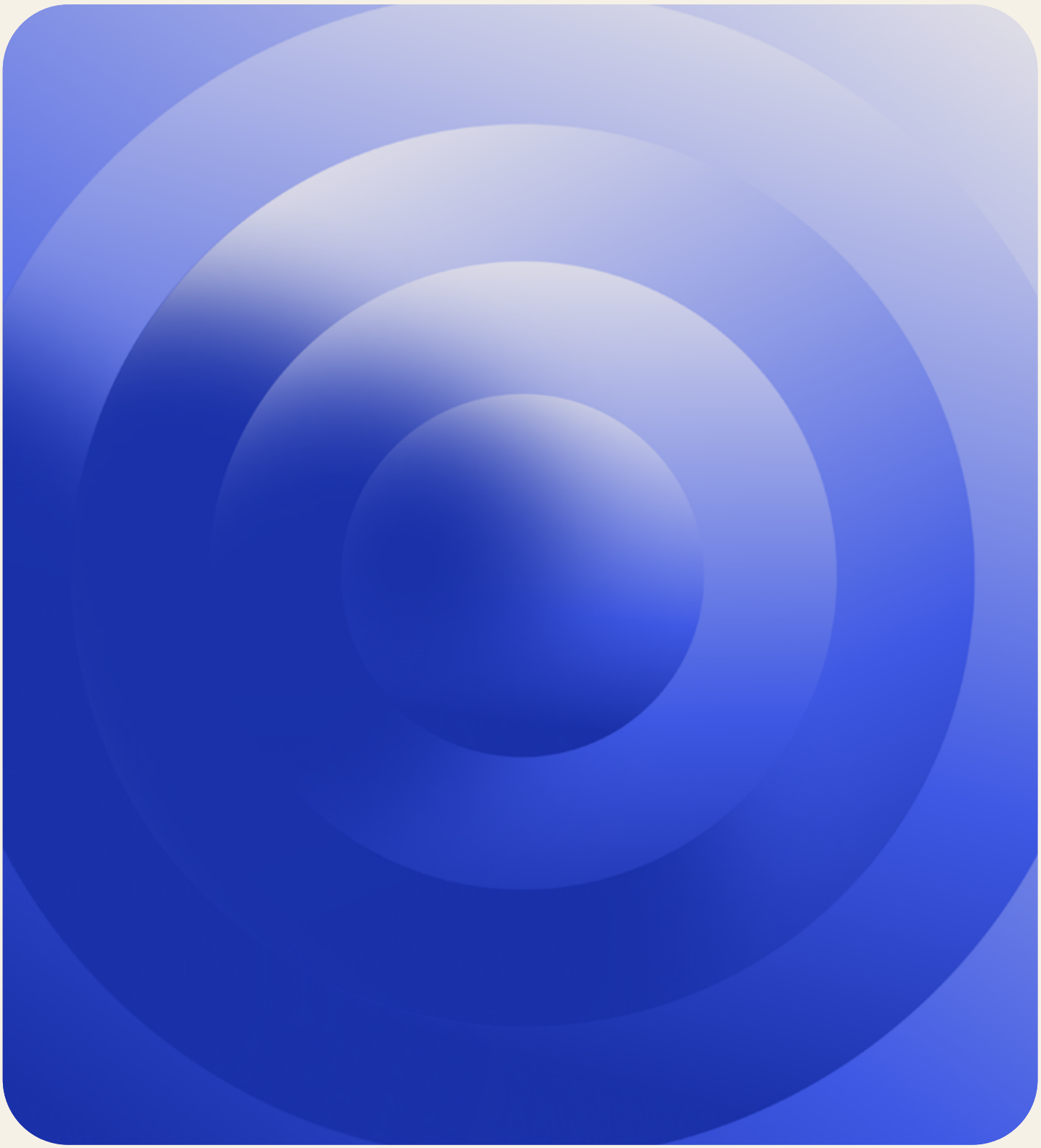
E-Book

Das Zeitalter der KI-Agenten: Ein neuer Ansatz für Identity-Management, Vertrauen und Kontrolle



Inhalt

Die neue Identity-Management-Front

Herkömmliches Identity and Access Management (IAM) ist auf menschliche Geschwindigkeit ausgelegt. KI bewegt sich jedoch mit Maschinentempo.

Wenn Unternehmen von statischen Chatbots zu autonomen Agenten wechseln, benötigen sie grundlegend anderes Identity-Management. Autonome Agenten liefern nicht nur Informationen, sondern führen auch mehrschichtige Aufgaben in komplexen Umgebungen aus. Dadurch entstehen großflächige tote Winkel, in denen unbefugte Aktionen stattfinden können, ohne dass Security-Teams davon etwas mitbekommen.

In diesem E-Book fassen wir die wichtigsten Highlights aus unserer umfassenden 7-teiligen Thought-Leadership-Reihe zur Absicherung KI-gestützter Unternehmen zusammen. Hier erhalten Sie hier tiefere strategische Einblicke in die architektonischen Veränderungen, die für autonome Agenten erforderlich sind.

91 %

aller Unternehmen setzen KI-Agenten bereits in Produktionsumgebungen ein.

10 %

aller Unternehmen verfügen über eine ausgereifte Strategie für die Verwaltung nicht-menschlicher Identitäten.

> 50 %

geben KI-Agenten-Governance und -Compliance als zentrales Anliegen an.

Quelle

IAM für KI-Agenten

Herkömmliche, auf den Menschen ausgerichtete Autorisierungsmodelle können mit KI-Agenten, die mit Maschinengeschwindigkeit bis zu 6.000 Operationen pro Minute ausführen, nicht mithalten. Sicherheit darf nicht mehr auf manuellen, zustimmungsbasierten Genehmigungen basieren, sondern muss während der Laufzeit automatisiert durchgesetzt werden. Nur so lässt sich verhindern, dass „verrückt gewordene“ Agenten in Sekundenschnelle katastrophale Datenverluste verursachen.

Empfehlungen

- **Richtliniengesteuerte Regeln:** Implementieren Sie automatisierte Regeln, die sich an die Geschwindigkeit der Agenten anpassen, damit Sicherheit mit der Ausführungsgeschwindigkeit von Maschinen Schritt halten kann.
- **Vorübergehend gültige Anmeldedaten:** Verwenden Sie Anmeldedaten, die nach wenigen Minuten ablaufen, anstatt dauerhaft zu bestehen. Dadurch lässt sich das Zeitfenster für die Angreifer erheblich begrenzen.
- **Beziehungsbasierter Zugriff:** Ermöglichen Sie innerhalb von Millisekunden Autorisierungsprüfungen, die feingranulare, beziehungsbasierte Zugriffskontrollen anwenden.
- **Kontinuierliche Bewertung:** Geben Sie das Konzept „Berechtigung gewähren und nie wieder prüfen“ auf und bewerten Sie stattdessen jede einzelne Agentenoperation neu.

Kann Ihre aktuelle Sicherheitstechnologie einen nicht autorisierten Agenten, der 100 Befehle pro Sekunde ausführt, schnell genug stoppen, bevor er Ihre Datenbank löscht? Wenn Sie sich bei Agentenaktionen weiterhin auf die Genehmigung durch Menschen verlassen, verlangsamen Sie nicht nur das Innovationstempo, sondern nehmen auch Sicherheitsverletzungen in Maschinengeschwindigkeit in Kauf.

Mehr erfahren: KI-Sicherheit: IAM für KI-Agenten

Autorisierung ohne Ablaufdatum

Von „Autorisierungsdrift“ spricht man, wenn digitale Schlüssel, die für eine bestimmte Aufgabe ausgestellt wurden, auch nach Erledigung der Aufgabe monatelang aktiv bleiben. Im Zeitalter von KI-Agenten sind diese ruhenden Token tickende Zeitbomben, die es Angreifern ermöglichen, legitime SaaS-to-SaaS-Verbindungen zu kapern, ohne je ein Passwort knacken zu müssen.

Empfehlungen

- **Dauerhafte delegierte Identität:** Jeder KI-Agent muss über eine eigene, von Benutzern getrennte Identität verfügen, die kontrolliert, geprüft und nachverfolgt werden kann.
- **Kontinuierlich erneuerbare Autorisierung:** Passen Sie den Zugriff automatisch an, wenn sich die Aufgabe, der Benutzer oder die Umgebung ändert, um Standing-Privilegien zu verhindern.
- **Sofortige systemübergreifende Deprovisionierung:** Erzwingen Sie den Widerruf der Zugriffsrechte in Echtzeit basierend auf gemeinsam genutzten Indikatoren, sodass eine in einer Anwendung widerrufene Berechtigung sofort überall ungültig wird.
- **Absichtsvalidierung in Echtzeit:** Überprüfen Sie jede Aktion anhand der aktuellen Richtlinien in dem Moment, in dem sie ausgeführt wird, und nicht nur zum Zeitpunkt der ursprünglichen Ausstellung der Anmeldedaten.

Verfügt Ihr Unternehmen über einen formalen, automatisierten Prozess, der Agenten-Anmeldedaten in dem Moment widerrufen kann, in dem eine Aufgabe durchgeführt wird? Wenn Ihre Token über ihren eigentlichen Zweck hinaus gültig bleiben, lassen Sie die Hintertür für „stille“ Verstöße weit offen.

Mehr erfahren: Sicherheit von KI-Agenten: Autorisierung ohne Ablaufdatum

Ihr Aktionsplan für KI-Sicherheit und -Compliance

Sind Sie bereit, die Kontrolle über Ihr KI-Agenten-Ökosystem zu übernehmen? Nutzen Sie unsere umfassende Checkliste, um Ihre nächsten Schritte für die Sicherheit und Compliance von KI-Identitäten zu planen.

Checkliste herunterladen

Systemübergreifende Vertrauensstellung von Agenten

Wenn KI-Agenten über die Grenzen des Unternehmens hinaus auf unabhängige Systeme zugreifen, verlieren sie oft ihr „Gedächtnis“ für Einschränkungen. Da die meisten Identity-Anbieter Token isoliert validieren, kann eine Kompromittierung in einer Vertrauensdomain unkontrolliert auf Hunderte andere übergreifen.

Empfehlungen

- **Verifizierbare Delegierung:** Implementieren Sie einen kryptografischen Delegierungsnachweis, der explizit zwischen menschlichen und agentenbasierten Identitäten unterscheidet, wenn diese sich zwischen verschiedenen Systemen bewegen.
- **Portable Einschränkungen:** Stellen Sie sicher, dass Sicherheitseinschränkungen (wie „nur Leserechte“) zusammen mit dem Token über Vertrauensbereiche hinweg übertragen und bei einer Übergabe nicht entfernt werden.
- **Koordinierter Widerruf:** Führen Sie föderierte Risikosignalisierung (wie IPSIE) ein, damit Sicherheitswarnungen in Echtzeit zwischen verschiedenen Service Providern ausgetauscht werden.
- **Kontrolle zum Abrufzeitpunkt:** Nutzen Sie feingranulare Autorisierung, um den Zugriff genau in dem Moment zu validieren, in dem ein Agent eine API aufruft, unabhängig davon, aus welcher Domain die Anfrage stammt.

Wer bürgt für die Sicherheit, wenn ein Agent auf das System eines Partners zugreift? Wenn Sie über ein statisches und dezentralisiertes Vertrauensmodell verfügen, fehlt ein gemeinsamer Schutz vor domainübergreifendem Token-Hijacking.

Mehr erfahren: KI-Sicherheit: Systemübergreifende Vertrauensstellung von Agenten

Sicheres Delegieren von KI-Agenten

Durch rekursive Delegierung, bei der Agenten Unteragenten für die Durchführung spezieller Aufgaben generieren, entstehen verschachtelte Sicherheitsrisiken. Ohne strikte Nachverfolgung der Delegierungsherkunft und Einschränkung des Berechtigungsbereichs kann ein einzelner schädlicher Prompt zu Agent Session Smuggling führen. Bei dieser Schwachstelle kann ein Unteragent unbemerkt nicht autorisierte Aktionen ausführen.

Empfehlungen

- **Out-of-Band-Verifizierung:** Nutzen Sie für Operationen mit potenziell schwerwiegenden Konsequenzen Push-Benachrichtigungen oder separate UIs, um den primären Chat-Kanal des Agenten zu umgehen.
- **Kontextverankerung:** Verankern Sie jede Agenten-Session mit der ursprünglichen Aufgabe und kennzeichnen Sie fortlaufend „semantische Abweichungen“, wenn das Verhalten des Agenten beginnt, vom beabsichtigten Ziel abzuweichen.
- **Verifizierte Identität und Fähigkeiten:** Setzen Sie eine Zero-Trust-Architektur durch, in der Agenten vor Interaktionen mit einer Systemressource als Nachweis für ihre Identität und spezifischen Berechtigungen kryptografische Anmeldedaten übermitteln müssen.
- **Benutzersichtbarkeit:** Minimieren Sie das Risiko eingeschleuster Anweisungen, indem Sie radikale Transparenz bieten und dem Benutzer in Echtzeit alle Tool-Aufrufe, Hintergrundlogiken und Ausführungsprotokolle anzeigen.

Können Sie die kryptografische Herkunft jeder Aktion beweisen, die von einem Unteragenten drei Schritte tief in einem Workflow ausgeführt wird? Ohne überprüfbare Delegierungsketten ist Ihr Multi-Agenten-Ökosystem leichte Beute für laterale Bewegungen.

Mehr erfahren: Agentensicherheit: Delegierungskette

Schließen der Autorisierungs-lücke

KI-Agenten rufen Daten oft mit den umfassenden Berechtigungen einer Führungskraft ab, stellen diese dann aber in gemeinsam genutzten Arbeitsbereichen wie Slack oder Teams bereit. Aufgrund dieser „Autorisierungslücke“ kann es dazu kommen, dass sensible Daten wie Vorstandsgehälter oder Sitzungsunterlagen unbeabsichtigt gegenüber unbefugten Empfängern offengelegt werden.

Empfehlungen

- **Zielgruppenorientierte Autorisierung:** Berechnen Sie die „Schnittmenge der Berechtigungen“ in Echtzeit, sodass ein Agent nur solche Daten abruft, die alle Benutzer im aktuellen Arbeitsbereich sehen dürfen.
- **Berechtigungsbereich-bezogene Abfrage:** Statt Daten nach dem Abrufen zu filtern, bestimmen Sie vor dem Abrufen den Gültigkeitsbereich der Anmeldedaten des Agenten, damit er gar nicht erst unbefugt auf sensible Dateien zugreifen kann.
- **Beziehungsbasierter Zugriff:** Wechseln Sie von statischen Rollen zu einem beziehungsbasierten Modell, das versteht, wer sich in welchem Kanal befindet und welche Berechtigungen aktuell bestehen.
- **Kontinuierliche Richtliniensynchronisierung:** Integrieren Sie Identity Governance mit der Autorisierungs-Engine, damit das Berechtigungsdiagramm auch dann eingehalten wird, wenn Benutzer gemeinsam genutzten Arbeitsbereichen beitreten oder diese verlassen.

Können Sie für jeden in einem gemeinsam genutzten Arbeitsbereich eingesetzten KI-Agenten nachweisen, dass seine Ausgabe auf den kleinsten gemeinsamen Nenner der Berechtigungen in der Umgebung beschränkt ist? Wenn Ihre Agenten die Zielgruppe ignorieren, werden sie zu Ihrem größten internen Datenleckrisiko.

Mehr erfahren: Autorisierungslücke bei KI-Agenten

Cyberphysische Sicherheit

Wenn KI-Agenten in physischen Branchen wie Gesundheitswesen oder Fertigung eingesetzt werden, entstehen aus Autorisierungsfehlern keine Datenlecks, sondern Sicherheitsrisiken, sodass physischer Schaden durch digitale Agenten verhindert werden muss.

Empfehlungen

- **End-to-End-Nachverfolgbarkeit:** Implementieren Sie Cross-App Access (XAA) mit Delegierungs-Token, um jede automatisierte Aktion sowohl einem bestimmten Agenten als auch dem ursprünglichen Benutzer zuzuordnen.
- **On-Demand-Anmeldedaten:** Verwenden Sie einen Token Vault, um statische, langlebige Anmeldedaten durch Token mit definiertem Berechtigungsbereich und kurzer Lebensdauer zu ersetzen, die erst zum Zeitpunkt der Ausführung abgerufen werden.
- **Human-in-the-Loop-Verifizierung:** Nutzen Sie Client-Initiated Backchannel Authentication (CIBA), um eine explizite menschliche Genehmigung anzufordern, wenn ein Agent versucht, Operationen mit schwerwiegenden Konsequenzen oder außergewöhnliche Operationen durchzuführen.
- **Richtliniendurchsetzung in Echtzeit:** Nutzen Sie feingranulare Autorisierung, die Sicherheitsbeschränkungen und betriebliche Grenzen zum Zeitpunkt der Entscheidung auswertet, anstatt sich auf statische Rollen zu verlassen.

Kann Ihr Team den Autorisierungsbereich und die aktiven Anmeldedaten für jeden Agenten, der auf kritische Systeme zugreift, klar definieren? Wenn Sie nicht genau feststellen können, was ein Agent tun darf und warum, wird diese fehlende Transparenz zu Ihrer primären Angriffsfläche.

Mehr erfahren: KI-Agenten: Cyberphysische IAM-Sicherheit

Zentrales Identity-Management und einheitliche Autorisierung für autonomes Vertrauen

Da KI-Agenten autonom agieren, können herkömmliche Identity-Management- und Autorisierungsmodelle ihren Zweck nicht mehr erfüllen. Systeme, die für die Gewährung von Zugriff für menschliche Benutzer ausgelegt sind, können Delegierung, indirekte Aktionen oder maschinengesteuerte Entscheidungsfindung nicht zuverlässig kontrollieren.
Für autonomes Vertrauen ist es notwendig, dass Identity-Management und Autorisierung als einheitliche Kontrollebene behandelt werden, die definiert, was Agenten tun dürfen und innerhalb welcher Grenzen.

Empfehlungen

- **Agenten mit eigener Identität:** Behandeln Sie KI-Agenten als eigenständige Identitäten und nicht als Erweiterungen von Benutzern oder Anwendungen.
- **Vereinheitlichte Identity- und Autorisierungsentscheidungen:** Vermeiden Sie fragmentierte Kontrollen, indem Sie konsistente Zugriffsentscheidungen für alle Agentenaktionen durchsetzen.
- **Explizite Berücksichtigung von Delegierungen:** Konzipieren Sie Autorisierungsmodelle, die erkennen, wenn Agenten im Namen von Benutzern, Systemen oder anderen Agenten handeln.
- **Reduzierung zu weit gefasster Zugriffe:** Beschränken Sie dauerhafte Berechtigungen, um zu verhindern, dass Agenten über ihren vorgesehenen Berechtigungsbereich hinaus agieren.
- **Definition von Identity-basierter Grenzen:** Nutzen Sie Identity-Management und Autorisierung als Mechanismen zum Einschränken von Agentenverhalten und nicht nur zur Beobachtung.

Wenn Identity-Management und Autorisierung als einheitliches System agieren, wird autonomes Vertrauen durchsetzbar, wenn Agenten handeln und delegieren.

Mehr erfahren: Identity-Management und Autorisierung als Basis für autonomes Vertrauen

Wie Okta Sie unterstützen kann

Okta setzt den Identity Security Fabric als einheitliche Kontrollebene um, die die Lücke zwischen menschlicher Absicht und maschineller Ausführungsgeschwindigkeit schließt. Unternehmen erhalten eine zentrale Ebene für die richtlinienbasierte Echtzeit-Durchsetzung, sodass jede Agentenaktion über jede Vertrauensdomain hinweg kontrollierbar, nachverfolgbar und sicher ist. Da das Identity-Management als Basis für die Infrastruktur dient, können Unternehmen beim Skalieren von KI-Agenten darauf vertrauen, dass Sicherheit und Autorisierung fest in jeden automatisierten Workflow integriert ist.