Produkte erfüllen die Kundenanforderungen
- Durch Tippen auf die Schaltfläche „Genehmigen“ genehmigt sie die Anfrage.

Token-Generierung und Übertragung

- Der Autorisierungsserver generiert ein Access Token mit einem festgelegten Berechtigungsbereich mit genehmigten Berechtigungen.
- Das Token enthält nur die von Sarah autorisierten Berechtigungen:

  `email:send`: Berechtigung zum Senden der Angebots-E-Mail

  `drive:write`: Berechtigung zum Speichern des Angebots in Google Drive

  `crm:update`: Berechtigung zum Protokollieren der Aktivität in Salesforce

  - Das Token wird über den CIBA-Callback-Endpoint an den Agenten übermittelt.
  - Der Agent empfängt das Token und setzt die Ausführung mit dem neu ausgestellten Token fort.

Vorteile für die Sicherheit

- Für sensible Aktionen ist eine explizite Genehmigung durch einen Benutzer erforderlich
- Zeitlich begrenztes Token (läuft nach Abschluss der Aktion ab)
- Token mit beschränktem Berechtigungsbereich (nur genehmigte Berechtigungen)
- Vollständiger Audit-Trail (wer hat wann was genehmigt)

Token-Flow für asynchrone Autorisierung (CIBA) – technische Details

Der CIBA-Flow (Client-Initiated Backchannel Authentication) ermöglicht die asynchrone Genehmigung von KI-Agentenaktionen durch einen Benutzer.

Wenn Sarah die Anfrage auf ihrem Mobilgerät genehmigt, passiert Folgendes:

- **Überprüfung der Genehmigung:** Der Autorisierungsserver überprüft, ob die Genehmigungsentscheidung von Sarahs authentifiziertem Gerät kam.
- **Token-Generierung:** Der Autorisierungsserver generiert ein neues Access Token mit Berechtigung für die zulässige Aktion.
- **Festlegung der Berechtigungen:** Das Token enthält nur die von Sarah autorisierten Berechtigungen (d. h. `email:send`, `drive:write`).
- **Sichere Bereitstellung:** Das Token wird über einen sicheren Callback-Endpoint an den Agenten übermittelt, der in der ursprünglichen CIBA-Anfrage spezifiziert wurde.
- **Agentenausführung:** Der Agent empfängt das Token und fährt mit der genehmigten Aktion fort.

Das gewährleistet Human-in-the-Loop-Autorisierung für sensible KI-Agentenaktionen, wobei explizite Genehmigungen durch einen Benutzer erforderlich sind, bevor der Agent aktiv werden kann.

Vorteile von CIBA-Flows:

- **Keine Blockierung:** Während der Agent wartet, muss er keine Verbindung aufrecht erhalten.
- **Benutzerfreundlich:** Sarah genehmigt den Prozess an ihrem Mobilgerät und nicht über einen Chatbot.
- **Sicher:** Token werden über sicheren Callback-Prozess statt über den Benutzerbrowser übermittelt.
- **Auditfähig:** Vollständige Aufzeichnung der Genehmigungsanforderungen, Benutzerentscheidungen und Token-Ausstellungen.
- **Flexibel:** Unterstützt verschiedene Genehmigungsmechanismen (Push-Benachrichtigung, SMS, E-Mail).

Phase 6

Multi-System-Ausführung (Token Vaulting)

Schritt 6.1: Speichern in Google Drive

- Der Agent ruft ein Google Workspace-Token aus dem Auth0 Token Vault ab.
- Das Token ist für den Zugriff auf Sarahs Google Drive berechtigt.
- Der Agent lädt das Angebot in dieses Verzeichnis hoch: `Sales/Proposals/2025/Acme-Corp-Q1.pdf`
- Dateiberechtigungen: Sarahs Teammitglieder und ihr Manager

Schritt 6.2: Versand der E-Mail

- Der Agent ruft ein Gmail-API-Token aus dem Token Vault ab.
- Der Agent erstellt eine E-Mail in Sarahs Account.
  - Empfänger: Acme Corp-CTO und -CFO
  - Text: Begleitschreiben für Geschäftsangebot (generiert per LLM)
  - Anhang: PDF-Datei mit Angebot aus Google Drive
- Die E-Mail wird mit Sarahs Signatur gesendet.

Schritt 6.3: Planung eines Follow-up-Termins

- Der Agent ruft ein Google Kalender-Token aus dem Token Vault ab.
- Der Agent überprüft Sarahs Verfügbarkeit in den nächsten 2 Wochen.
- Der Agent schlägt den Acme Corp-Kontakten mehrere Meeting-Termine vor.
- Der Agent fügt ein Kalenderereignis hinzu: „Vorstellung des Acme Corp-Angebots – 30 Min.“.

Vorteile des Token Vaults:

- Der Agent kommt nie mit den eigentlichen OAuth-Token in Berührung.
- Alle Token werden vor dem Ablaufdatum automatisch aktualisiert.
- Die Anmeldedaten werden niemals im Agentencode oder in Protokollen gespeichert.
- Die Auth0-Anmeldedaten und die Okta-Agenten-Identity bleiben vollständig isoliert.

Phase 7

Governance und Überwachung (Kontrollebene – Governance)

Schritt 7.1: Audit-Trail

Alle Aktivitäten werden im Systemprotokoll der Plattform mit umfangreichen Details protokolliert:

- Agentenauthentifizierungsereignisse
- Token-Austausch-Operationen und Berechtigungsgewährungen
- Ressourcenzugriffsversuche bei allen Systemen
- Autorisierungsentscheidungen (genehmigt/verweigert)
- Benutzerdelegierungsereignisse
- API-Aufrufe im Namen von Benutzern
- Alle Zeitstempel und kontextbezogenen Metadaten für forensische Analysen

Schritt 7.2: Vierteljährliche Zugriffsprüfung

- Governance-Workflow ausgelöst: Zertifizierung des Zugriffs im 1. Quartal 2025
- E-Mail an John Smith (zuständig für den Agenten): „Zugriff für sales-agent-prod-001 überprüfen“
- Die Zugriffsprüfung zeigt folgende Zugriffsrechte des Agenten:
  - Zugriff auf Salesforce CRM
  - Zugriff auf Google Workspace
  - Zugriff auf Preisdatenbank des Unternehmens
  - Zugriff auf E-Mail-System
  - Zugriff auf Kalender-System
- John bestätigt, dass alle Zugriffsrechte weiterhin erforderlich sind.
- Die Zertifizierung wird in einem Okta-Audit-Protokoll erfasst.

Schritt 7.3: Zertifizierung des Zugriffs

- Die vierteljährliche Überprüfung des Zugriffs zeigt, dass der Agent angemessene Berechtigungen für seine Rolle hat.
- Alle Zugriffe sind gerechtfertigt und von der für den Agenten zuständigen Person genehmigt.
- Die Zertifizierung wird zu Compliance-Zwecken im Audit-Protokoll erfasst.

Phase 8

Bedrohungserkennung (Kontrollebene – Überwachung)

Schritt 8.1: Erkannte Anomalie

- **Tag 45:** Der Agent greift plötzlich innerhalb von 10 Minuten auf 500 Kundendatensätze zu.
- Die Verhaltensanalyse entdeckt eine Abweichung vom Basiswert
- Der Risiko-Score schießt in die Höhe: NORMAL → HOCH
- Anomalietyp: „Ungewöhnliche Anzahl an Datenzugriffen“

Schritt 8.2: Automatisierte Reaktionen

- Die Plattform blockiert automatisch den Agentenzugriff auf Salesforce.
- **Globaler Token-Widerruf:** Alle aktiven Token werden sofort für alle Systeme gesperrt.
- Security-Team erhält in Echtzeit eine Warnmeldung.
- Sarah (die Benutzerin) erhält eine Benachrichtigung: „Vertriebsagent temporär gesperrt“
- Vollständige Zugriffssperrung verhindert nicht autorisierte Aktivitäten.

Schritt 8.3: Untersuchung und Behebung

- Das Security-Team überprüft das Systemprotokoll, um das Ausmaß des Vorfalls zu verstehen.
- Die Ursache wird gefunden und behoben.

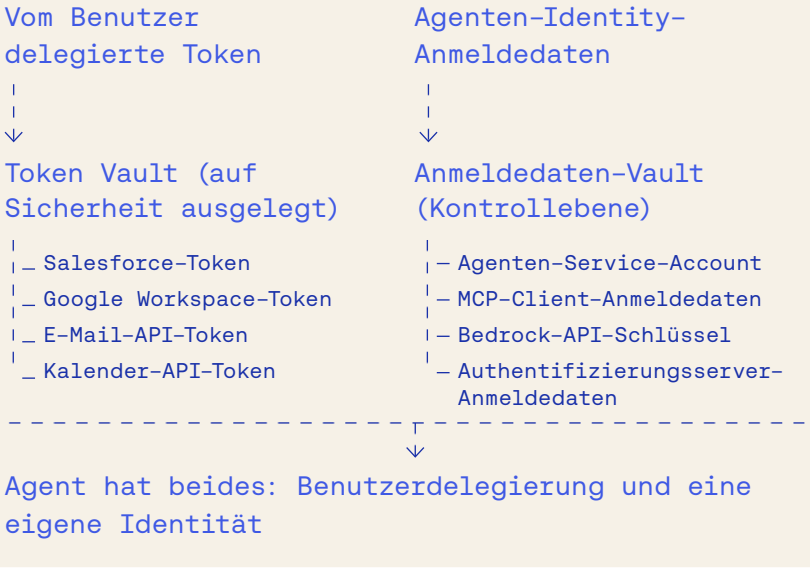
Wichtige Metrik
Der Angriff wurde dank automatisierter Bedrohungserkennung und -abwehr erkannt und blockiert.

Integrations-punkte: Komponenten-vernetzung der einheitlichen Plattform

1. Authentifizierungsablauf

```
Benutzer ---> universelle Authentifizierung --->
ID-Token ---> Plattform validiert Token --->
Plattform validiert Token ---> Kombinierter
Benutzer- und Agentenkontext ---> Zugriff mit
vollständigem Kontext gewährt
```

2. Token-Lebenszyklus



3. Autorisierungsebenen

Ebene 1: Datensicherheit (feingranulare Autorisierung)

Berechtigungen auf Dokumentebene für RAG
„Kann Benutzerin Sarah proposal-acme-2024 aufrufen?“

Ebene 2: Token Vaulting

Berechtigungen auf API-Ebene für SaaS-Tools
„Kann das Token von Benutzerin Sarah auf Salesforce zugreifen?“

Ebene 3: Zugriffskontrolle (Kontrollebene)

Systemweite Berechtigungen für Unternehmensressourcen
„Kann Agent sales-agent-prod-001 auf die Preisdatenbank zugreifen?“

Ebene 4: Token-Austausch (ID-JAG)

Domainübergreifendes Vertrauen mit Benutzerkontext
„Kann Sarah (per Agent) auf das lokale Preissystem zugreifen?“

Ergebnis: Mehrstufiger Schutz mit mehreren Autorisierungs-Checkpoints

4. MCP-Autorisierungsmuster

1. Kontext für Agentenanfragen per MCP-Client
2. MCP-Server erhält Anfrage
3. MCP-Server überprüft Gültigkeit des Agenten-Tokens

Token-Überprüfung ---> Auth0-Autorisierungsdienst

- Validiert Agenten-Identity
- Überprüft Token-Geltungsbereiche
- Überprüft Berechtigungen

4. MCP-Server-Prüfungen: Besitzt der Benutzer Berechtigungen für die Daten?

Berechtigungsprüfung ---> feingranulare Autorisierung

- Überprüft Beziehungs-Tupel
- Gibt nur autorisierte Dokumente zurück

5. MCP-Server gibt autorisierten Kontext an Agenten zurück

Wichtige Architektur-prinzipien

1. Aufgabentrennung

- Auth0 übernimmt die Benutzerauthentifizierung und vom Benutzer delegierte Zugriffe.
- Okta übernimmt Agenten-Identity und Lebenszyklusverwaltung.
- MCP übernimmt Abruf von standardisiertem Kontext.
- Jedes System spielt seine Stärken aus.

2. Mehrstufiger Schutz

- Kein Single Point of Failure dank mehrerer Autorisierungsebenen.
- FGA filtert Dokumente, Token Vault sichert APIs ab, Okta kontrolliert Systeme.
- Selbst wenn eine Ebene überwunden wird, bieten die anderen Schutz.

3. Least-Privilege-Prinzip

- Agenten erhalten die minimal erforderlichen Berechtigungen für jede Aufgabe.
- Token gelten für bestimmte APIs und Aktionen.
- Zeitgebundene Zugriffe mit automatischem Ablaufzeitpunkt.
- Just-in-Time-Provisionierung minimiert Standing-Privilegien.

4. Bewahrung des Benutzerkontexts

- ID-JAG-Token-Austausch etabliert die Benutzer-Identity über mehrere Vertrauensbereiche hinweg.
- Agentenaktionen lassen sich stets zu einem konkreten Benutzer zurückverfolgen.
- Autorisierungsentscheidungen berücksichtigen den Benutzerkontext, nicht nur die Agenten-Identity.
- Audit-Trails zeigen „Agent X“ und „im Namen von Benutzer Y“.

5. Kontinuierliche Überwachung

- Anomalien lassen sich anhand von Basiswerten für das Verhalten entdecken.
- Echtzeit-Bedrohungsreaktion blockt Angriffe.
- Umfassende Protokollierung ermöglicht umfassende forensische Analysen.
- Automatisierte Behebung verkürzt Reaktionszeit.

Architekturvergleich: Klassischer Ansatz und einheitliche Plattform

Funktion	Ohne einheitliche Plattform	Mit einheitlicher Plattform
Agentenerkennung	Manuelle Tabellen ohne Schatten-KI-Erkennung	Automatisierte Erkennung, vollständige Transparenz
Anmeldedaten-Management	Festkodierte Schlüssel im Code ohne Rotation	Vault-Speicherung und Rotation
Benutzerauthentifizierung	Benutzerdefinierter Authentifizierungscode und Passwortspeicherung	Universelle Authentifizierung, Social SSO
API-Zugriff	Gespeicherte Token in Konfigurationsdateien	Token Vaulting mit automatischer Aktualisierung
Domainübergreifender Zugriff	Separate Authentifizierung für jedes System	ID-JAG-Token-Austausch mit Benutzerkontext
Genehmigung durch einen Benutzer	Benutzerdefinierte Abfrage ohne Mobilgeräte-Unterstützung	CIBA mit Mobilgerätebenachrichtigung oder E-Mail
Dokumentberechtigungen	Inkonsistente Prüfungen auf Anwendungsebene	Feingranulare Autorisierung mit beziehungsbasierter Kontrolle
Zugriffsprüfungen	Vierteljährlich manuell mithilfe von Tabellen	Automatisierte Zertifizierungs-Workflows
Bedrohungserkennung	Protokollanalyse nach Vorfällen	Echtzeit-Verhaltensanalysen
Audit-Trail	Über mehrere Systeme verteilt	Einheitliches Systemprotokoll
Reaktion auf Zwischenfälle	Manuelle Untersuchung und Behebung	Automatisierte Blockierung und Token-Widerruf
MCP-Autorisierung	Individuelle Authentifizierungslogik bei jedem MCP-Server	Standardisiertes OAuth2 mit Plattformvalidierung

Diese Architektur demonstriert, wie eine einheitliche Identity-Plattform die KI-Agentensicherheit umfassend angeht. Sie deckt sowohl sichere Entwicklung (Authentifizierung, Token-Management, Autorisierung, Überwachung durch Menschen) sowie unternehmensweite Lebenszyklusverwaltung (Erkennung, Registrierung, Governance, Bedrohungserkennung) ab. Dabei stellt MCP standardisierte Mechanismen für den Kontextabruf zur Verfügung, die durchgehende Identity- und Autorisierungskontrollen berücksichtigen.

Fazit: Eine einheitliche Plattform für vollständige KI-Agentensicherheit

Die KI-Agenten-Revolution ist bereits angekommen. 91 % der Unternehmen nutzen bereits KI-Agenten – und diese Zahl wird weiter wachsen. Die Nutzung hat jedoch Sicherheit und Governance längst überflügelt und führt zu erheblichen Risiken, die den geschäftlichen Mehrwert von KI-Agenten zu untergraben drohen.

Um die Herausforderung zu bewältigen, müssen parallel zwei miteinander verbundene Probleme behoben werden:

Standardmäßige Absicherung aller Agenten während der Entwicklung, sodass ordnungsgemäße Authentifizierung, Autorisierung, Token-Management und Datenzugriffskontrollen von Anfang an integriert werden.

Absicherung aller Agenten über eine gemeinsame Kontrollebene während des gesamten Lebenszyklus mit Erkennung, Provisionierung, Governance und Bedrohungserkennung für den gesamten Agentenbestand.

Unternehmen, die beiden Dimensionen Rechnung tragen, können umfassende Sicherheit erreichen:

- **Sichere Entwicklungsmethoden** mit Authentifizierung, Token Vaulting, Autorisierung und Überwachung durch Menschen
- **Vollständiger Überblick** über alle KI-Agenten, einschließlich Schatten-KI
- **Zuverlässige Lebenszyklusverwaltung** von der Registrierung bis zur Deprovisionierung
- **Umfassende Governance** mit Zugriffsprüfungen und Zertifizierungen
- **Bedrohungserkennung in Echtzeit** mit Verhaltensanalysen und automatisierten Reaktionen
- **Einhaltung von Vorschriften** mit vollständigen Audit-Trails und Richtliniendurchsetzung

Mehr erfahren

Dieser einheitliche Plattformansatz schließt die kritische Lücke, die im „AI at Work 2025“-Bericht erwähnt wurde: Für 85 % der Führungskräfte ist IAM für die sichere KI-Nutzung unverzichtbar, dennoch verfügen nur 10 % über gut ausgearbeitete Strategien zur Verwaltung nicht-menschlicher Identitäten.

Warten Sie mit der Implementierung zuverlässiger KI-Agenten-Sicherheit nicht, bis es zu einer Sicherheitsverletzung oder einem Compliance-Verstoß kommt. Beginnen Sie noch heute mit der Absicherung von Agenten während der Entwicklung und implementieren Sie zentrale Kontrollen für Ihren Agentenbestand.

Mit den Okta-Identity-Lösungen erhalten Sie eine einheitliche Plattform, die beides abdeckt. Unternehmen auf der ganzen Welt nutzen diesen Ansatz bereits zur sicheren Bereitstellung von KI-Agenten und kombinieren dazu Auth0 for GenAI für sichere Entwicklung mit der Okta Identity Platform für die unternehmensweite Lebenszyklusverwaltung.

Sichere Erstellung von KI-Agenten
Dokumentation und QuickStarts für die sichere Agentenentwicklung

Oktas Ansatz zur Absicherung des KI-Agenten-Lebenszyklus
Erfahren Sie, wie Okta unternehmensweite Governance- und Kontrollfunktionen zur Verwaltung von KI-Agenten im großen Maßstab bereitstellt.

Whitepaper

Schutz von KI-Agenten von der Entwicklung bis zum unternehmensweiten Einsatz

okta
The World's Identity Company™

Okta GmbH
Salvatorplatz 3
80333 München, Germany
info_germany@okta.com
+49 (89) 2620 3329