

• WHITE PAPER

KI-Agenten mit Zugriff, aber ohne Identität

KI-Agenten, Servicekonten und technische Benutzer in Banken und Versicherern: Eigentümerschaft, Laufzeitkontrolle und was DORA verlangt

AUSGABE

Oktober 2026

Fassung 2.3

AUTOR

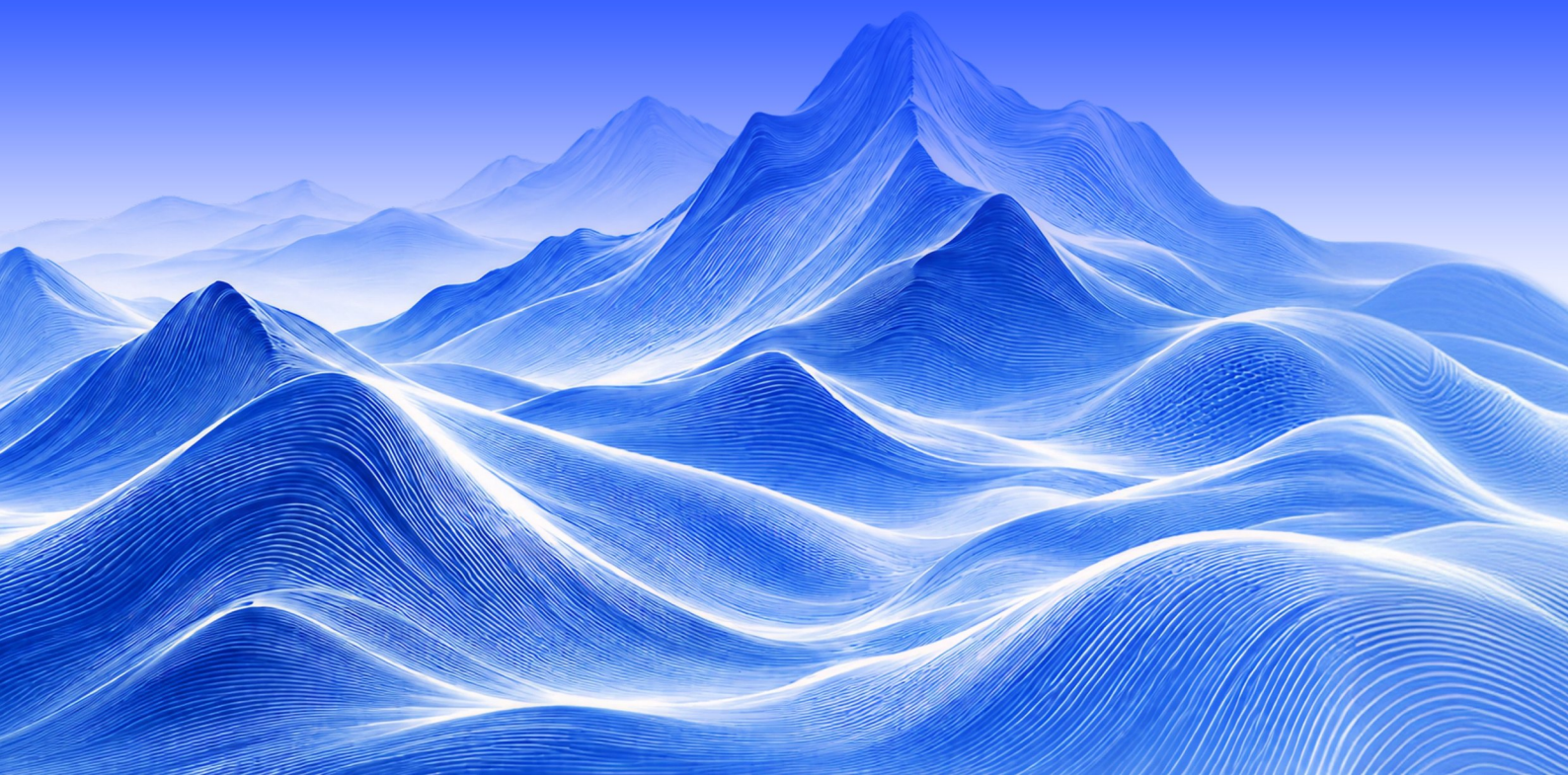
Parsa Jandaghi

CISSP

ZIELGRUPPE

CISO und IT-Leitung

IAM-Architektur



Inhalt

	Das Wichtigste in Kürze	3
1	Nicht-menschliche Identitäten als Kontrolllücke	5
2	Die Anforderungen der DORA-RTS	7
3	Ein KI-Agent ist kein weiteres Servicekonto	11
4	Das Kontrollmodell im Überblick	14
5	Identität und Stellvertretung	17
6	Vollmacht und Laufzeitkontrolle	21
7	Verantwortung und Lebenszyklus	25
8	Microsoft Entra und SailPoint im September 2026	31
9	Vorgehen in fünf Phasen	35
	Glossar	38
	Quellen	41

Das Wichtigste in Kürze

Servicekonten, Arbeitslasten und seit kurzem KI-Agenten greifen mit ihren Schlüsseln, Zertifikaten und Tokens auf dieselben Systeme zu wie Mitarbeitende. Oft haben sie weitergehende Rechte, und oft kann niemand sagen, wer für sie verantwortlich ist.

Die KI-Agenten sind dabei die neue Klasse, und sie passen in keines der bestehenden Verfahren. Ein Servicekonto tut, was bei seiner Anlage festgelegt wurde. Ein Agent bekommt ein Ziel und entscheidet selbst, welche Werkzeuge er dafür aufruft. Ein Berechtigungskonzept nach dem Muster der Servicekonten beschreibt für ihn etwas, das es nicht gibt.

Die BaFin hat im Juli 2026 ihre Auswertung der DORA-Vorfallmeldungen für 2025 veröffentlicht. 733 Vorfälle wurden gemeldet, rund 1.200 Finanzunternehmen waren betroffen. Bei der Hälfte lag die Ursache bei einem Dritten, und rund 40 Prozent der gemeldeten Angriffe richteten sich gegen einen Dienstleister statt gegen das Institut selbst.¹

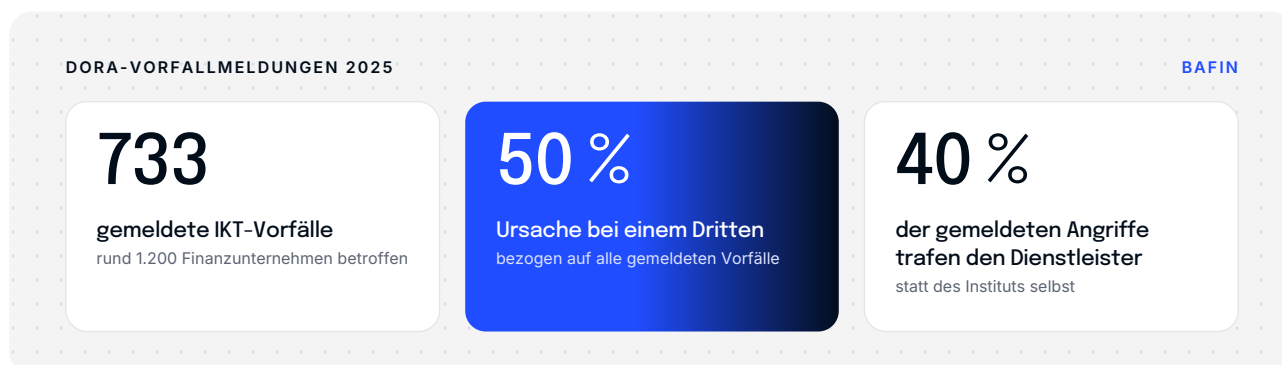


Abbildung 1 BaFin-Auswertung der DORA-Vorfallmeldungen für 2025 · Quelle: BaFin, Fachartikel vom 16. Juli 2026

Die EZB-Bankenaufsicht zählt Identity and Access Management zu den Bereichen, die bei einem Teil der beaufsichtigten Banken unzureichend entwickelt sind.² Am 7. Juli 2026 hat sie alle bedeutenden Institute angeschrieben und bis zum 31. Oktober 2026 einen Aktionsplan gegen KI-getriebene Angriffe verlangt. Least Privilege, Protokollierung und menschliche Aufsicht nennt sie als Schwerpunkte.³

Dieses Papier beschreibt zwei Lücken. Die erste liegt im Regelwerk. Der technische Regulierungsstandard zu DORA erfasst Maschinen, aber er erfasst sie ungleich: Die eindeutige Identifizierung verlangt er für Personen und Systeme, die Pflicht zu einem eindeutigen Nutzerkonto je Identität dagegen nur für Mitarbeitende.⁴ Die zweite Lücke liegt in der Technik. Eine Anmeldung bestätigt,

wer handelt. Ob die nächste Buchung eines Agenten zulässig ist, sagt sie nicht, und genau diese Entscheidung fällt bei einem Agenten bei jeder einzelnen Aktion.

WertFels begegnet beiden Lücken mit einem Kontrollmodell aus sechs Ebenen. Zwei davon wirken vor dem Einsatz eines Agenten, drei bei jeder Aktion und eine laufend und in festen Abständen. Jede Ebene ist einer Fundstelle in den DORA-RTS zugeordnet, und jede lässt sich mit heute verfügbaren Plattformen und Standards umsetzen, zum Teil mit Eigenbau.

Dieses Papier belegt die folgenden Aussagen:

1. Die DORA-RTS verlangen im Regelrahmen, Zugangsrechte mindestens jährlich zu aktualisieren, bei Systemen kritischer oder wichtiger Funktionen halbjährlich. Eine Ausnahme für technische Konten enthält der Text nicht.⁵
2. Generische und gemeinsam genutzte Konten sind so weit wie möglich einzuschränken, und jede Handlung muss einem Nutzer zuordenbar bleiben. Ein geteiltes Servicekonto, dessen Aktionen im Protokoll keinem Nutzer zuzuordnen sind, verfehlt diese Anforderung.
3. In den geprüften Texten von EZB, EBA, EIOPA, ESMA und BaFin kommt der Begriff Maschinenidentität bisher nicht vor. Die Anforderungen entstehen durch Auslegung allgemeiner Formulierungen.
4. Ein KI-Agent ist kein weiteres Servicekonto, weil seine Aktionsmenge erst zur Laufzeit feststeht. Das NIST NCCoE listet Least Privilege für Agenten unter den offenen Fragen.⁶
5. Für Agenten braucht es eine Entscheidung je Aktion an vier Stellen: beim Auftrag, bei den Daten, beim Werkzeug und beim Ergebnis. Die Regeln dafür gehören an eine zentrale Stelle, die Durchsetzung nah an das Werkzeug.
6. Ob ein Agent unter eigener Identität mit mitgeführter Vollmacht handelt oder unter der Identität eines Menschen, entscheidet über die Nachvollziehbarkeit. Der Standard dafür existiert seit 2020.⁷
7. Jeder Agent braucht genau einen menschlichen Eigentümer, eine Vertretung und eine Verankerung an einer Anwendung. Geht der Eigentümer, wird der Agent entlang einer Nachfolgekette übertragen, die nach oben eskaliert, wenn niemand übernimmt. Der Weggang des Eigentümers ist kein Grund, den Agenten anzuhalten oder stillzulegen.
8. Microsoft Entra führt Agenten seit 2026 als eigene Objektklasse mit verpflichtendem Sponsor. SailPoint hat Agentensicherheit für die Identity Security Cloud gebaut. Wer SailPoint IdentityIQ betreibt, muss Befristung, Verhaltensabgleich, Eskalation und kurzlebige Zugangsnachweise selbst ergänzen.
9. Die Rezertifizierung eines Agenten prüft seine Vollmacht: ob er gebraucht wird, ob sein Katalog dem tatsächlichen Verhalten entspricht und ob sein Anhalten geübt wurde.

Nicht-menschliche Identitäten als Kontrolllücke

Viele Angriffe auf Finanzinstitute beginnen heute mit einem gültigen Zugangsnachweis. Bei nicht-menschlichen Konten fehlt fast jede Kontrolle, die bei Menschen längst selbstverständlich ist.

Die Hälfte der gemeldeten Vorfälle hat ihre Ursache bei Dritten

Die BaFin-Auswertung für 2025 umfasst alle gemeldeten IKT-Vorfälle, auch Ausfälle ohne Angriff. Sie zeigt trotzdem ein Muster, das mit der gewohnten Vorstellung von Angriffen schlecht zusammenpasst. Zuerst getroffen wird häufig ein Dienstleister, der Zugang zu den Systemen des Instituts hat. Nach unserer Beobachtung läuft dieser Zugang selten über einen persönlichen Benutzernamen. Er läuft über eine Schnittstelle, ein Zertifikat, einen Schlüssel oder ein Token.

Ein gültiges Token braucht keine Anmeldung

Zwischen dem 8. und 18. August 2025 nutzte eine Angreifergruppe gestohlene OAuth-Tokens der Vertriebsanwendung Salesloft Drift, um Daten aus den Salesforce-Umgebungen zahlreicher Unternehmen abzuziehen. Laut Google Threat Intelligence waren die Tokens kompromittiert, aber gültig. Damit war keine Anmeldung nötig und keine Mehrfaktorabfrage, die hätte auffallen können. In den abgezogenen Daten suchten die Angreifer gezielt nach weiteren Zugangsdaten, nach Schlüsseln für Cloudkonten, Kennwörtern und Datenbanktokens.⁸ Die amerikanische Selbstregulierungsorganisation FINRA hielt den Vorfall für so relevant, dass sie ihre Mitgliedsfirmen eigens warnte.⁹



Abbildung 2 Ablauf des Salesloft-Drift-Vorfalles, August 2025 · Darstellung WertFels nach Google Threat Intelligence

Für die Bewertung zählt, was der Fall nicht zeigt. Bei den betroffenen Salesforce-Kunden wurde keine Schwachstelle ausgenutzt und kein Kennwort erraten. Ein gültiger maschineller Zugangsnachweis wurde an einer Stelle angewendet und an anderer Stelle eingesetzt.

Zugangsdaten sind eine Handelsware

Das BSI beschreibt in seinem Lagebericht 2025 den Markt der Access Broker, die legitime Zugangsdaten zu Firmennetzen sammeln und an andere Angreifer weiterverkaufen. Im Berichtszeitraum wurden 461 Datenleaks bekannt.¹⁰ Die ENISA berichtet für 2023 bis 2024 von einem Zuwachs von 25 Prozent bei Schlüsseln, Tokens und anderen Geheimnissen, die in Coderepositorys entdeckt wurden.¹¹

In der Bitkom-Erhebung zum Wirtschaftsschutz 2026 nennen geschädigte Unternehmen als häufigste Ursachen die unzureichende Erkennung von Sicherheitsvorfällen mit 59 Prozent, Fehlkonfigurationen mit 57 Prozent und unzureichendes Identitäts- und Zugriffsmanagement mit 55 Prozent. Technische Schwachstellen folgen mit 50 Prozent erst danach. Befragt wurden 1.003 Unternehmen ab zehn Beschäftigten und einer Million Euro Jahresumsatz.¹²

Agenten vergrößern die Reichweite jedes Zugangs

Mit KI-Agenten kommt eine Klasse hinzu, die einen gültigen Zugang nutzt und dabei selbst entscheidet, wofür. Gartner erwartet, dass bis 2028 jede vierte Sicherheitsverletzung in Unternehmen auf den Missbrauch von KI-Agenten zurückgeht, durch externe Angreifer ebenso wie durch Innentäter.¹³ Die Risikoliste des OWASP GenAI Security Project für agentische Anwendungen nennt ausdrücklich den Missbrauch von Identitäten, Rechten und Werkzeugen sowie die Übernahme des Ziels.¹⁴

Die Anforderungen der DORA-RTS

Der technische Regulierungsstandard zu DORA erfasst technische Konten an mehr Stellen, als viele Institute annehmen. Eine Stelle nimmt sie aus, und genau dort beginnen die Diskussionen in Prüfungen.

Dieses Kapitel bezieht sich auf die Delegierte Verordnung (EU) 2024/1774, den technischen Regulierungsstandard zum IKT-Risikomanagement nach DORA. Die Artikel 20 und 21 gelten für Finanzunternehmen im Regelrahmen. Für kleine und nicht verflochtene Wertpapierfirmen und einige weitere Gruppen gilt der vereinfachte Rahmen in Titel III. Dort ist Artikel 33 die Norm zur Zugangskontrolle.

Artikel 20 erfasst Systeme, eindeutige Konten verlangt er nur für Mitarbeitende

Absatz 1 verlangt Richtlinien und Verfahren, die die eindeutige Identifizierung und Authentifizierung der natürlichen Personen und Systeme sicherstellen, die auf Informationen des Finanzunternehmens zugreifen. Absatz 2 Buchstabe a verengt dann: Jedem Mitarbeiter des Finanzunternehmens und jedem Mitarbeiter eines IKT-Drittdienstleisters ist eine eindeutige Identität zuzuweisen, die einem eindeutigen Nutzerkonto entspricht. Der Buchstabe steht unter dem Vorbehalt, dass Artikel 21 Buchstabe c unberührt bleibt. Die Pflicht, über diese Zuweisungen Aufzeichnungen zu führen und sie auch nach Umstrukturierungen oder Vertragsende aufzubewahren, bezieht sich im selben Absatz ausdrücklich auf Buchstabe a.⁴

Hier liegt der Kern des Problems. Ein Servicekonto, das mehrere Teams gemeinsam nutzen, verstößt nicht gegen Absatz 2 Buchstabe a, weil dort nur von Mitarbeitenden die Rede ist. Einen benannten Eigentümer kann ein Prüfer aber über Artikel 21 Buchstabe e Ziffer i verlangen, dazu weiter unten mehr.

Der Lebenszyklus gilt für alle Konten

Artikel 20 Absatz 2 Buchstabe b verlangt einen Prozess für den gesamten Lebenszyklus von Identitäten und Konten: Erstellung, Änderung, Überprüfung und Aktualisierung, vorübergehende Deaktivierung und Beendigung. Der Text spricht von allen Konten. Wo es angemessen und möglich ist, soll dieser Prozess automatisiert laufen.

Rechte nach Erforderlichkeit, generische Konten nur mit Zurechenbarkeit

Artikel 21 Buchstabe a verlangt, Zugangsrechte nach Erforderlichkeit und nach dem Grundsatz der minimalen Berechtigung zu vergeben, ausdrücklich auch für Fern- und Notfallzugänge. Buchstabe c verlangt, generische und gemeinsam genutzte Nutzerkonten so weit wie möglich einzuschränken und sicherzustellen, dass die Handlungen in den IKT-Systemen jederzeit einem Nutzer zugeordnet werden können.⁵ Buchstabe d verlangt, den Zugang zu IKT-Assets durch Kontrollen und Werkzeuge gegen unbefugten Zugriff zu beschränken. Das ist die Grundlage der Durchsetzung zur Laufzeit.

Für Servicekonten ist Buchstabe c die praktisch wichtigste Stelle. Generische Konten sind zulässig, aber nur unter der Bedingung erhaltener Zurechenbarkeit. Ein Konto, das drei Teams gemeinsam nutzen und dessen Aktionen im Protokoll nicht unterscheidbar sind, erfüllt diese Bedingung nicht.

Aktualisierung jährlich, bei kritischen Funktionen halbjährlich

Buchstabe e legt die Verfahren zur Kontoverwaltung fest und nennt dabei Nutzer- und generische Konten ausdrücklich. Ziffer i verlangt, Rollen und Verantwortlichkeiten für Vergabe, Überprüfung und Entzug von Zugangsrechten zuzuweisen. Das ist die naheliegende Norm für einen benannten Eigentümer, auch bei technischen Konten. Ziffer iii verlangt den umgehenden Entzug von Zugangsrechten, wenn sie nicht mehr notwendig sind. Ziffer iv verlangt, Zugangsrechte zu aktualisieren, wenn Änderungen notwendig sind, mindestens aber einmal jährlich für alle IKT-Systeme und mindestens alle sechs Monate für Systeme, die kritische oder wichtige Funktionen unterstützen. In der Praxis setzen Institute diese Pflicht über die Rezertifizierung um. Eine Unterscheidung nach Kontoart enthält der Text an dieser Stelle nicht.

Die Protokollierung nennt Zugangskontrolle ausdrücklich

Artikel 12 Absatz 2 Buchstabe c Ziffer i verlangt die Aufzeichnung von Ereignissen der logischen und physischen Zugangskontrolle nach Artikel 21 und des Identitätsmanagements. Buchstabe d verlangt, die Aufzeichnungssysteme selbst vor Manipulation, Löschung und unbefugtem Zugriff zu schützen. Buchstabe f verlangt, die Uhren aller IKT-Systeme gegen eine dokumentierte, verlässliche Referenzzeitquelle zu synchronisieren.¹⁵

Ohne synchrone Zeitstempel lässt sich eine Handlungskette über mehrere Systeme hinweg nicht rekonstruieren. Bei Maschinenkonten, die in Millisekunden agieren, ist das keine Formalie.

DORA-RTS		DELEGIERTE VERORDNUNG (EU) 2024/1774	
ANFORDERUNG	FUNDSTELLE	Mitarbeitende	System, Agent
Eindeutige Identifizierung und Authentifizierung	Art. 20 Abs. 1	✓	✓
Eindeutiges Nutzerkonto je Identität	Art. 20 Abs. 2 Buchst. a	✓	⊖
Lebenszyklus aller Konten	Art. 20 Abs. 2 Buchst. b	✓	✓
Rechte nach Erforderlichkeit und minimaler Berechtigung	Art. 21 Buchst. a	✓	✓
Generische Konten einschränken, Zurechenbarkeit	Art. 21 Buchst. c	✓	✓
Aktualisierung der Rechte, jährlich, kritisch halbjährlich	Art. 21 Buchst. e Ziff. iv	✓	✓
Protokollierung von Zugangskontrolle und Identitätsmanagement	Art. 12 Abs. 2 Buchst. c	✓	✓
⊖ im Verordnungstext nur für Mitarbeitende formuliert		✓ gilt dem Wortlaut nach ohne Unterscheidung nach Kontoart	

Abbildung 3 Anforderungen der DORA-RTS nach Identitätsart · Darstellung WertFels nach Delegierter Verordnung (EU) 2024/1774

Für die Prüfung

Wer ein Servicekonto nicht rezertifiziert, weil es kein Mensch ist, findet dafür im Verordnungstext kein Argument. Artikel 21 Buchstabe e nennt Nutzer- und generische Konten ausdrücklich. Zwischen Mitarbeitenden und anderen Identitäten unterscheiden nur die Pflicht zu einem eindeutigen Nutzerkonto je Identität und die Aufzeichnung dieser Zuweisungen.

Lücken im Verordnungstext

Die Begriffe Maschinenidentität, Servicekonto, technischer Benutzer und Non-Human Identity kommen in der Delegierten Verordnung (EU) 2024/1774 an keiner Stelle vor. Auch in den von uns geprüften Veröffentlichungen von EZB, EBA, EIOPA, ESMA und BaFin kommen sie nicht vor. Die früheren BaFin-Rundschreiben BAIT und VAIT enthielten Vorgaben zu technischen Nutzern. Für Unternehmen unter DORA gelten sie seit dem 17. Januar 2025 nicht mehr.¹⁶

Darin liegt die eigentliche Schwierigkeit in Prüfungen. Es gibt keine Norm, die technische Konten beim Namen nennt. Die Pflichten ergeben sich aus allgemein formulierten Artikeln, und beide Seiten argumentieren über diese Formulierungen.

Weitere Anforderungen

Für Hochrisiko-KI-Systeme verlangt die KI-Verordnung eine automatische Protokollierung, eine wirksame menschliche Aufsicht und angemessene Cybersicherheit (Artikel 12, 14 und 15). Artikel 14 bemisst die Aufsichtsmaßnahmen ausdrücklich am Risiko, am Grad der Autonomie und am Einsatzkontext. Soweit angemessen und verhältnismäßig, muss die beaufsichtigende Person in den Betrieb eingreifen und das System über eine Stopptaste oder ein vergleichbares Verfahren anhalten können.¹⁷ Diese drei Artikel richten sich an die Gestaltung des Systems. Als Betreiber eines Systems, ob selbst gebaut oder eingekauft, ist ein Institut vor allem an Artikel 26 (menschliche Aufsicht durch befähigte Personen, Aufbewahrung der Protokolle) und Artikel 27 (Grundrechte-Folgenabschätzung) gebunden. Für Banken und Versicherer liegen die naheliegenden Fälle in Anhang III: die Kreditwürdigkeitsprüfung natürlicher Personen sowie Risikobewertung und Preisbildung in der Lebens- und Krankenversicherung. Nach der Verschiebung durch den Digital Omnibus gelten diese Pflichten ab dem 2. Dezember 2027.¹⁸

Für bedeutende Institute kommt der Aktionsplan hinzu, den die EZB-Bankenaufsicht bis zum 31. Oktober 2026 verlangt. Das Schreiben nennt Least Privilege, Protokollierung und menschliche Aufsicht, also Kontrollen, die dieses Papier für Agenten konkretisiert. Es richtet sich nur an bedeutende Institute. Versicherer betrifft es nicht, für sie gelten die Anforderungen aus DORA unverändert.

Ein KI-Agent ist kein weiteres Servicekonto

Das gesamte Berechtigungswesen beruht auf der Annahme, dass man bei der Vergabe weiß, was ein Konto tun wird. Für KI-Agenten trifft diese Annahme nicht mehr zu.

Die Aktionsmenge steht erst zur Laufzeit fest

Ein klassisches Servicekonto hat eine Eigenschaft, auf der alles aufbaut: Man kennt bei der Vergabe seine Aktionen. Der Nachlauf liest aus zwei Tabellen und schreibt in eine dritte. Die Schnittstelle zum Zahlungsverkehr ruft vier Endpunkte auf. Least Privilege ist in diesem Fall eine Rechenaufgabe: alle benötigten Rechte, sonst keine.

Bei einem KI-Agenten gilt das nicht. Er bekommt ein Ziel und keine Abfolge. Welche Werkzeuge er aufruft, in welcher Reihenfolge und wie oft, ergibt sich aus dem Ziel, dem Zustand der Systeme und dem, was er unterwegs findet. Die Aktionsmenge entsteht bei jeder Ausführung neu.

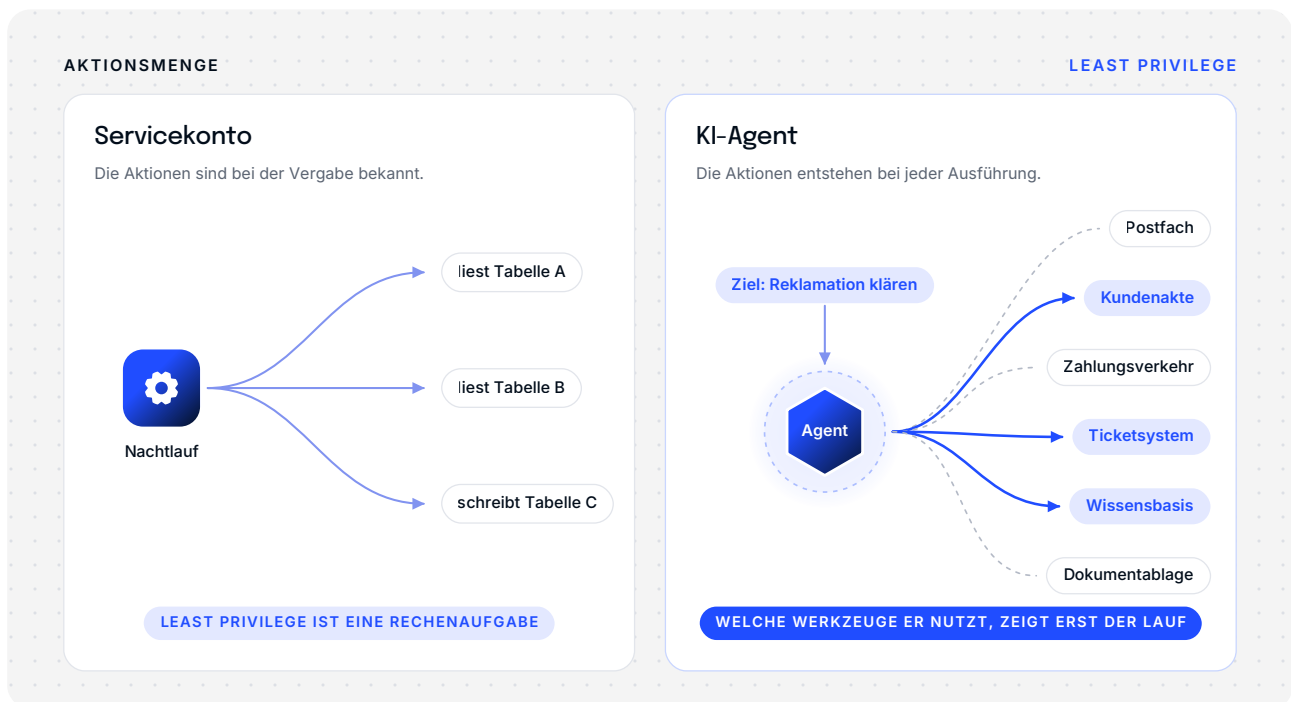


Abbildung 4 Servicekonto und KI-Agent: wann die Aktionsmenge feststeht · Darstellung WertFels

Das NIST National Cybersecurity Center of Excellence beschreibt genau diese Lage als ungelöst. In seinem Konzeptpapier vom Februar 2026 steht die Frage, wie sich Least Privilege bestimmen lässt, wenn die nötigen Aktionen bei der

Bereitstellung unbekannt sind, in einer Liste offener Punkte. Daneben stehen die Fragen, wie ein Agent seine Handlungsvollmacht belegt und wie Nichtabstreitbarkeit für seine Handlungen entsteht.⁶

Wir ziehen daraus eine praktische Konsequenz. Ein Berechtigungskonzept, das für Agenten dieselbe Form hat wie für Servicekonten, listet Rechte auf, die zur Laufzeit entweder nicht reichen oder zu weit gehen.

Drei Arten nicht-menschlicher Identität

Die Diskussion leidet darunter, dass alles, was kein Mensch ist, in einen Topf fällt. Für die Praxis sind drei Arten zu trennen, weil sie in Ausstellung, Prüfung und Entzug nichts gemeinsam haben.

Das statische Servicekonto ist langlebig, hat ein Kennwort oder einen Schlüssel, der irgendwo hinterlegt ist, und einen festen Zweck. In großen Häusern gibt es solche Konten oft in vierstelliger Zahl. Die Kontrolle besteht aus Inventar, Eigentümer, Rotation und Entzug.

Die Arbeitslast ist kurzlebig und erhält ihre Identität bei jedem Start neu, etwa ein Container, der über Workload Identity Federation ein Token bekommt, das nach Minuten verfällt. Kontrolliert wird hier die Ausstellung selbst: wer sich als diese Arbeitslast ausweisen darf, gegen welchen Vertrauensanker und mit welcher Gültigkeitsdauer.

Der Agent unterscheidet sich von beiden dadurch, dass er im Auftrag handelt und dabei wählt. Er hat eine eigene Identität. Seine Berechtigung leitet er von einem Auftraggeber ab. Deshalb liegt die Kontrolle bei ihm auf der Vollmacht.

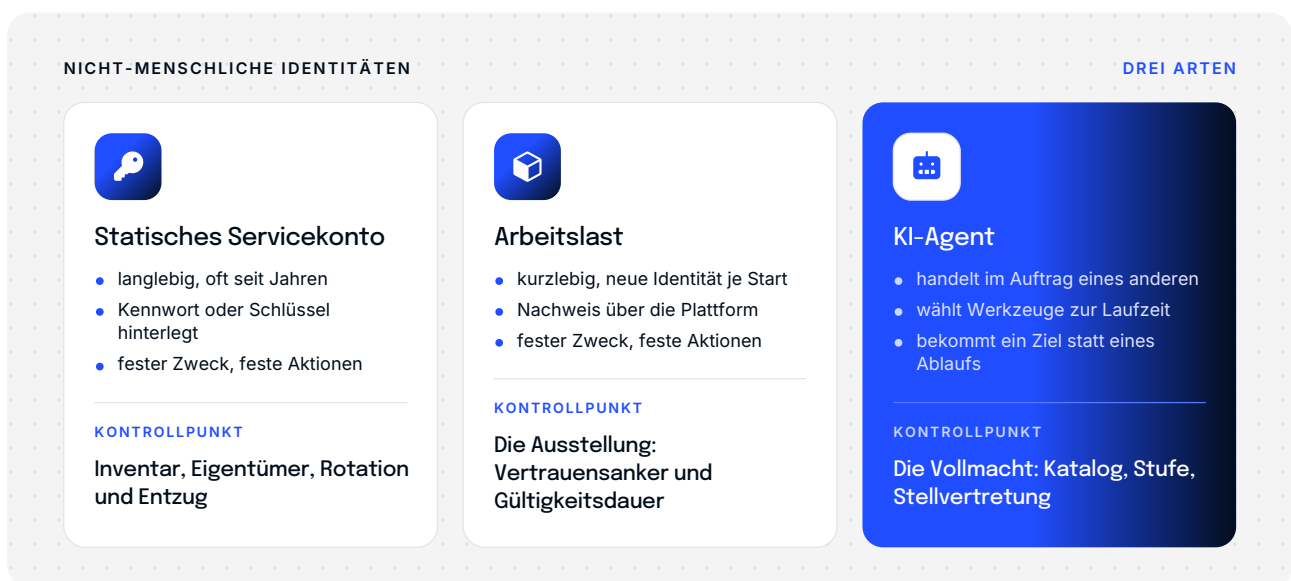


Abbildung 5 Drei Arten nicht-menschlicher Identität und ihre Kontrollpunkte · Darstellung WertFels

Erst die Anbindung macht aus einem Modell einen Akteur

Ein Sprachmodell für sich erzeugt Text. Handeln kann es erst, wenn es an Werkzeuge angebunden ist, an Postfächer, Kundenakten, Ticketsysteme und interne Schnittstellen. Das Model Context Protocol ist derzeit einer der verbreitetsten Wege für diese Anbindung. Jeder neue Server erweitert, was ein Agent erreichen kann, und zwar oft schneller, als die Informationssicherheit davon erfährt.

Die Autorisierungsspezifikation des Protokolls behandelt einen Server als OAuth-Ressourcenserver. Sie verlangt, dass der Client in Autorisierungs- und Tokenanfragen die Zielressource nach RFC 8707¹⁹ benennt, dass der Server nur Tokens akzeptiert, die für ihn selbst ausgestellt wurden, und dass er empfangene Tokens nie an nachgelagerte Dienste weiterreicht.²⁰ So kann ein Token nur bei dem Dienst verwendet werden, für den es ausgestellt wurde.

Zwei Einschränkungen gehören zur ehrlichen Darstellung. Die Autorisierung ist im Protokoll optional, und bei lokaler Anbindung über die Standardein- und -ausgabe stammen die Zugangsdaten aus der Umgebung des Prozesses. Außerdem schreibt der Standard keine Prüfung der Vollmacht vor, er stellt nur die Mittel bereit. Eine Richtlinie muss das Institut selbst mitbringen.

Die Entscheidung fällt bei jeder Aktion

Daraus folgt der wichtigste Unterschied zum Servicekonto. Eine erfolgreiche Anmeldung klärt, wer der Agent ist. Danach beginnen die eigentlichen Fragen: ob er diese Daten lesen, dieses Werkzeug mit diesen Parametern aufrufen und dieses Ergebnis ausgeben darf. Jede dieser Fragen stellt sich bei jeder Aktion neu. Ein Kontrollmodell für Agenten muss deshalb vor dem Einsatz, bei jeder Aktion sowie laufend und in festen Abständen wirken.

Das Kontrollmodell im Überblick

Die Governance fragt, wer für einen Agenten verantwortlich ist. Die Laufzeitkontrolle fragt, was er in diesem Moment tun darf. Beide Fragen gehören in ein Modell.

Wir ordnen die Kontrollen für Agenten in sechs Ebenen. Sichtbarkeit und Verantwortung müssen stehen, bevor ein Agent produktiv arbeitet. Identität, Vollmacht und Durchsetzung wirken bei jeder einzelnen Aktion. Der Nachweis schließt den Kreis, laufend über das Protokoll und in festen Abständen über die Rezertifizierung.

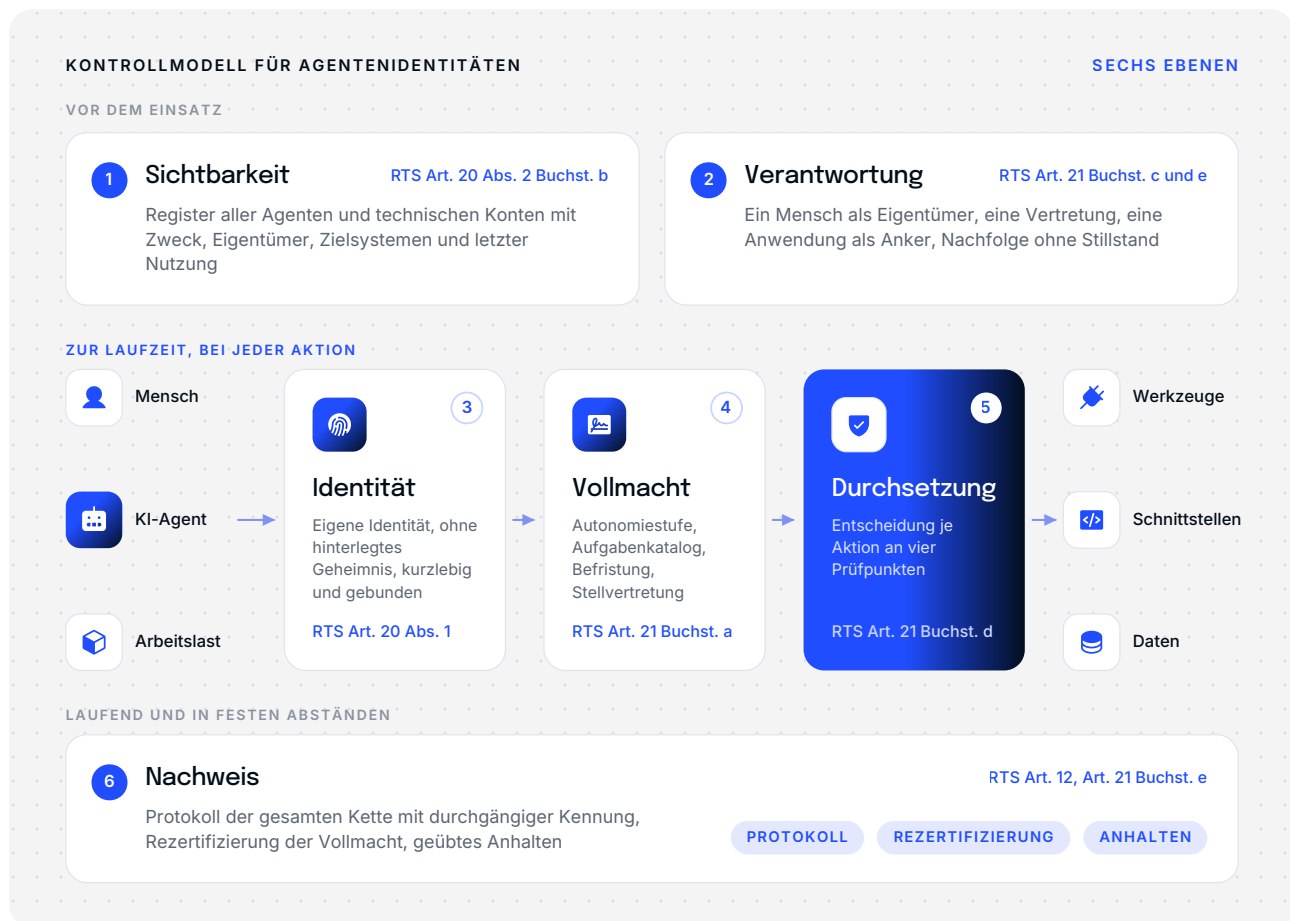


Abbildung 6 Das WertFels Kontrollmodell für Agentenidentitäten · Darstellung WertFels

Die erste Ebene ist die Sichtbarkeit. Ein Register führt jeden Agenten und jedes technische Konto mit Zweck, Eigentümer, Zielsystemen und letzter Nutzung. Es zeigt auch die Beziehungen: welcher Agent welche Werkzeuge, Schnittstellen und Datenbestände erreicht.

Auf der zweiten Ebene steht die Verantwortung. Jeder Agent hat genau einen Menschen als Eigentümer, eine Vertretung und eine Anwendung als Anker. Verlässt der Eigentümer das Haus, greift eine Nachfolgekette, und der Agent läuft weiter.

Dann folgt die Identität. Jeder Agent hat eine eigene Identität, ausgestellt ohne hinterlegtes Geheimnis, kurzlebig und an einen Schlüssel gebunden, sodass ein entwendetes Token anderswo nichts nützt.

Die Vollmacht bildet die vierte Ebene. Sie legt fest, auf welcher Autonomiestufe ein Agent arbeitet, welche Werkzeuge und Operationen in seinem Aufgabenkatalog stehen, wie lange die Vollmacht gilt und in wessen Auftrag er handelt.

Auf der fünften Ebene liegt die Durchsetzung. An vier Prüfpunkten entscheidet eine zentrale Richtlinie über jede Aktion: beim Auftrag, bei den Daten, beim Werkzeug und beim Ergebnis.

Den Abschluss bildet der Nachweis. Das Protokoll hält die gesamte Kette mit einer durchgängigen Kennung fest, die Rezertifizierung prüft die Vollmacht, und das Anhalten ist geübt.

Sechs Grundsätze

Die Technik in diesem Feld ändert sich in Monaten. Die Grundsätze sollten stabil genug sein, um wechselnde Plattformen und Protokolle zu überdauern.

1. Identität zuerst, aber nicht allein. Jede Entscheidung beginnt bei der Identität des Agenten und bezieht die Person, den Zweck und die Daten ein.
2. Keine Dauerberechtigung. Die Vollmacht legt den Rahmen für Monate fest. Die Rechte selbst erhält der Agent je Aufgabe und für Minuten, nie auf Vorrat.
3. Stellvertretung statt Identitätsübernahme. Im Protokoll stehen immer der Agent und sein Auftraggeber, also die Person oder die Organisationseinheit, für die er handelt.
4. Richtlinie zentral, Durchsetzung dezentral. Die Regeln werden an einer Stelle gepflegt und geprüft, angewendet werden sie nah am Werkzeug und an den Daten.
5. Nachweis ab dem ersten Aufruf. Was nicht protokolliert ist, lässt sich weder prüfen noch rekonstruieren.
6. Anhalten leicht, Reaktivieren schwer. Anhalten darf der Eigentümer ohne weitere Zustimmung, dazu die für das Anhalten benannte Stelle. Reaktivieren darf der technische Betreuer nur mit Zustimmung des Eigentümers.

Unser vierter Grundsatz entspricht der Trennung von Entscheidungspunkt und Durchsetzungspunkt, die das NIST für Zero-Trust-Architekturen beschreibt.²¹ Neu ist bei Agenten nur, wie oft entschieden werden muss.

Das Modell an einem Beispiel

Ein Versicherer setzt in der Kreditorenbuchhaltung einen Agenten ein. Der Agent liest Eingangsrechnungen aus einem Sammelpostfach, gleicht sie mit den Lieferantenstammdaten ab und legt Zahlungen im Buchhaltungssystem an. Jede Zahlung wird erst wirksam, wenn eine Sachbearbeiterin sie freigibt. Ab 10.000 Euro sind zwei Freigaben nötig. Die Zahlen sind Beispielwerte.

EBENE	AUSPRÄGUNG IM BEISPIEL
Sichtbarkeit	Registereintrag mit Zweck Rechnungsbearbeitung, Zielsystemen Postfach, Stammdaten und Buchhaltung, letzter Nutzung
Verantwortung	Eigentümerin ist die Leiterin der Kreditorenbuchhaltung, Vertretung der Teamleiter Zahlungsverkehr, Anker das Buchhaltungssystem
Identität	eigene Agentenidentität, Token mit 15 Minuten Gültigkeit, an den Schlüssel des Agenten gebunden
Vollmacht	Stufe 3 für Zahlungen und Antworten an Dritte, Stufe 1 für Stammdaten, Katalog mit vier Operationen, Befristung auf zwölf Monate
Durchsetzung	Zahlung nur an hinterlegte Bankverbindungen bekannter Lieferanten, Freigabe je Zahlung, ab 10.000 Euro zwei Freigaben, keine Bankverbindung im Ergebnis an Dritte
Nachweis	jede Buchung mit Kennung des Laufs protokolliert, halbjährliche Rezertifizierung, Anhalten einmal je Periode geübt

Identität und Stellvertretung

Ein Agent braucht eine eigene Identität, die ohne hinterlegtes Geheimnis entsteht, nach Minuten verfällt und im Zielsystem erkennen lässt, in wessen Auftrag er handelt.

Ausstellung ohne hinterlegtes Geheimnis

Jeder API-Schlüssel, jedes Clientgeheimnis und jedes einfache Bearer-Token muss irgendwo gespeichert, verteilt und rotiert werden. Bei Agenten, deren Instanzen schneller entstehen und verschwinden als jeder Rotationszyklus, geht diese Rechnung nicht auf.

Die Alternative leitet die Identität aus einer Eigenschaft der Laufzeitumgebung ab, die ein Dritter bestätigen kann. Die Plattform bescheinigt etwa die Prüfsumme des Abbilds, das Servicekonto im Kubernetes-Cluster oder den Zustand eines TPM. Ein Vertrauensanker prüft diese Bescheinigung gegen eine Richtlinie und stellt erst dann eine kurzlebige Identität aus.

SPIFFE beschreibt dieses Muster herstellerübergreifend. Die Identität eines Dienstes heißt dort SPIFFE ID und wird in einem kurzlebigen, kryptografisch prüfbaren Dokument ausgestellt, dem SVID, als X.509-Zertifikat oder als JWT. Die produktionsreife Implementierung SPIRE attestiert Knoten und Arbeitslasten und stellt das Dokument nur aus, wenn die bestätigten Eigenschaften zur Richtlinie passen.²² Die großen Cloudanbieter bieten dasselbe Prinzip als Workload Identity Federation an. Auch Microsoft rät für Agent Identity Blueprints in Entra ausdrücklich davon ab, im Produktivbetrieb Clientgeheimnisse zu verwenden, und empfiehlt föderierte Anmeldeinformationen mit verwalteten Identitäten.²³



Abbildung 7 Ausstellung einer Agentenidentität ohne hinterlegtes Geheimnis · Darstellung WertFels nach SPIFFE, RFC 9449 und RFC 8705

Kurzlebig und an einen Schlüssel gebunden

Die Lebensdauer eines Agenten bestimmt das Modell seiner Zugangsnachweise. Langlebige Agenten, etwa ein dauerhaft laufender Servicebot, melden sich über föderierte Identitäten an und werden in festen Abständen neu attestiert. Kurzlebige Agenten entstehen für eine Aufgabe, erhalten Tokens mit einer Gültigkeit von wenigen Minuten und verlieren ihre Identität, sobald der Container endet.

Ein kurzes Ablaufdatum allein genügt nicht. Ein entwendetes Token ist bis zu seinem Ablauf so gut wie das Original. Demonstrating Proof of Possession (DPoP) nach RFC 9449 bindet ein Zugangstoken an einen Schlüssel, den nur der Client besitzt, sodass ein abgefangenes Token allein anderswo nicht verwendbar ist.²⁴ Zertifikatsgebundene Tokens über gegenseitiges TLS nach RFC 8705 leisten dasselbe auf der Transportebene.²⁵ Im Drift-Vorfall hätte diese Bindung die gestohlenen Tokens für die Angreifer wertlos gemacht, sofern der private Schlüssel nicht mit entwendet worden wäre.

Stellvertretung statt Identitätsübernahme

Wenn ein Agent im Auftrag eines Sachbearbeiters eine Buchung anlegt, gibt es zwei technische Wege. Nutzt der Agent das Token des Sachbearbeiters, steht im Protokoll des Zielsystems dessen Name. Hat der Agent eine eigene Identität und führt die Vollmacht des Sachbearbeiters mit, steht im Protokoll, dass Agent A im Auftrag von Person B gehandelt hat.

Der Unterschied ist seit Januar 2020 standardisiert. RFC 8693 zu OAuth 2.0 Token Exchange trennt beide Fälle ausdrücklich. Bei der Identitätsübernahme erhält der Handelnde alle Rechte des Auftraggebers und ist von ihm nicht unterscheidbar. Bei der Stellvertretung behält er seine eigene Identität, und erkennbar bleibt, dass er für einen anderen auftritt. Technisch geschieht das über zwei Angaben in der Anfrage: Das Subject-Token bezeichnet die Partei, in deren Auftrag gehandelt wird, das Actor-Token die handelnde Partei. Das ausgestellte Token trägt dann einen Actor-Claim, der die Delegation festhält.⁷

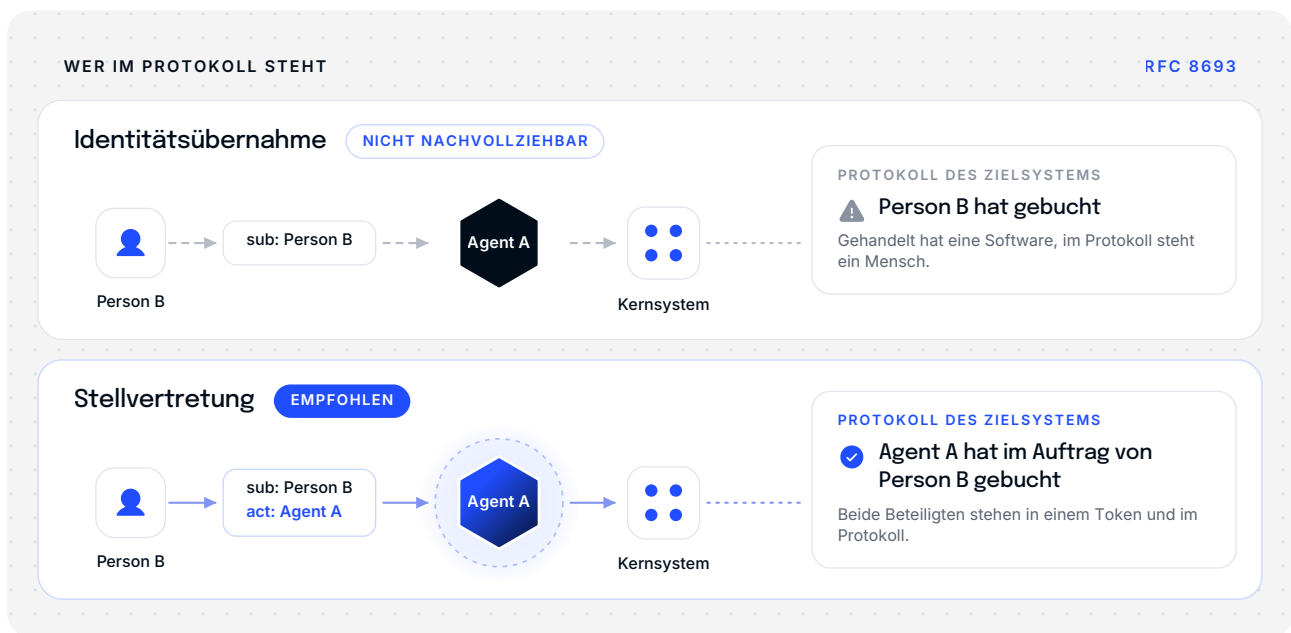


Abbildung 8 Identitätsübernahme und Stellvertretung im Protokoll des Zielsystems · Darstellung WertFels nach RFC 8693

Wir halten diesen Punkt für die wichtigste Einzelentscheidung beim Einsatz von Agenten in beaufsichtigten Häusern. Artikel 21 Buchstabe c der DORA-RTS verlangt, dass Handlungen jederzeit einem Nutzer zugeordnet werden können. Handelt ein Agent unter der Identität eines Menschen, ist diese Anforderung dem Wortlaut nach erfüllt und in der Sache verfehlt. Im Protokoll steht ein Mensch, gehandelt hat eine Software. Bei einer Prüfung lässt sich das nicht erklären, bei einem Vorfall nicht rekonstruieren.

Plattformen setzen die Stellvertretung unterschiedlich um. Microsoft Entra bildet sie für Agenten über einen eigenen On-Behalf-Of-Fluss ab, der auf dem JWT-Bearer-Verfahren aufsetzt und nicht auf dem Token Exchange nach RFC 8693.²³ Welche Angaben das Token im Zielsystem über den handelnden Agenten enthält, sollte ein Institut im eigenen Mandanten prüfen, bevor es seine Protokollierung darauf stützt.

Bei Ketten zählen Auftraggeber und letztes Glied

Sobald ein Agent einen weiteren Agenten beauftragt, entsteht eine Kette. RFC 8693 bildet sie durch Verschachtelung ab: Der äußerste Actor-Claim bezeichnet den aktuellen Akteur. Die Vorgabe des Standards ist eindeutig: Wer ein Token auswertet, darf für die Zugriffsentscheidung nur die obersten Angaben und den aktuellen Akteur berücksichtigen. Frühere Akteure in verschachtelten Einträgen sind reine Information.⁷



Abbildung 9 Delegationskette und Auswertung des Actor-Claims · Darstellung WertFels nach RFC 8693, Abschnitt 4.1

Für ein Institut folgen daraus zwei Regeln, die in kein Standardwerkzeug eingebaut sind. Die Länge der Kette ist zu begrenzen, und jedes Glied ist mit derselben Kennung zu protokollieren, weil das Zielsystem die Kette bei seiner Entscheidung nicht auswertet. Bleibt die Tiefe offen, lässt sich der Ursprung einer Handlung nur noch bestimmen, indem man mehrere Protokolle zusammenführt.

Standards, die gerade entstehen

Zwei Entwürfe der IETF erweitern das Muster über die Grenze eines einzelnen Dienstes hinaus. Transaction Tokens tragen Nutzeridentität, Identität der Arbeitslast und Autorisierungskontext als kurzlebige, signierte Tokens durch die Aufrufkette innerhalb einer Vertrauensdomäne.²⁶ Der Identity Assertion JWT Authorization Grant erlaubt einer Anwendung, über den gemeinsamen Identitätsanbieter ein Zugangstoken für die Schnittstelle einer anderen Anwendung zu erhalten. Er ist die Grundlage des Musters, das unter dem Namen Cross App Access bekannt geworden ist.²⁷

Beide Entwürfe sind nicht verabschiedet. Institute sollten sie als Richtung behandeln und ihre Tokenausstellung so gestalten, dass sie später übernommen werden können. Bereits verabschiedet sind dagegen Rich Authorization Requests nach RFC 9396, mit denen sich in einer Tokenanfrage genau beschreiben lässt, welcher Zugriff wofür und in welchem Umfang benötigt wird.²⁸ Vordringlich ist die Reihenfolge: Attestierte Anmeldung ohne Geheimnis und Stellvertretung gehören in den Agenten, bevor er produktiv geht. Nachträglich in Agenten eingebaut, die bereits mit geteilten Geheimnissen laufen, kosten sie ein Vielfaches.

Vollmacht und Laufzeitkontrolle

Die Vollmacht legt fest, was ein Agent grundsätzlich darf. Die Laufzeitkontrolle entscheidet, ob die konkrete Aktion in diesem Moment zulässig ist.

Rechte nach Autonomiestufe

Gartner unterscheidet vier Autonomiestufen: Beobachten, Beraten, Handeln mit Freigabe und autonomes Handeln.²⁹ Ein Institut legt je Stufe fest, welche Systeme, Datenklassen und Operationen zulässig sind. Ein Agent auf Stufe 1 darf nur lesen. Auf Stufe 2 darf er Vorschläge anlegen, die ein Mensch übernimmt, aber nichts endgültig schreiben. Auf Stufe 3 darf er schreiben, jede schreibende Operation wartet aber auf eine menschliche Freigabe. Wir empfehlen, Stufe 4 für Systeme kritischer oder wichtiger Funktionen bis auf Weiteres nicht zu genehmigen.

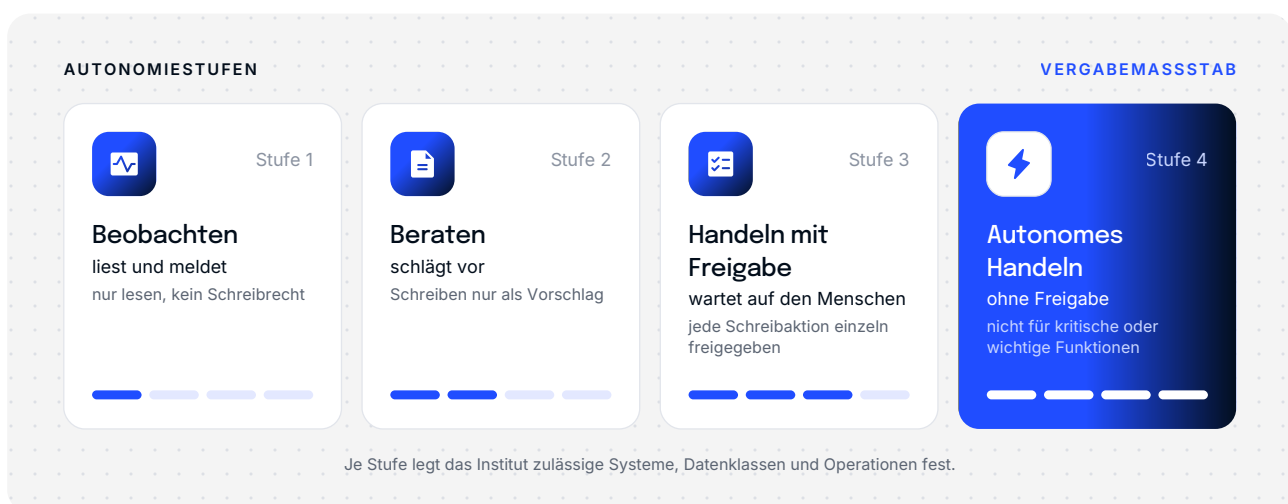


Abbildung 10 Vier Autonomiestufen als Vergabemaßstab · Stufen nach Gartner, Zuordnung WertFels

Rechte je Aufgabe statt auf Vorrat

Die Vollmacht legt fest, was ein Agent für einen Zeitraum von Monaten überhaupt darf. Die konkrete Berechtigung bekommt er nur für die Dauer einer Aufgabe und gibt sie danach zurück. Technisch sind das kurzlebige, an ein Ziel gebundene Tokens mit engem Umfang. Access Packages mit Ablaufdatum in Entra begrenzen zusätzlich die Zuweisung selbst, sie wirken aber in Tagen und nicht je Aufgabe. Das Model Context Protocol sieht dafür eine Nachforderung zur Laufzeit vor: Fehlt einem Token ein Umfang, benennt der Server die nötigen Berechtigungen, und der Client holt sie gezielt nach.²⁰

Freigabe je Aktion für alles, was Geld bewegt

Für Aktionen, die Geld bewegen, Rechte ändern oder Daten löschen, ist die Freigabe je Aktion Pflicht, unabhängig von der Stufe des Agenten. Die OpenID Foundation arbeitet mit dem Access Request and Approval Profile an einem Standard, der eine Anfrage in einen Wartezustand versetzt, bis eine benannte Person zugestimmt hat. Das Beispiel im Entwurf ist ein Agent, der Lieferantenzahlungen oberhalb eines Schwellwerts auslöst.³⁰ Bis der Standard verabschiedet ist, bleibt die Freigabe je Aktion herstellerspezifisch, etwa als Eskalation in einem Autorisierungsgateway.

Vier Prüfpunkte auf dem Weg einer Aktion

Eine Agentenaktion durchläuft vier Stellen, an denen entschieden werden muss. Am Auftrag wird geprüft, ob die Aufgabe zum Zweck und zur Stufe des Agenten passt und ob sie überhaupt von einer berechtigten Stelle kommt. Anweisungen, die in gelesenen Dokumenten oder Mails stecken, sind kein Auftrag. Bei den Daten wird geprüft, welche Quellen der Agent lesen darf und was davon in seinen Kontext gelangt, mit Filtern je Zeile, Spalte und Dokument. Am Werkzeug wird geprüft, ob die Operation im Katalog steht und ob Parameter und Beträge in den Grenzen liegen. Am Ergebnis wird geprüft, was den Agenten verlässt und an wen.



Abbildung 11 Vier Prüfpunkte auf dem Weg einer Agentenaktion, entschieden an einer zentralen Stelle · Darstellung WertFels

Im Beispiel aus der Kreditorenbuchhaltung heißt das: Ein Auftrag, der aus dem Text einer Rechnung stammt, wird verworfen. Gelesen werden nur das Sammelpostfach und die Stammdaten der Lieferanten. Eine Zahlung an eine Bankverbindung, die nicht beim Lieferanten hinterlegt ist, wird abgewiesen. Jede andere Zahlung wartet auf die Freigabe, ab 10.000 Euro auf zwei. Eine Antwort an einen externen Empfänger enthält keine Bankdaten.

Gemeinsam mit den Ebenen Identität und Vollmacht mindern die vier Prüfpunkte die ersten drei Risiken, die das OWASP GenAI Security Project für agentische Anwendungen nennt: die Übernahme des Ziels, den Missbrauch von Werkzeugen und den Missbrauch von Identitäten und Rechten.¹⁴

Person und Agent zählen gemeinsam

Für jede Entscheidung gibt es drei Muster. Zählt nur der Agent, sammelt er mit der Zeit die Rechte aller seiner Nutzer an. Zählt nur die Person, darf der Agent alles, was die Person darf, auch das, was nie zu seinem Zweck gehörte. Belastbar ist die zusammengesetzte Entscheidung: Erlaubt ist nur, was die Person darf und was zugleich im Katalog des Agenten steht.



Abbildung 12 Zusammengesetzte Entscheidung aus Nutzerrecht und Agentenvollmacht · Darstellung WertFels

Das Prinzip ist aus delegierten Berechtigungen bekannt. Eine Anwendung, die im Auftrag eines Nutzers zugreift, erreicht dort nichts, was der angemeldete Nutzer selbst nicht erreichen könnte.³¹ Bei Agenten kommt die zweite Grenze hinzu, der Aufgabenkatalog.

Zentrale Entscheidung, verteilte Durchsetzung

Die Regeln für diese Entscheidungen gehören an eine Stelle. Wird die Berechtigungslogik in jeden Agenten einzeln eingebaut, baut jedes Team sein eigenes Modell, und niemand kann mehr sagen, welche Regel wo gilt. Die Durchsetzung dagegen gehört nah an das Geschehen: in die Laufzeitumgebung des Agenten, in ein Gateway vor den MCP-Servern, in das API-Gateway, in die Datenzugriffsschicht und an den Ausgang.

Die Verbindung zwischen beiden ist seit Januar 2026 standardisiert. Die OpenID Foundation hat die AuthZEN Authorization API 1.0 als endgültige Spezifikation verabschiedet. Sie beschreibt, wie ein Durchsetzungspunkt eine Entscheidung beim Entscheidungspunkt anfragt, unabhängig vom Hersteller beider Seiten.³² Die Richtlinien selbst sollten als Code gepflegt, versioniert und vor jeder Änderung getestet werden. Jede Entscheidung wird mit Anfrage, Regel und Ergebnis protokolliert.

Für die Prüfung

Artikel 21 Buchstabe a verlangt Rechte nach Erforderlichkeit. Bei einem Agenten ist Erforderlichkeit eine Eigenschaft der einzelnen Aktion. Ein Institut, das die Entscheidung je Aktion zentral trifft und jede Entscheidung protokolliert, kann für jede Handlung belegen, nach welcher Regel sie erlaubt wurde. Die Regeln selbst sind regelmäßig auf Erforderlichkeit zu prüfen.

Verantwortung und Lebenszyklus

Ein Agent kann nicht haften. Jede Kontrolle, die im Ernstfall greifen soll, braucht am Ende einen Menschen, der entscheidet, bestätigt und verantwortet.

Ohne benannten Menschen gibt es keine Verantwortung

Wir empfehlen als erste und einfachste Festlegung, dass jeder Agent genau einen menschlichen Eigentümer hat. Ein Team, eine Abteilung oder ein Funktionspostfach genügt dafür nicht. Der Grund ist derselbe wie bei Servicekonten, nur schärfer: Ein Konto, das einer Gruppe gehört, gehört niemandem. Bei einem Agenten kommt hinzu, dass er Entscheidungen trifft, die vorher niemand im Einzelnen genehmigt hat. Wer sie nachträglich verantwortet, muss vorher feststehen.

Ein genehmigungsfähiger Agent hat deshalb drei Bindungen: einen Eigentümer mit Vertretung, eine Anwendung in der Konfigurationsdatenbank als Anker und eine Vollmacht mit Katalog, Stufe und Befristung.

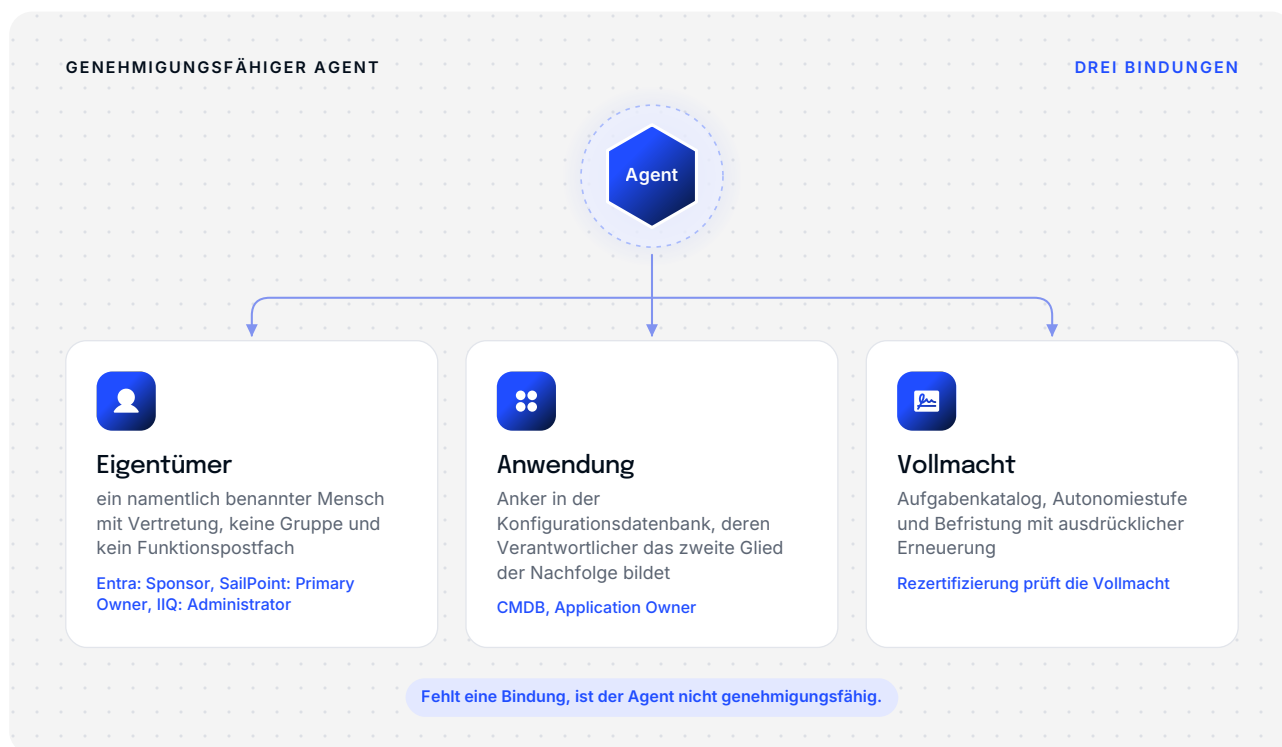


Abbildung 13 Die drei Bindungen eines genehmigungsfähigen Agenten · Darstellung WertFels

Sieben Angaben je Agent

In Bestandsaufnahmen prüfen wir je Agent sieben Angaben. Fehlt eine davon, ist der Agent nicht genehmigungsfähig.

1. Der Eigentümer ist ein namentlich benannter Mensch mit Vertretung.
2. Festgelegt ist, in wessen Auftrag der Agent handelt: dauerhaft für eine Organisationseinheit, dann ist die Vollmacht generisch und muss eng sein, oder jeweils für die auslösende Person, dann muss die Stellvertretung technisch abgebildet sein.
3. Ein Katalog beschreibt die erlaubten Werkzeuge und Operationen. Systemrechte beschreiben, was möglich wäre. Der Katalog beschreibt, was gewollt ist.
4. Die Vollmacht ist befristet und muss ausdrücklich erneuert werden. Eine Vollmacht ohne Ablaufdatum ist bei Servicekonten der Normalfall und der Grund, warum sie zur Altlast werden.
5. Die Identität stammt von einem kontrollierten Vertrauensanker und gilt Minuten, nicht Jahre.
6. Protokolliert wird die Kette: auslösende Person, Agent, nachgelagerte Agenten, Werkzeug, Zielsystem und Ergebnis, mit durchgängiger Kennung und synchronisierten Uhren.
7. Eine benannte Stelle kann den Agenten anhalten, mit einem geübten Verfahren und einer gemessenen Zeit vom Entschluss bis zum Stillstand.

Der Lebenszyklus in sechs Stationen

Der Lebenszyklus eines Agenten entspricht dem, was Institute für Menschen als Eintritt, Wechsel und Austritt kennen, mit zwei Unterschieden: Der Agent hat von Anfang an einen Eigentümer, und er hat ein Ablaufdatum.



Abbildung 14 Lebenszyklus eines Agenten in sechs Stationen · Darstellung WertFels

Bei der Registrierung wird der Agent als eigenes Objekt angelegt, mit Zweck, Eigentümer, Vertretung, Katalog und Befristung. Ein Agent, der nur in einer Entwicklungsumgebung existiert und nie registriert wurde, ist Schatten-IT. Die Genehmigung bestätigen Eigentümer, Fachbereich und Informationssicherheit. Greift der Agent auf Systeme kritischer oder wichtiger Funktionen zu, gehört die Genehmigung in denselben Prozess wie privilegierte Rechte. Bei der Ausstellung erhält der Agent kurzlebige Zugangsnachweise. Neue statische Schlüssel und Kennwörter sind für ihn nicht zulässig. Im Betrieb wird jede Handlung protokolliert, und auffälliges Verhalten führt zu einem Eingriff. Bei der Rezertifizierung bestätigt der Eigentümer Bedarf, Katalog und Verhalten. Die Stilllegung schließlich deaktiviert den Agenten, widerruft seine Nachweise, entzieht seine Rechte und bewahrt sein Protokoll auf. Auslöser sind das Ende der Befristung ohne Erneuerung, der Wegfall des Zwecks oder ein Vorfall, nach dem der Agent nicht wieder in Betrieb gehen soll. Bei einem Vorfall wird zuerst angehalten, stillgelegt wird nach der Bewertung. Läuft für den Agenten gerade die Nachfolgekette, wird eine ablaufende Befristung nicht zum Auslöser. Die Vollmacht bleibt im bisherigen Umfang eingefroren, bis ein neuer Eigentümer bestätigt.

Eigentümerschaft vergeben: eine Person, eine Vertretung, eine Frist

Aus den Plattformmodellen und den Standards lässt sich ein Vergabemuster ableiten, das überall funktioniert. Der fachliche Eigentümer ist eine natürliche Person, in Entra der Sponsor, in der SailPoint Identity Security Cloud der Primary Owner, in SailPoint IdentityIQ der Administrator. Eine Vertretung ist benannt. Die

technische Betreuung ist von der fachlichen Verantwortung getrennt: Der fachliche Eigentümer darf anhalten und übertragen, der technische Betreuer darf mit Zustimmung des Eigentümers reaktivieren und wiederherstellen.

Die Eigentümerschaft selbst sollte befristet sein. Die Plattformen unterstützen Ablaufdaten für Berechtigungen, für die Eigentümerschaft nicht. Wir empfehlen, beides über die Rezertifizierung zu verbinden: Wer den Agenten nicht fristgerecht bestätigt, verliert die Eigentümerschaft an das nächste Glied der Nachfolgekette. Der Agent läuft weiter, die Frage nach seinem Eigentümer wandert nach oben.

Wenn der Eigentümer geht: übertragen und eskalieren statt anhalten

Microsoft Entra und SailPoint reagieren auf den Austritt des Eigentümers mit einer Übertragung, Entra allerdings nur für den Sponsor. Keine der beiden Plattformen legt den Agenten deshalb still. Wir halten das für richtig und gehen einen Schritt weiter: Auch am Ende der Nachfolgekette nach abgelaufenen Fristen ist das Anhalten oder die Stilllegung eines produktiven Agenten allein wegen fehlender Eigentümerschaft keine zulässige Maßnahme. Ein Austritt oder eine Suspendierung kommt oft ohne Vorlauf. Ein Agent, der dann deaktiviert wird, hält einen Geschäftsprozess an, den niemand anhalten wollte. Ein Zahlungslauf oder eine Schnittstelle zum Kernbanksystem darf nicht davon abhängen, ob die Personalabteilung rechtzeitig einen Nachfolger benennt.

Die Begründung liefert ein Grundsatz der Security Governance: Informationssicherheit soll Geschäftsprozesse ermöglichen. Eine Kontrolle, die den Betrieb lahmlegt, um eine Verwaltungslücke zu schließen, hat ihre Verhältnismäßigkeit verloren. Ein Sicherheitsvorfall ist ein Grund, einen Agenten anzuhalten, ein fehlender Eigentümer ist es nicht.

Den Druck, einen Nachfolger zu finden, erzeugt deshalb eine Nachfolgekette mit drei Gliedern, an deren Ende eine Eskalation steht. Erstes Glied ist die benannte Vertretung, ohne die sich ein Agent gar nicht anlegen lässt. Zweites Glied ist der Verantwortliche der Anwendung laut Konfigurationsdatenbank. Er kennt den Zweck des Agenten, weil er die Anwendung kennt, und er wird bei einem Wechsel über den Änderungsprozess ersetzt statt über die Personalabteilung. Die Umsetzungshinweise des BSI zum Baustein ORP.4 lassen für Gruppenkennungen einen Rolleninhaber als Verantwortlichen ausdrücklich zu³³, und die CIS Controls verlangen in Safeguard 5.5 ein Inventar der Servicekonten mit verantwortlicher Abteilung, Prüfdatum und Zweck.³⁴ CIS genügt eine Abteilung. Wir verlangen eine Person. Drittes Glied ist der Linienvorgesetzte der ausgeschiedenen Person als Übergangseigentümer mit Frist. Läuft sie ab, geht die Eigentümerschaft an dessen Vorgesetzten, mit derselben Frist, und so weiter nach oben. Ein Agent ohne Eigentümer landet so innerhalb weniger Wochen bei einer Bereichsleitung, die genau wissen will, wer sich darum kümmert.

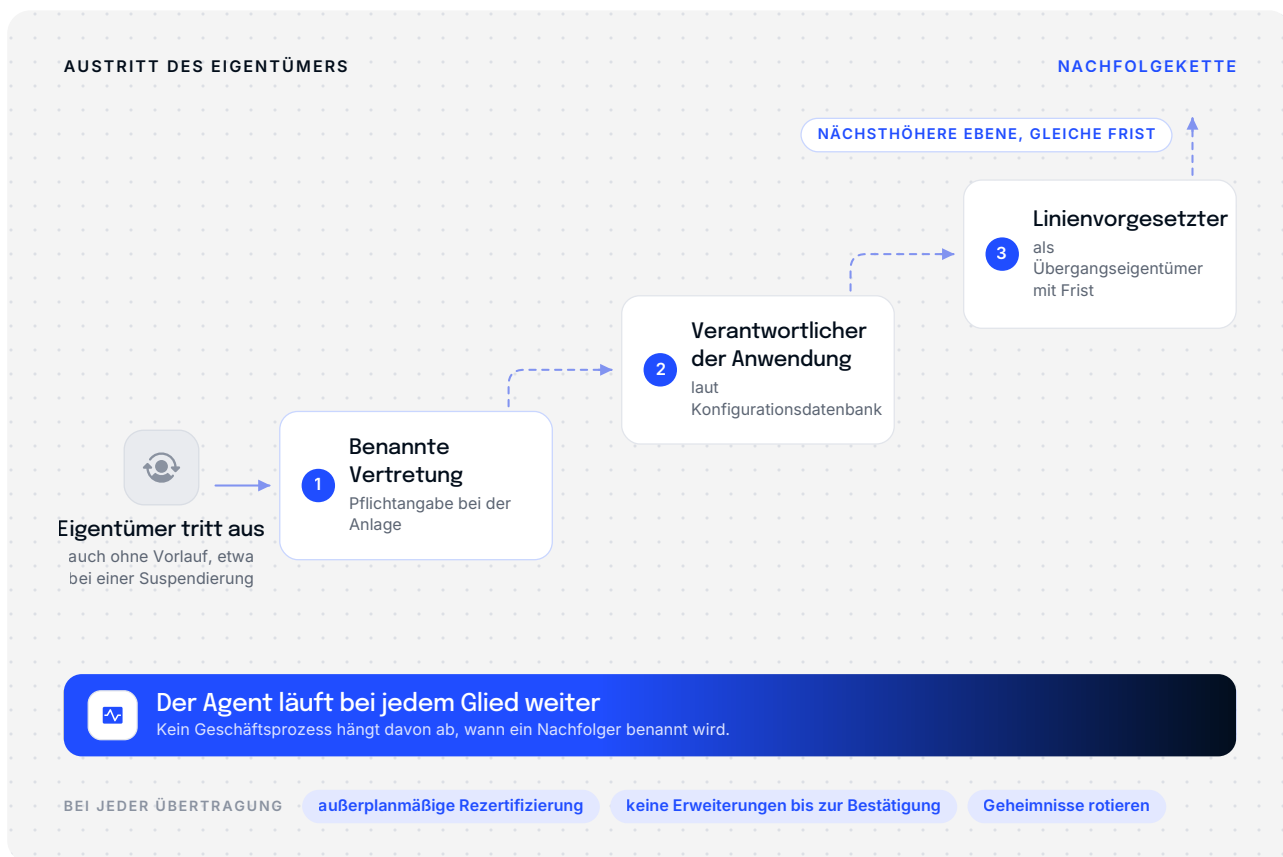


Abbildung 15 Nachfolgekette beim Austritt des Eigentümers · Darstellung WertFels

Jede Übertragung löst drei Maßnahmen aus. Der neue Eigentümer bestätigt in einer außerplanmäßigen Rezertifizierung Zweck, Katalog und Verhalten der letzten Periode. Bis dahin sind keine Erweiterungen genehmigungsfähig, also keine neuen Werkzeuge, keine neuen Zielsysteme und keine Verlängerung der Befristung. Zuletzt werden Geheimnisse, die der ausgeschiedene Eigentümer kannte oder besaß, rotiert, und seine Rechte an Registrierung oder Blueprint werden entfernt.

Dass diese Sorgfalt nötig ist, zeigt ein Angriffspfad, den Ridgeline Cyber im Mai 2026 für Entra beschrieben hat. Wird ein Benutzerkonto deaktiviert, bleibt seine Eigentümerschaft an Service Principals bestehen. Wer das Konto später reaktiviert oder übernimmt, erbt sie und kann neue Zugangsnachweise anlegen.³⁵ Verwaiste Eigentümerschaft ist damit ein Einstiegspunkt. Die Cloud Security Alliance stellt die Beendigung einer Agentenidentität auf eine Stufe mit ihrer Erstellung.³⁶ Microsoft berichtet von zehn Gesprächsrunden mit rund 150 Fachleuten für Identitätssicherheit auf der Identiverse 2026. In neun davon wurde der eigentümerlose Agent als Problem genannt.³⁷ Gartner erwartet, dass bis 2027 vier von zehn Unternehmen autonome Agenten wegen Lücken in der Governance zurückstufen oder abschalten.²⁹

Die Rezertifizierung prüft die Vollmacht

Die halbjährliche Frist für kritische Funktionen trifft Agenten genauso wie menschliche Konten. In der gewohnten Form läuft sie bei ihnen aber ins Leere, denn eine abgehakte Liste von Systemrechten sagt über einen Agenten wenig. Wir schlagen fünf Prüffragen vor, die der Eigentümer beantwortet:

1. Wird der Agent noch gebraucht?
2. Entspricht sein Katalog dem, was er tatsächlich getan hat?
3. Ist die Vollmacht auf das Nötige begrenzt?
4. Enthalten die Läufe der letzten Periode nur Handlungen, die der Eigentümer verantworten würde?
5. Wurde das Anhalten in der Periode mindestens einmal geübt?

Die zweite Frage ist die aufschlussreichste. Sie setzt voraus, dass die tatsächlichen Aufrufe ausgewertet und dem Katalog gegenübergestellt werden. Wer das tut, findet in der Regel zwei Dinge: Werkzeuge, die genehmigt und nie benutzt wurden, und Aufrufe, die niemand vorhergesehen hat.

Das Anhalten wie eine Notfallübung behandeln

Der gemeinsame Leitfaden von CISA und fünf Partnerbehörden vom Mai 2026 zur Einführung agentischer KI empfiehlt menschliche Kontrollpunkte im gesamten Ablauf eines Agenten, ausdrücklich auch die Möglichkeit, eine laufende Aufgabe zu unterbrechen.³⁸ Ein festes Verfahren zum Anhalten mit Fristen beschreibt er nicht. Wir empfehlen, das Anhalten vor dem Produktivgang zu üben und danach einmal je Rezertifizierungsperiode an jedem produktiven Agenten, in einem geplanten Wartungsfenster. Die Zeit vom Entschluss bis zum Stillstand wird gemessen und dokumentiert. Solange das Verfahren nie geübt wurde, weiß niemand, ob es funktioniert.

Microsoft Entra und SailPoint im September 2026

In deutschen Banken und Versicherern stehen meist Microsoft Entra ID als Verzeichnis und SailPoint als Governance-Schicht nebeneinander. Beide haben seit 2025 Funktionen für Agenten geliefert, mit deutlichen Unterschieden.

Weil sich dieser Bereich in Monaten verändert, geben die Angaben den Stand vom September 2026 wieder. Sie stammen aus der Herstellerdokumentation und sind mit Datum belegt, wo der Hersteller eines nennt. Für weitere Plattformen wie Keycloak, One Identity Manager oder Omada Identity Suite gilt dasselbe Prüfschema. Bewertet werden sie in diesem Papier nicht.

Microsoft Entra Agent ID

Microsoft führt Agenten seit dem Frühjahr 2026 als eigene Objektklasse im Verzeichnis. Die allgemeine Verfügbarkeit meldet Microsoft auf der Seite zu den Neuerungen mit Stand vom 1. Mai 2026.³⁹ Seit Mai 2026 erzeugt Copilot Studio für jeden neuen Agenten zwingend eine solche Identität. Abschalten lässt sich das auf Umgebungsebene nicht mehr.⁴⁰

Das Objektmodell kennt vier Typen. Der Agent Identity Blueprint ist die Vorlage, vergleichbar einer App-Registrierung. Die Agent Identity ist die laufende Instanz, vergleichbar einem Service Principal. Der Agent User ist ein Nutzerkonto für Agenten, die Postfach, Lizenz oder Gruppenmitgliedschaft brauchen. Der Blueprint Principal vertritt den Blueprint im jeweiligen Mandanten, bei mandantenfähigen Agenten auch in weiteren Mandanten.⁴¹ Im Audit-Protokoll erscheinen die Typen unter den Basistypen Application, Service Principal und User, kenntlich am Attribut agentType.⁴²

Die Eigentümerschaft ist zweigeteilt. Der Owner ist die technische Rolle, optional, als Nutzer oder Service Principal, aber nie als Gruppe. Der Sponsor ist die fachliche Rolle und bei der Anlage verpflichtend. Sponsoren dürfen deaktivieren, den Sponsor wechseln und löschen. Reaktivieren dürfen nur Owner. Wer einen Agenten im Namen eines Benutzers anlegt und keinen Sponsor angibt, wird automatisch selbst Sponsor. Für Administratoren mit Agent-ID-Rollen gilt das nicht, und legt eine Anwendung den Agenten ohne Benutzerkontext an, muss sie den Sponsor selbst setzen.⁴³ Diese Trennung ist die praktisch wertvollste Entscheidung im Modell von Entra, weil sie die Verantwortung für den Zweck beim Fachbereich und die für den Betrieb bei der IT belässt. Die Zustimmungsregel aus Grundsatz 6 erzwingt Entra nicht: Ein Entra-Owner, in unseren Begriffen der technische Betreuer, kann allein reaktivieren.

Conditional Access wird bei jeder Tokenausstellung für Agenten ausgewertet. Richtlinien können auf einzelne Agent Identities, auf Blueprints und über Custom Security Attributes zielen. Die Grenzen nennt dieselbe Dokumentation: Richtlinien auf alle Nutzer schließen Agent User nicht ein, und Zugriffe über API-Schlüssel umgehen die Steuerung vollständig.⁴⁴ Ob Continuous Access Evaluation für Agent Identities greift, ist nicht belegt. Für Service Principals ist sie auf Microsoft Graph und Szenarien mit einem Mandanten beschränkt.⁴⁵ Ein Institut sollte deshalb annehmen, dass ein ausgestelltes Agententoken bis zum Ablauf gültig bleibt, und die Gültigkeit kurz wählen.

Identity Protection führt einen Bericht über riskante Agenten mit acht Erkennungen, etwa ungewohntem Ressourcenzugriff und verdächtiger Nutzung von Zugangsnachweisen. Alle laufen nachgelagert. Handelt ein Agent im Auftrag eines Nutzers, wird das Risiko dem Nutzer zugerechnet.⁴⁶

Entra ID Governance stellt Access Packages für Agent Identities bereit, mit Ablaufdatum je Zuweisung und Benachrichtigung des Sponsors.⁴⁷ Microsoft empfiehlt eine Rezertifizierung durch den Sponsor alle sechs bis zwölf Monate.⁴⁸ Beim Austritt überträgt ein Lifecycle Workflow die Sponsorschaft an den Vorgesetzten.⁴⁹ Für den technischen Owner gibt es keinen solchen Automatismus. Microsoft sieht mehrere Owner zur Absicherung vor.⁴³ Eine Befristung der Agentenidentität selbst erzwingt Entra nicht, nur die der Zuweisungen, und Agenten, die keine Entra-Tokens verwenden, sieht es nicht.

SailPoint Identity Security Cloud und SailPoint IdentityIQ

SailPoint hat Agentensicherheit in zwei Schritten gebaut, beide für die Identity Security Cloud. Agent Identity Security ist seit Herbst 2025 verfügbar und führt KI-Agenten als Art der Maschinenidentität.⁵⁰ Agentic Fabric ist seit dem 4. August 2026 allgemein verfügbar und bringt ein Register nicht-menschlicher Identitäten, Erkennung über Sensoren an Endpunkt und Browser, Aktivieren und Deaktivieren mit Genehmigungslauf, Eigentümerzuordnung mit Nachfolge und Zugriffsüberprüfungen.⁵¹

Das Eigentümermodell der Cloud-Suite ist das ausgereifteste der hier betrachteten Plattformen. Der Primary Owner muss ein Mensch sein. Bis zu zehn Additional Owners sind möglich. Wird der Primary Owner inaktiv, geht die Eigentümerschaft an den ersten aktiven Additional Owner, sonst an seinen Vorgesetzten. Fehlt auch dieser, hat der Agent keinen Primary Owner, bis jemand von Hand einen einträgt.⁵²

Für SailPoint IdentityIQ, die in vielen deutschen Banken und Versicherern installierte Suite, gibt es keines dieser Produkte. SailPoint bietet ein Plugin an, das Identitäten aus IdentityIQ als Eigentümer und Nutzer nicht-menschlicher Identi-

täten an Agentic Fabric liefert.⁵³ Die Release Notes der Version 8.5 vom Juli 2025 enthalten keine Funktion für Agenten.⁵⁴ Auch die Ankündigung von IdentityIQ 9.0 im Oktober 2026 nennt keine.⁵⁵

Mit Bordmitteln kann IdentityIQ mehr, als die fehlenden Produkte vermuten lassen. Das Identitätsattribut Type kennt Werte wie Service Account und RPA/Bots und lässt sich um einen Wert für KI-Agenten erweitern. Das Attribut Administrator benennt Eigentümer und Zertifizierer eines nicht-menschlichen Kontos. Targeted Certifications lassen sich auf einen Typ filtern.⁵⁶ Beim Austritt kann Rapid Setup die Identitäten, deren Eigentümer die Person war, neu zuweisen, an den Vorgesetzten, per Regel oder an eine Alternatividentität.⁵⁷ Eine automatische Eskalation ohne Nachfolger ist nicht vorgesehen und muss über Lifecycle Events gebaut werden.

Für Institute auf IdentityIQ ist unsere Folgerung klar. Agenten als eigener Identitätstyp mit Eigentümer, gezielte Zertifizierung und Neuzuweisung beim Austritt sind mit vorhandenen Mitteln in der Regel in wenigen Wochen umsetzbar. Selbst bauen muss das Institut vier Dinge: die Befristung mit Erneuerungspflicht, den Abgleich von Katalog und Verhalten, die Eskalation bei ausbleibender Nachfolge und die Anbindung kurzlebiger Zugangsnachweise. Agentenfunktionen in IdentityIQ selbst erwarten wir nicht, weil SailPoint mit Agentic Fabric auf die Cloud setzt.

PLATTFORMEN

STAND SEPTEMBER 2026

FUNKTION	Microsoft Entra Agent ID	SailPoint Identity Security Cloud mit Agentic Fabric	SailPoint IdentityIQ
Agent als eigenes Objekt	●	●	◐
Eigentümer verpflichtend	●	◐	◐
Vertretung und Nachfolge	◐	●	◐
Befristung der Zuweisung	●	●	◐
Übertragung bei Austritt	◐	●	●
Verhaltensabgleich mit dem Katalog	○	◐	○
Eskalation bei ausbleibender Nachfolge	○	○	○

● verfügbar

◐ teilweise, Vorschau oder mit Bordmitteln

○ Eigenbau erforderlich

Abbildung 16 Abdeckung der Lebenszyklusfunktionen je Plattform · Stand September 2026, Herstellerdokumentation

Betrieb und Nachweis

Für die Nachvollziehbarkeit liefert Entra in den Audit-Protokollen das Attribut `agentType` und die Kennung des Blueprints, in den Anmeldeprotokollen den Agententyp als Filter.⁴² Microsoft Purview erfasst für Agent-365-Instanzen Eingabe, Ausgabe und referenzierte Dateien der Interaktionen zwischen Agent, Mensch und Werkzeug.⁵⁸

Ein herstellerübergreifender Standard, der eine Handlungskette über Entra, SailPoint und die Zielsysteme mit einer Kennung verbindet, fehlt noch. Mögliche Bausteine sind Shared Signals, AuthZEN und Transaction Tokens. Bis dahin stellt das Institut die Korrelation selbst her: mit einer eigenen Kennung, die der Agent bei jedem Aufruf mitführt, und mit synchronisierten Uhren.

Vorgehen in fünf Phasen

Die Reihenfolge ist entscheidend. Wer mit dem Berechtigungskonzept beginnt, ohne zu wissen, welche Agenten es gibt, schreibt Regeln für eine unbekannte Menge.



Abbildung 17 Vorgehen in fünf Phasen · Darstellung WertFels

Phase 1: Bestandsaufnahme

Die erste Frage lautet, wie viele Agenten und technische Konten es gibt und wem sie gehören. Wer sie zum ersten Mal systematisch stellt, findet in der Regel deutlich mehr Konten als erwartet, und für einen erheblichen Teil lässt sich kein Eigentümer benennen. Belastbare branchenweite Zahlen gibt es nicht. Alle veröffentlichten Verhältniszahlen zwischen Maschinen- und Menschenidentitäten stammen von Anbietern mit Verkaufsinteresse oder wurden von ihnen beauftragt. Sie liegen je nach Methode zwischen 20 zu 1 und über 100 zu 1.⁵⁹

Die Bestandsaufnahme erfasst die drei Arten aus Kapitel 3 und ihre Zugangs-nachweise an fünf Fundorten: Servicekonten in Verzeichnissen und Datenbanken, Schlüssel und Zertifikate in Konfigurationen und Tresoren, OAuth-Anwendungen mit ihren Zustimmungen, Arbeitslasten in Cloud- und Containerumgebungen sowie Agenten, auch solche, die Mitarbeitende in Werkzeugen wie Copilot Studio ohne Wissen der IT angelegt haben. Ein reines Inventar reicht nicht. Das Register zeigt, welcher Agent welche Werkzeuge, Schnittstellen und Daten erreicht, denn erst diese Beziehungen sind die tatsächlichen Zugriffspfade. Ein Eintrag ohne Nutzung in 90 Tagen ist ein Kandidat für die Stilllegung. Ein Eintrag ohne Eigentümer geht an den Verantwortlichen der Anwendung laut Konfigurationsdatenbank, ohne Anwendung an die zuständige Bereichsleitung.

Phase 2: Eigentümerschaft herstellen

Erst mit dem Register lässt sich Verantwortung verteilen. Jedes Konto bekommt einen menschlichen Eigentümer, eine Vertretung und eine Verankerung an einer Anwendung. Diese Phase ist organisatorisch und die unbeliebteste, weil sie Fachbereiche zwingt, Konten anzunehmen, die sie nie bewusst bestellt haben. Nach unserer Erfahrung steigt die Bereitschaft, sobald ein Prüfbericht ein eigentümerloses Konto mit Zugriff auf ein kritisches System benennt. Wer diese Phase vor der Prüfung durchläuft, bestimmt den Ton des Gesprächs.

Phase 3: Zielbild und Berechtigungskonzept

Mit bekannten Agenten und benannten Eigentümern legt das Institut fest, wie Agenten gesteuert werden. Das Zielbild beschreibt, welche Bausteine des Kontrollmodells als gemeinsame Dienste bereitstehen: Ausstellung der Identitäten, Richtlinie und Entscheidungspunkt, Durchsetzungspunkte und Protokollierung. Stehen sie früh bereit, baut kein Entwicklungsteam sein eigenes Identitätsmodell. Das Berechtigungskonzept legt Autonomiestufen, Aufgabenkatalog und Stellvertretung fest. Wo Zielsysteme die Stellvertretung noch nicht unterstützen, wird die Lücke im Register vermerkt und in der Risikobewertung berücksichtigt.

Phase 4: Pilot und Betrieb

Die erste Umsetzung soll das Zielbild prüfen. Der Pilot ist klein genug, um schnell zu laufen, und typisch genug, um für spätere Agenten zu stehen, etwa der Agent aus der Kreditorenbuchhaltung. Er deckt alle sechs Ebenen des Kontrollmodells ab und durchläuft den Lebenszyklus von der Registrierung bis zur ersten Rezertifizierung. Was sich bewährt, wird als Muster für die nächsten Agenten übernommen. Die Rezertifizierung läuft in derselben Frist wie für menschliche Konten, und jeder Eigentümerwechsel löst eine außerplanmäßige aus.

Phase 5: Einordnung und Nachweis

Zuletzt wird das Vorgehen gegen die Anforderungen gelegt, die für das Institut gelten: die DORA-RTS mit den Artikeln 12, 20 und 21; für Häuser mit Hochrisiko-KI-Systemen ab Dezember 2027 die KI-Verordnung, als Betreiber in jedem Fall die Artikel 26 und 27, bei eigenen Systemen zusätzlich die Artikel 12, 14 und 15; für bedeutende Institute den Aktionsplan der EZB-Bankenaufsicht. Weil dieser bis zum 31. Oktober 2026 fällig ist, ist er parallel zu den Phasen 1 und 2 aufzusetzen. Die Einordnung zeigt, wo das Institut die Anforderungen erfüllt, und sie zeigt, wo der Verordnungstext allgemein bleibt und das Institut seinen Maßstab selbst begründen muss. Bei Maschinen- und Agentenidentitäten ist das fast überall der Fall. Die meisten Agenten in Instituten sind keine

Hochrisiko-Systeme, auch der Beispielagent aus der Kreditorenbuchhaltung ist keines. Diese Einstufung gehört mit Begründung in die Dokumentation, bevor die Aufsicht danach fragt.

Zehn Fragen, die ein CISO heute beantworten können sollte

Wir stellen in Erstgesprächen die folgenden Fragen. Jede Frage, die Sie nicht sicher und belegt beantworten können, ist eine Aufgabe.

1. Kennen Sie die Zahl der nicht-menschlichen Konten mit Zugriff auf Systeme kritischer oder wichtiger Funktionen?
2. Hat jedes dieser Konten einen namentlich benannten Eigentümer und eine Vertretung?
3. Wissen Sie, wie viele KI-Agenten in Ihrem Haus laufen, auch die von Mitarbeitenden selbst angelegten?
4. Handeln Ihre Agenten unter eigener Identität oder unter der Identität der auslösenden Person?
5. Können Sie für eine beliebige Handlung eines Agenten in unter einer Stunde die auslösende Person, die Vollmacht und alle Zwischenschritte rekonstruieren?
6. Was geschieht mit den Agenten eines Mitarbeitenden, der morgen suspendiert wird?
7. Wann wurde zuletzt ein produktiver Agent bewusst angehalten, und wie lange hat es gedauert?
8. Werden Servicekonten in derselben Frist rezertifiziert wie menschliche Konten?
9. Sind alle Zugangsnachweise Ihrer Agenten nur Minuten gültig?
10. Steht in Ihrer Dokumentation, welchen Maßstab Sie für Agentenidentitäten anlegen und warum?

Glossar

Die Begriffe sind so definiert, wie dieses Papier sie verwendet. Wo ein deutscher Fachbegriff nicht eingeführt ist, steht der englische daneben.

Actor-Claim

Angabe in einem Sicherheitstoken nach RFC 8693, die den handelnden Akteur benennt und vom Auftraggeber unterscheidet. Technische Grundlage der Stellvertretung.

Agent, KI-Agent

Software, die ein Ziel erhält und selbst entscheidet, welche Werkzeuge sie in welcher Reihenfolge aufruft. Ihre Aktionsmenge steht erst zur Laufzeit fest.

Anhalten

Sofortiges Unterbrechen eines Agenten, etwa bei einem Vorfall, mit Widerruf seiner laufenden Zugangsnachweise. Vor dem Produktivgang und danach je Rezertifizierungsperiode zu üben und zu messen.

Arbeitslast (Workload)

Kurzlebige Ausführungseinheit wie ein Container oder eine Funktion, die ihre Identität bei jedem Start neu erhält.

Attestierung

Bestätigung von Eigenschaften einer Laufzeitumgebung durch einen Dritten, etwa der Prüfsumme eines Abbilds. Grundlage einer Ausstellung ohne hinterlegtes Geheimnis.

Aufgabenkatalog

Liste der Werkzeuge und Operationen, die ein Agent für seinen Zweck verwenden darf. Ersetzt für Agenten die Rechteliste und ist die Grundlage der Rezertifizierung.

Autonomiestufe

Einteilung von Agenten nach dem Grad ihrer Selbständigkeit: Beobachten, Beraten, Handeln mit Freigabe, autonomes Handeln.

DPoP

Demonstrating Proof of Possession nach RFC 9449. Bindet ein Token an einen Schlüssel des Clients, sodass ein entwendetes Token allein nicht verwendbar ist.

Durchsetzungspunkt (Policy Enforcement Point)

Stelle, die eine Aktion zulässt oder abweist, etwa ein Gateway vor Werkzeugen oder Schnittstellen. Sie fragt die Entscheidung beim Entscheidungspunkt an.

Eigentümer

Natürliche Person, die fachlich für Zweck, Katalog und Verhalten einer nicht-menschlichen Identität verantwortlich ist. In Entra Agent ID heißt die fachli-

che Rolle Sponsor, in der SailPoint Identity Security Cloud Primary Owner, in SailPoint IdentityIQ Administrator.

Entscheidungspunkt (Policy Decision Point)

Zentrale Stelle, die Anfragen gegen die Richtlinie prüft und entscheidet. Die AuthZEN Authorization API standardisiert die Anfrage an ihn.

Identitätsübernahme (Impersonation)

Ein Agent handelt mit dem Token einer Person und ist im Zielsystem von ihr nicht unterscheidbar.

Konfigurationsdatenbank (CMDB)

Verzeichnis der Anwendungen und Systeme eines Hauses mit ihren Verantwortlichen. Der dort eingetragene Verantwortliche der Anwendung ist das zweite Glied der Nachfolgekette.

Kontrollmodell

Die sechs Ebenen, in denen dieses Papier die Kontrollen für Agenten ordnet: Sichtbarkeit, Verantwortung, Identität, Vollmacht, Durchsetzung und Nachweis.

Least Privilege

Grundsatz, dass eine Identität nur die Rechte erhält, die sie für ihren Zweck braucht. Bei Agenten nur über Aufgabenkatalog, Rechte je Aufgabe und Entscheidung je Aktion umsetzbar.

Model Context Protocol (MCP)

Offenes Protokoll, über das Agenten Werkzeuge und Datenquellen ansprechen. Seine Autorisierungsspezifikation behandelt Server als OAuth-Ressourcenserver und verbietet das Weiterreichen von Tokens.

Nachfolgekette

Reihenfolge, in der die Eigentümerschaft eines Agenten übergeht, wenn der Eigentümer ausfällt: benannte Vertretung, Verantwortlicher der Anwendung, Linienvorgesetzter mit Frist, danach Eskalation nach oben. Der Agent läuft dabei weiter.

Nicht-menschliche Identität (Non-Human Identity)

Sammelbegriff für Identitäten, die keine natürliche Person repräsentieren. Dieses Papier unterscheidet drei Arten: Servicekonten, Arbeitslasten und Agenten. Schlüssel, Zertifikate, Tokens und OAuth-Zustimmungen sind ihre Zugangsnachweise.

Rezertifizierung

Regelmäßige Bestätigung durch den Eigentümer, dass eine Identität noch gebraucht wird und ihre Rechte angemessen sind. Der übliche Weg, die Pflicht der DORA-RTS zur Aktualisierung von Zugangsrechten zu erfüllen, mindestens jährlich, bei kritischen oder wichtigen Funktionen mindestens halbjährlich.

SPIFFE

Herstellerübergreifender Standard für kurzlebige, kryptografisch prüfbare Identitäten von Arbeitslasten, mit SPIRE als verbreiteter Implementierung.

Stellvertretung (Delegation)

Ein Agent handelt unter eigener Identität mit mitgeführter Vollmacht einer Person. Im Protokoll steht, dass Agent A im Auftrag von Person B gehandelt hat.

Stilllegung

Dauerhafte Außerbetriebnahme eines Agenten: deaktivieren, Nachweise widerrufen, Rechte entziehen, Protokoll aufbewahren. Auslöser sind Fristende ohne Erneuerung, Wegfall des Zwecks oder ein Vorfall, nie allein ein fehlender Eigentümer. Während der Nachfolgekette löst das Fristende nichts aus.

Technischer Betreuer

Rolle für den Betrieb eines Agenten, getrennt vom fachlichen Eigentümer. Reaktiviert und stellt wieder her, nur mit Zustimmung des Eigentümers. In Entra entspricht ihr der Owner.

Transaction Token

Entwurf der IETF für kurzlebige, signierte Tokens, die Identität und Kontext durch eine Aufrufkette innerhalb einer Vertrauensdomäne tragen.

Vertrauensanker (Trust Anchor)

Stelle, gegenüber der eine Arbeitslast oder ein Agent die eigene Identität nachweist, um Zugangsnachweise zu erhalten.

Vollmacht

Rahmen, in dem ein Agent handeln darf: Autonomiestufe, Aufgabenkatalog, Auftraggeber und Befristung. Die konkreten Rechte entstehen je Aufgabe innerhalb dieses Rahmens.

Widerruf (Revocation)

Außerkraftsetzen der Zugangsnachweise einer Identität. Teil des Anhaltens und der Stilllegung.

Zusammengesetzte Entscheidung

Zugriffsentscheidung, die die Rechte der Person und die Vollmacht des Agenten zugleich prüft. Erlaubt ist nur die Schnittmenge.

Quellen

1. BaFin, Vernetzung vervielfacht IT-Risiken, Auswertung der DORA-Vorfalldmeldungen 2025, Fachartikel vom 16. Juli 2026.
2. EZB-Bankenaufsicht, Supervisory Newsletter vom 13. November 2024.
3. EZB-Bankenaufsicht, Schreiben SSM-2026-0301 an alle bedeutenden Institute, 7. Juli 2026.
4. Delegierte Verordnung (EU) 2024/1774, Artikel 20 Absatz 1 und Absatz 2.
5. Delegierte Verordnung (EU) 2024/1774, Artikel 21 Buchstaben a, c, d und e.
6. NIST NCCoE, Accelerating the Adoption of Software and AI Agent Identity and Authorization, Konzeptpapier, Entwurf vom Februar 2026, Kommentierung ab 5. Februar 2026.
7. RFC 8693, OAuth 2.0 Token Exchange, Januar 2020, Abschnitte 1.1, 2.1 und 4.1.
8. Google Threat Intelligence, Widespread Data Theft Targets Salesforce Instances via Salesloft Drift, 26. August 2025.
9. FINRA, Cybersecurity Alert: Salesloft Drift AI Supply Chain Attack, ohne Datumsangabe.
10. BSI, Die Lage der IT-Sicherheit in Deutschland 2025, Kapitel Datenleaks.
11. ENISA Threat Landscape 2025, Version 1.2, Seite 11.
12. Bitkom, Wirtschaftsschutz 2026, Pressemitteilung vom 26. August 2026.
13. Gartner, Top Predictions for IT Organizations and Users in 2025 and Beyond, Pressemitteilung vom 22. Oktober 2024.
14. OWASP GenAI Security Project, OWASP Top 10 for Agentic Applications for 2026, 9. Dezember 2025, Risiken ASI01 bis ASI03.
15. Delegierte Verordnung (EU) 2024/1774, Artikel 12 Absatz 2 Buchstaben c, d und f.
16. BaFin, DORA kommt: Änderungen bei den aufsichtlichen Anforderungen an die IT, Meldung vom 10. Januar 2025.
17. Verordnung (EU) 2024/1689 (KI-Verordnung), Artikel 12, 14, 15, 26 und 27 sowie Anhang III Nummer 5.
18. Verordnung (EU) 2026/1744 zur Änderung der KI-Verordnung (Digital Omnibus), in Kraft seit 27. Juli 2026.
19. RFC 8707, Resource Indicators for OAuth 2.0, Februar 2020.
20. Model Context Protocol, Specification 2026-07-28, Authorization, einschließlich Scope Challenge Handling.
21. NIST Special Publication 800-207, Zero Trust Architecture, August 2020.
22. SPIFFE und SPIRE, Projektdokumentation, spiffe.io, abgerufen im September 2026.
23. Microsoft Learn, Agent OAuth flows: On-behalf-of flow, Microsoft Entra Agent ID, aktualisiert 10. August 2026.
24. RFC 9449, OAuth 2.0 Demonstrating Proof of Possession (DPoP), September 2023.
25. RFC 8705, OAuth 2.0 Mutual-TLS Client Authentication and Certificate-Bound Access Tokens, Februar 2020.
26. IETF, Transaction Tokens, draft-ietf-oauth-transaction-tokens-08, 2. März 2026, inzwischen Revision 11 vom 30. Juli 2026.
27. IETF, Identity Assertion JWT Authorization Grant, draft-ietf-oauth-identity-assertion-authz-grant-04, 21. Mai 2026.
28. RFC 9396, OAuth 2.0 Rich Authorization Requests, Mai 2023.
29. Gartner, Gartner Says Applying Uniform Governance Across AI Agents Will Lead to Enterprise AI Agent Failure, Pressemitteilung vom 26. Mai 2026.
30. OpenID Foundation, AuthZEN Working Group Drafts: Access Request and Approval Profile, 15. Juni 2026.
31. Microsoft Learn, Overview of permissions and consent in the Microsoft identity platform, aktualisiert 18. März 2025.
32. OpenID Foundation, Authorization API 1.0, Final Specification, im Januar 2026 verabschiedet, bekanntgegeben am 12. Januar 2026.

33. BSI, Umsetzungshinweise zum Baustein ORP.4 Identitäts- und Berechtigungsmanagement, Edition 2022.
34. Center for Internet Security, CIS Controls v8.1, Safeguard 5.5, Establish and Maintain an Inventory of Service Accounts, abgerufen im September 2026.
35. Ridgeline Cyber, Service Principal Ownership Is the Attack Path Nobody Governs, 12. Mai 2026.
36. Cloud Security Alliance, Agentic AI Identity and Access Management: A New Approach, 18. August 2025.
37. Microsoft Entra Blog, AI agents are everywhere, are your access controls ready, 15. Juli 2026.
38. CISA und Partnerbehörden, Careful Adoption of Agentic AI Services, 1. Mai 2026, Seite 15.
39. Microsoft Learn, What's new in Microsoft Entra Agent ID, Stand 1. Mai 2026.
40. Microsoft Learn, Use Microsoft Entra agent identities in Copilot Studio, aktualisiert 15. September 2026.
41. Microsoft Learn, Microsoft Entra Agent ID governance overview, 5. Juni 2026.
42. Microsoft Learn, Microsoft Entra Agent ID logs, aktualisiert 3. September 2026.
43. Microsoft Learn, Agent owners, sponsors, and managers, 16. April 2026, aktualisiert 21. Juli 2026.
44. Microsoft Learn, Conditional Access for agents, aktualisiert 30. September 2026.
45. Microsoft Learn, Continuous access evaluation for workload identities, 22. April 2026.
46. Microsoft Learn, Risky agents in Microsoft Entra ID Protection, 17. Juni 2026.
47. Microsoft Learn, Access packages for agent identities, 1. Mai 2026.
48. Microsoft Learn, Best practices for Microsoft Entra Agent ID, abgerufen im September 2026.
49. Microsoft Learn, Lifecycle workflow tasks for agent sponsors, aktualisiert 3. September 2026.
50. SailPoint Developer Community, New capability: Agent Identity Security, 6. Oktober 2025.
51. SailPoint, Pressemitteilung vom 4. August 2026, sowie Developer Community, New product: SailPoint Agentic Fabric, 4. August 2026.
52. SailPoint Documentation, Managing AI agents, Identity Security Cloud, abgerufen im September 2026.
53. SailPoint Developer Community, Agentic Fabric, Abschnitt zur Anbindung von IdentityIQ über ein Plugin, 4. August 2026.
54. SailPoint Community, IdentityIQ 8.5 Release Notes, Juli 2025.
55. SailPoint, Ankündigung von IdentityIQ 9.0 auf der Navigate 2026, 6. Oktober 2026.
56. SailPoint Documentation, IdentityIQ, RPA/Bot identities, abgerufen im September 2026.
57. SailPoint Documentation, IdentityIQ Rapid Setup, Leaver configuration, abgerufen im September 2026.
58. Microsoft Learn, Microsoft Purview for Agent 365, 1. Mai 2026.
59. Palo Alto Networks, 2026 Identity Security Landscape, Mai 2026 (109 zu 1, Befragung); Entro Security Labs, 2025 State of Non-Human Identities (92 zu 1, eigene Daten und Umfragen); Cloud Security Alliance im Auftrag von Astrix, State of Non-Human Identity Security, September 2024 (20 zu 1, Befragung). Alle Angaben von Herstellern erhoben oder beauftragt.

Die Fragen aus Kapitel 9 beantworten wir mit Ihnen.

Die WertFels GmbH berät Banken, Versicherer und andere regulierte Unternehmen herstellerneutral bei Identity und Access Management, von der Bestandsaufnahme nicht-menschlicher Identitäten über Kontrollmodelle und Berechtigungskonzepte für KI-Agenten bis zur Einordnung gegen DORA und die KI-Verordnung. Wir arbeiten mit Keycloak, Microsoft Entra ID, Omada, One Identity Manager, SailPoint IdentityIQ, SailPoint Identity Security Cloud und Saviynt sowie mit PAM-Lösungen wie CyberArk.



WertFels GmbH
info@wertfels.com
+49 89 244102321
wertfels.com

Dieses Papier gibt den Stand vom September 2026 wieder, ergänzt um Korrekturen vom Oktober 2026. Rechtsaussagen sind Auslegungen der WertFels GmbH und ersetzen keine Rechtsberatung. Produktangaben stammen aus der Herstellerdokumentation mit dem jeweils genannten Stand.

