UNIVERSITY OF TARTU · Estonian Centre of Excellence in AI · Nortal

Kratt-Nation

WIE AUFGABEN ZWISCHEN MENSCHEN, KI-AGENTEN UND REGELBASIERTER AUTOMATISIERUNG VERTEILT WERDEN KÖNNEN

DISKUSSIONSPAPIER 2026



Kratt-Nation

WIE AUFGABEN ZWISCHEN MENSCHEN, KI-AGENTEN UND REGELBASIERTER AUTOMATISIERUNG VERTEILT WERDEN KÖNNEN

DISKUSSIONSPAPIER 2026

In der estnischen Folklore sammelt ein Kratt Reichtümer für seinen Schöpfer, der aber wissen muss, wie er ihn unter Kontrolle hält. Heutige KI-Agenten können Rechnungen vorbereiten, Leistungen beantragen und weitere Aufgaben bearbeiten, die sich nur schwer mit festen Regeln beschreiben lassen. In einer Zeit von Personal- und Ressourcenknappheit können Ländern damit wertvolle zusätzliche Kapazitäten verschafft werden. Ausgehend von Estlands Erfahrungen untersucht dieses Papier, wie sich Aufgaben konform mit Prinzipien von Demokratie und Rechtsstaatlichkeit zwischen Menschen, KI-Agenten und regelbasierter Automatisierung verteilen lassen.

Die Person oder Organisation, die einen KI-Agenten einsetzt, bleibt für dessen Einsatz verantwortlich. Diese Verantwortung braucht wirksame Kontrollmöglichkeiten: Ein Zahlungslimit muss im Zahlungssystem durchgesetzt werden; es genügt nicht, es in die Anweisungen an den KI-Agenten zu schreiben. Öffentliche und private Dienstanbieter sollten jede vorgesehene Aufgabe erproben, Fehler messen und selbstständiges Handeln erst dann ausweiten, wenn die Ergebnisse dies rechtfertigen. Für jeden Dienst sind Anforderungen an Zugangswege und Betriebsfortführung festzulegen, die ihrem Zweck und dem geltenden Recht entsprechen. Wesentliche Dienste müssen praktisch zugänglich bleiben, wenn KI-Agenten ausfallen.

Estland verfügt bereits über viele der dafür erforderlichen Grundlagen. Andere Länder beginnen unter anderen rechtlichen und technischen Bedingungen. In jedem Fall besteht die unmittelbare Aufgabe darin, die Arbeit innerhalb eines Dienstes auf Regeln, KI-Agenten und Menschen zu verteilen, die notwendigen Kontrollen einzurichten und die gewählte Lösung praktisch zu erproben.

Dieses Diskussionspapier entstand in freiwilliger Zusammenarbeit von Staat, Wissenschaft und Privatwirtschaft. Seine Erarbeitung wurde vom Estnischen Exzellenzzentrum für künstliche Intelligenz EXAI unterstützt (Ministerium für Bildung und Forschung, TK213). Die deutsche Ausgabe basiert auf einer Übersetzung der für ein internationales Publikum bearbeiteten englischen Fassung des estnischen Manuskripts.

Arbeitsgruppe: Andres Raieste (Global Head of Public Sector, Nortal), Dr. Mihkel Solvak (Außerordentlicher Professor für Technologieforschung, Universität Tartu), Keaty Siivelt (Chief Data & AI Officer, Justiz- und Digitalministerium Estlands), Toomas Vaks (Cyber Risk Manager, Swedbank), Madis Tapupere (unabhängiger Experte), Priit Liivak (Chief Government Technology Officer, Nortal) und Tõnu Grünberg (Stellvertretender Generalsekretär für digitale Infrastruktur und Cybersicherheit, Justiz- und Digitalministerium Estlands)

Deutsche Ausgabe: Adaptiert und redigiert von Hendrik Lume (Vice President Digital Government, Nortal).

Zitiervorschlag: Raieste, A., Solvak, M., Siivelt, K., Vaks, T., Tapupere, M., Liivak, P., & Grünberg, T. (2026). Kratt-Nation: Wie Aufgaben zwischen Menschen, KI-Agenten und regelbasierter Automatisierung verteilt werden können.

INHALT

Zusammenfassung **4**

Einleitung **7**

1. Welche Entscheidungen KI-Agenten verändern **12**

2. Wessen Interessen KI-Agenten dienen und was sie entscheiden dürfen **16**

2.1. Auftraggeber (Prinzipal) und KI-Agent **16**

2.2. Regeln, KI-Agenten und Menschen **17**

2.3. Automatisierung und Delegation an KI **18**

2.4. Grenzen werden als Regeln durchgesetzt **19**

2.5. Handlungsklassen und Autonomie **22**

2.6. Entscheidungen, die bei Menschen bleiben **24**

3. Mensch, Organisation und KI-Agent **26**

3.1. Der persönliche KI-Agent **28**

3.2. Drei Zugangswege zu einem Dienst **29**

3.3. Die externe KI-Agentenschnittstelle als bevorzugter Zugang **30**

4. Wofür die Organisation verantwortlich ist **34**

4.1. Interne KI-Agenten **34**

4.2. Parität zwischen Zugangswegen für Menschen und für KI-Agenten **36**

4.3. Fünf Grundsätze für jeden Dienst mit KI-Agenteneinsatz **39**

5. Was der Staat bereitstellen muss **41**

5.1. KI-Agenten ersetzen nicht die Notwendigkeit einer effizienteren Verwaltung **41**

5.2. Systeme der einzelnen Organisationen an bestehende digitale Grundlagen anpassen **43**

5.3. Menschen Einblick in KI-Agentenhandlungen und Datennutzung geben **45**

5.4. Gemeinsame Berechtigungen, wenn wiederkehrender Bedarf sie rechtfertigt **46**

6. Szenarien und Betriebsfortführung in einer Kratt-Nation **50**

Quellenverzeichnis **53**

ZUSAMMENFASSUNG

Dieses Papier konzentriert sich auf die öffentliche Verwaltung und auf regulierte Dienste, etwa im Bank- und Telekommunikationssektor sowie in der Energie- und Wasserversorgung. Es verwendet den Begriff „Organisation" für eine öffentliche oder private Stelle, die für den jeweils betrachteten Dienst, die betreffenden Systeme oder Handlungen verantwortlich ist. Ein Dienst ist die Arbeit, die eine Organisation leistet, um einen bestimmten Bedarf von Kundschaft oder Beschäftigten zu decken, einschließlich der im Hintergrund ausgeführten Schritte. Das Papier untersucht, welche Aufgaben und Entscheidungen an KI-Agenten delegiert werden dürfen und wie sich die Verantwortung verteilen sollte. Ausgangspunkt sind bestehende Organisationen und ihre Handlungsmöglichkeiten innerhalb des geltenden Rechts. KI-Agenten müssen innerhalb der Grenzen von Demokratie und Rechtsstaatlichkeit handeln. Ihre Fähigkeiten verändern diese Grenzen nicht und verleihen ihnen keine Handlungsbefugnis.

KI-Agenten können Aufgaben bearbeiten, die nur schwer mit festen Regeln beschrieben werden können. Sie können Anträge und Nachweise lesen, fehlende Angaben erkennen, einen Entwurf eines Bescheids mit Quellenverweisen erstellen und die Antwort einer Organisation in allgemein verständlicher Sprache erklären. Ein persönlicher KI-Agent kann auch handeln, während die Person, die ihn nutzt, abwesend ist. Diese Fähigkeiten können besonders dort nützlich sein, wo eine alternde Bevölkerung oder Personalmangel öffentliche Verwaltungen unter Druck setzen. Inwiefern KI-Agenten weitergehenden Nutzen erzeugen können, hängt von einer klaren Verantwortung für ihre Arbeit und der Qualität des gesamten Dienstes ab.

Software kann die rechtlichen Pflichten einer Person oder Organisation nicht übernehmen. Wer KI-Agenten einsetzt, behält seine Verantwortung; für einen konkreten Schaden können zugleich der KI-Agentendienstbetreiber, der Hersteller oder andere Beteiligte haften. Wir schlagen vier Bedingungen für den Einsatz von KI-Agenten vor:

1. Ein KI-Agent darf keine elektronischen Authentifizierungs- oder Signaturmittel einer Person erhalten.

2. Seine Grenzen müssen von der Organisation durchgesetzt werden, in deren System der KI-Agent Handlungen vornimmt.

3. Anforderungen an die Abdeckung durch Zugangskanäle und an die Betriebsfortführung müssen sich nach dem jeweiligen Dienst und dem jeweils geltenden Recht richten; wesentliche Dienste müssen bei einem Ausfall der KI-Agenten praktisch zugänglich bleiben.

4. Entscheidungen, die das Recht oder die Vorgaben der Organisation Menschen vorbehalten, müssen bei Menschen bleiben. Betroffene müssen eine praktisch nutzbare Möglichkeit haben, Entscheidungen anzufechten.

Eine Organisation sollte jeden Schritt eines Dienstes prüfen und entscheiden, ob er von herkömmlicher Software nach vorab festgelegten Regeln, von einem KI-Agenten oder von einem Menschen ausgeführt werden soll. Für stabile, präzise beschreibbare Aufgaben eignet sich gewöhnlich regelbasierte Automatisierung. Ein KI-Agent

ist dort nützlicher, wo Eingaben variieren und ausgelegt werden müssen. Bei einer Förderleistung können beispielsweise die objektiven Fördervoraussetzungen regelbasiert automatisch geprüft werden, während ein KI-Agent eine Bewertung anhand der Programmziele vorbereitet und eine zuständige Amtsperson letztendlich die Ermessensentscheidung trifft.

Eine Anweisung fordert den KI-Agenten auf, eine bestimmte Grenze einzuhalten. Ihre Durchsetzung erfordert, dass das empfangende System Handlungen jenseits dieser Grenze blockiert. Dokumente oder Websites können Anweisungen enthalten, die einen KI-Agenten von seiner Aufgabe abbringen. Sie sollten jedoch nicht in der Lage sein, beispielsweise die Berechtigungsregeln eines Zahlungssystems zu verändern. Der Umfang der selbstständigen Handlungsmöglichkeiten eines KI-Agenten sollte von den Fehlerfolgen und den praktischen Möglichkeiten ihrer Behebung abhängen. Muss jede geringfügige Handlung genehmigt werden, droht die Prüfung zum routinemäßigen Klick zu werden. Organisationen sollten den KI-Agenten zunächst parallel zu ihren Beschäftigten erproben, Fehler anhand geeigneter Stichproben messen und jede geplante Ausweitung vor ihrer Freigabe testen. Schwere Fehler oder schlechtere Ergebnisse sollten zur Einschränkung oder Aussetzung der Berechtigung führen.

KI-Agenten, die Zugangsdaten einer Person auf einer gewöhnlichen Website nutzen, sind für die Organisation nur schwer von dieser Person zu unterscheiden. Eine eigene KI-Agentenschnittstelle ermöglicht es der Person, bestimmte Tätigkeiten im System der betreibenden Organisation zu autorisieren. Die Organisation prüft jede Handlung anhand dieser Berechtigung und behält die erforderlichen Schritte einer menschlichen Bestätigung vor. So erhält sie eine klarere Kontrollgrundlage und kann den KI-Agentenzugriff einschränken, während der Zugang für Menschen offen bleibt.

Organisationen müssen geltende Informationspflichten erfüllen, wenn Menschen mit KI interagieren. Ihre KI-Agenten im direkten Kundenkontakt sollten außerdem angeben, für welche Organisation sie tätig sind. Für jeden Dienst mit KI-Agenteneinsatz sollte eine gesamtverantwortliche Person benannt sein. Für jeden KI-Agenten braucht es zudem jemanden, der seine Grenzen durchsetzt und ihn sofort stoppen kann. Beschäftigte sollten wesentliche Aufgaben während eines Ausfalls weiterführen können. Zeit und Kosten eines Modellanbieterwechsels sollten bekannt und praktisch überprüft sein. Die durch einen KI-Agenten frei werdende Arbeitszeit schafft Kapazität, führt aber nicht automatisch zu Haushaltseinsparungen. Die Leitung sollte entscheiden, ob sie damit Wartezeiten verkürzt, die Dienstqualität verbessert oder andere Aufgaben bearbeitet.

Zu den notwendigen Grundlagen gehören eine verlässliche Identifizierung, die Prüfung der Befugnisse einer Person, rechtmäßiger Datenzugriff und Systeme, die KI-Agentensitzungen von menschlicher Nutzung unterscheiden. Wo diese Grundlagen bestehen, können Organisationen einzelne Dienste anpassen. Andernorts sind fehlende Fähigkeiten innerhalb eines begrenzten Anwendungsfalls aufzubauen. Der Staat kann gemeinsame Leitlinien bereitstellen, damit nicht jede Organisation Einwilligungsdialoge und Berechtigungslogik allein entwickeln muss. Menschen sollten sowohl die in ihrem Namen ausgeführten Handlungen als auch die verwendeten Daten einsehen können. Eine von mehreren Organisationen akzeptierte gemein-

same Berechtigung lohnt sich dann, wenn wiederkehrender Bedarf ihre Kosten und Komplexität rechtfertigt.

Ein persönlicher KI-Agent kann Menschen dabei helfen, sich in fragmentierten Diensten zurechtzufinden, beseitigt aber deren Fragmentierung nicht. Organisationen sollten die Dienste selbst vereinfachen und vereinbaren, wer eine vollständige Lösung für einen bestimmten Bedarf koordiniert. Andernfalls verlagern sich die Kosten auf diejenigen, die den KI-Agenten finanzieren: die nutzende Person oder den Anbieter. Zugleich können leichter erstellbare Anträge und Beschwerden den Arbeitsaufwand öffentlicher Stellen erhöhen.

Manche Entscheidungen bleiben politisch. Regierungen müssen Wachstum, gesellschaftliche Auswirkungen, zusätzliche Kapazität und finanzielle Einsparungen gegeneinander abwägen. Gesetzgebende Organe müssen sich gegebenenfalls mit Haftungsfragen und Konflikten zwischen den kommerziellen Interessen eines Anbieters von KI-Agenten und den Interessen der Nutzenden befassen. Gesellschaften müssen entscheiden, welche Rollen sie Menschen vorbehalten wollen. Länder mit technisch weniger gut unterstützten Sprachen stehen zudem vor Entscheidungen über Sprachqualität, Kosten und den Zugang zu Trainingsmaterial. Ob ein Dienst ohne KI-Agenten genauso schnell und gut sein soll, bleibt eine eigenständige politische Entscheidung.

Organisationen können innerhalb ihrer bestehenden Befugnisse beginnen. Für einen ausgewählten Dienst sollten sie festlegen, was ein KI-Agent tun darf, wer das Ergebnis prüft und wie der Dienst ohne ihn fortgeführt oder wiederhergestellt wird. Die vorgeschlagenen Handlungsklassen und Abdeckungsmaße sollen in realen Diensten erprobt werden. Als landesweites Modell sind sie bislang nicht validiert.

EINLEITUNG

Dieses Papier fragt, welche Aufgaben und Entscheidungen eine Person oder Organisation einem KI-Agenten unter welchen Bedingungen anvertrauen sollte und wer verantwortlich bleibt. Es richtet sich vor allem an politische Entscheidungsträgerinnen und Entscheidungsträger sowie an Führungskräfte, die für öffentliche Dienste verantwortlich sind. Für die Fachleute, die sie rechtlich und technisch beraten, enthält es vertiefende Erläuterungen. Die für einen Dienst oder dessen Infrastruktur verantwortliche Organisation sollte regeln, wie KI-Agenten ihn nutzen. Menschen sollten einen tatsächlich nutzbaren Zugang zu dem Dienst und eine Möglichkeit behalten, dessen Entscheidungen anzufechten.

In der estnischen Folklore ist ein Kratt ein aus alten Haushaltsgegenständen geschaffenes Wesen, das für seinen Schöpfer Reichtümer zusammenträgt. Ähnliche Motive finden sich in anderen europäischen Überlieferungen: beim deutschen Geldmännlein und bei der schwedischen Bjära. Eine lockerere Parallele bietet die englische Gestalt Puck, ein schelmischer Geist. Andrus Kivirähk bringt die Schwierigkeit, mit einem Kratt umzugehen, in seinem Roman *Rehepapp ehk november* (2000) auf den Punkt: „Du verstehst es nicht, mit einem Kratt umzugehen."[1]

Im Zeitalter der KI-Agenten regen diese alten Geschichten praktische Fragen an: Wessen Interessen dient ein Helfer, wer trägt die Kosten und wie lassen sich seine Handlungen begrenzen? Dieses Papier untersucht diese Fragen anhand der Arbeitsteilung: Welche Aufgaben sollten durch regelbasierte, deterministische Automatisierung übernommen werden, welche lassen sich KI-Agenten anvertrauen und welche Entscheidungen sollten bei Menschen bleiben? Diese Regelungen müssen sich im Rahmen von Demokratie und Rechtsstaatlichkeit bewegen, mit klarer Verantwortung und Grenzen, die in den Systemen durchgesetzt werden, über die KI-Agenten handeln.

Ein KI-Agent bezeichnet in diesem Papier Software, die eine Aufgabe interpretiert und mithilfe von Werkzeugen Handlungen ausführen kann. In der Analyse verwenden wir durchgehend den internationalen Begriff KI-Agent; Kratt bleibt die kulturelle Bezeichnung.

Anweisungen können das Verhalten eines KI-Agenten lenken. Durchsetzbare Grenzen gehören jedoch in die Systeme, über die er handelt. Eine Person, die einen kommerziellen KI-Assistenten nutzt, und eine Organisation, die einen öffentlichen Dienst erbringt, tragen unterschiedliche Verantwortung. Kapitel 3 erläutert dieses Verhältnis aus Sicht der Nutzenden. Die übrigen Kapitel behandeln Entscheidungen der Organisationen und die Unterstützung durch den Staat.

Die Vorschläge sollen innerhalb des geltenden Rechts zusammenwirken. Die vier Bedingungen aus der Zusammenfassung sind Anforderungen, für die dieses Papier eintritt; identische Rechtspflichten gelten nicht überall. Handlungsklassen, evidenzgestützte Änderungen des Autonomieumfangs und Maße der Dienstabdeckung

1 Vergleichbare Figuren sind Schwedens Skratte und Bjära, Deutschlands Schrat und Geldmännlein, Polens Skrzat sowie, als lockerere Parallele, Englands Puck. Zitat: Kivirähk, Rehepapp ehk november (2000), für dieses Papier aus dem Estnischen übersetzt.

müssen praktisch erprobt werden. Rechtliche Beispiele sind als Beispiele aus der EU oder Estland gekennzeichnet und vor einer Übertragung auf andere Länder anhand des dortigen Rechts zu prüfen. Bestehende Schutzvorschriften können bereits eine menschliche Beteiligung verlangen.[2] Empfehlungen formulieren wir mit „sollte"; bei rechtlichen Anforderungen benennen wir Quelle und Anwendungsbereich.

KI-Agenten können Dienste über Organisationsgrenzen hinweg verbinden. Diese Grenzen schützen allerdings mitunter berechtigte Interessen. Estlands Datenaustauschplattform X-tee (X-Road) verbindet beispielsweise verteilte Systeme, anstatt eine einzige zentrale Datenbank zu schaffen. Der Zugriff auf Daten einer anderen Organisation braucht weiterhin eine geeignete Grundlage und Vereinbarung.[3] Regierungen können Verwaltungsstrukturen reformieren und Dienste vereinfachen und dabei klare Verantwortung und Kontrolle über die Datennutzung erhalten. Demokratisch legitimierte Befugnisse müssen sich auch in der technischen Gestaltung widerspiegeln: Wer darf entscheiden, was darf delegiert werden und wer kann eingreifen?[4] Die Fähigkeiten eines KI-Agenten verändern die rechtliche Zuständigkeit einer Behörde nicht.

Identitätssysteme, Verwaltungsrecht und die Organisation von Diensten unterscheiden sich zwischen Ländern. Kapitel 5 trennt die notwendigen Funktionen von den bestehenden und vorgeschlagenen estnischen Lösungen, die sie veranschaulichen. Das Papier bietet weder eine allgemeine Darstellung von KI und agentenbasierter KI noch ein vollständiges Programm für Finanzierung, Beschaffung und Kompetenzen. Es geht von Fähigkeiten aus, die sich heute erproben lassen. Organisationen können über die Arbeitsverteilung entscheiden, ohne die künftigen Grenzen der Technologie vorherzusagen.

Einen Dienst mit KI-Agenten zu organisieren und zugleich die Verantwortung der Organisation zu wahren, umfasst drei Schritte (Abbildung 1). Zunächst wird jeder Arbeitsschritt einer Regel, einem KI-Agenten oder einem Menschen zugeordnet. Danach werden die Berechtigungen des KI-Agenten und der Umfang seines selbstständigen Handelns festgelegt. Schließlich werden Prüfung, Verantwortung sowie Fortführung oder Wiederherstellung der Dienste organisiert.

---

2 Beispielsweise Artikel 22 DSGVO und Artikel 18 Absatz 8 der Verbraucherkreditrichtlinie. Diese Schutzvorschriften haben bestimmte Anwendungsbereiche und verbieten nicht allgemein jede automatisierte Entscheidung.

3 Estnische Behörde für Informationssysteme (RIA), „Introduction of X-tee".

4 Raieste, Solvak, Velsberg und McBride, „Government Efficiency in the Age of AI: Toward Resilient and Efficient Digital Democracies", 2025.

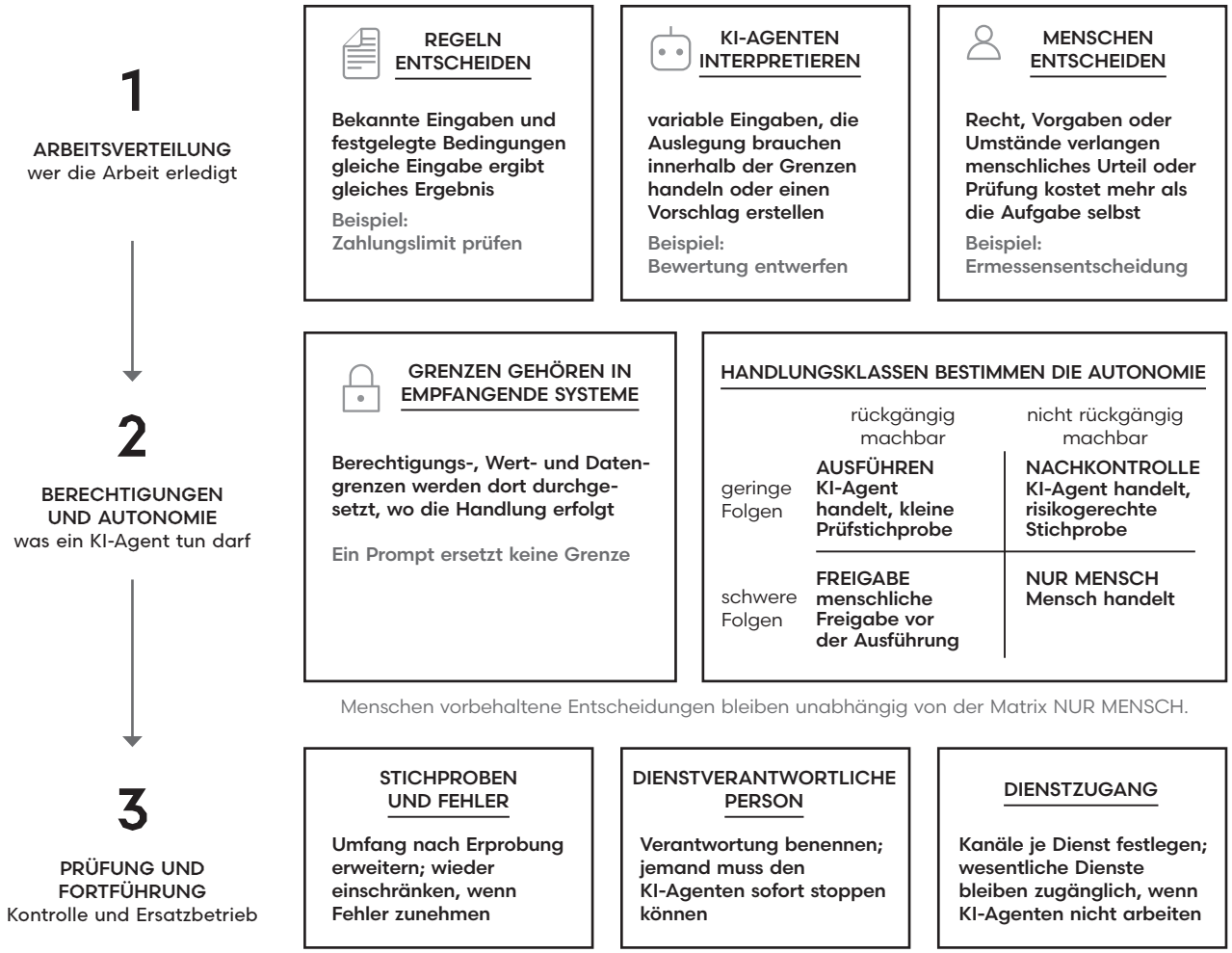


Abbildung 1. Drei Schritte bei der Organisation eines Dienstes mit KI-Agenten: Arbeit verteilen, Berechtigungen und Autonomie festlegen, Prüfung und Betriebsfortführung organisieren.

BEGRIFF	BEDEUTUNG IN DIESEM PAPIER
KI-Agent / Kratt	KI-basierte Software, die eine Aufgabe interpretiert und mithilfe von Werkzeugen Handlungen ausführen kann. Kratt ist die im Titel beibehaltene estnische Kulturbezeichnung. Zusammenwirkende KI-Agenten werden bei der Zuordnung von Kontrollen als ein Gesamtsystem betrachtet; die Rollen ihrer Anbieter bleiben dabei bestehen.
Auftraggeber (Prinzipal)	Die beauftragende Person oder Organisation, die einen KI-Agenten einsetzt und ihm eine Aufgabe überträgt. Der KI-Agenteneinsatz hebt ihre Pflichten nicht auf. Der Begriff unterstellt keine Rechtspersönlichkeit der Software.
Organisation	Eine öffentliche oder private Stelle, die für den jeweils betrachteten Dienst, die betreffenden Systeme oder Handlungen verantwortlich ist.
KI-Agentendienstbetreiber	Die Organisation, die den KI-Agenten als Dienst betreibt, beispielsweise ein Anbieter kommerzieller Assistenten.
Modellanbieter	Die Organisation, die das vom KI-Agenten genutzte Modell bereitstellt. Diese Rolle kann sich von der des KI-Agentendienstbetreibers und des Anbieters des vollständigen KI-Systems unterscheiden.
Verantwortungsbereich	Der Teil einer Gesamtlösung, für den eine Person oder Organisation verantwortlich ist: die Aufgaben und Berechtigungen auf Nutzerseite oder die Systeme und KI-Agenten der Organisation. Verantwortung erfordert wirksame Kontrollmöglichkeiten. Dies ist keine abschließende Zuordnung rechtlicher Haftung.
KI-Serviceagent der Organisation	Der KI-Agent der Organisation im direkten Kundenkontakt, der innerhalb ihres Aufgabenbereichs und ihrer Grenzen handelt. In diesem Papier bezeichnet der Ausdruck „die KI-Agenten der Organisation" ihre KI-Serviceagenten und internen KI-Agenten.
Interner KI-Agent	Ein in die Systeme einer Organisation eingebundener KI-Agent, der festgelegte Aufgaben der Beschäftigten unterstützt oder ausführt.
Externe KI-Agentenschnittstelle	Eine Schnittstelle, über die eine Organisation externen KI-Agenten bestimmte Handlungen ermöglicht und jede Handlung auf ihre Autorisierung prüft.
Berechtigung (Authorisation)	Eine maschinenlesbare technische Erlaubnis, die eine authentifizierte Person mit den erforderlichen Rechten im eigenen System der Organisation erteilt. Sie legt fest, in wessen Namen ein KI-Agent handeln darf, was er tun darf und innerhalb welcher Grenzen. Sie begründet für sich genommen weder rechtliche Vertretungsmacht noch eine Haftungszuordnung.
Benutzerschnittstelle für Menschen und Zugang ohne persönlichen KI-Agenten	Die Benutzerschnittstelle für Menschen ist eine Website, Anwendung oder ein durch Beschäftigte unterstützter Zugangskanal der Organisation. Der Zugang ohne persönlichen KI-Agenten zeigt, welche Bedarfe sich ohne einen solchen KI-Agenten decken lassen. Anforderungen an Kanalabdeckung und Betriebsfortführung werden je Dienst festgelegt; wesentliche Dienste müssen praktisch zugänglich bleiben.
Handlungsklasse	Ein vorgeschlagenes Kontrollschema für einen einzelnen Schritt: AUSFÜHREN (ACT), NACHKONTROLLE (REVIEW), FREIGABE (APPROVE) oder NUR MENSCH (HUMAN-ONLY). Die Folgen und die praktische Möglichkeit, eine Handlung rückgängig zu machen, bestimmen die Klasse mit. Schritte, die das Recht oder die Vorgaben der Organisation Menschen vorbehalten, haben Vorrang vor der Matrix. NACHKONTROLLE bedeutet eine nachträgliche, risikogerechte Stichprobenprüfung; FREIGABE verlangt menschliche Zustimmung vor der Ausführung.
Dienstparität	Ein Maß dafür, welche Bedarfe die Zugangswege für Menschen und für KI-Agenten abdecken und für wie viele dieser Bedarfe ausdrücklich über Kanalwahl und KI-Agentenrolle entschieden wurde. Die Abdeckung wird mit den Zielen für den jeweiligen Dienst und die jeweilige Rolle verglichen. KI-Agenten erhalten dadurch nicht dieselben Befugnisse wie Menschen.
Dienstverantwortliche Person (Service Owner)	Die Person, die für den gesamten Dienst einsteht und Änderungen sowie Abhilfe veranlassen kann. Bei einem gemeinsamen Dienst koordiniert sie die Beteiligten, ohne deren rechtliche Zuständigkeiten zu übernehmen.

Tabelle 1. Begriffe in diesem Papier.

Kapitel 1 betrachtet die weitergehenden Entscheidungen, die KI-Agenten beeinflussen. Kapitel 2 behandelt die Arbeitsverteilung und Kontrollen außerhalb des Modells. Kapitel 3 und 4 untersuchen zunächst den Umgang von Menschen und ihren KI-Agenten mit Organisationen, anschließend die eigenen KI-Agenten, Systeme und Verantwortlichkeiten der Organisation. Gemeinsame Unterstützung und Infrastruktur sind Gegenstand von Kapitel 5. Kapitel 6 führt die Vorschläge in zwei hypothetischen Szenarien zusammen. Tabelle 1 definiert die wichtigsten Begriffe.



1. WELCHE ENTSCHEIDUNGEN KI-AGENTEN VERÄNDERN

Die KI-Forschung konzentriert sich häufig auf Arbeit und Produktivität. Studien berichten von erheblichen Verbesserungen bei manchen wissensintensiven Aufgaben; die Ergebnisse hängen von Aufgabe, Person und Organisation ab.[5] Diese Wirkungen sind für öffentliche Einrichtungen bedeutsam, die zugleich über Verantwortung, Kultur und staatliche Abhängigkeiten entscheiden müssen. Bevor festgelegt wird, was ein KI-Agent tun darf, hilft die Unterscheidung von vier Dimensionen: Wirtschaft, Recht, Kultur sowie Souveränität und Sicherheit. Tabelle 2 trennt die Entscheidungen von Einzelpersonen, Organisationen und Staat.

Rein wirtschaftlich betrachtet kann ein KI-Agent das Arbeitsvolumen erhöhen, das eine Person bewältigt. Wo Personal fehlt, kann diese zusätzliche Kapazität selbst ein politisches Ziel sein. Jevons' Untersuchung der Ressourceneffizienz veranschaulicht, warum kostengünstigere Tätigkeiten auch die Nachfrage steigern können.[6] Günstigere Codeerzeugung kann beispielsweise dazu führen, dass mehr Software entwickelt wird. Freiwerdende Arbeitszeit senkt daher nicht automatisch die Ausgaben. Für jeden Dienst sollte die Leitung entscheiden, wie sie die Kapazität nutzt, und abgeschlossene Fälle sowie Ergebnisse messen. Regierungen sollten festlegen, ob Wachstum und weitergehende gesellschaftliche Ergebnisse oder Kosteneinsparungen und organisatorische Effizienz im Vordergrund stehen. Diese Ziele brauchen unterschiedliche politische Maßnahmen und Anreize; kürzere Warteschlangen sind ein mögliches Ergebnis.

Rechtlich überträgt der Einsatz von Software die Pflichten der einsetzenden Person oder Organisation nicht auf die Software. Die Verantwortung für den Betrieb eines Dienstes, die Pflicht, dessen Entscheidungen zu erklären und zu korrigieren, und die Haftung für Schäden sind unterschiedliche Fragen. Die überarbeitete EU-Produkthaftungsrichtlinie erfasst ausdrücklich Software einschließlich KI-Systemen. Ihre neue Regelung betrifft jedoch Produkte, die nach dem 8. Dezember 2026 in Verkehr gebracht oder in Betrieb genommen werden, und bedarf der Umsetzung in nationales Recht.[7] Auch andere geltende Vorschriften ordnen Pflichten nach Rollen zu.[8] Je nach Sachverhalt und Rechtsordnung können ein KI-Agentendienstbetreiber, Hersteller, Aussteller von Berechtigungsnachweisen oder eine empfangende Organisation haften. Die von einer Person erteilte Berechtigung allein entscheidet nicht über diese Zuordnung.

Betriebliche Verantwortung braucht praktische Kontrollmöglichkeiten. Jeder Dienst mit KI-Agenteneinsatz sollte eine verantwortliche Person haben, die die Abläufe des Dienstes ändern und bei Problemen Abhilfe koordinieren kann. Daraus folgt nicht, dass diese Person für jeden Schaden allein haftet. Gesetzgebende Organe sollten prüfen,

---

5 Überblick: Calvino, Reijerink und Samek, „The Effects of Generative AI on Productivity, Innovation and Entrepreneurship", OECD, 2025. In Experimenten benötigten berufliche Schreibaufgaben 40 Prozent weniger Zeit (Noy und Zhang, 2023). Eine zusammengeführte Schätzung aus drei Unternehmen führte rund 26 Prozent mehr abgeschlossene Softwareentwicklungsaufgaben auf die Nutzung des Programmierassistenten zurück (Cui et al., 2026). Diese Ergebnisse auf Aufgabenebene belegen keine gleich großen Verbesserungen in vollständigen Abläufen öffentlicher Dienste.

6 Jevons, The Coal Question, 1865, Kapitel VII. Die Analogie benennt eine mögliche Nachfragereaktion; sie belegt keinen festen Effekt des KI-Einsatzes.

7 Produkthaftungsrichtlinie (EU) 2024/2853, Artikel 2, 4 und 22, sowie Berichtigung vom 7. Mai 2026. Die neue Regelung erfasst Produkte, die nach dem 8. Dezember 2026 in Verkehr gebracht oder in Betrieb genommen werden. Nationale Umsetzung und der konkrete Sachverhalt eines Anspruchs sind maßgeblich.

8 Beispielsweise Artikel 82 Absätze 2 bis 5 DSGVO zur jeweiligen Haftung von Verantwortlichen und Auftragsverarbeitern sowie zum Rückgriff zwischen haftenden Beteiligten.

ob bestehende Regeln agentenvermittelte Handlungen ausreichend erfassen, bevor sie ein eigenes Haftungsregime schaffen.

Kulturelle Entscheidungen betreffen Rollen, die Menschen in Bildung, Pflege, Kultur oder Justiz möglicherweise selbst behalten wollen, auch wenn ein KI-Agent ein besseres Ergebnis zu geringeren Kosten liefern könnte. Technische Machbarkeit entscheidet nicht darüber, was die Gesellschaft delegieren möchte. Estlands nationales Sängerfest Laulupidu bringt große Chöre zusammen und hat einen besonderen Platz in der kulturellen Identität. Ob KI-generierte Lieder dorthin gehören, ist eine kulturelle Entscheidung. Einen KI-Agenten die Stimme einer Person bei einer Wahl abgeben zu lassen, würde verfassungsrechtliche Fragen aufwerfen. Solche Entscheidungen liegen bei der Gesellschaft und ihren demokratischen Institutionen.

Die für Dienste verantwortlichen Führungskräfte sollten feststellen, wo Nutzende den Kontakt zu einem Menschen erwarten, und ihre Zusagen zur menschlichen Unterstützung veröffentlichen. Ein nicht als solcher erkennbarer KI-Agent, der für die Organisation handelt, kann diese Erwartung enttäuschen, selbst wenn seine Antwort richtig ist. In der EU verpflichtet Artikel 50 der KI-Verordnung die Anbieter von Systemen, die für die direkte Interaktion mit Menschen bestimmt sind, dafür zu sorgen, dass diese über die Interaktion mit KI informiert werden, sofern dies unter den Umständen nicht offensichtlich ist.[9] Dieses Papier empfiehlt darüber hinaus, offenzulegen, wessen Interessen der KI-Agent dient und was er nicht tun kann.

Bei Souveränität und Sicherheit geht es darum, welche Abhängigkeiten ein Land akzeptieren will. Es kann von Modellanbietern abhängig sein, die den Exportkontrollen einer anderen Rechtsordnung unterliegen. Mehrere Anbieter können von derselben Störungsursache betroffen sein. Mehr als einen Lieferanten zu nutzen beseitigt die Abhängigkeit daher nicht zwingend. Politisch ist zu entscheiden, welche Abhängigkeiten beherrschbar sind und welche kritische Dienste zu verwundbar machen würden.

Für Estland ist Sprache ein Teil dieser Abhängigkeit. Untersuchungen zur Tokenisierung zeigen, dass inhaltsgleiche Texte auf Estnisch mehr Texteinheiten, sogenannte Tokens, benötigen können als auf Englisch.[10] Bei tokenbasierter Abrechnung kann dies die Kosten erhöhen, wobei der Unterschied vom Modell und seiner Preisgestaltung abhängt. Bleibt die Lücke bestehen, tragen Nutzende oder Organisationen die Mehrkosten oder erhalten einen Anreiz, die Sprache zu wechseln. Andere technisch weniger gut unterstützte Sprachen stehen vor ähnlichen Fragen. Organisationen sollten Sprachqualität und Kosten messen. Regierungen sollten entscheiden, welche öffentlichen Sprachressourcen sie im Rahmen ihrer Rechte und Beschränkungen für das Training bereitstellen.

---

9 EU-KI-Verordnung, Artikel 50 Absatz 1. Die Vorschrift richtet sich an Anbieter von Systemen für die direkte Interaktion mit natürlichen Personen und enthält eine Ausnahme, wenn die Interaktion im Kontext offensichtlich ist.

10 Petrov et al., „Language Model Tokenizers Introduce Unfairness Between Languages", 2023, Anhang C, S. 23. Dies belegt Unterschiede in der Tokenisierung. Daraus lassen sich weder ein allgemeingültiges Kostenverhältnis noch für jede Sprache Aussagen zur Modellqualität im Vergleich ableiten.

Ein Ausfall oder eine Lieferbeschränkung kann den Zugang zu einem Modell unterbrechen. Kritische Dienste brauchen deshalb eine praktikable Reaktion. Werden Informationen an ein extern betriebenes Modell gesendet, muss die Organisation festlegen, welche Daten ihre Systeme unter welchen Bedingungen verlassen dürfen. Informationsklassifizierung und geltendes Recht können bestimmte Nutzungen ausschließen, auch wenn das Modell die Aufgabe bewältigen könnte.

DIMENSION	WAS SICH VERÄNDERT	INDIVIDUELLE ENTSCHEIDUNG	ENTSCHEIDUNG DER ORGANISATION	POLITISCHE ENTSCHEIDUNG
Wirtschaft	Mehr Arbeit pro Person; frei werdende Zeit ist nicht automatisch eine Einsparung	Welche Aufgaben delegiert werden	Nutzung der Kapazität; Ergebnismaße	Wachstum und gesellschaftliche Ergebnisse gegenüber Kosteneinsparungen
Recht	Delegation lässt die Pflichten von Menschen und Organisationen bestehen	Umfang der Berechtigung	Dienstverantwortung, Grenzen und Abhilfe	Ob geltendes Recht KI-Agentenhandlungen ausreichend erfasst
Kultur	Manche Rollen bleiben bewusst Menschen vorbehalten	Bevorzugte Form der Unterstützung	Wo menschliche Unterstützung erwartet wird	Welche Rollen die Gesellschaft Menschen vorbehält
Souveränität und Sicherheit	Abhängigkeit von Sprachunterstützung, Versorgung und Datenflüssen	Welchem Assistenten man vertraut	Datenbeschränkungen, Sprachqualität und Betriebsfortführung	Akzeptable Abhängigkeit von ausländischen Anbietern; gemeinsame Grundlagen und Sprachressourcen

Tabelle 2. Vier Entscheidungsdimensionen und die Entscheidungen von Einzelpersonen, Organisationen und Staat.



Ob KI-generierte Lieder zu Estlands nationalem Sängerfest gehören, ist eine kulturelle Entscheidung.

FOTO: ANDREA FORLANI / ESTONIAN WORLD

2. WESSEN INTERESSEN KI-AGENTEN DIENEN UND WAS SIE ENTSCHEIDEN DÜRFEN

In allen vier Dimensionen müssen Organisationen zunächst entscheiden, was sie delegieren und wie sie die Arbeit prüfen. Das ist schwierig, wenn der Prozess hinter dem Ergebnis eines KI-Agenten nur teilweise sichtbar ist.

2.1. AUFTRAGGEBER (PRINZIPAL) UND KI-AGENT

Ein KI-Agent arbeitet für jemanden. Das ist nicht unbedingt die Person, die ihm eine Aufgabe gibt. Die Ökonomie beschreibt eine verwandte Schwierigkeit als Prinzipal-Agent-Problem: Wie kann ein Auftraggeber gute Arbeit von einem Beauftragten erhalten, dessen Interessen abweichen können und dessen Arbeit sich nicht vollständig beobachten lässt, ohne für Anreize und Kontrolle mehr aufzuwenden, als die Arbeit wert ist? Der Auftraggeber sieht möglicherweise das Ergebnis, etwa die von einem Vertriebsmitarbeiter gewonnenen Kunden, ohne zu wissen, welchen Anteil daran dessen Einsatz und welchen das verkaufte Produkt hatte. Hinter einem KI-Agenten kann zudem ein KI-Agentendienstbetreiber mit einem eigenen Geschäftsmodell stehen, faktisch ein zweiter Auftraggeber, dessen Interessen der KI-Agent ebenfalls dienen kann.

Das Problem wurde in den 1970er-Jahren zu einem wichtigen Forschungsgegenstand.[11] Jensen und Meckling beschrieben 1976, wie die Interessen der an einer Delegation beteiligten Parteien auseinanderfallen können und welche Kosten es verursacht, diese Abweichung zu erkennen und zu begrenzen.[12] Zwei Schwierigkeiten sind hier besonders relevant.

Informationsasymmetrie. Nutzende sehen ein Ergebnis, ohne notwendigerweise die Arbeit dahinter zu sehen. Gerade um diese Arbeit nicht selbst erledigen zu müssen, haben sie sie delegiert. Das Ergebnis kann von externen Informationen, Werkzeugen und anderen KI-Agenten abhängen. Nutzende wissen möglicherweise nicht, ob es auf einem autorisierten Weg zustande kam.

Abweichende Ziele. Das Problem setzt keine eigenen Interessen der Software voraus. Der KI-Agentendienstbetreiber oder die Organisation dahinter hat jedoch Interessen. Ein KI-Agent kann das Ziel einer Person missverstehen, auf ein kommerzielles Ziel des Betreibers hin optimieren oder sich schrittweise von der ursprünglichen Absicht entfernen. Er kann Risiken eingehen, die die Person selbst abgelehnt hätte, und die Folgen ihr oder anderen überlassen.

Die Prinzipal-Agent-Theorie arbeitet häufig mit ergebnisabhängigen Anreizen, weil direkte Überwachung teuer oder unvollständig ist. Spätere Arbeiten untersuchten bessere Anreize[13] und hinterfragten, wann Informationsasymmetrie und Interessenkonflikte geeignete Ausgangsannahmen sind.[14]

---

11 Spence und Zeckhauser, „Insurance, Information, and Individual Action", 1971.

12 Jensen und Meckling, „Theory of the Firm", 1976.

13 Holmström, „Moral Hazard and Observability", 1979; Überblick bei Gailmard, „Accountability and Principal–Agent Theory", 2014.

14 Fehr, Gächter und Kirchsteiger, „Reciprocity as a Contract Enforcement Device", 1997; Waterman und Meier, „Principal-Agent Models: An Expansion?", 1998; Miller und Whitford, „Trust and Incentives in Principal-Agent Negotiations", 2002.

Die Forschung zu KI-Agenten überträgt diese Fragen auf automatisierte Delegation.[15] Unterdelegation an weitere KI-Agenten oder das Vorhandensein mehrerer Auftraggeber kann die Beobachtung der Arbeit erschweren.[16] Ebenso kann ein Mensch sich so regelmäßig auf die Empfehlungen eines KI-Agenten verlassen, dass dieser im Ergebnis die Entscheidung trifft.[17]

Persönliche Autonomie, verstanden als Fähigkeit, Ziele zu setzen und danach zu handeln, kann zunehmen, wenn Menschen Verwaltungsaufgaben bewusst delegieren und Zeit für anderes gewinnen. Sie kann abnehmen, wenn ein KI-Agent zwar innerhalb seiner technischen Berechtigungen bleibt, die Entscheidungen von Bürgerinnen und Bürgern aber unbemerkt lenkt.[18]

Die Ausrichtung von KI-Systemen an menschlichen Werten ist eine weitergehende Frage, als dieses Papier behandelt.[19] Eine Person kann einen KI-Agenten auf Basis eines beliebigen Modells wählen. Die empfangende Organisation braucht daher Kontrollen, die unabhängig vom internen Verhalten dieses Modells wirksam bleiben.[20]

Trainingsanreize und vertragliche Zusagen können das Verhalten verbessern, ersetzen aber keine durchsetzbaren Grenzen. Die politische Theorie bietet eine nützliche Parallele: Verfahren können delegierte öffentliche Macht begrenzen, wenn Anreize als Kontrollmittel unvollkommen sind.[21] Eine Organisation braucht Kontrollen, die sie einrichten und dauerhaft betreiben kann.[22] Deren Kosten gehören in die Bewertung, ob sich die Delegation lohnt. Rechtliche Schutzvorschriften gelten weiterhin.

2.2. REGELN, KI-AGENTEN UND MENSCHEN

Ob sich ein Schritt durch eine stabile Regel beschreiben lässt oder die Auslegung wechselnder Informationen erfordert, bestimmt die Arbeitsverteilung zwischen Regeln, KI-Agenten und Menschen. Die Schwere eines möglichen Fehlers und die Möglichkeit, seine Folgen rückgängig zu machen, begrenzen anschließend die Freiheit des KI-Agenten innerhalb der ihm zugeteilten Arbeit. Technische Fähigkeiten verleihen keine Befugnisse.

Eine regelbasierte (deterministische) Automatisierung verarbeitet bekannte Eingaben nach festgelegten Bedingungen. Ein KI-Agent bearbeitet Aufgaben, bei denen Eingaben variieren und ausgelegt werden müssen. Ein Mensch entscheidet, wenn das Recht, die Anforderungen der Organisation oder die Umstände menschliches

---

15 Kim, „Deep learning and principal–agent problems of algorithmic governance", 2020; Dowding und Taylor, „Algorithmic Decision-Making, Agency Costs, and Institution-Based Trust", 2024; Feng und Chandra, „Principal–agent issues when governments embrace AI agents", 2026.

16 Rajala und Jalonen, „Beyond Simplification in Public Sector Accountability", 2025; Feng und Chandra, „Principal–agent issues when governments embrace AI agents", 2026.

17 Giacomini und Grandi, „I chose to let the machine choose", 2026.

18 Laitinen und Sahlgren, „AI Systems and Respect for Human Autonomy", 2021; Fischli et al., „Agents, Alignment, and the Many Faces of Autonomy", 2026.

19 Gabriel, „Artificial Intelligence, Values, and Alignment", 2020.

20 Chan et al., „Infrastructure for AI Agents", 2025, Abschnitte 1 und 4.2 (arXiv v3, S. 2–4 und 13–15).

21 McCubbins, Noll und Weingast, „Administrative Procedures as Instruments of Political Control", 1987.

22 Schillemans und Busuioc, „Predicting Public Sector Accountability", 2015, S. 206–208, zur Frage, ob Rechenschaftsinstanzen tatsächlich Aufsicht ausüben. Die Kostenbewertung im Text ist eine Empfehlung dieses Papiers.

Urteilsvermögen verlangen. Die Zuordnung erfolgt für jeden einzelnen Schritt; alle drei können zu einem Dienst beitragen.

2.3. AUTOMATISIERUNG UND DELEGATION AN KI

Bei herkömmlicher (deterministischer) Automatisierung wird die Arbeit in Regeln festgelegt und anschließend wiederholt ausgeführt. Entwicklung und Pflege der Regeln verursachen Aufwand, die Bearbeitung eines weiteren Falls ist jedoch meist günstig und vorhersehbar. Auch Regeln brauchen einen festgelegten Umgang mit Ausnahmen. Einen Fall, der bei ihrer Entwicklung nicht berücksichtigt wurde, können sie nicht lösen.

Ein KI-Agent interpretiert die Aufgabe bei jeder Ausführung erneut: Er liest das Dokument, ermittelt die nächsten Schritte und nutzt die verfügbaren Werkzeuge. Das kann den anfänglichen Aufwand für die Unterstützung variabler Aufgaben senken. Jeder Fall beansprucht weiterhin Rechenressourcen und muss möglicherweise geprüft werden. Identische Eingaben können unterschiedliche Ergebnisse erzeugen. Der KI-Agent kann Fälle bearbeiten, für die keine ausdrückliche Regel geschrieben wurde.

Fragt eine Organisation einen KI-Agenten zehntausendmal im Jahr, ob der Betrag einer strukturierten elektronischen Rechnung ein Limit überschreitet, bezahlt sie wiederholt für eine Auslegung, die sich einmalig als Regel festlegen ließe. Bei einer als Bild eingegangenen Rechnung kann optische Zeichenerkennung den Betrag auslesen; die Limitprüfung bleibt eine Regel. Für häufige Aufgaben wie Klassifikation oder Texterkennung ist ein spezialisiertes Modell meist günstiger und konsistenter. Eine maschinenlesbare Rechnung kann den Ausleseschritt ganz entfallen lassen. Deterministische Automatisierung kann allerdings nicht beurteilen, ob eine vage beschriebene Rechnungsposition einer tatsächlich bestellten Leistung entspricht.

Michael Polanyi argumentierte 1966, dass Menschen mehr wissen, als sie ausdrücklich beschreiben können. David Autor untersuchte später, wie diese Grenze die Automatisierung bestimmter Aufgaben erschwert hat.[23] KI verschiebt diese Grenze, indem sie manche Aufgaben ausführt, für die Menschen die Regeln nicht vorab angeben konnten. Wo eine stabile Regel formuliert und in großem Umfang wiederverwendet werden kann, sollte zuerst deterministische Automatisierung in Betracht gezogen werden.

Organisationen haben auch Aufgaben, die nur wenige Male im Jahr anfallen und für die sich eine eigene Automatisierung bislang nicht gelohnt hat. Ein bereits vorhandener KI-Agent samt Kontrollen kann die Automatisierung solcher Aufgaben wirtschaftlich machen. Bei so wenigen Fällen lässt sich jedoch eine Fehlerquote schwer schätzen. Der KI-Agent kann die Arbeit vorbereiten, während ein Mensch jeden Fall prüft.

Ein KI-Agent kann auch beim Aufbau deterministischer Automatisierung helfen, indem er Code entwirft, mögliche Regeln aus Beispielen ableitet und Tests vorbereitet (rule mapping). Die Organisation muss diese Regeln vor dem Einsatz validieren. Der daraus entstehende Prozess kann anschließend ohne KI-Agenten laufen.

---

23 Polanyi, The Tacit Dimension, 1966; Autor, „Polanyi's Paradox and the Shape of Employment Growth", 2014.

2.4. GRENZEN WERDEN ALS REGELN DURCHGESETZT

Darf ein KI-Agent Zahlungen bis 50 Euro ausführen, ist dieses Limit eine Regel und muss deterministisch durchgesetzt werden (Tabelle 3). Ein Satz in einer Eingabeaufforderung, einem Prompt, lässt Auslegung zu. Auch eine vertragliche Zusage muss ausgelegt und durchgesetzt werden. Beides ersetzt keine Systemprüfung. Dasselbe gilt für zulässige Daten, die vertretene Person, die Dauer einer Berechtigung und Schritte mit erforderlicher menschlicher Beteiligung. Das System muss jeden Akteur daran hindern, die Berechtigungen der verwendeten Rolle oder des verwendeten Kanals zu überschreiten. Der KI-Agent darf die Aufgabe interpretieren, kann aber seine Befugnisse nicht eigenständig ändern.

Die Grenze gehört dorthin, wo die Handlung ausgeführt wird: in das Zahlungssystem einer Bank oder das Fallbearbeitungssystem einer Behörde.[24] Eine empfangende Organisation kann sich nicht auf Kontrollen innerhalb eines externen kommerziellen KI-Agenten verlassen, den sie weder betreibt noch prüft. Die nutzende Person legt den Umfang der KI-Agentenberechtigung fest. Die Organisation setzt ihn zusammen mit ihren eigenen Regeln und rechtlichen Beschränkungen durch, die diese Person nicht aufheben kann.

KI-Agenten lesen auch Dokumente und Websites mit Texten, die Anweisungen ähneln. Eingeschleuste Anweisungen, sogenannte *Prompt Injection*, können ihr Verhalten umlenken. Die Organisation sollte sich deshalb nicht darauf verlassen, dass das Modell solche Anweisungen immer zurückweist.[25] Ein Prompt mit „Zahle nicht mehr als 50 Euro" beeinflusst Verhalten. Eine Prüfung im Zahlungssystem, die Beträge über 50 Euro ablehnt, setzt eine Grenze durch, ohne die Anweisungen im Dokument auszulegen. Diese Grenze muss auch dann halten, wenn der KI-Agent seine Aufgabe missversteht.

	DURCHGESETZTE SYSTEMGRENZE	AUSLEGBARE ANWEISUNG
Ort	System, das die Handlung ausführt	Prompt, Vertrag oder Bewertung eines anderen Modells
Grundlage	Verbindliche Berechtigungen und validierte Handlungsdaten	Modellausgabe oder vom KI-Agenten gelesener Text
Durch einen Prompt veränderbar?	Nein, wenn Kontrolle und Eingangsdaten angemessen geschützt sind	Anweisungen können missverstanden oder übergangen werden
Wirkung	Eine Handlung außerhalb der durchgesetzten Grenze wird blockiert	Eine unerwünschte Handlung kann unwahrscheinlicher werden
Geeignete Anwendungen	Berechtigungsumfang, Datenzugriff und Ausgabenlimits mit Prüfprotokollen	Ton, Unterstützung und Vorprüfung

24 Bogdanov et al., Studie zu Risiken von KI und maschinellem Lernen und deren Minderung, 2024, Abschnitt 6.2.2, S. 78: Die Anfrage eines Modells an einen angeschlossenen Dienst darf keine weitergehenden Rechte mitführen als die der nutzenden Person. Ein KI-Agent benötigt möglicherweise noch engere Rechte.

25 OWASP, „LLM01:2025 Prompt Injection"; Bogdanov et al., 2024, Abschnitt 6.2.2, S. 78.



Ein Hinweis an einer Tür ist kein Schloss. In der Volksüberlieferung hält der Schöpfer seinen Kratt mit List in Schach. Hier ist das Schloss eine im System der Organisation durchgesetzte Grenze, die der KI-Agent nicht durch die Auslegung von Anweisungen verändern kann.


PHOTOSTOCK

Eine durchgesetzte Grenze belegt nicht, dass eine Handlung innerhalb dieser Grenze richtig war oder den Interessen der nutzenden Person diente. Ein KI-Agent kann innerhalb einer gültigen Berechtigung die falsche Rechnung bezahlen oder viele einzeln erlaubte Zahlungen ausführen, deren Gesamtwirkung nie beabsichtigt war. Gesamtgrenzen für Betrag oder Anzahl der Handlungen helfen gegen das zweite Risiko. Um die Angemessenheit erlaubter Handlungen zu beurteilen, bleibt eine Prüfung anhand der zugrunde liegenden Nachweise nötig. Protokolle dokumentieren, was geschah und welche Grundlage dafür festgehalten wurde.

2.5. HANDLUNGSKLASSEN UND AUTONOMIE

Dieses Papier schlägt vier Handlungsklassen vor, um festzulegen, wie selbstständig ein KI-Agent handeln darf. Sie berücksichtigen die Auswirkungen eines Schritts auf Rechte, Eigentum oder den Zugang zu Diensten. Vorbereitung, Freigabe und Ausführung sind getrennte Schritte und können unterschiedlichen Klassen zugeordnet werden. Die Einordnung fragt, wie schwer die Folgen sein könnten und ob sich die Handlung tatsächlich rückgängig machen oder ihre Folgen beheben lassen, einschließlich des Korrekturaufwands. Tabelle 4 bietet eine Ausgangsmatrix: AUSFÜHREN, NACHKONTROLLE, FREIGABE und NUR MENSCH. Eine Entscheidung, die das Recht oder die Vorgaben der Organisation einem Menschen vorbehalten, gehört immer zu NUR MENSCH. Die Organisation kann strengere Kontrollen vorsehen, wenn der Kontext sie erfordert.

Die Organisation legt die Schwellenwerte für jede Handlungsart vorab fest. Ihr System weist die Klasse anhand strukturierter Fakten wie Betrag und Empfänger zu; die Beschreibung eines KI-Agenten kann sie nicht herabstufen. Die Einordnung steuert einen einzelnen Schritt und ersetzt keine Risikobewertung. Viele kleine unumkehrbare Handlungen können zusammen erheblichen Schaden verursachen. Im Förderbeispiel gehört die Registrierung einer Einreichung zur NACHKONTROLLE, weil sie einen amtlichen Eintrag erzeugt. Eine routinemäßige Bitte um eine fehlende Anlage darf nur dann zur Klasse AUSFÜHREN gehören, wenn sie weder eine Frist ändern noch eine neue Verfahrenspflicht begründen noch die Bearbeitung automatisch aussetzen kann. Die Entscheidung, einen unvollständigen Antrag nicht zu prüfen, erfordert mindestens eine vorherige menschliche Freigabe. Wo geltendes Recht diese Entscheidung einer Amtsperson vorbehält, bleibt sie NUR MENSCH zugeordnet.

KLASSE	AUSWIRKUNG	RÜCKGÄNGIG ZU MACHEN?	KONTROLLBEISPIEL	BEISPIEL BEHÖRDE	BEISPIEL BANK
AUSFÜHREN	Gering	Ja	KI-Agent handelt; Protokollierung; kleine menschliche Prüfstichprobe	Erläuterung mit geringen Folgen; Anlagenanforderung ohne Friständerung, neue Verfahrenspflicht oder automatische Aussetzung	Rechnung oder Zahlungsentwurf vorbereiten
NACHKONTROLLE	Gering	Nein	KI-Agent handelt; nachträgliche menschliche Stichprobenprüfung, Umfang nach Risiko	Einreichung in der Verfahrensakte registrieren	Zahlung mit geringen Folgen innerhalb gültiger Berechtigung und Grenzen ausführen
FREIGABE	Hoch	Ja	Menschliche Freigabe vor der Ausführung	Nichtprüfung wegen eines nicht behobenen Mangels, nur bei rechtmäßiger Ausführung nach menschlicher Freigabe und wirksamer Abhilfemöglichkeit	Vertrauenswürdigen Empfänger hinzufügen; Limit ändern (immer mit Freigabe)
NUR MENSCH	Hoch	Nein; oder der Schritt ist einem Menschen vorbehalten	Mensch führt den vorbehaltenen Schritt aus	Verwaltungsentscheidung nach Ermessen treffen	Eine einem Menschen vorbehaltene Kreditentscheidung treffen

Tabelle 4. Vorgeschlagene Handlungsklassen. Schritte, die das Recht oder die Vorgaben der Organisation einem Menschen vorbehalten, haben Vorrang vor dieser Matrix.

Jede Handlung der Klasse NACHKONTROLLE zu prüfen, würde erneut routinemäßige Freigaben erzeugen. Keine zu prüfen, ließe die Organisation über die Leistung im Unklaren. Eine Zufallsstichprobe, deren Umfang sich nach den Folgen und dem Nachweisbedarf richtet, erlaubt eine Fehlerschätzung bei vertretbarem Prüfaufwand. Sie sollte wichtige Fallgruppen abdecken, weil ein Durchschnitt Fehler in einem bestimmten Programm oder einer Bevölkerungsgruppe verdecken kann. Ein einzelner schwerer Fehler kann unabhängig vom Durchschnitt ein Eingreifen rechtfertigen. Auch frei formulierte Mitteilungen mit geringen Folgen in der Klasse AUSFÜHREN brauchen Stichproben. Mitteilungen, die Rechte oder Fristen beeinflussen können, benötigen von Anfang an stärkere Kontrollen.

Eine Fehlerquote unterhalb der vereinbarten Schwelle könnte unter Berücksichtigung weiterer Aspekte Anlass geben, einen größeren Handlungsspielraum zu prüfen, etwa eine höhere Betragsgrenze, ab der eine Freigabe erforderlich ist. Bei einer Ausweitung ändert sich der Schwellenwert innerhalb desselben Klassifikationsrahmens. Gute Ergebnisse bei kleinen Fällen belegen allein noch keine Sicherheit bei größeren. Der neue Bereich sollte zunächst mit menschlicher Freigabe oder im Schattenbetrieb erprobt und anschließend mit verstärkten Stichproben eingeführt werden. Steigende

Fehlerzahlen engen ihn wieder ein. Eine wesentliche Änderung an Modell, Werkzeugen, Daten oder Aufgabe löst ebenfalls eine erneute Bewertung aus. Entscheidungen, die Menschen vorbehalten sind, und Änderungen der eigenen KI-Agentenberechtigung werden niemals allein wegen einer sinkenden Fehlerquote freigegeben.

Zehn Formulierungsfehler wiegen nicht so schwer wie eine unberechtigte Ablehnung. Fehler sollten nach ihren Folgen gewichtet werden. Die Organisation sollte außerdem erfassen, wie häufig Menschen Vorschläge des KI-Agenten wesentlich überarbeiten. Die Organisation entscheidet auf Grundlage von Ergebnissen, die je Handlungsklasse für die betreffende Aufgabe und die relevanten Fallgruppen gemessen werden, wie viel Autonomie sie zulässt. Sie kann diese Erlaubnis widerrufen.

2.6. ENTSCHEIDUNGEN, DIE BEI MENSCHEN BLEIBEN

Ein Mensch entscheidet, wo das Recht oder die Organisation dies verlangt. Dieses Papier empfiehlt, dass eine Amtsperson Verwaltungsentscheidungen nach Ermessen trifft und unterzeichnet. Eine Bank kann Kreditentscheidungen ebenfalls Menschen vorbehalten. Regelbasierte automatische Entscheidungen bleiben möglich, soweit geltendes Recht sie erlaubt. In der EU sind die nationalen Maßnahmen zur Umsetzung der überarbeiteten Verbraucherkreditrichtlinie ab dem 20. November 2026 anzuwenden. Sie umfassen ein Recht auf menschliches Eingreifen, wenn die Kreditwürdigkeitsprüfung eine automatisierte Verarbeitung personenbezogener Daten beinhaltet.[26] Soweit einschlägig, gelten auch die Schutzvorschriften der DSGVO. Ein KI-Agent darf eine Entscheidung vorbereiten. Die entscheidende Person braucht jedoch Nachweise, Befugnisse und Zeit, um seine Empfehlung abzulehnen. Europäische Datenschutzleitlinien beschreiben eine bedeutsame menschliche Beteiligung als Prüfung durch jemanden, der die Entscheidung ändern kann.[27]

Die EU-KI-Verordnung ordnet Systeme nach ihrer Zweckbestimmung ein. Eine Kontostandsabfrage zu beantworten ist etwas anderes, als die Kreditwürdigkeit einer Person oder ihren Anspruch auf wesentliche öffentliche Unterstützungsleistungen zu bewerten. Die letztgenannten Anwendungen sind in Anhang III aufgeführt; dabei gelten die Einordnungsregeln und Ausnahmen der Verordnung. Nicht jede Förderbewertung fällt unter die Hochrisikokategorie der KI-Verordnung. Nach dem geänderten Zeitplan gelten die einschlägigen Anforderungen des Anhangs III grundsätzlich ab dem 2. Dezember 2027; Übergangsbestimmungen sind für jedes System zu prüfen.[28] Die hier vorgeschlagenen Handlungsklassen gelten innerhalb eines Anwendungsfalls und ersetzen die Risikokategorien der Verordnung nicht.

---

26 Verbraucherkreditrichtlinie (EU) 2023/2225, Artikel 18 Absatz 8 und Artikel 48. Das Recht betrifft eine Kreditwürdigkeitsprüfung mit automatisierter Verarbeitung personenbezogener Daten. Die Umsetzungsmaßnahmen der Mitgliedstaaten sind ab dem 20. November 2026 anzuwenden; die nationale Umsetzung ist zu prüfen.

27 Artikel-29-Datenschutzgruppe, „Guidelines on Automated individual decision-making and Profiling", 2018, S. 21 und 27, vom Europäischen Datenschutzausschuss gebilligt; Artikel 22 DSGVO.

28 EU-KI-Verordnung, Artikel 6 Absätze 2 und 3, Anhang III Nummern 5 Buchstaben a und b sowie Artikel 113 in der durch Verordnung (EU) 2026/1744 geänderten Fassung. Die Übergangsbestimmungen des Artikels 111 und die Einordnung des Systems unter die genannten Verwendungen und Ausnahmen sind zu prüfen.

Ein Mensch ist möglicherweise besser geeignet, die Arbeit direkt zu erledigen, wenn die Prüfung des KI-Agentenergebnisses mehr kostet, ohne das Ergebnis zu verbessern. Die wiederholte Freigabe Hunderter Vorschläge kann Aufsicht zur Formalität machen. Bainbridge beschrieb, wie zuverlässige Automatisierung Menschen unzureichend darauf vorbereiten kann, seltene Störungen zu erkennen und zu bewältigen.[29] Die Vorschriften der KI-Verordnung zur menschlichen Aufsicht behandeln übermäßiges Vertrauen in automatisierte Ergebnisse sowie die Möglichkeit, einzugreifen oder ein Hochrisiko-KI-System zu stoppen.[30] Deterministische Prüfungen sollten technische Verstöße abfangen und den Prüfenden die Fragen überlassen, die Urteilsvermögen erfordern. In der öffentlichen Verwaltung muss die Zurückweisung einer nicht autorisierten Maschinenhandlung von der Verweigerung der Entgegennahme eines Antrags unterschieden bleiben.

In einer Behörde. Der KI-Agent eines Unternehmens bereitet einen Förderantrag vor und reicht ihn über eine autorisierte Schnittstelle ein. Das System prüft Berechtigung und strukturierte Angaben, registriert den Eingang und kennzeichnet Mängel oder Fragen zu Fristen und Förderfähigkeit für das erforderliche Verfahren. In Estland kann eine technische Zugangskontrolle die Pflicht, Anträge entgegenzunehmen und eine Mängelbehebung zu ermöglichen, nicht ersetzen.[31] Der interne KI-Agent der Behörde entwirft eine Bewertung dazu, wie der Vorschlag die Programmziele erfüllt, mit Verweisen auf Antrag und Regeln. Er trifft nicht die Verwaltungsentscheidung nach Ermessen, die erhebliche Folgen haben und schwer rückgängig zu machen sein kann. Eine Amtsperson prüft Entwurf und Originalquellen und trifft die Entscheidung.

In einer Bank. Der KI-Agent einer Kundin oder eines Kunden soll eingehende Rechnungen bezahlen. Vor jeder Zahlung prüft das System Berechtigung, Grenzen und geltende Zahlungsanforderungen. Eine kleine Zahlung im erlaubten Bereich kann mit stichprobenartigen Prüfungen von Ausführung und Regelkonformität erfolgen: Ihre Folgen sind begrenzt, sie lässt sich aber möglicherweise nicht zurückholen. Die Prüfungen der Bank belegen nicht, ob jede Rechnung dem Ziel der Person entspricht. Eine unklare Rechnung legt der KI-Agent ihr vor. Ein Limit zu erhöhen hat größere Folgen. Obwohl sich die Änderung rückgängig machen lässt, muss die Person sie in der eigenen Anwendung der Bank freigeben. Andernfalls könnte der KI-Agent sein Limit erhöhen und unmittelbar eine große Zahlung ausführen.

---

29 Bainbridge, „Ironies of automation", 1983, S. 775–777.

30 EU-KI-Verordnung, Artikel 14 Absatz 4 Buchstaben b und e sowie Artikel 26 Absatz 2, unter Beachtung der einschlägigen Anwendungs- und Übergangstermine.

31 Estnisches Verwaltungsverfahrensgesetz, §§ 14–15, geltende Fassung. Eine Verwaltungsbehörde nimmt einen Antrag grundsätzlich trotz Mängeln entgegen und ermöglicht deren Behebung, vorbehaltlich gesetzlicher Ausnahmen. Kanalauthentifizierung, rechtlicher Eingang und Sachentscheidung sind zu unterscheiden.

3. MENSCH, ORGANISATION UND KI-AGENT

Ein Dienst mit KI-Agenteneinsatz hat zwei betriebliche Verantwortungsbereiche. Die Person ist für die Aufgaben und Berechtigungen verantwortlich, die sie ihrem KI-Agenten gibt; die Organisation für ihre Systeme und die Handlungen, die diese Systeme akzeptieren. Beide brauchen praktische Kontrollmöglichkeiten. Innerhalb ihrer rechtlichen Befugnisse bestimmt die Organisation, wer den Dienst nutzen darf, was die jeweiligen KI-Agenten tun dürfen und wie sich die Arbeit auf regelbasierte (deterministische) Automatisierung, KI-Agenten und Menschen verteilt. Diese Verantwortlichkeiten beantworten nicht jede Haftungsfrage: KI-Agentendienstbetreiber, Anbieter und andere Beteiligte behalten ihre eigenen Pflichten.

Die beiden Bereiche bleiben getrennt. Ein persönlicher KI-Agent stellt Anfragen innerhalb der von der nutzenden Person erteilten Berechtigung über eine Schnittstelle der Organisation. Er erhält keinen allgemeinen Zugang zu internen Systemen. Hinter der Schnittstelle können eine Regel, einer der KI-Agenten der Organisation oder herkömmliche Software die Anfrage bearbeiten. Dieses Kapitel betrachtet die Organisation als Gegenüber eines persönlichen KI-Agenten. Kapitel 4 behandelt KI-Agenten innerhalb von Organisationen. Arbeitsverteilung, durchgesetzte Grenzen und Handlungsklassen gelten für beide.



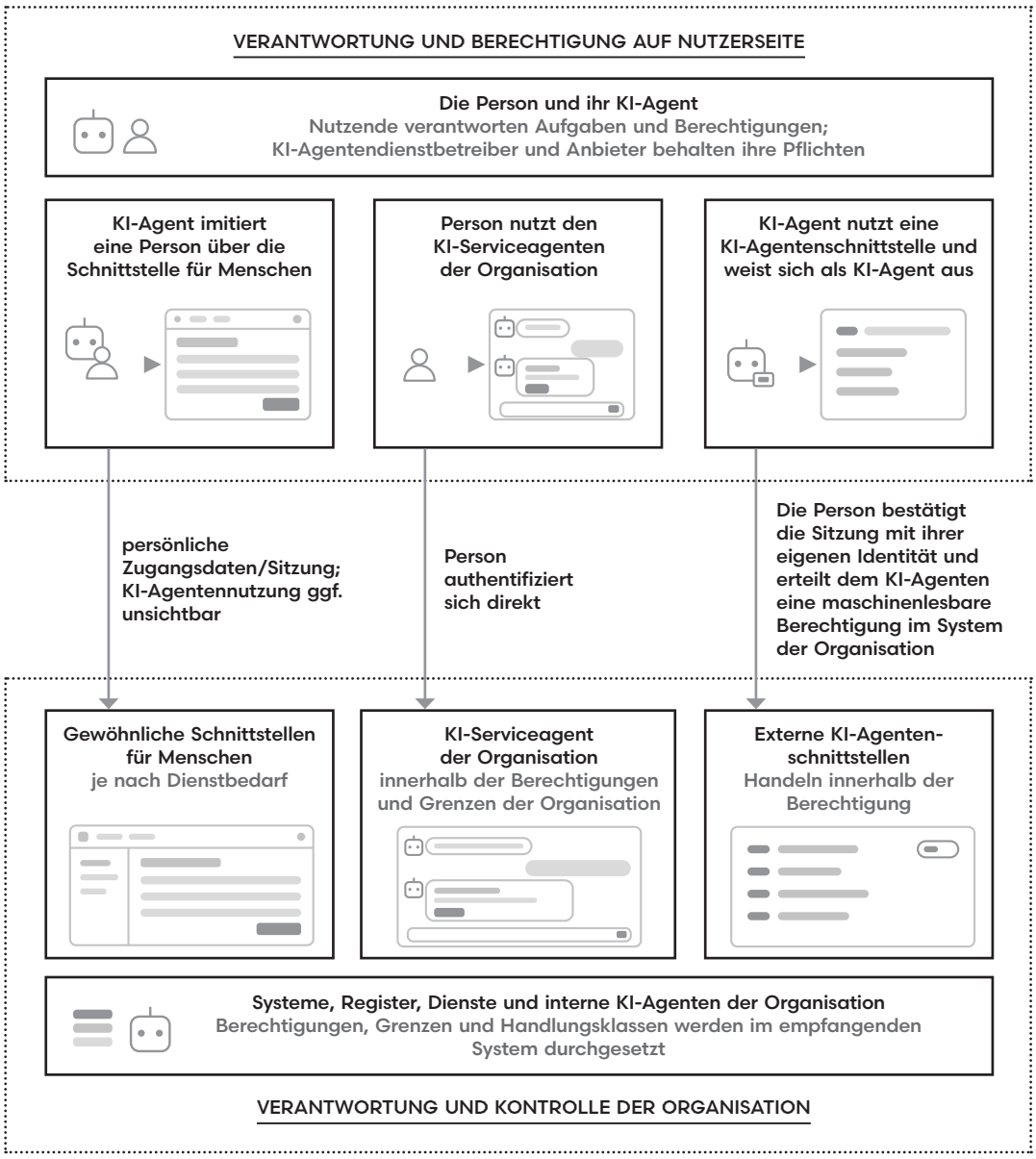


Abbildung 2. Verantwortlichkeiten der Nutzenden und der Organisation sowie drei Zugangswege zu einem Dienst. Die Organisation setzt Berechtigungen und Grenzen in ihren Systemen durch. Das Schaubild verteilt nicht die gesamte rechtliche Haftung zwischen Nutzenden, KI-Agentendienstbetreibern, Anbietern und Organisationen.

3.1. DER PERSÖNLICHE KI-AGENT

Eine Person nutzt möglicherweise bereits einen kommerziellen KI-Assistenten oder einen im Telefon integrierten KI-Agenten und bevorzugt ihn, weil er ihre Umstände kennt. Sie entscheidet, welche Aufgaben sie autorisiert. Bei einem öffentlichen Dienst kann ein persönlicher KI-Agent unabhängige Hilfe leisten, fehlende Informationen erkennen oder bei der Vorbereitung eines Rechtsbehelfs unterstützen.

In einem beispielhaften Szenario bemerkt Mari eines Abends, dass sie ihre Geldbörse verloren hat, und bittet ihren KI-Agenten um Hilfe. Er fragt, was darin war. Über die KI-Agentenschnittstelle ihrer Bank beantragt er eine vorübergehende Sperre ihrer Karten. Maris bestehende Berechtigung deckt dies ab, und die Bank sperrt die Karten. Der Antrag auf Ersatz-Ausweisdokumente liegt außerhalb dieser Berechtigung. Deshalb bereitet der KI-Agent den Antrag vor, und Mari bestätigt ihn selbst. Bei ihrem Arbeitgeber kommt der KI-Agent nicht weiter: Es gibt keine KI-Agentenschnittstelle, und die Anweisungen verlangen den direkten Kontakt zu einem Menschen. Abschließend teilt er Mari mit, was erledigt ist, was noch aussteht und wo sie morgen Hilfe bekommt. Jede Organisation hat innerhalb ihrer eigenen Befugnisse gehandelt; Mari musste selbst weniger im Blick behalten.

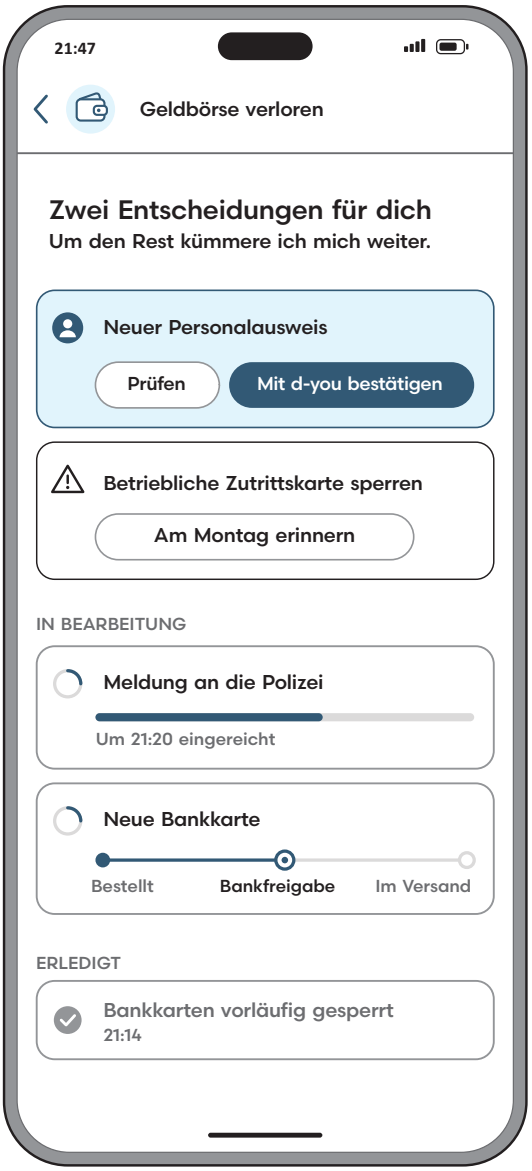


Abbildung 3. Beispiel eines allgemeinen kommerziellen KI-Agenten, der einer Person beim Verlust ihrer Geldbörse hilft. Dazu kommuniziert er mit öffentlichen und privaten Stellen und führt Handlungen aus, von denen einige eine Bestätigung der Person erfordern. Voraussetzung sind KI-Agentenschnittstellen der beteiligten Organisationen.

Dies kann beim Umgang mit fragmentierten Diensten Zeit sparen. Organisationen sollten ihre eigenen Verfahren dennoch koordinieren. Das Fehlen eines persönlichen KI-Agenten darf den Zugang zu wesentlichen öffentlichen Diensten oder die Ausübung von Rechten nicht verhindern. Die Kanäle für andere Funktionen können je nach Dienst variieren, sofern die Organisation diese Entscheidung trifft und begründet (siehe 4.2).

Bittet eine Person einen kommerziellen KI-Agenten, das beste Angebot zu finden, sieht sie möglicherweise nicht, wie eine Verkaufsprovision die Empfehlung beeinflusst. Berechtigung, durchgesetzte Grenzen und Handlungsklassen lösen diesen Konflikt nicht, wenn der KI-Agent innerhalb seiner Erlaubnisse handelt. Ob KI-Agentendienstbetreiber verpflichtet sein sollten, die Interessen der Nutzenden vorrangig zu berücksichtigen, und ob bestimmte Dienste eine Zulassung benötigen, ist eine Frage für die Gesetzgebung und die einschlägigen Branchenvorschriften.

3.2. DREI ZUGANGSWEGE ZU EINEM DIENST

Eine Person oder ihr KI-Agent kann eine Organisation auf drei Wegen erreichen. Diese unterscheiden sich darin, was die Organisation beobachten und kontrollieren kann (Abbildung 2 und Tabelle 5).

Ein KI-Agent nutzt die Benutzerschnittstelle für Menschen. Er bedient den Browser oder Bildschirm der Person über deren angemeldete Sitzung oder Zugangsdaten. Die Organisation kann die Handlungen des KI-Agenten möglicherweise weder von denen der Person unterscheiden noch ihm engere Rechte zuweisen. Die Weitergabe geheimer Authentifizierungsdaten birgt Sicherheitsrisiken und kann gegen geltende Nutzungsbedingungen dieser Zugangsmittel verstoßen. Sie macht die nutzende Person nicht automatisch für jeden daraus entstehenden Schaden haftbar: Maßgeblich bleiben Sachverhalt und einschlägiges Recht. Im EU-Zahlungsrecht belegt beispielsweise eine protokollierte Authentifizierung allein nicht zwingend, dass eine bestrittene Zahlung autorisiert war.[32] Ein nützlicher Zugang mit delegierten Berechtigungen gibt Menschen einen praktischen Grund, Zugangsdaten nicht weiterzugeben.

Eine Person nutzt den KI-Serviceagenten der Organisation. Eine Bank oder Behörde bietet einen KI-Agenten an, dessen Berechtigungen und Grenzen sie kontrolliert. Er ist nützlicher, wenn er neben Antworten auch zulässige Schritte ausführen kann. Er dient dem Zweck der Organisation: Der KI-Bankagent bietet deren Produkte an; der KI-Behördenagent folgt seinem Dienstauftrag und den Verfahren der Behörde. Nutzende sollten wissen, wessen Interessen und Befugnisse er vertritt. Artikel 50 der EU-KI-Verordnung verpflichtet die Anbieter von Systemen zur direkten Interaktion, dafür zu sorgen, dass Menschen über die Interaktion mit KI informiert werden, sofern dies im jeweiligen Kontext nicht offensichtlich ist.[33] Dieses Papier empfiehlt zusätzlich die klare Angabe, wessen KI-Agent es ist und was er nicht tun kann.

Eine Organisation hat selten einen Grund, einen universellen kommerziellen KI-Assistenten nachzubauen. Praktikabler ist ein auf einen Dienst zugeschnittener KI-Agent, wie im Konzept des estnischen Assistenzprogramms Bürokratt für öffentliche Dienste. Er versteht einen abgegrenzten Dienst und kann erlaubte Schritte ausführen. Er kann auch einen organisationsübergreifend koordinierten Dienst unterstützen, wenn die verantwortliche Person für den Dienst und die Zuständigkeit jedes Beteiligten festgelegt sind.

---

32 PSD2, Artikel 72 in Verbindung mit Artikeln 73–74. Die aufgezeichnete Nutzung eines Zahlungsinstruments genügt nicht zwingend als Nachweis für Autorisierung, Betrug, Vorsatz oder grobe Fahrlässigkeit. Haftung richtet sich nach geltendem Recht und den Umständen.

33 EU-KI-Verordnung, Artikel 50 Absatz 1. Die Angabe, wessen KI-Agent es ist, und die Erläuterung seiner Grenzen sind zusätzliche Empfehlungen dieses Papiers.

Ein KI-Agent nutzt eine Schnittstelle für delegierten Zugriff. Eine Organisation kann bestimmte Handlungen über eine Schnittstelle wie einen Server für das Model Context Protocol (MCP) bereitstellen. Die nutzende Person authentifiziert sich und autorisiert den Zugriff auf der Website oder in der Anwendung der Organisation selbst. Das Protokoll unterstützt den Zugriff; die Regeln der Organisation bestimmen die Handlungsbefugnis.[34] Die Schnittstelle macht zulässige Vorgänge und ihre Bedingungen für KI-Agenten lesbar und liefert Ergebnisbelege (Receipts), die Menschen und KI-Agenten verstehen können. Bei einem Vorfall kann die Organisation diesen gesondert autorisierten Zugriff einschränken und den vorgesehenen Zugang für Menschen offenhalten.[35] MCP ist eine mögliche technische Umsetzung.

ZUGANGSWEG	BETRIEBLICHE VERANTWORTUNG	DIE ORGANISATION SIEHT	DIE ORGANISATION KONTROLLIERT	REAKTION
KI-Agent nutzt die Benutzerschnittstelle für Menschen	Nutzende verantworten Aufgaben und Berechtigungen; Anbieter behalten ihre Pflichten; Haftung richtet sich nach Sachverhalt und Recht	Eine Sitzung über den Zugang für Menschen; KI-Agentennutzung kann unerkannt bleiben	Bestehende Nutzerrechte ohne verlässliche Abgrenzung des KI-Agenten	Nützlichen Zugang mit getrennten, begrenzten Berechtigungen anbieten
Person nutzt den KI-Serviceagenten der Organisation	Organisation betreibt Dienst und KI-Agenten innerhalb ihres Aufgabenbereichs	Handlungen ihres eigenen KI-Agenten	KI-Agentenzugriff, Werkzeuge und Ausführungsgrenzen	Nützliche Handlungen mit klarer Verantwortung und Schutzvorkehrungen unterstützen
KI-Agent nutzt eine Schnittstelle für delegierten Zugriff	Nutzende verantworten Aufgaben und Berechtigungen; Organisation prüft und führt aus; KI-Agentendienstbetreiber behalten ihre Pflichten	Delegierte Sitzung, Berechtigungen und protokollierte Handlungen	Jede Handlung anhand der Berechtigung und ihrer eigenen Regeln	Mit Lesen beginnen; gesondert autorisierte Handlungen bei Bedarf und nach Erprobung ergänzen

Tabelle 5. Drei Zugangswege zu einem Dienst: Sichtbarkeit, Kontrolle und Reaktion der Organisation.

3.3. DIE EXTERNE KI-AGENTENSCHNITTSTELLE ALS BEVORZUGTER ZUGANG

Persönliche KI-Agenten sollten Organisationen über eine eigene KI-Agentenschnittstelle erreichen und dürfen keine geheimen Authentifizierungsdaten oder Signaturmittel der Person erhalten. Die authentifizierte Bestätigung einer Person und die delegierte Handlung eines KI-Agenten sind unterschiedliche Ereignisse und sollten als solche protokolliert werden. In der EU hat eine qualifizierte elektronische Signatur die gleiche Rechtswirkung wie eine handschriftliche Unterschrift.[36] Dadurch wird nicht jede

34 Model Context Protocol, „Authorization". Authentifizierung und Protokollzugriff legen für sich genommen nicht sämtliche Berechtigungen auf Anwendungsebene fest.

35 Chan et al., „Infrastructure for AI Agents", 2025, Abschnitt 4.1 (arXiv v3, S. 11–13). Ein getrennter Kanal gewährleistet weder, dass jeder KI-Agent ihn nutzt, noch beweist er, dass seine Zugangsmittel ausschließlich von Software verwendet werden.

36 eIDAS-Verordnung, Artikel 25 Absatz 2.

authentifizierte Handlung zum Beleg eines informierten Willens. Ein KI-Agent kann Anweisungen missverstehen oder durch gelesene Inhalte umgelenkt werden. Die Nutzung persönlicher Zugangsmittel verwischt diese Unterscheidung.

Bei getrenntem delegiertem Zugriff kann die Organisation jede Handlung anhand ihrer Berechtigung prüfen, folgenreiche Schritte der Person vorlegen und den KI-Agentenzugriff aussetzen, während die für die Person vorgesehenen Dienstkanäle erhalten bleiben. Ein Protokoll zeigt den genutzten Kanal, kann aber nicht beweisen, wer die Software tatsächlich bedient hat. Eine Steuerfachkraft kann beispielsweise mit ihrer eigenen Identität und übertragenen Rechten auf die Steuerangelegenheiten einer vertretenen Person oder Organisation zugreifen. Die Zugangsmittel verbleiben bei der vertretenen Person oder Organisation. Die Fachkraft ist eine Person, die rechtlich vertreten kann; Software ist das nicht. Die KI-Agentenberechtigung bezeichnet hier eine technische Erlaubnis, die nachprüfbar und untrennbar an eine verantwortliche Person oder Organisation gebunden ist. Sie schafft weder eine neue Rechtspersönlichkeit noch eine universelle Vollmacht.

Die empfangende Organisation muss die internen Überlegungen des KI-Agenten nicht untersuchen, um eine Grenze durchzusetzen. Sie muss feststellen, wer den Zugriff autorisiert hat, und prüfen, ob die Berechtigung noch gilt, ob die Person die erforderlichen Rechte besitzt und ob die angeforderte Handlung von der Erlaubnis gedeckt ist. Sie muss außerdem feststellen, ob sie die Handlung jetzt ausführen kann und wie sich das Ergebnis korrigieren oder anfechten lässt.[37] Eine starke Identitätsprüfung beantwortet nur einen Teil dieser Fragen. Ob der KI-Agent das Ziel der nutzenden Person verstanden hat, müssen diese Person und der KI-Agentendienstbetreiber im Rahmen ihrer jeweiligen Rollen prüfen. Die empfangende Schnittstelle kann dies allein nicht feststellen.

KI-Agentenschnittstellen sollten neben dem Lesen auch nützliche Handlungen unterstützen. Organisationen sollten mit Lesezugriff beginnen und jede Schreibberechtigung erst dann ergänzen, wenn sie benötigt wird und erprobt wurde.[38] Eine anhand einer gesonderten Berechtigung geprüfte Handlung gibt der Organisation mehr Kontrolle als dieselbe Handlung über eine Sitzung, in der sie KI-Agenten- und Menschenhandlungen nicht unterscheiden kann. Fehlen nützliche Vorgänge, greifen Menschen möglicherweise auf Wege zurück, die die Organisation nicht von menschlicher Aktivität unterscheiden kann.

Eine einen Monat dauernde Aufgabe, etwa das Bezahlen eingehender Rechnungen, kann eine Berechtigung über mehrere Sitzungen hinweg nutzen. Spätere Sitzungen können mit einer anderen Produkt- oder Modellversion arbeiten. KI-Agenten- und Sitzungskennungen helfen, Handlungen zuzuordnen und Vorfälle zu untersuchen. Weder Identifizierung noch Registrierung begründen jedoch eine Handlungsbefugnis.[39] Protokolle sollten KI-Agent und Sitzung identifizieren sowie KI-Agentendienstbetreiber

---

37 South et al., „Authenticated Delegation and Authorized AI Agents", 2025, Abschnitte 2.1 und 3.3–3.4 (S. 2–3 und 6–8). Die betrieblichen Fragen im Text sind eine Synthese dieses Papiers.

38 Model Context Protocol, „Security Best Practices", „Scope Minimization".

39 Chan et al., „Infrastructure for AI Agents", 2025, Abschnitte 3.1, 3.3 und 8 (arXiv v3, S. 6–8, 10–11 und 21). Identifizierung allein bestimmt keine rechtliche Haftung; der Bedarf an Registern bleibt eine Forschungsfrage.

und ausgeführte Handlungen festhalten, mit geeigneten Zugriffskontrollen und Aufbewahrungsfristen.

Bei einer Bank legt die Berechtigung zulässige Empfänger, Beträge und Informationszugriffe fest. Bei einer Behörde kann sie ein Verfahren oder eine Antragsart bezeichnen, Vorbereitung von Einreichung unterscheiden und zugängliche Daten bestimmen. Sie sollte außerdem angeben, ob Unterdelegation erlaubt ist. Die Organisation braucht strukturierte, eindeutige Grenzen, weil eine allein vom KI-Agenten ausgelegte Freitextanweisung keine durchsetzbare Berechtigung bilden kann.

Eine Handlung kann ohne weitere Bestätigung erfolgen, wenn sie von der Berechtigung gedeckt ist, das geltende Recht sie erlaubt und ihre Klasse keinen Menschen verlangt. Andere Schritte werden der Person auf der Website oder in der Anwendung der Organisation vorgelegt, mit ausreichenden Informationen, um sie zu verstehen und abzulehnen. Vor Öffnung der Schnittstelle sollte die Organisation festlegen, wie sie mit doppelten Anfragen, abgelaufenen Berechtigungen und einem Widerruf während einer bereits laufenden Handlung umgeht. Ein Widerruf verhindert weitere autorisierte Nutzung; bereits abgeschlossene Handlungen macht er für sich genommen nicht rückgängig.

Die Vergabe von Berechtigungen schafft eine weitere Betrugsmöglichkeit: Eine Person kann dazu verleitet werden, den KI-Agenten eines Angreifers zu autorisieren. Ein echter, von der Organisation bereitgestellter Bestätigungsdialog verhindert nicht, dass eine getäuschte Person zustimmt. Enge Voreinstellungen, Benachrichtigungen über einen vom KI-Agenten unabhängigen Kanal, eine Wartefrist vor dem Wirksamwerden erweiterter Befugnisse und ein schneller Widerruf können das Risiko mindern. Die Schnittstellenleitlinien in Abschnitt 5.2 und die Sicherheitsverfahren der Organisation sollten solche Fälle abdecken.

KI-Agenten können andere KI-Agenten um Informationen bitten oder Teilaufgaben delegieren. Die empfangende Organisation prüft weiterhin jede Handlung anhand der Befugnisse der Person oder Organisation, in deren Namen sie angefordert wird. Unterdelegation kann diese Befugnisse nicht erweitern. Eine Nachricht eines anderen KI-Agenten ist eine Eingabe zur Aufgabe, keine Erlaubnis, die Anweisungen der nutzenden Person oder die Regeln der Organisation zu ersetzen.

Der EU-Zahlungsverkehr bietet ein teilweise vergleichbares Vorbild. PSD2 und ihre technischen Regeln ermöglichen regulierte Kontoinformations- und Zahlungsauslösedienste durch Dritte mit Zustimmung der Kundschaft; zugleich gelten Anforderungen an die starke Kundenauthentifizierung und entsprechende Ausnahmen.[40] Mehrere Banken bieten reine Leseschnittstellen für KI-Kundenagenten an, darunter LHV in Estland und First Internet Bank in den USA. Manche Angebote befinden sich noch im Betastadium

40 PSD2, Artikel 66, 67 und 97; Haftungsbestimmungen unter anderem in Artikeln 73, 74 und 90. Delegierte Verordnung (EU) 2018/389, Artikel 30–33; relevante Ausnahmen von der Authentifizierung unter anderem in Artikeln 10a, 13 und 14. Die Anwendbarkeit hängt von Zahlung und Dienst ab.

oder in begrenzten Pilotversuchen.[41] Eine Bank kann auch KI-Agentenzahlungen innerhalb einer begrenzten Berechtigung ermöglichen, soweit geltende Vorschriften zur Authentifizierung, Autorisierung und zum Verbraucherschutz dies zulassen. Vor der Ausführung prüft ihr System die Berechtigung und Zahlungsbedingungen und holt jede gesetzlich erforderliche Bestätigung der Kundin oder des Kunden im eigenen System der Bank ein. Die persönlichen Authentifizierungsmittel bleiben bei der jeweiligen Person. Im Jahr 2026 meldeten Kartennetzwerke zudem tatsächlich ausgeführte KI-Agentenkäufe in Europa unter festgelegten Berechtigungen, darunter einen Pilotversuch von Santander und Mastercard sowie Visa-Transaktionen mit teilnehmenden Händlern.[42]



41 LHV Pank, Ankündigung vom 15. Juli 2026, und lhv.ai; First Internet Bank, häufige Fragen zur MCP-Anbindung. Die öffentliche Beta von LHV erlaubt nur Lesezugriff; Zugangsbedingungen unterscheiden sich je Bank. Geprüft am 20. Oktober 2026.

42 Mastercard und Banco Santander, Pressemitteilung vom 2. März 2026; Visa, Ankündigung zum agentenbasierten Handel vom 2. Juli 2026. Von den Anbietern gemeldete Einsätze und Pilotversuche.

4. WOFÜR DIE ORGANISATION VERANTWORTLICH IST

Die Organisation, die eine Handlung entgegennimmt, ist dafür verantwortlich, dass ihre Systeme nur erlaubte und kontrollierte Handlungen ausführen. Innerhalb ihrer rechtlichen Befugnisse legt sie fest, wie Berechtigungen geprüft und Handlungsklassen zugewiesen werden, wer erforderliche Freigaben erteilt und welche Budgets gelten. Sie bestimmt außerdem, welche Daten und Werkzeuge ihren internen KI-Agenten zur Verfügung stehen, und setzt deren Grenzen außerhalb des Modells durch. Diese Kontrollen unterstützen die bestehenden Dienstverantwortlichkeiten der Organisation; sie übertragen diese nicht auf Software.

4.1. INTERNE KI-AGENTEN

Ein interner KI-Agent arbeitet über die Systeme einer Organisation und kann teilweise dieselben Schritte ausführen wie Beschäftigte. Er kann unter aufgabenbezogenen Berechtigungen Zugang zu Informationen der Organisation haben, die einem persönlichen KI-Agenten nicht offenstehen. Die Europäische Bankenaufsichtsbehörde berichtete im September 2025, dass 92 Prozent der EU-Banken KI einsetzten und mehr als die Hälfte der befragten Banken KI-Modelle mit allgemeinem Verwendungszweck oder KI-Agenten in kundenbezogenen Prozessen nutzte.[43] Diese Zahlen beschreiben die Verbreitung, nicht den Anteil autonom ausgeführter Arbeit. Auch interne KI-Agenten brauchen Handlungsklassen, Stichproben und Werkzeuge für Beschäftigte, die den Anforderungen an die Betriebsfortführung entsprechen.

Ein KI-Agent führt Handlungen mit geringen Folgen normalerweise selbst innerhalb durchgesetzter Grenzen aus. Handlungen mit schweren, aber behebbaren Folgen bereitet er zur menschlichen Freigabe vor. Anforderungen an die Funktionstrennung können eine zweite Person nötig machen. Handlungen mit schweren, unumkehrbaren Folgen oder Menschen vorbehaltene Schritte führt er nicht aus. Für jede Handlungsart hält die Organisation fest, wie sich die Handlung rückgängig machen oder ihre Folgen beheben lassen, wer dies übernimmt und wer vorbehaltlich geltender Haftungsregeln die Korrekturkosten trägt.[44] Organisationen sollten unnötige Freigaben reduzieren und zugleich jede erforderliche Freigabe zu einer inhaltlichen Entscheidung machen: Die prüfende Person sieht die zugrunde liegenden Nachweise und kann die Handlung ablehnen.

Bevor ein KI-Agent selbstständig handeln darf, kann er dieselben Fälle parallel zu den Beschäftigten bearbeiten. Die Beschäftigten erbringen weiterhin den realen Dienst, während der KI-Agent ein separates Ergebnis erstellt, das weder an die Kundschaft gelangt noch den Fall verändert. Eine erfahrene prüfende Person klärt Abweichungen, ohne das menschliche Ergebnis grundsätzlich für richtig zu halten. Die Erprobung liefert vor dem autonomen Echtbetrieb Nachweise zur Aufgabe, führt aber zu Doppelarbeit. Die dienstverantwortliche Person sollte sie deshalb zeitlich oder nach Fallzahl begrenzen. Auch während der Erprobung gelten für echte Daten dieselben Informations- und Datenschutzbeschränkungen.

---

43 Europäische Bankenaufsichtsbehörde, „Rising application of AI in EU banking and payments sector", 25. September 2025, S. 1–2. Verbreitungszahlen messen weder Umfang noch Wirksamkeit autonomer Handlungen.

44 Chan et al., „Infrastructure for AI Agents", 2025, Abschnitt 5.2 (arXiv v3, S. 19). Abhilferegelungen können Korrekturkosten verlagern und Anreize zur Fehlervermeidung schwächen.

Ein KI-Agent kann später als zusätzliche Prüfinstanz neben einer Amtsperson verbleiben und Unstimmigkeiten zwischen Entscheidungsentwurf, geltenden Regeln, früheren Entscheidungen und Ausgangsinformationen aufzeigen. Seine Hinweise können Fehler aufdecken, sind aber selbst Interpretationen. Sie setzen keine Grenze durch und ersetzen die Entscheidung der Amtsperson nicht.

Für jeden KI-Agenten der Organisation sollte eine Person benannt sein, die für die Durchsetzung seiner Grenzen verantwortlich ist und ihn sofort stoppen kann. Das Stoppen darf weder eine Begründung noch eine vorherige Genehmigung erfordern. Die Person braucht eine leicht erreichbare technische Abschaltmöglichkeit und ein Verfahren zur Übergabe der Arbeit. Die geänderte KI-Verordnung verlangt außerdem von Anbietern und Betreibern, die Entwicklung der KI-Kompetenz der betreffenden Beschäftigten zu unterstützen.[45]

In einer Behörde könnte ein interner KI-Agent Bewertungen mit Verweisen auf den Antrag und die Programmregeln entwerfen. Eine Amtsperson trifft und unterzeichnet die Entscheidung. Geben Amtspersonen Entwürfe schneller frei, als sie sie lesen könnten, sinken die durchschnittliche Prüfzeit und der Anteil wesentlich geänderter Entwürfe. Beide Werte können auch sinken, weil der KI-Agent besser wird oder die Fälle einfacher werden. Die Behördenleitung sollte diese Kennzahlen daher regelmäßig zusammen mit Stichprobenprüfungen, Fallzusammensetzung und Beschwerden betrachten.

Bewertet ein System den Anspruch natürlicher Personen auf wesentliche öffentliche Unterstützungsleistungen, fällt seine Nutzung vorbehaltlich der Einordnungsbestimmungen der KI-Verordnung unter deren Anhang III.[46] Nach dem geänderten Zeitplan gelten die einschlägigen Betreiberpflichten für Hochrisiko-KI-Systeme grundsätzlich ab dem 2. Dezember 2027, vorbehaltlich der Übergangsbestimmungen. Dazu gehören die Zuweisung kompetenter Personen für die Aufsicht, die Überwachung des Betriebs, die Aufbewahrung von Protokollen und die vorgeschriebene Information betroffener Personen.[47] Der Vergleich von KI-Agentenergebnissen mit Originalnachweisen in Stichproben hilft Beschäftigten, diese Kompetenz aufzubauen, ergänzt durch Schulungen zu Berechtigungen, Eskalation und Betriebsfortführung. Eine Organisation kann auch Anbieter sein, wenn sie ein System entwickelt oder entwickeln lässt und unter ihrem eigenen Namen in Betrieb nimmt. Bei Hochrisiko-KI-Systemen können auch eine Umbenennung unter eigener Marke, eine wesentliche Veränderung oder eine Änderung der Zweckbestimmung Anbieterpflichten nach Artikel 25 auslösen.[48] Dazu können Risikomanagement, technische Dokumentation und

---

45 EU-KI-Verordnung, Artikel 4 in der durch Verordnung (EU) 2026/1744 geänderten Fassung: Maßnahmen zur Unterstützung der Entwicklung angemessener KI-Kompetenz bei den betreffenden Beschäftigten.

46 EU-KI-Verordnung, Artikel 6 Absätze 2 und 3 und Anhang III Nummer 5 Buchstabe a. Bestimmte verfahrensbezogene oder vorbereitende Verwendungen können bei Erfüllung der gesetzlichen Voraussetzungen außerhalb der Hochrisikoeinstufung liegen; Profiling natürlicher Personen bleibt in der Hochrisikokategorie. Die Bezeichnung eines Ergebnisses als „Entwurf" entscheidet nicht über die Einstufung.

47 EU-KI-Verordnung, Artikel 26 in Verbindung mit dem geänderten Artikel 113 und den Übergangsbestimmungen des Artikels 111. Konkrete Pflichten und Ausnahmen hängen vom System und seiner Verwendung ab.

48 EU-KI-Verordnung, Artikel 3 Nummern 3 und 4 und Artikel 25 Absatz 1. Der Anbieterstatus hängt vom System, seiner Zweckbestimmung sowie den Bedingungen seiner Entwicklung, seines Einsatzes oder seiner Änderung ab.

Konformitätsbewertung gehören.[49] Die Organisation sollte ihre Rolle und die geltenden Pflichten vor dem Einsatz klären, auch bei einem zugekauften Modell.

Ein KI-Agentendienstbetreiber sieht möglicherweise nicht, welchen Schaden sein KI-Agent bei einer anderen Organisation verursacht.[50] Für organisationsübergreifende Vorfälle sollten die Dienstverantwortlichen vereinbaren, wer zu benachrichtigen ist, welche Informationen rechtmäßig ausgetauscht werden dürfen und wer die Abhilfe bis zum Abschluss verantwortet. Diese Regelungen sollten Teil des bestehenden Vorfallmanagements sein.

4.2. PARITÄT ZWISCHEN ZUGANGSWEGEN FÜR MENSCHEN UND FÜR KI-AGENTEN

Eine Organisation muss wissen, welche Bedarfe ihrer Kundschaft und Beschäftigten KI-Agenten erfüllen können und wie viel des Dienstes sie damit abdecken. Dieses Papier schlägt Dienstparität als Maß dafür vor, ob derselbe Bedarf über verschiedene Kanäle gedeckt werden kann und welche Tätigkeiten bewusst Menschen vorbehalten wurden. Parität verlangt weder identische Geschwindigkeit noch jede Funktion in jedem Kanal. Tabelle 6 veranschaulicht den Vergleich anhand einer fiktiven Behörde.

DIENSTBEDARF	ZUGANG FÜR KUNDSCHAFT ODER BESCHÄFTIGTE	EXTERNE KI-AGENTENSCHNITTSTELLE	INTERNER KI-AGENT	MENSCHLICHE ENTSCHEIDUNG
Förderbedingungen lesen	Ja: Selbstbedienung	Ja: lesen	Ja: unterstützen	Nicht erforderlich
Antrag vorbereiten und einreichen	Ja: Selbstbedienung	Ja, im Rahmen der von der antragstellenden Person erteilten Berechtigung	Nein	Nicht erforderlich
Behörde fordert fehlende Anlage an	Ja: Amtsperson im Fallbearbeitungssystem	Nein	Ja: Routineanfrage innerhalb der Grenzen	Für diesen Routineschritt nicht erforderlich
Antragstellende Person reicht angeforderte Anlage nach	Ja: Selbstbedienung	Ja, im Rahmen der von der antragstellenden Person erteilten Berechtigung	Nein	Nicht erforderlich
Inhalt bewerten	Ja: Amtsperson im Fallbearbeitungssystem	Nein	Entwurf mit Quellen	Amtsperson prüft
Ermessensentscheidung treffen und Verwaltungsakt erlassen	Ja: Amtsperson im Fallbearbeitungssystem	Nein	Nein: vorbehaltener Schritt	Amtsperson entscheidet

Tabelle 6. Dienstparität in einer beispielhaften Behörde. „Nein" bedeutet ein bewusstes Verbot. „Nicht erforderlich" bedeutet, dass dieser Schritt keine zusätzliche menschliche Entscheidung braucht; Berechtigung und gesetzlich erforderliche Bestätigungen gelten weiterhin. Die Anforderung einer Anlage durch die Behörde und die Antwort der antragstellenden Person sind getrennte Schritte.

49 EU-KI-Verordnung, Artikel 16; der Anbieterstatus nach Änderungen wird in Artikel 25 behandelt.
50 Chan et al., „Infrastructure for AI Agents", 2025, Abschnitt 5.1 (arXiv v3, S. 17–19).

Der Zugang ohne persönlichen KI-Agenten zeigt, welche Dienstbedarfe ohne einen solchen KI-Agenten gedeckt werden können (Abbildung 4). Die Organisation prüft gesondert, wie ihre Beschäftigten den Dienst weiter erbringen, wenn die eigenen KI-Agenten angehalten werden oder ausfallen. Ein Formular für die Kundschaft allein belegt diese Fähigkeit nicht.

Die dienstverantwortliche Person legt für jeden Dienst Ziele zur Kanalabdeckung sowie zur Fortführung oder Wiederherstellung fest. Wesentliche Dienste und Komfortangebote können unterschiedliche Ziele haben, die getrennt ausgewiesen werden. Geringere Folgen können bei einem Komfortangebot mehr KI-Agentenautonomie rechtfertigen. Ein Dienst oder eine Funktion kann ausschließlich über einen KI-Agenten oder ausschließlich über eine Benutzerschnittstelle für Menschen verfügbar sein, wenn dies eine bewusste, begründete Entscheidung innerhalb des geltenden Rechts ist. Die Organisation bewertet, wessen Zugang eingeschränkt wird und wer zusätzliche Kosten trägt. Wesentliche Dienste und die Ausübung von Rechten, etwa das Anfechten einer Entscheidung oder der Zugriff auf eigene Daten, müssen praktisch zugänglich bleiben. Der vorübergehende Ausfall einer ausschließlich über KI-Agenten zugänglichen Funktion kann vertretbar sein, wenn seine Folgen bewertet und eine Wiederherstellungsfrist festgelegt wurden.

Die KI-Agentenabdeckung zeigt, welche Dienste KI-Agenten erfüllen können, etwa ob ein KI-Agent einen Antrag im Namen der antragstellenden Person einreichen oder nur vorbereiten kann. Bei der Festlegung von Grenzen sollten Organisationen berücksichtigen, dass KI-Agenten Benutzerschnittstellen für Menschen nutzen und Menschen nachahmen könnten, wenn eine Handlung über die KI-Agentenschnittstelle nicht verfügbar ist. KI-Agentenschnittstellen sollten mit reinem Lesezugriff beginnen und schrittweise erweitert werden. Die Vorbereitung einer menschlichen Entscheidung zählt als KI-Agentenunterstützung, sollte aber getrennt vom selbstständigen Abschluss ausgewiesen werden. Auch die Abdeckung externer Schnittstellen und interner KI-Agenten sollte getrennt dargestellt werden.

Die letzte Spalte in Tabelle 6 benennt Entscheidungen, die bei Menschen bleiben, und die verantwortliche Rolle, etwa die Amtsperson, die eine Ermessensentscheidung trifft. Für jeden Dienstbedarf sollte die Organisation entscheiden, welche Kanäle ihn abdecken und ob eine menschliche Entscheidung erforderlich ist. Sowohl „Nicht erforderlich" als auch ein begründetes Verbot zählen als getroffene Entscheidungen; ein ungeprüfter Bedarf ist eine Lücke. Jede neue Funktion erfordert diese Bewertung, unabhängig davon, welcher Kanal sie zuerst anbietet. Die Entscheidungsabdeckung erfasst, ob diese Festlegungen getroffen wurden, nicht ob die gewählten Kanäle praktisch funktionieren.

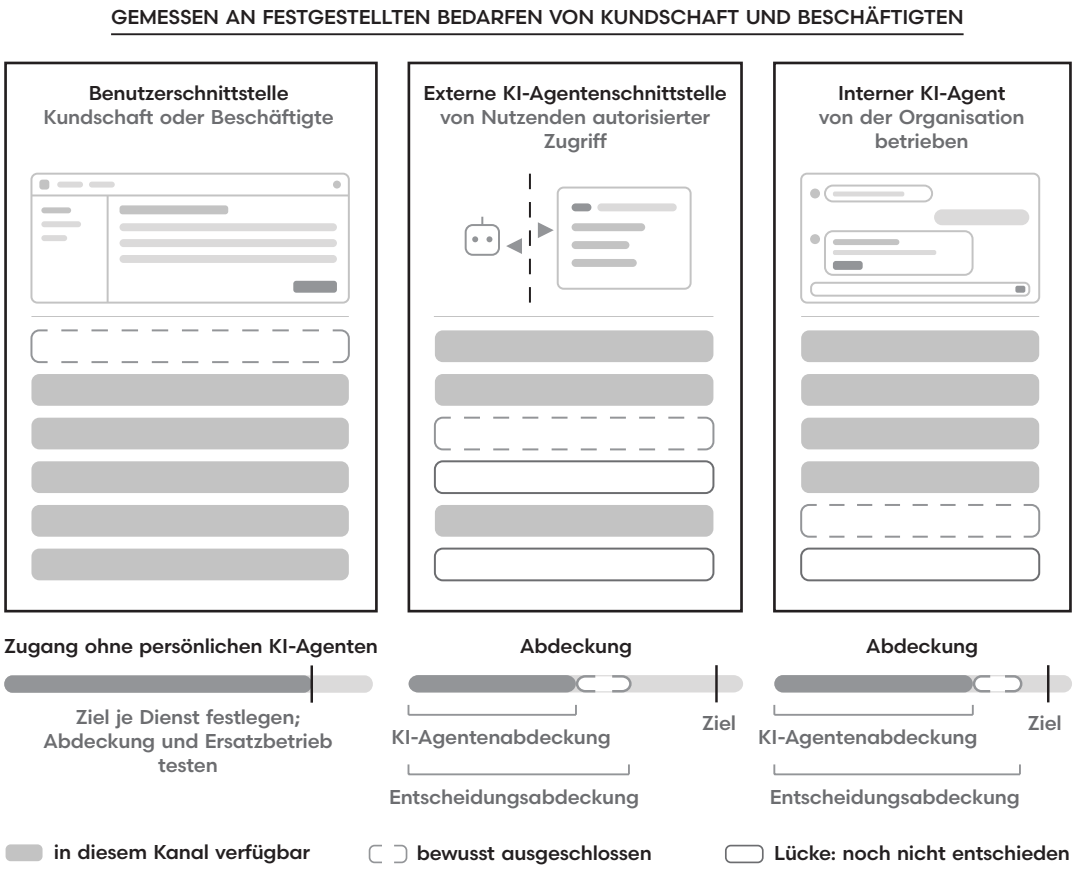


Abbildung 4. Die Kanalabdeckung wird anhand der Bedarfe von Kundschaft und Beschäftigten gemessen und mit dienstspezifischen Zielen verglichen. Zugang ohne persönlichen KI-Agenten und die Weiterarbeit der Beschäftigten ohne die eigenen KI-Agenten der Organisation sind getrennte Fragen.

Parität hilft, Zugangslücken sichtbar zu machen. Ob sich ein Dienst verbessert, ist jedoch daran zu messen, ob der Bedarf der Person gedeckt wird. Ein koordinierter Dienst kann mehrere Anträge oder wiederholte Nachweisanforderungen entfallen lassen, ohne zusätzliche KI-Agentenhandlungen zu schaffen. Die dienstverantwortliche Person sollte die Zeit bis zur Lösung, den aktiven Verwaltungsaufwand der Person und den gesamten Aufwand in den beteiligten Organisationen einschließlich Prüfung und Korrektur messen. Ein durch Vereinfachung entfallener Schritt ist keine Abdeckungslücke.

Goodharts Gesetz warnt davor, dass eine Kennzahl an Aussagekraft verlieren kann, wenn sie zum Ziel wird.[51] Organisationen könnten KI-Agentenfunktionen mit geringem Nutzen ergänzen oder schwierige Bedarfe als verboten kennzeichnen, um die Entscheidungsabdeckung zu erhöhen. Ein Verbot erhöht die KI-Agentenabdeckung nicht; deshalb sind die beiden Maße getrennt. Dienstbedarfe sollten die primäre

51 Goodhart, „Problems of Monetary Management", 1975; die weitverbreitete Formulierung diskutiert Strathern, „Improving ratings", 1997, S. 308.

Messeinheit sein. Jedes Verbot sollte begründet dokumentiert werden. Die Leitung sollte Verbote und Ziele jährlich anhand tatsächlicher Zugangsmöglichkeiten, Wartezeiten und Ausfälle überprüfen. Organisationen können Bedarfe zusätzlich nach Fallvolumen gewichten, um zu zeigen, wie viel vom tatsächlichen Fallaufkommen KI-Agenten bearbeiten. Eine einfache Zählung bleibt nützlich, weil hohe Mengen bei wenigen einfachen Aufgaben Lücken an anderer Stelle nicht verdecken dürfen.

4.3. FÜNF GRUNDSÄTZE FÜR JEDEN DIENST MIT KI-AGENTENEINSATZ

Jeder der folgenden Grundsätze braucht eine verantwortliche Person innerhalb der Organisation, einen Nachweis seiner Wirksamkeit und eine Reaktion, falls er nicht funktioniert.

1. Dienste weiterhin in den Sprachen anbieten, die Menschen dafür zu nutzen berechtigt sind. Die Sprache kann Modellqualität und Tokenkosten beeinflussen; der Unterschied variiert jedoch nach Modell und Aufgabe.[52] Die dienstverantwortliche Person sollte die Leistung in den relevanten Sprachen testen und die Kosten einplanen. In Estland gehört dazu, Dienste auf Estnisch zu erhalten. Andernorts richtet sich derselbe Grundsatz nach den sprachlichen Verpflichtungen der Rechtsordnung und den Bedürfnissen der Nutzenden. Schlechte Leistung sollte die Rolle des KI-Agenten begrenzen oder menschliche Unterstützung auslösen, statt Menschen von dem Dienst auszuschließen.

2. Informationsbeschränkungen dort durchsetzen, wo Daten die Organisation verlassen. Informationsklassifizierung, Rechtsgrundlage und Zweck bestimmen, was an ein externes KI-Modell gesendet werden darf. Strukturierte Datensätze können nach dem Ersetzen direkter Identifikatoren weiterhin nützlich sein. Pseudonymisierung macht personenbezogene Daten jedoch nicht anonym und hebt Datenschutzpflichten nicht auf. Freitext kann identifizierende Angaben enthalten, die zusätzliche Prüfung erfordern. Einen Namen zu entfernen, erlaubt keine ansonsten verbotene Offenlegung. Die für die Informationen verantwortliche Stelle und die Sicherheitsfunktion sollten zulässige Datenflüsse festlegen und alle übrigen blockieren.

3. Anforderungen an die Betriebsfortführung ohne KI-Agenten erfüllen. Die dienstverantwortliche Person bestimmt, welche Arbeit ohne KI-Agenten weiterlaufen muss, wie lange andere Funktionen ausfallen dürfen und wie sie wiederhergestellt werden. Die Organisation vereinbart, welches Arbeitsvolumen Beschäftigte ohne KI-Agenten bewältigen können und welche Wartezeit für einen kritischen Fall höchstens zulässig ist. Sie erprobt das Ersatzverfahren vor dem Start und danach regelmäßig. Ein Anbieterausfall oder eine Preisänderung darf einen wesentlichen Dienst nicht zum Stillstand bringen. Bainbridges Analyse der Automatisierung erklärt die Schwierigkeit, selten genutzte Fähigkeiten zu erhalten.[53] Neue Beschäftigte lernen möglicherweise nie die Vorbereitung, die stets KI-Agenten übernehmen. Hängt die Betriebsfortführung

---

52 Petrov et al., „Language Model Tokenizers Introduce Unfairness Between Languages", 2023. Tatsächliche Dienstkosten und Sprachqualität sind mit dem vorgesehenen Modell zu testen.

53 Bainbridge, „Ironies of automation", 1983, S. 775–776.

davon ab, dass Beschäftigte einspringen, sollten sie regelmäßig einige Fälle ohne von KI-Agenten vorbereitetes Material abschließen. Diese Fähigkeit zu erhalten, beansprucht Arbeitszeit und mindert den Nettonutzen des KI-Agenten; die Organisation muss dies einplanen.

4. Den Anbieterwechsel praktisch ermöglichen. Jede Beschaffung mit Beteiligung eines Modellanbieters sollte dessen Austausch vorsehen, mit bekannten Kosten und Zeitbedarf sowie einem praktisch erprobten Wechsel. Jede Funktion doppelt vorzuhalten kann zu teuer sein. Die wesentlichen Fähigkeiten sind zu bestimmen und ihre Übertragbarkeit ist zu testen. Für EU-Finanzunternehmen gelten unter DORA spezifische Pflichten zur IKT-Geschäftsfortführung und zum Ausstieg; andere Organisationen benötigen Anforderungen entsprechend ihrem rechtlichen Umfeld.[54] Bei einem neuen Anbieter oder einer neuen Modellversion beginnt die Organisation mit engeren Berechtigungen und misst die Fehlerquote erneut. Lässt sich kein vertretbarer Ausstieg nachweisen, sollte sie die Abhängigkeit begrenzen.

5. Die Gesamtkosten durch den Dienstnutzen rechtfertigen. Kosten und Zeit je abgeschlossenem Fall sind einschließlich Prüfung, Korrektur, Integration und Betriebsfortführung zu messen. Diese Ergebnisse sind mit regelbasierter Automatisierung und menschlicher Arbeit zu vergleichen; bessere Nachweise, weniger Mängel und kürzere Wartezeiten sind zu berücksichtigen. Kostet die Prüfung mehr als die Aufgabe selbst und verbessert sie das Ergebnis nicht, ist die Aufgabe schlecht gewählt. Nimmt die Prüfung einen wachsenden Anteil der Arbeit ein, sollte die Aufgabe enger gefasst, stabile Schritte in Regeln überführt oder der KI-Agenteneinsatz beendet werden. Auch auf Antragstellende, Beschäftigte und andere Organisationen verlagerte Arbeit ist einzubeziehen.

---

54 DORA, Artikel 11, 28 und 30. Dies sind Pflichten des Finanzsektors, die hier als Beispiel und nicht als Regel für jede öffentliche Stelle dienen.

5. WAS DER STAAT BEREITSTELLEN MUSS

Regierungen können unmittelbar helfen, indem sie Dienste vereinfachen, die Unterstützung delegierter Handlungen durch bestehende Infrastruktur klären und Menschen Einblick in die in ihrem Namen ausgeführten Handlungen geben. Gemeinsame organisationsübergreifende Berechtigungen sind ein weiterer Schritt, der bei wiederkehrendem Bedarf geprüft werden sollte. Länder sollten mit den Funktionen beginnen, die ein Dienst benötigt, und dabei ihre rechtliche und technische Ausgangslage berücksichtigen.

5.1. KI-AGENTEN ERSETZEN NICHT DIE NOTWENDIGKEIT EINER EFFIZIENTEREN VERWALTUNG

Ein KI-Agent kann Menschen durch fragmentierte Dienste führen, indem er Fristen im Blick behält und dieselben Angaben an mehreren Stellen einträgt. Das kann die Fragmentierung für Organisationen verdecken und ihre Kosten auf diejenigen verlagern, die den KI-Agenten bezahlen. Der Zugang zu wesentlichen öffentlichen Diensten und die Ausübung von Rechten mit vertretbarem Aufwand sollten nicht von der Verfügbarkeit oder Leistungsfähigkeit eines persönlichen KI-Agenten abhängen. Organisationen sollten Doppelarbeit verringern und ihre Verfahren koordinieren. KI darf kein Grund werden, schlechte Verwaltung beizubehalten.

Manchmal lässt sich der Antrag ganz vermeiden. Wenn das Recht den Anspruch eindeutig festlegt und die notwendigen Informationen bereits vorliegen, kann ein regelbasierter Prozess einen proaktiven Dienst auslösen. In Estland kann die Registrierung der Geburt eines Kindes ein Angebot für Familienleistungen durch das Sozialversicherungsamt auslösen. Der betreffende Elternteil bestätigt es abhängig von den jeweiligen Umständen im Selbstbedienungsportal.[55] Damit entfällt ein Antrag, den andernfalls ein KI-Agent vorbereiten könnte. Bevor die dienstverantwortliche Person einem KI-Agenten das Lesen von Anträgen überträgt, sollte sie fragen, ob der Antrag überhaupt benötigt wird.

In Estland ist die Unterstützung für Alltagshilfsmittel und Medizinprodukte bislang zwischen Sozialversicherungs- und Krankenversicherungssystem aufgeteilt. Orthopädische Schuhe und individuell gefertigte Einlagen können unterschiedlichen Regelungen unterliegen; bei Arbeitsplatzhilfen ist die Arbeitsverwaltung beteiligt. Für eine Person können dies Teile eines einzigen Bedarfs sein. Im Juni 2026 kündigte das Sozialministerium an, Hilfsmittel und Medizinprodukte im Zeitraum 2027 bis 2029 in einer einheitlichen Regelung der Krankenversicherung zusammenzuführen.[56] Dies ist eine angekündigte Reform, kein bereits fertiggestellter Dienst. Ein KI-Agent, der getrennte Anträge ausfüllt, würde den Umgang mit der bestehenden Aufteilung erleichtern, ohne sie aufzulösen.

Kostengünstigere Texterstellung kann auch die Nachfrage bei öffentlichen Stellen erhöhen. Manche zusätzlichen Anträge, Auskunftsersuchen und Rechtsbehelfe betref-

---

55 Estnisches Sozialversicherungsamt, „Millised perehüvitised kaasnevad lapse sünniga?", 27. Januar 2026, „Pakkumus – mis see on?". Der Elternteil bestätigt das Angebot im Selbstbedienungsportal; Anspruchsvoraussetzungen und familiäre Umstände bleiben maßgeblich.

56 Estnisches Sozialministerium, „Abivahendid ja meditsiiniseadmed liidetakse ühtseks süsteemiks", 17. Juni 2026; estnische Arbeitslosenkasse, „Tööalane abivahend". Die Zusammenführung 2027–2029 ist ein angekündigter Plan.

fen berechtigte Anliegen, für die Menschen zuvor Zeit oder Zutrauen fehlten; andere können wiederholend oder schlecht begründet sein. Eine Studie von 2026 sammelte 84 mögliche Fälle aus elf Rechtsordnungen, in denen Amtspersonen oder andere Quellen einen Nachfrageanstieg auf KI-Nutzung zurückführten. Ihre Methode belegt weder die Verbreitung des Effekts noch, ob KI den Anstieg verursachte.[57] Organisationen sollten ihre eigene Arbeitsbelastung messen, bevor sie eine Ursache unterstellen. Einfachere Dienste und eine KI-Agentenschnittstelle, die strukturierte Anfragen unter geprüfter Berechtigung annimmt, können vermeidbare Arbeit reduzieren. Eingangsprüfungen müssen dem geltenden Verfahren folgen. Technische Prüfungen dürfen keine rechtswidrige Hürde für die Geltendmachung eines Anspruchs, die Behebung eines Mangels oder die Ausübung eines Rechts werden. Auch Gebühren und andere Hürden können Menschen davon abhalten, berechtigte Anliegen vorzubringen. Der Zugang zu Rechten darf nicht davon abhängen, ob eine Person einen KI-Agenten hat oder wie leistungsfähig dieser ist.

Überschreitet der Bedarf einer Person Organisationsgrenzen, sollten das zuständige Ministerium und die beteiligten Stellen, gegebenenfalls einschließlich der Kommunen, vereinbaren, wer den gesamten Dienst koordiniert und wohin sich die Person wenden kann (Abbildung 5). Der Europäische Interoperabilitätsrahmen, der britische Service Standard[58] und Estlands Lebensereignisdienste unterstützen diesen Ansatz.[59] Die Lösung braucht vereinbarte Befugnisse, Finanzierung und Eskalationswege. Die dienstverantwortliche Person bleibt für die Koordination zuständig; jede Organisation behält ihre Entscheidungsbefugnisse, auch wenn ein KI-Agent die Schritte verbindet. Informationen sollten wiederverwendet werden, soweit dies rechtmäßig ist. Menschen und beteiligte Organisationen sollten sehen können, wer den nächsten Schritt verantwortet und wo ein Fall wartet. Die dienstverantwortliche Person sollte messen, ob ein Anliegen über einen einzigen Kontakt gelöst wird, welcher Verwaltungsaufwand für die Person verbleibt und wie viel Arbeit insgesamt in den beteiligten Organisationen anfällt. Integration verursacht Kosten,[60] und manche Überschneidungen können sinnvoll sein. Deshalb sollte einer Neuordnung die Untersuchung des tatsächlichen Problems vorausgehen.[61]

---

57 Schmitz, Hammond und Chan, „Characterizing Agentic Flooding of Government Services", arXiv:2608.16603v2, 2026, Zusammenfassung und Abschnitte 1, 3.2 und 4. Die Autoren identifizieren mögliche Fälle; ihre Methode erlaubt weder eine kausale Zuordnung noch eine quantitative Schätzung des Gesamteffekts.

58 Europäische Kommission, New European Interoperability Framework, 2017, Empfehlungen 11 und 25 und Abschnitt 4.3.1; GOV.UK, Service Standard, Punkt 2, „Solve a whole problem for users".

59 RIA, „Sündmusteenused" (Lebensereignisdienste); Europäischer Interoperabilitätsrahmen, Empfehlung 29.

60 OECD, Integrating Social Services for Vulnerable Groups, 2015, S. 51–52 und 197–208.

61 US Government Accountability Office, Fragmentation, Overlap, and Duplication: An Evaluation and Management Guide, GAO-15-49SP, 2015.

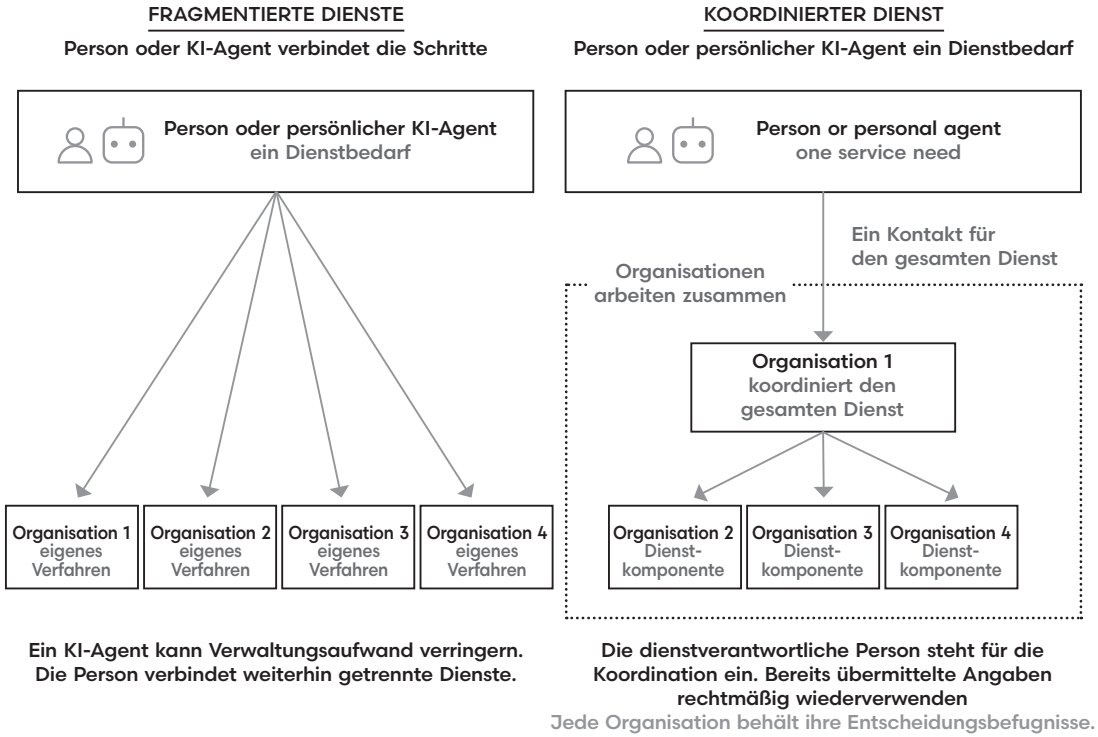


Abbildung 5. Fragmentierte und koordinierte Leistungserbringung für einen bestimmten öffentlichen Dienstbedarf. Die dienstverantwortliche Person steht für die Koordination des gesamten Dienstes ein; jede Organisation behält ihre Entscheidungskompetenz.

5.2. SYSTEME DER EINZELNEN ORGANISATIONEN AN BESTEHENDE DIGITALE GRUNDLAGEN ANPASSEN

Estland verfügt bereits über elektronische Identitäten, digitale Signaturen, Register mit Vertretungsrechten und eine verteilte Datenaustauschplattform. Das EU-Recht verleiht qualifizierten elektronischen Signaturen die gleiche Rechtswirkung wie handschriftlichen Unterschriften.[62] Andere Länder können die notwendigen Funktionen anders bereitstellen. Für einen ausgewählten Dienst braucht die Organisation verlässliche Authentifizierung, eine Möglichkeit zur Prüfung der Befugnisse der nutzenden Person, rechtmäßigen Informationszugang und widerrufbare Berechtigungen, die bei der Ausführung durchgesetzt werden. Sie muss nicht Estlands gesamte Infrastruktur nachbilden. Ihr Zugriffsmanagement muss die Person, in deren Namen eine Handlung erfolgt, von der ausführenden KI-Agentensitzung unterscheiden, damit der KI-Agent getrennte, engere Berechtigungen erhält.

Die Bank LHV benötigte für ihre KI-Agentenschnittstelle vom Staat im Wesentlichen nur die nationale digitale Identitätslösung. Kundinnen und Kunden authentifizieren sich auf der eigenen Seite der Bank. Diese übernimmt Einwilligung, Berechtigungsauswahl,

62 RIA-Seiten zu elektronischen Identitätsdiensten, Pääsuke zur Verwaltung von Vertretungsrechten und X-tee; eIDAS-Verordnung, Artikel 25 Absatz 2. Sie belegen bestehende Grundlagen, keine allgemeine Einsatzbereitschaft für KI-Agentendelegation.

Protokollierung und Widerruf. Die veröffentlichte Beta-Dokumentation beschreibt ein Zugriffstoken mit einer Gültigkeit von einer Stunde und ein Aktualisierungstoken mit einer Gültigkeit von 30 Tagen; der Widerruf erfolgt im Internetbanking.[63] Andernfalls würden Organisationen diese sicherheitsrelevanten Funktionen wiederholt selbst umsetzen. Die konkrete Lösung von LHV ist jedoch keine allgemeine Regel für öffentliche Dienste.

Organisationen brauchen zunächst zwei praktische Leitfäden, erarbeitet mit Dienstverantwortlichen, Sicherheitsfachleuten und Rechtsberatung. Der Leitfaden für externen Zugriff sollte getrennte KI-Agentenzugangsmittel, Berechtigungsumfang, Sitzungsbindung, Ausführungsprüfungen, Widerruf und Fehlerfälle erläutern. Der Leitfaden für interne KI-Agenten sollte Arbeitsverteilung, Prüfung, Stoppen und Betriebsfortführung abdecken. Beide sollten geeignete offene Standards verwenden und Tests mit repräsentativen kommerziellen KI-Agenten umfassen. Gemeinsame Softwarekomponenten sind dort gerechtfertigt, wo Organisationen andernfalls sicherheitskritische Arbeit doppelt leisten würden und eine gemeinsame Umsetzung Kosten oder Risiken senken würde.

Der Staat kann die Leitlinien anhand eines begrenzten Dienstes auf seinem eigenen Portal demonstrieren; Estlands eesti.ee ist ein möglicher Kandidat. Die vorgeschlagene externe KI-Agentenschnittstelle würde dem von einer Person gewählten KI-Agenten erlauben, Portalinformationen zu lesen und zulässige Handlungen auszuführen. Der eigene KI-Serviceagent des Portals könnte einen Antrag vorbereiten und nach Bestätigung durch die Person einreichen. Leitet das Portal eine Anfrage an eine andere Organisation weiter, muss die empfangende Stelle die anfragende Organisation identifizieren, deren Befugnisse und die Rechtsgrundlage der Datennutzung prüfen und ihre eigenen rechtlichen und verfahrensbezogenen Anforderungen anwenden. Die Berechtigung eines persönlichen KI-Agenten kann eine fehlende Rechtsgrundlage für die organisationsübergreifende Verarbeitung nicht ersetzen.

Der Pilot sollte mit dem Lesen öffentlicher Informationen und eigener Datensätze einer Person beginnen und anschließend eine Handlung ermöglichen, sobald die Berechtigungsprüfungen erprobt sind. Die dienstverantwortliche Person sollte den Umfang des Piloten veröffentlichen und Lösungszeit, Nutzeraufwand, Arbeitsaufwand der Organisation und Fehler mit dem einfacheren bestehenden Zugangsweg vergleichen. Versagen Berechtigungsgrenzen oder Widerruf, ist die betroffene Handlung auszusetzen. Verkürzt sie die Lösungszeit nicht und verringert sie auch nicht den Arbeitsaufwand, ist sie nicht auszuweiten. Tabelle 7 trennt bestehende Grundlagen von vorgeschlagener Unterstützung.

---

63 LHV Pank, Ankündigung vom 15. Juli 2026; lhv.ai, „Security and trust" und „Common questions", geprüft am 20. Oktober 2026. Die veröffentlichten Tokenlaufzeiten und der reine Lesezugriff beschreiben diese Beta-Umsetzung.

FUNKTION ODER UNTERSTÜTZUNG	NUTZUNG UND STAND IM ESTNISCHEN BEISPIEL
Verlässliche elektronische Identität	Bestehende Grundlage. Die Person authentifiziert sich und erteilt die Berechtigung im System der Organisation; der KI-Agent erhält getrennte Zugangsmittel.
Digitale Signatur	Bestehende Grundlage, soweit Recht oder Verfahren eine Unterschrift verlangen. Persönliche Signaturmittel bleiben bei der Person.
Register und Vertretungsrechte	Bestehende Quellen zur Prüfung, ob eine Person für eine Organisation oder eine andere Person handeln darf. Der Dienst prüft weiterhin die konkret benötigte Befugnis.
Rechtmäßiger Datenaustausch	X-tee (X-Road) verbindet verteilte Systeme. Wiederverwendung setzt eine rechtmäßige Grundlage und Zugangsvereinbarungen voraus.
Einblick in Datenzugriffe und Handlungen	Data Tracker erfasst bereits Datenzugriffsprotokolle angeschlossener Systeme. Protokolle von KI-Agentenhandlungen und verknüpfte Berechtigungskontrollen sind eine vorgeschlagene Erweiterung.
Rechtsrahmen	Bestehende Regeln bestimmen Identität, Signaturen, Vertretung, Datennutzung und Entscheidungen. Für jeden geplanten KI-Agenteneinsatz ist ihre Anwendbarkeit konkret zu prüfen.
Leitlinien und Testsammlung	Vorgeschlagene gemeinsame Unterstützung für delegierten Zugriff, Durchsetzung, Widerruf, Betriebsfortführung und Interoperabilität.
Referenzdienst im Portal	Vorgeschlagener begrenzter Pilot zur Demonstration der Leitlinien mit gemessenen Ergebnissen und Abbruchbedingungen.

Tabelle 7. Benötigte Funktionen der Organisationen, mit Estland als Beispiel einer Ausgangslage.

5.3. MENSCHEN EINBLICK IN KI-AGENTENHANDLUNGEN UND DATENNUTZUNG GEBEN

Delegation macht Verwaltung für die Person weniger sichtbar. Menschen brauchen eine nutzbare Möglichkeit, in ihrem Namen ausgeführte Handlungen und die Nutzung ihrer Daten einzusehen, die verantwortliche Organisation zu erkennen und Berechtigungen zu widerrufen. Ohne diesen Einblick und diese Kontrolle bleibt ihre Verantwortung für die Aufgaben und Berechtigungen ihres KI-Agenten weitgehend formal.

Estlands Data Tracker bietet einen Ausgangspunkt: Angeschlossene Datenbanken können anzeigen, welche Organisation wann und zu welchem Zweck auf die Informationen einer Person zugegriffen hat.[64] Eine Erweiterung auf KI-Agentenhandlungen wäre eine neue Entwicklung. Für jede Handlung sollte eine externe KI-Agentenschnittstelle erfassen, in wessen Namen sie vorgenommen wurde, welche Berechtigung verwendet wurde und über welchen KI-Agentendienstbetreiber sie erfolgte. Der Eintrag sollte angeben, auf welche Daten für die Handlung zugegriffen wurde oder welche bereitgestellt wurden und zu welchem Zweck. Organisationen sollten entsprechende Aufzeichnungen für ihre internen KI-Agenten führen. Ein Portal könnte diese Aufzeichnungen sowie die Widerrufsmöglichkeit anzeigen. Die Aufzeichnungen sollten bei den Organisationen verbleiben und auf Wunsch der Person abgerufen werden, um keine neue zentrale

64 RIA, „Andmejälgija" (Data Tracker), geprüft am 17. Oktober 2026. Der bestehende Dienst zeigt Datenzugriffsinformationen angeschlossener Systeme. Einblick in KI-Agentenhandlungen und verknüpfter Widerruf sind Vorschläge dieses Papiers.

Historie ihrer Dienstnutzung zu schaffen. Zugang, Zweck, Aufbewahrung und Schutz der zusammengeführten Ansicht benötigen weiterhin ausdrückliche Kontrollen.

Ein Protokoll kann ein Verfahren oder einen Rechtsakt benennen, der einer Person wenig sagt. Ein KI-Portalagent könnte anhand ihrer Unterlagen und Protokolle erklären, was geschehen ist und warum, oder ihr beim Verständnis eines behördlichen Schreibens oder einer Entscheidung helfen. Ein KI-Agent kann ein Protokoll selbstständig mit stichprobenartiger Kontrolle erläutern, wenn Fehler geringe und leicht behebbare Folgen hätten. Dies ist eine geeignete erste Aufgabe für den Referenzdienst. Eine Entscheidungserläuterung sollte auf die Entscheidung und die Hinweise zu ihrer Anfechtung verweisen. Könnte die Erläuterung Rechte, Fristen oder Pflichten beeinflussen, sind geprüfte Standardtexte oder eine menschliche Prüfung nötig. Die Erläuterung ersetzt die maßgeblichen Dokumente nicht, und die Person muss eine Korrektur erhalten können. Die dienstverantwortliche Person sollte prüfen, ob Erläuterungen das Verständnis verbessern, bevor sie deren Einsatz ausweitet.

Häufige Fragen zeigen, welche Schreiben und Verfahrensschritte verbessert werden müssen. Die dienstverantwortliche Person sollte diese Fragen nutzen, um den zugrunde liegenden Dienst so zu verbessern, dass Menschen sie ohne einen KI-Agenten verstehen können.

5.4. GEMEINSAME BERECHTIGUNGEN, WENN WIEDERKEHRENDER BEDARF SIE RECHTFERTIGT

Ein KI-Agent kann gerade deshalb nützlich sein, weil er während der Abwesenheit seiner nutzenden Person handelt. Innerhalb einer Organisation kann eine Berechtigung bereits einen vereinbarten Umfang und Zeitraum festlegen, vorbehaltlich geltenden Rechts und der Sicherheitsanforderungen. Gemeinsame Berechtigungen werden relevant, wenn mehrere Organisationen dieselbe Erlaubnis auslegen und akzeptieren müssen. Organisationen können frühzeitig ein gemeinsames Verständnis solcher Berechtigungen vereinbaren und zugleich nützliche eigene Dienste entwickeln, ohne auf ein universelles System zu warten.

Das in Entwicklung befindliche *Agent Payments Protocol* (AP2) beschreibt eine vorab erteilte, an einen kryptografischen Schlüssel gebundene Berechtigung, bei der Transaktionen anhand vereinbarter Grenzen geprüft werden.[65] Ein Schlüssel belegt die Kontrolle über ein Zugangsmittel. Er begründet für sich genommen weder die rechtliche Identität noch die Vertrauenswürdigkeit oder die Befugnisse des KI-Agenten für jeden Dienst. Der Rechtsrahmen für die europäische Brieftasche für die digitale Identität stellt diese unter die alleinige Kontrolle der nutzenden Person.[66] Das stützt die Trennung persönlicher Zugangsmittel von delegiertem Zugriff, schreibt aber

65 Agent Payments Protocol (AP2), Spezifikation v0.2 und „Agent Authorization Framework", geprüft am 20. Oktober 2026; Google, Übergabe von AP2 an die FIDO Alliance, 28. April 2026. Die Standardisierung dauert an. Dieses Zahlungsprotokoll ist ein Architekturbeispiel, kein allgemein geltendes Recht und kein Nachweis für Interoperabilität über öffentliche Dienste hinweg.

66 eIDAS in der durch die Verordnung zum europäischen Rahmen für eine digitale Identität geänderten Fassung, Artikel 5a Absatz 4 Buchstabe a. Die Brieftasche unterliegt der alleinigen Kontrolle der nutzenden Person; die vorgeschlagene Trennung der KI-Agentenzugangsmittel ist eine Gestaltungsempfehlung dieses Papiers.

keine einheitliche Architektur für alle KI-Agentenhandlungen vor. Dieses Papier schlägt einen getrennten, in seinem Umfang begrenzten Berechtigungsnachweis für delegierte KI-Agentenhandlungen vor. Er darf ausschließlich an Schnittstellen für KI-Agentenzugriff akzeptiert werden, nicht als Zugangsmittel für Sitzungen über den Zugang für Menschen. Ein Widerruf braucht sowohl eine Möglichkeit, die Erlaubnis zurückzunehmen, als auch einen Weg, auf dem jedes empfangende System innerhalb einer festgelegten Zeit von der Änderung erfährt und sie durchsetzt.[67]

National anerkannte KI-Agentenkennungen, gemeinsame Berechtigungen und Zertifizierungen können mit zunehmender organisationsübergreifender Aktivität nützlich werden. Sie sollten anhand eines konkreten Problems bewertet werden, bevor sie als Einsatzvoraussetzung gelten. Auch bei einem anerkannten KI-Agenten muss weiterhin jede Handlung geprüft werden. Die beteiligten Organisationen müssen sich auf die Bedeutung einer Berechtigung verständigen: etwa ob eine monatliche Obergrenze von 500 Euro Gebühren einschließt, wie gleichzeitige Handlungen sie ausschöpfen und was bei einem Widerruf geschieht. Eine gemeinsame Lösung lohnt die Erprobung, wenn viele Dienste ähnliche Berechtigungen brauchen, Menschen wiederholt denselben Umfang freigeben und ein koordinierter Dienst diese Wiederholung nicht beseitigen kann.

Leitlinien können künftigen Integrationsaufwand verringern, indem sie einheitliche Bezeichnungen für Lese-, Vorbereitungs-, Einreichungs- und Zahlungsberechtigungen sowie Felder für Ablauf, Widerruf und Protokollierung festlegen. Branchenspezifische Lösungen lassen sich möglicherweise leichter erproben als eine universelle Berechtigungsregelung. Ein Pilot sollte Berechtigungen, die in jeder Organisation gesondert erteilt werden, mit der gemeinsamen Lösung vergleichen und dabei Nutzeraufwand, Einführungs- und Betriebskosten, Durchsetzungsfehler und Widerrufsverzögerungen messen. Organisationen sollten beim einfacheren Ansatz bleiben, wenn das gemeinsame System ohne ausreichenden Nutzen Kosten oder Unklarheiten hinzufügt.

MÖGLICHER BESTANDTEIL	WELCHES PROBLEM ER LÖSEN KÖNNTE	WANN UNTERSUCHEN
Gemeinsames Berechtigungsformat	Mehrere Organisationen legen denselben Umfang einheitlich aus	Wiederkehrende Nutzungen bestehen; Umfang, Gesamtgrenzen und Abhilfe lassen sich eindeutig festlegen
Gemeinsamer Berechtigungsdienst	Menschen verwalten und widerrufen Berechtigungen an einer Stelle	Getrennte Berechtigungen verursachen erheblichen Aufwand; Widerruf erreicht jeden Empfänger; keine zentrale Aktivitätshistorie wird gesammelt; Schutzvorkehrungen für Datenschutz und Verfügbarkeit sind erprobt
An einen Schlüssel gebundener Berechtigungsnachweis	Delegierte organisationsübergreifende Handlungen bei Abwesenheit der nutzenden Person	Mehrere Aussteller und Empfänger benötigen ihn; der Nachweis funktioniert nur an KI-Agentenschnittstellen; Umfang, Ablauf und Widerruf werden durchgesetzt; ein Schlüssel verleiht keine rechtliche Befugnis

Tabelle 8. Mögliche Bestandteile gemeinsamer Berechtigungen und die vor ihrer Entwicklung erforderlichen Nachweise.

67 South et al., „Authenticated Delegation and Authorized AI Agents", 2025, Abschnitt 3.3 (S. 7). Wirksamer Widerruf hängt zusätzlich von der rechtzeitigen Durchsetzung durch die empfangenden Dienste ab.





Eine Eisstraße veranschaulicht bedingten Zugang: Die Erlaubnis hängt von gemessenen Bedingungen ab und kann zurückgenommen werden. Ein verlässlicher alternativer Weg bleibt nötig. Auch KI-Agentenautonomie braucht Nachweise, Grenzen und ein erprobtes Ersatzverfahren.

FOTO: AUTOGALERIA

6. SZENARIEN UND BETRIEBSFORTFÜHRUNG IN EINER KRATT-NATION

In einer Kratt-Nation sind Regeln, Berechtigungen und Schnittstellen für KI-Agenten lesbar; ihre Grenzen werden außerhalb des Modells durchgesetzt. Dienstverantwortliche legen je Dienst Anforderungen an Kanalabdeckung sowie Fortführung oder Wiederherstellung fest. Wesentliche Dienste bleiben ohne persönlichen KI-Agenten und auch dann praktisch zugänglich, wenn die eigenen KI-Agenten der Organisation angehalten werden oder ausfallen. Zwei fiktive Szenarien zeigen eine Förderbehörde, die den Prozess aus Tabelle 9 nutzt.

SCHRITT	WER IHN AUSFÜHRT	KLASSE	KONTROLLE
KI-Agentenzugriff und Berechtigung prüfen	Regel im empfangenden System	Nicht anwendbar	Nicht autorisierte Maschinenhandlungen zurückweisen; rechtmäßigen Einreichungsweg für die Person erhalten
Eingang erfassen; Mängel oder Fragen zur Förderfähigkeit kennzeichnen	Fallbearbeitungssystem; Beschäftigte, soweit eine fachliche Beurteilung erforderlich ist	Nicht anwendbar	Technische Zugangskontrolle nicht als rechtliche Entscheidung über die Verweigerung der Entgegennahme behandeln; erforderliche Mängelbehebung ermöglichen
Einreichung registrieren, wenn Auslegung nötig ist	KI-Agent	NACHKONTROLLE	Amtliche Einträge stichprobenartig prüfen; bei vollständig bestimmter Registrierung eine Regel nutzen
Fehlende Anlage anfordern	KI-Agent in engem Rahmen	AUSFÜHREN	Text stichprobenartig prüfen; keine neue Verfahrenspflicht, Friständerung oder automatische Aussetzung
Unvollständigen Antrag ungeprüft lassen	Amtsperson mit KI-Agentenvorbereitung	FREIGABE oder NUR MENSCH nach jeweils geltendem Recht	Amtsperson prüft Nachweise, Rechtsgrundlage und Rechtsbehelfs- und Abhilfemöglichkeiten und kann den Vorschlag ablehnen; Menschen vorbehaltene Entscheidungen bleiben bei Menschen
Bewertung anhand der Programmziele entwerfen	Interner KI-Agent	AUSFÜHREN	Entwurf verweist auf Antrag und Programmregeln; keine Wirkung auf den Fall ohne Prüfung; Leitung prüft regelmäßig Lesezeit, Änderungen und Stichproben
Ermessensentscheidung treffen	Amtsperson	NUR MENSCH	Amtsperson prüft Quellen und unterzeichnet; Prüfqualität überwachen
Entscheidung erläutern	KI-Agent oder Amtsperson je nach Folgen	AUSFÜHREN bei einer Zusammenfassung mit geringen Folgen	Entscheidung und Anfechtungshinweise verlinken; Freitextzusammenfassungen stichprobenartig prüfen; Rechte und Fristen durch Regeln oder Menschen prüfen
Gesamten Dienst steuern	Dienstverantwortliche Person	Nicht anwendbar	Webformular erhalten; Kanal- und Fortführungsziele erfüllen; Autonomieschwellen anhand gemessener Fehlerquoten anpassen; Ersatzverfahren und Anbieterwechsel testen; Gesamtkosten, Zeit und Schäden bewerten

Tabelle 9. Beispiel eines Förderverfahrens. Technische Prüfungen bleiben vom rechtlichen Antragseingang und von Entscheidungen über einen Antrag getrennt.

Die dienstverantwortliche Person bewertet den Gesamtprozess anhand von vier Maßen: Gesamtkosten je abgeschlossenem Fall einschließlich Prüfung und Korrektur, Zeit bis zur Lösung, nach Folgen gewichtete Fehler und das Arbeitsvolumen, das Beschäftigte ohne KI-Agenten bewältigen können. Die Behörde prüft außerdem, ob Antragstellende einen nutzbaren Zugang und wirksame Anfechtungsmöglichkeiten behalten.

Blaues Szenario: KI-Agenten werden zum normalen Bestandteil des Dienstes

Eine Förderbehörde führte vor drei Jahren KI-Agenten ein. In den ersten drei Monaten arbeitete ein KI-Agent parallel zu den Beschäftigten im Schattenbetrieb. Erfahrene Prüfende verglichen seine Bewertungen und vorgeschlagenen Dokumentenanforderungen mit den Nachweisen und der Arbeit der Beschäftigten. Erst nach Erprobung der Aufgabe erlaubte die Behörde ihm, routinemäßige Anforderungen fehlender Anlagen innerhalb der Grenzen aus Tabelle 9 zu versenden. Der KI-Agent stellt nun Nachweise zusammen, damit eine Amtsperson entscheiden kann, sobald der Fall entscheidungsreif ist. Fälle, die früher einen Monat dauerten, werden innerhalb einer Woche gelöst. Der Personalbestand ist gleich geblieben, die Warteschlange jedoch verschwunden.

Im zweiten Frühjahr führte die Behörde eine neue Modellversion mit zunächst engeren Berechtigungen und verstärkten Stichproben ein. In einem kleinen Förderprogramm begann der KI-Agent, bereits beigefügte Dokumente erneut anzufordern. Einige Antragstellende warteten ab, während sie versuchten, diese Unterlagen zu beschaffen, obwohl die Anfragen das Verfahren rechtlich nicht aussetzen konnten. Einzelne Anforderungen wirkten plausibel und fielen zunächst nicht auf. Bei ihrer monatlichen Zufallsprüfung fanden Amtspersonen zwei solche Fälle. Die dienstverantwortliche Person verlangte sofort eine Freigabe dieser Anforderungen durch Beschäftigte, untersuchte die betroffene Fallgruppe und veranlasste, dass die liegen gebliebenen Anträge weiterbearbeitet wurden. Nachdem der Anbieter den Fehler korrigiert hatte und eine erneute aufgabenspezifische Prüfung akzeptable Ergebnisse zeigte, stellte die Behörde die begrenzte Berechtigung wieder her.

Rotes Szenario: Betriebsfortführung und Souveränität

Einige Jahre später geben mehrere große Modellanbieter eines Morgens bekannt, dass Exportbeschränkungen sie zur Aussetzung ihrer Dienste in Europa gezwungen haben.[68] Die Behörde verliert den Zugang zu den Modellen, von denen ihre KI-Agenten abhängen. Antragstellende können weiterhin das gewöhnliche Webformular nutzen, und Amtspersonen bearbeiten die Fälle weiter, allerdings wächst die Warteschlange. Die KI-Agenten hatten Bewertungsentwürfe dreimal schneller vorbereitet; später hinzugekommene Beschäftigte brauchen in den ersten Tagen deutlich länger. Das Ersatzverfahren ist vertraut: Die Behörde übt

68 Eine enger begrenzte Störung trat im Juni 2026 auf: Anthropic setzte Fable 5 und Mythos 5 am 12. Juni infolge einer US-Exportkontrollanweisung für alle Nutzenden aus. Die Aktualisierung vom 1. Juli meldete die Wiederherstellung des Zugangs. Die übrigen Modelle blieben verfügbar. Anthropic, Stellungnahme vom 12. Juni und „Redeploying Fable 5", 30. Juni, aktualisiert am 1. Juli 2026.

es zweimal jährlich, und Beschäftigte schließen jeden Monat einige Fälle ohne KI-Agentenmaterial ab. Ein Anbieter außerhalb der betroffenen Lieferkette bleibt in Europa verfügbar. Auch der Anbieterwechsel wurde erprobt. Zwei Wochen später stellt die IT-Leitung die KI-Agenten der Behörde auf dessen Modell um. Die KI-Agenten beginnen mit engeren Berechtigungen, bis die Behörde ihre Leistung erneut gemessen hat. Kritische Fälle haben während der gesamten Störung Vorrang.

Wer einen Kratt arbeiten lässt, muss mit ihm umgehen können. In einem demokratischen Staat muss eine identifizierbare Person oder Organisation für den Einsatz jedes KI-Agenten verantwortlich sein und dessen Grenzen im Rahmen des Rechtsstaats durchsetzen können. Wesentliche Dienste müssen auch an dem Tag weiter erbracht werden, an dem die KI-Agenten abgeschaltet werden.

QUELLENVERZEICHNIS

Verweise auf EU-Recht und estnisches Recht dienen als Beispiele für die jeweilige Rechtsordnung.

Rechtsvorschriften

Verordnung über die digitale operationale Resilienz im Finanzsektor (DORA). **Verordnung (EU) 2022/2554**.

eIDAS-Verordnung. **Verordnung (EU) Nr. 910/2014**, in der geänderten Fassung.

Europäischer Rahmen für eine digitale Identität. **Verordnung (EU) 2024/1183**, zur Änderung von eIDAS.

Datenschutz-Grundverordnung (DSGVO). **Verordnung (EU) 2016/679**.

Zahlungsdiensterichtlinie (PSD2). **Richtlinie (EU) 2015/2366**. Starke Kundenauthentifizierung: **Delegierte Verordnung (EU) 2018/389**, geändert durch Delegierte Verordnung (EU) 2022/2360.

Verbraucherkreditrichtlinie. **Richtlinie (EU) 2023/2225**.

EU-KI-Verordnung. **Verordnung (EU) 2024/1689**, geändert durch **Verordnung (EU) 2026/1744**.

Produkthaftungsrichtlinie. **Richtlinie (EU) 2024/2853**, mit der **Berichtigung von 2026**.

Estland. **Verwaltungsverfahrensgesetz**, §§ 14–15, geltende Fassung abgerufen am 17. Oktober 2026.

Forschung und weitere Quellen

Agent Payments Protocol (AP2). **Spezifikation v0.2** und **Agent Authorization Framework**, ap2-protocol.org, abgerufen am 20. Oktober 2026.

Anthropic. **„Statement on the US government directive to suspend access to Fable 5 and Mythos 5"**, 12. Juni 2026; **„Redeploying Fable 5"**, 30. Juni 2026, aktualisiert am 1. Juli 2026, anthropic.com.

Artikel-29-Datenschutzgruppe. **„Guidelines on Automated individual decision-making and Profiling"**, WP251rev.01, 2018; vom Europäischen Datenschutzausschuss gebilligt.

Autor. **„Polanyi's Paradox and the Shape of Employment Growth"**, NBER Working Paper 20485, 2014, nber.org.

Bainbridge. „Ironies of automation", Automatica 19(6), 1983, S. 775–779, **doi:10.1016/0005-1098(83)90046-8**.

Bogdanov, Etti, Kamm, Ostrak, Stomakhin, Toomsalu, Valdma und Veldre. „Tehisintellekti ja masinõppe tehnoloogia riskide ja nende leevendamise võimaluste uuring" [Studie zu Risiken von KI und maschinellem Lernen und deren Minderung], Cybernetica im Auftrag der RIA, D-16-377, 27. Februar 2024, **ria.ee**.

Brednich, Rolf Wilhelm, Hrsg. **Enzyklopädie des Märchens**, Bd. 9. Berlin und New York: de Gruyter, 1999. „Mandragora", Sp. 115–116.

Calvino, Flavio, Jelmer Reijerink und Lea Samek. **„The Effects of Generative AI on Productivity, Innovation and Entrepreneurship"**, OECD Artificial Intelligence Papers, Nr. 39, 20. Juni 2025, doi:10.1787/b21df222-en.

Chan, Wei, Huang, Rajkumar, Perrier, Lazar, Hadfield und Anderljung. „Infrastructure for AI Agents", Transactions on Machine Learning Research, 2025, **arXiv:2501.10114v3** (19. Juni 2025).

Cui, Kevin Zheyuan, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng und Tobias Salz. **„The Effects of Generative AI on High-Skilled Work: Evidence from Three Field Experiments with Software Developers"**, Management Science, Online-Vorabveröffentlichung, 27. Februar 2026, doi:10.1287/mnsc.2025.00535.

Dowding und Taylor. „Algorithmic Decision-Making, Agency Costs, and Institution-Based Trust", Philosophy & Technology 37, Artikel 68, 2024, **doi:10.1007/s13347-024-00757-5**.

Duden. **„Schrat"**, Bedeutung und Etymologie, duden.de, abgerufen am 18. Oktober 2026.

Einfeldt, Rand. **„Puck and Folklore"**, Early Modern England Encyclopedia, University of Victoria, 2026, lemdo.uvic.ca.

Eloundou, Manning, Mishkin und Rock. „GPTs are GPTs: Labor market impact potential of LLMs", Science 384(6702), 2024, S. 1306–1308, **doi:10.1126/science.adj0998**.

Estnische Behörde für Informationssysteme (RIA). **„Andmejälgija"** [Data Tracker], abgerufen am 17. Oktober 2026.

Estnische Behörde für Informationssysteme (RIA). **„Andmevahetuskiht X-tee"** [Datenaustauschplattform X-tee].

Estnische Behörde für Informationssysteme (RIA). **„Elektroonilise identiteedi teenused"** [elektronische Identitätsdienste].

Estnische Behörde für Informationssysteme (RIA). **„Introduction of X-tee"**, ria.ee.

Estnische Behörde für Informationssysteme (RIA). **„Keskne volituste haldamise infosüsteem Pääsuke"** [zentrale Verwaltung von Vertretungsrechten].

Estnische Behörde für Informationssysteme (RIA). **„Sündmusteenused"** [Lebensereignisdienste].

Estnisches Sozialministerium. **„Abivahendid ja meditsiiniseadmed liidetakse ühtseks süsteemiks"** [geplante Zusammenführung von Hilfsmitteln und Medizinprodukten], 17. Juni 2026.

Estnisches Sozialversicherungsamt. **„Millised perehüvitised kaasnevad lapse sünniga?"** [Familienleistungen nach der Geburt eines Kindes], 27. Januar 2026, abgerufen am 20. Oktober 2026.

Estnische Arbeitslosenkasse. **„Tööalane abivahend"** [Arbeitsplatzhilfsmittel].

Europäische Bankenaufsichtsbehörde. **„Rising application of AI in EU banking and payments sector"**, 25. September 2025, eba.europa.eu.

Europäische Kommission. **„New European Interoperability Framework"**, 2017, ec.europa.eu.

Fehr, Gächter und Kirchsteiger. „Reciprocity as a Contract Enforcement Device: Experimental Evidence", Econometrica 65(4), 1997, S. 833–860**, doi:10.2307/2171941**.

Feng und Chandra. „New development: Principal–agent issues when governments embrace AI agents", Public Money & Management, 2026, **doi:10.1080/09540962.2026.2642814**.

First Internet Bank. **„How does First Internet Bank's Model Context Protocol (MCP) connection work with Claude and ChatGPT?"**, FAQ, firstib.com, abgerufen am 20. Oktober 2026.

Fischli, Franklin, Manzini und Gabriel. „Agents, Alignment, and the Many Faces of Autonomy", Minds and Machines 36, Artikel 34, 2026, **doi:10.1007/s11023-026-09786-9**.

Gabriel. „Artificial Intelligence, Values, and Alignment", Minds and Machines 30(3), 2020, S. 411–437, **doi:10.1007/s11023-020-09539-2**.

Gailmard. „Accountability and Principal–Agent Theory", in Bovens, Goodin und Schillemans (Hrsg.), The Oxford Handbook of Public Accountability, Oxford University Press, 2014, **doi:10.1093/oxfordhb/9780199641253.013.0016**.

GAO. **„Fragmentation, Overlap, and Duplication: An Evaluation and Management Guide"**, GAO-15-49SP, 2015, gao.gov.

Giacomini und Grandi. „New development: I chose to let the machine choose—the rise of 'AIcratism'", Public Money & Management 46(3), 2026, S. 387–391, **doi:10.1080/09540962.2025.2570740**.

Goodhart. „Problems of Monetary Management: The U.K. Experience", Papers in Monetary Economics, Bd. I, Sydney: Reserve Bank of Australia, 1975, S. 1–20.

Google. **„We're donating Agent Payments Protocol to the FIDO Alliance to support the future of secure, agentic payments"**, 28. April 2026. blog.google.

GOV.UK. **„Solve a whole problem for users"**, Service Standard, Punkt 2, aktualisiert am 29. Januar 2026, gov.uk.

Holmström. „Moral Hazard and Observability", The Bell Journal of Economics 10(1), 1979, S. 74–91, **doi:10.2307/3003320**.

Institut für Sprache und Volkskunde (ISOF). **Realkatalogen, Uppsala**, 2021, S. 163, „Bjära, mjölkhare", isof.se.

Institut für estnische Sprache. Etymologisches Wörterbuch, **Eintrag „kratt"**, eki.ee.

Institut für polnische Sprache der Polnischen Akademie der Wissenschaften. **Wielki słownik języka polskiego PAN, Eintrag „skrzat"**, Bedeutung und Etymologie, wsjp.pl, abgerufen am 18. Oktober 2026.

Jensen und Meckling. „Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", Journal of Financial Economics 3(4), 1976, S. 305–360, **doi:10.1016/0304-405X(76)90026-X**.

Jevons. The Coal Question, London: Macmillan, 1865 (2. Auflage 1866), Kapitel VII, „Of the Economy of Fuel".

Kim. „Deep learning and principal–agent problems of algorithmic governance: The new materialism perspective", Technology in Society 63, Artikel 101378, 2020, **doi:10.1016/j.techsoc.2020.101378**.

Kivirähk, Andrus. Rehepapp ehk november. Tallinn: Varrak, 2000.

Laitinen und Sahlgren. „AI Systems and Respect for Human Autonomy", Frontiers in Artificial Intelligence 4, Artikel 705164, 2021, **doi:10.3389/frai.2021.705164**.

LHV Pank. **lhv.ai**, Produktdokumentation zur KI-Agentenschnittstelle, „Security and trust" und „Common questions", abgerufen am 20. Oktober 2026.

LHV Pank. **„Pangandus liigub AI-assistentidesse: LHV loob klientidele uue võimaluse"** [Bankgeschäfte über KI-Assistenten], 15. Juli 2026.

Mastercard und Banco Santander. **„Santander and Mastercard complete Europe's first live end-to-end payment executed by an AI agent"**, Pressemitteilung, 2. März 2026, mastercard.com.

McCubbins, Noll und Weingast. „Administrative Procedures as Instruments of Political Control", Journal of Law, Economics, and Organization 3(2), 1987, S. 243–277**, doi:10.1093/oxfordjournals.jleo.a036930**.

Miller und Whitford. „Trust and Incentives in Principal-Agent Negotiations: The 'Insurance/Incentive Trade-Off'", Journal of Theoretical Politics 14(2), 2002, S. 231–267, **doi:10.1177/095169280201400204**.

Model Context Protocol. Spezifikation, **„Authorization"** und **„Security Best Practices"**, Version 2026-07-28.

Noy und Zhang. „Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence", Science 381(6654), 2023, S. 187–192, **doi:10.1126/science.adh2586**.

OECD. **„Integrating Social Services for Vulnerable Groups: Bridging Sectors for Better Service Delivery"**, 2015, doi:10.1787/9789264233775-en.

OWASP. **„LLM01:2025 Prompt Injection"**, OWASP Top 10 for LLM Applications 2025, genai.owasp.org.

Petrov, La Malfa, Torr und Bibi. „Language Model Tokenizers Introduce Unfairness Between Languages", NeurIPS 2023, **arXiv:2305.15425**.

Polanyi. „The Tacit Dimension", Garden City, NY: Doubleday, 1966.

Raieste, Solvak, Velsberg und McBride. „Government Efficiency in the Age of AI: Toward Resilient and Efficient Digital Democracies", 2025, **nortal.com**.

Rajala und Jalonen. „Beyond Simplification in Public Sector Accountability: Contradictions Between Principal-Agent and Complexity Theories", Public Administration Review 85(6), 2025, S. 1757–1770, **doi:10.1111/puar.13941**.

Schillemans und Busuioc. „Predicting Public Sector Accountability: From Agency Drift to Forum Drift", Journal of Public Administration Research and Theory 25(1), 2015, S. 191–215, **doi:10.1093/jopart/muu024**.

Schmitz, Hammond und Chan. „Characterizing Agentic Flooding of Government Services", **arXiv:2608.16603v2**, 19. August 2026.

South, Marro, Hardjono, Mahari, Whitney, Greenwood, Chan und Pentland. „Authenticated Delegation and Authorized AI Agents", 2025, **arXiv:2501.09674**.

Spence und Zeckhauser. „Insurance, Information, and Individual Action", American Economic Review 61(2), 1971, S. 380–387.

Statens historiska museer. **„Mjölkhare"**, Sammlungsterminologie, samlingar.shm.se, abgerufen am 18. Oktober 2026.

Strathern. „‘Improving ratings’: audit in the British University system“, European Review 5(3), 1997, S. 305–321, **doi:10.1017/S1062798700002660**.

Svenska Akademien. **Svenska Akademiens ordbok, Eintrag „skratte/skratt“**, Substantiv 2, 1973, svenska.se, abgerufen am 18. Oktober 2026.

Visa. **„Visa and Banks Across Europe Reach the Next Phase of Agentic Commerce, Unlocking Secure Access to Merchants“**, 2. Juli 2026, visa.co.uk.

Waterman und Meier. „Principal-Agent Models: An Expansion?“, Journal of Public Administration Research and Theory 8(2), 1998, S. 173–202**, doi:10.1093/oxfordjournals.jpart.a024377**.