Ein souveräner Blueprint für Agentic AI in Behördenprozessen: Von der Vision zur Wertschöpfung

27.10.2025, Olaf Dunkel, https://www.olafdunkel.de

Executive Summary

Die öffentliche Verwaltung steht an der Schwelle zu einem Paradigmenwechsel, der weit über die bisherige Digitalisierung hinausgeht. Agentic AI, die nächste Evolutionsstufe der künstlichen Intelligenz, markiert den Übergang von reaktiven Assistenzsystemen zu autonomen, zielorientierten KI-Agenten, die komplexe, mehrstufige Prozesse eigenständig ausführen können. Dieser Bericht legt einen konkreten, machbaren und souverän gedachten Blueprint vor, der Entscheidern, Führungskräften und CIOs in Behörden einen realistischen Pfad von der Pilotphase bis zur produktiven Implementierung aufzeigt.

Das Kernversprechen von Agentic AI liegt in der fundamentalen Neugestaltung von Verwaltungsarbeit. Durch die Automatisierung ganzer Arbeitsabläufe – von der Antragsbearbeitung bis zur Betrugserkennung – können signifikante Effizienzsteigerungen realisiert, die Qualität von Bürgerdiensten nachhaltig verbessert und der Mangel an Fachkräften abgefedert werden.[1] Gleichzeitig ermöglicht die datengestützte Analysefähigkeit dieser Systeme eine fundiertere, vorausschauende Politikgestaltung.

Der hier vorgestellte Blueprint basiert auf einem praxiserprobten Drei-Phasen-Modell:

1. **Strategie & Entdeckung:** Identifikation von wertstiftenden Anwendungsfällen und Bewertung der organisatorischen Reife.
2. **Konzeption & Implementierung:** Sorgfältiges Design des Zielprozesses, Entwicklung des KI-Agenten und Test in einer kontrollierten Pilotumgebung.
3. **Skalierung & Optimierung:** Überführung des erfolgreichen Piloten in den produktiven Betrieb und Etablierung eines kontinuierlichen Verbesserungszyklus.

Dieser Rahmen ist auf den zentralen Säulen der Souveränität, einer robusten Governance, der Menschzentrierung und einer messbaren Wertschöpfung aufgebaut. Er bietet nicht nur eine technologische Roadmap, sondern auch die notwendigen ethischen, rechtlichen und organisatorischen Leitplanken, um das transformative Potenzial von Agentic AI verantwortungsvoll zu heben und die Verwaltung zukunftsfähig zu gestalten.

Die nächste Stufe der digitalen Verwaltung: Agentic

AI verstehen

Jenseits von Chatbots: Was Agentic AI für Behörden bedeutet

Die Einführung von Agentic AI in der öffentlichen Verwaltung stellt mehr als nur eine inkrementelle Verbesserung dar; sie ist ein fundamentaler Wandel in der Art und Weise, wie staatliche Aufgaben konzipiert und ausgeführt werden. Um dieses Potenzial zu verstehen, ist eine klare Abgrenzung zu bisherigen KI-Anwendungen wie einfachen Chatbots oder reaktiver generativer KI erforderlich.

Agentic AI bezeichnet autonome Systeme, die nicht nur auf Befehle reagieren, sondern proaktiv Ziele verfolgen, sich in Echtzeit an veränderte Bedingungen anpassen und komplexe, mehrstufige Arbeitsabläufe ausführen.[3] Der entscheidende Unterschied liegt im intelligent abgeleiteten Ausführungspfad: Während traditionelle Software einem fest programmierten Ablauf folgt, leiten KI-Agenten ihre Handlungsschritte mithilfe von großen Sprachmodellen (Large Language Models, LLMs) selbstständig ab, um ein definiertes Ziel zu erreichen.[5] Diese Autonomie ermöglicht es ihnen, als „digitale Teamkollegen" zu agieren, die Aufgaben eigenständig managen, anstatt nur als „intelligente Werkzeuge" zu fungieren, die auf menschliche Anweisungen warten.[4]

Dieser technologische Sprung ist die logische Konsequenz einer Entwicklung, die KI in der Verwaltung in drei Reifestufen sieht: Zuerst als Assistent, der bei Einzelaufgaben unterstützt, dann als digitaler Kollege, der in Prozesse integriert ist, und schließlich als autonomer Agent, der komplexe Abläufe unter menschlicher Aufsicht durchführt.[7] Agentic AI repräsentiert diese dritte Stufe und ermöglicht es, von der reinen Digitalisierung bestehender Prozesse zu einer vollständigen Neugestaltung von Verwaltungsabläufen überzugehen.[8]

Die Kernfähigkeiten eines solchen Systems basieren auf einer Kombination aus drei Elementen:

1. **Planung und Logik:** LLMs dienen als „Gehirn" des Agenten, um komplexe Ziele in logische, ausführbare Unterschritte zu zerlegen (Task Decomposition).[9]
2. **Kontext und Gedächtnis:** Agenten verfügen über ein Gedächtnis, das es ihnen erlaubt, Informationen aus früheren Interaktionen zu speichern und für zukünftige Entscheidungen zu nutzen.
3. **Werkzeuge und Handlungsfähigkeit:** Über Programmierschnittstellen (APIs) können Agenten mit externen Systemen interagieren, um Informationen abzurufen (z.B. aus Datenbanken), Aktionen auszuführen (z.B. eine E-Mail versenden) oder andere Dienste zu nutzen. Diese Fähigkeit, Werkzeuge autonom zu verwenden, ist die Quelle ihrer Flexibilität und Macht.[5]

Wertschöpfungspotenziale: Konkrete Anwendungsfälle und Nutzenversprechen

Die Implementierung von Agentic AI in Behördenprozessen verspricht einen weitreichenden Nutzen, der sich in vier zentralen Bereichen manifestiert:

- **Operationelle Effizienz:** Der größte Hebel liegt in der Automatisierung vollständiger End-to-End-Prozesse, nicht nur einzelner Aufgaben.[3] Anwendungsfälle wie die Bearbeitung von Genehmigungen, die Verwaltung von Sozialleistungen oder das Fallmanagement können signifikant beschleunigt werden.[11] Dies führt zu einer Entlastung der Mitarbeiter von repetitiven Routineaufgaben, sodass sie sich auf Tätigkeiten konzentrieren können, die menschliche Stärken wie Empathie, Kreativität und komplexes Urteilsvermögen erfordern.[2]
- **Verbesserte Dienstleistungsqualität:** KI-Agenten ermöglichen die Bereitstellung von personalisierten, proaktiven und rund um die Uhr verfügbaren Bürgerdiensten.[11] Sie können Bürger aktiv durch komplexe Antragsverfahren führen, auf relevante, bisher unbekannte Leistungen hinweisen und die Interaktion mit der Verwaltung reibungsloser gestalten.[4]
- **Fundiertere Entscheidungsfindung:** Durch die Analyse von Echtzeitdaten können KI-Agenten prädiktive Einblicke generieren und datengestützte Empfehlungen für die strategische Planung und Politikgestaltung liefern.[14] Dies stärkt die Fähigkeit der Verwaltung, vorausschauend zu agieren.
- **Erhöhte Resilienz und Sicherheit:** In Bereichen wie der Cybersicherheit können Agenten proaktiv Bedrohungen erkennen und abwehren. In der Finanzverwaltung ermöglichen sie eine effektivere Erkennung von Betrug, Verschwendung und Missbrauch durch die autonome Überwachung und Analyse von Transaktionsdaten.[12]

Die folgende Tabelle illustriert diese Potenziale anhand konkreter Anwendungsfälle und zeigt auf, wie Agentic AI die öffentliche Verwaltung transformieren kann.

Tabelle 1: Anwendungsfälle von Agentic AI in der öffentlichen Verwaltung

Verwaltungsbereich	Konkreter Anwendungsfall	Aufgabe des KI-Agenten	Erwarteter Nutzen	Relevante Quellen
Soziales & Leistungen	Antragsbearbeitung für Sozialleistungen	Proaktive Prüfung der Anspruchsberechtigung, Anforderung fehlender Dokumente,	Schnellere Bearbeitung, Reduzierung von Rückständen, Entlastung der Mitarbeiter, verbesserte	[3]

		automatisierte Genehmigung einfacher Fälle, Eskalation von Grenzfällen an Sachbearbeiter.	Bürgererfahrung.	
Bau & Genehmigungen	Baugenehmigungsverfahren	Automatisierte Prüfung von Anträgen auf Vollständigkeit und grundlegende Konformität, Kommunikation mit Antragstellern, Terminierung von Inspektionen.	Beschleunigung der Verfahren, höhere Transparenz, Reduzierung bürokratischen Aufwands.	[3]
Innere Sicherheit & Justiz	Betrugserkennung & Ermittlungsunterstützung	Autonome Überwachung von Transaktionen, Analyse von Verbindungsdaten zur Identifizierung von Risiken (z.B. sanktionierte Entitäten), Erstellung von Ermittlungsberichten.	Erhöhte Aufklärungsquote, schnellere Fallbearbeitung, strategische Allokation von Personalressourcen.	[12]
Personal & Beschaffung	Onboarding neuer Mitarbeiter	Koordination von IT-Zugängen, Bereitstellung von Schulungsmaterial, Beantwortung von Standardfragen, Planung von Einarbeitungsterminen.	Effizienteres Onboarding, Entlastung der Personalabteilung, verbesserte Erfahrung für neue Mitarbeiter.	[6]

Das Fundament der Souveränität: Ein ethischer und

rechtlicher Rahmen

Die erfolgreiche und nachhaltige Implementierung von Agentic AI in der öffentlichen Verwaltung erfordert mehr als nur technologische Exzellenz. Sie muss auf einem soliden Fundament aus ethischen Prinzipien und rechtlicher Konformität ruhen, das Vertrauen bei Bürgern und Mitarbeitern schafft und die digitale Souveränität des Staates gewährleistet.
Ein souveräner Ansatz für KI in der Verwaltung lässt sich nicht allein auf die Frage des Datenhostings reduzieren. Vielmehr stützt er sich auf eine Triade aus drei untrennbaren Säulen:

1. **Rechtliche Souveränität:** Die strikte Einhaltung der geltenden Gesetzgebung, insbesondere des EU AI Acts, der für europäische Verwaltungen einen verbindlichen Rahmen setzt.[18]
2. **Ethische Souveränität:** Die Verankerung demokratischer Werte wie Fairness, Transparenz, Nichtdiskriminierung und menschliche Kontrolle im Design und Betrieb der KI-Systeme, wie sie in Rahmenwerken wie dem US "Blueprint for an AI Bill of Rights" formuliert sind.[20]
3. **Technologische Souveränität:** Die strategische Entscheidung für modulare, offene Architekturen, die eine Abhängigkeit von einzelnen Anbietern (Vendor Lock-in) vermeiden und die Kontrolle über die eingesetzte Technologie sicherstellen.[6]

Nur das Zusammenspiel dieser drei Säulen ermöglicht es der Verwaltung, KI-Systeme zu schaffen, die nicht nur effizient, sondern auch vertrauenswürdig, resilient und im Einklang mit den Grundwerten einer demokratischen Gesellschaft sind.

Leitplanken für Vertrauen: Kernprinzipien für den verantwortungsvollen Einsatz

Um einen ethisch und rechtlich robusten Rahmen zu schaffen, müssen internationale Best Practices und regulatorische Vorgaben zu einem kohärenten Set von Leitprinzipien synthetisiert werden.
Der **EU AI Act** bildet hierfür die rechtliche Grundlage. Sein risikobasierter Ansatz ist für die Verwaltung von zentraler Bedeutung, da viele behördliche Anwendungen in die Kategorie der „Hochrisikosysteme“ fallen werden.[19] Dies zieht strenge Verpflichtungen nach sich, die bereits in der Konzeptionsphase berücksichtigt werden müssen. Dazu gehören unter anderem eine umfassende Risikobewertung, die Sicherstellung hoher Datenqualität zur Vermeidung von Diskriminierung, die lückenlose Protokollierung von Systemaktivitäten (Traceability), eine angemessene menschliche Aufsicht und ein hohes Niveau an Cybersicherheit.[19]
Ergänzend zu diesen rechtlichen Vorgaben bietet der US-amerikanische **"Blueprint for an AI**

Bill of Rights" eine wertvolle ethische Orientierung. Seine fünf Kernprinzipien stellen den Menschen in den Mittelpunkt und sollten als Designgrundsätze für jedes KI-Projekt in der Verwaltung dienen [20]:

1. **Sichere und effektive Systeme:** Schutz der Bürger vor unsicheren oder fehlerhaften KI-Anwendungen.
2. **Schutz vor algorithmischer Diskriminierung:** Gewährleistung einer fairen und gerechten Behandlung aller Bürger.
3. **Datenschutz:** Schutz vor missbräuchlicher Datennutzung und Stärkung der informationellen Selbstbestimmung.
4. **Benachrichtigung und Erklärung:** Transparenz darüber, wann ein KI-System eingesetzt wird und wie es zu seinen Ergebnissen kommt.
5. **Menschliche Alternativen und Eskalationspfade:** Die Möglichkeit, eine menschliche Überprüfung zu verlangen und Entscheidungen anzufechten.

Die Kombination aus der rechtlichen Verbindlichkeit des EU AI Acts und der ethischen Tiefe des AI Bill of Rights schafft ein starkes Fundament für den Aufbau vertrauenswürdiger und souveräner KI-Systeme.

Governance in der Praxis: Rollen, Verantwortlichkeiten und Aufsichtsgremien

Die Verankerung dieser Prinzipien im Verwaltungsalltag erfordert klare Strukturen und Verantwortlichkeiten. Ohne eine institutionalisierte Governance bleiben ethische Leitlinien wirkungslos. Die Einrichtung zentraler Aufsichtsgremien ist daher ein entscheidender erster Schritt. Nach dem Vorbild fortschrittlicher Behörden wie der US-amerikanischen General Services Administration (GSA) empfiehlt sich die Etablierung folgender Rollen und Gremien [23]:

- **Chief AI Officer (CAIO):** Eine zentrale, hochrangige Rolle, die für die KI-Strategie, die Einhaltung von Richtlinien und die Überwachung des gesamten KI-Portfolios der Behörde verantwortlich ist.
- **AI Governance Board:** Ein interdisziplinäres Gremium aus Führungskräften verschiedener Abteilungen (z.B. IT, Recht, Datenschutz, Fachbereiche), das strategische Entscheidungen trifft, Pilotprojekte genehmigt und die Einhaltung der Governance-Richtlinien überwacht.
- **AI Safety Team:** Ein operatives Team, das für die technische Risikobewertung, die Überwachung der Systemleistung und das Management von KI-Vorfällen zuständig ist.

Um die Zusammenarbeit dieser Gremien und weiterer Stakeholder zu strukturieren und Verantwortlichkeiten klar zuzuweisen, ist die Verwendung einer **RACI-Matrix** ein bewährtes Instrument.[25] Sie definiert für jede wesentliche Governance-Aufgabe, wer die Durchführung verantwortet (Responsible), wer die Gesamtverantwortung trägt (Accountable), wer konsultiert werden muss (Consulted) und wer informiert wird (Informed). Dies schafft

Transparenz und vermeidet Zuständigkeitslücken, die in komplexen Technologieprojekten häufig zu Problemen führen.

Tabelle 2: RACI-Matrix für KI-Governance

Aktivität / Aufgabe	Chief AI Officer (CAIO)	AI Governance Board	Fachabteilungs- leiter	Datenschutz-beauftragter (DPO)	IT-Sicherheit	Rechtsabteilung
Definition der KI-Strategie & Use Cases	A	R	C	I	I	C
Risikobewertung & Klassifizierung (gem. EU AI Act)	R	A	C	C	C	A
Sicherstellung der Datenqualität & Bias-Prüfung	R	A	R	C	I	C
Genehmigung von Pilotprojekten	A	R	C	I	I	I
Überwachung der Systemleistung & KPIs	R	A	R	I	C	I
Management von KI-Vorfällen	R	A	C	C	R	C
Einhaltung von Transparenz- & Erklärungspflichten	A	R	R	C	I	A
Schulung & Befähigung	R	A	R	I	I	I

der Mitarbeiter						

Legende: R = Responsible (führt aus), A = Accountable (verantwortlich), C = Consulted (wird konsultiert), I = Informed (wird informiert)

Der Blueprint in Aktion: Ein schrittweiser Pfad von Pilot zu Produktion

Die Umsetzung von Agentic AI in der Verwaltung ist keine einmalige technologische Einführung, sondern eine transformative Reise. Um diese Komplexität zu beherrschen, ist ein strukturiertes, phasenbasiertes Vorgehen unerlässlich. Der folgende Blueprint, angelehnt an das Value-Chain-Modell von PwC.[7] gliedert den Prozess in drei logische Phasen und bietet einen klaren, nachvollziehbaren Weg von der ersten Idee bis zur skalierten, wertschöpfenden Lösung.

Tabelle 3: Phasenmodell des Agentic-AI-Blueprints

Phase	Zielsetzung	Kernaktivitäten	Wichtigste Ergebnisse
I: Strategie & Entdeckung	Identifikation von wertstiftenden, machbaren und risikoarmen Anwendungsfällen.	Use-Case-Workshops, Prozessanalyse (Process Mining), Reifegradbewertung, Erstellung des Business Case, Stakeholder-Alignment.	Priorisierte Liste von Pilotkandidaten, Wirkungslogik-Modell, vorläufige Risikoanalyse, genehmigter Business Case.
II: Konzeption & Implementierung	Entwicklung und Test eines funktionsfähigen, sicheren und wertschöpfenden Piloten in einer kontrollierten Umgebung.	Detaillierte Prozesszerlegung (BPMN), Aufbau der Datenpipeline, Auswahl/Entwicklung des KI-Agenten, Implementierung von "Human-in-the-Loop"-Kontrollen, Pilottest.	Detailliertes Prozessmodell, trainierter KI-Agent, Testprotokolle, evaluierter Pilot mit Lessons Learned.
III: Skalierung & Optimierung	Überführung des erfolgreichen Piloten in den produktiven	Entwicklung des Skalierungsplans, Aufbau der	Skalierte Lösung im Produktivbetrieb, geschulte Mitarbeiter,

	Betrieb und Etablierung eines kontinuierlichen Verbesserungszyklus.	Produktionsinfrastruktur, Change Management & Mitarbeiterschulung, Implementierung von Monitoring & KPIs, kontinuierliche Optimierung.	etablierter Governance-Prozess, Dashboard zur Erfolgsmessung.

Phase I: Strategie & Entdeckung

In dieser grundlegenden Phase wird das Fundament für den Erfolg gelegt. Das Ziel ist es, nicht technologiegetrieben, sondern problemorientiert zu agieren und die Anwendungsfälle mit dem größten Hebel für die Organisation zu identifizieren.

Use-Case-Identifikation und Priorisierung

Der Ausgangspunkt für jedes erfolgreiche KI-Projekt sind klar definierte Geschäfts- oder Verwaltungsziele, nicht die Technologie selbst.[28] Anstatt zu fragen „Wo können wir KI einsetzen?“, lautet die richtige Frage: „Welche unserer Kernprozesse sind am ineffizientesten, fehleranfälligsten oder bürgerunfreundlichsten?“. Interaktive Workshops mit Fachexperten und Mitarbeitern, die Methoden wie Empathy und Journey Mapping nutzen, sind ideal, um Prozesse mit hohem Reibungsverlust, vielen repetitiven Aufgaben und einem gleichzeitig überschaubaren Risiko zu identifizieren – die perfekten Kandidaten für erste Pilotprojekte.[29] Die identifizierten Anwendungsfälle werden anschließend anhand einer Matrix bewertet, die das potenzielle Wertschöpfungspotenzial gegen die technische und organisatorische Machbarkeit abwägt.

Reifegradanalyse und Aufbau des Business Case

Parallel zur Use-Case-Identifikation muss eine ehrliche Bestandsaufnahme der organisatorischen Reife erfolgen. Diese Analyse bewertet die Bereitschaft der Behörde in kritischen Dimensionen wie der vorhandenen Dateninfrastruktur, den digitalen Kompetenzen der Belegschaft und den etablierten Governance-Prozessen.[3] Das Ergebnis dieser Analyse fließt in einen umfassenden Business Case ein. Dieser sollte über eine reine Kosten-Nutzen-Rechnung hinausgehen und auch qualitative Verbesserungen wie eine höhere

Dienstleistungsqualität, schnellere Entscheidungswege und eine gesteigerte Mitarbeiterzufriedenheit berücksichtigen.[7] Ein **Wirkungslogik-Modell (Logic Model)** ist ein hervorragendes Instrument, um die erwartete Wirkungskette von den investierten Ressourcen (Inputs) bis hin zu den langfristigen strategischen Zielen (Impact) transparent und nachvollziehbar darzustellen (siehe Abschnitt 5.1).[30]

Phase II: Konzeption & Implementierung

Nachdem ein vielversprechender Pilotkandidat ausgewählt wurde, beginnt die Phase der konkreten Umsetzung. Hierbei ist ein methodisches Vorgehen entscheidend, um die Komplexität zu beherrschen und ein robustes, sicheres System zu entwickeln.

Prozessanalyse und -zerlegung für KI-Agenten

Bevor ein Prozess automatisiert werden kann, muss er in seiner Tiefe verstanden werden. Die bloße Automatisierung eines ineffizienten oder fehlerhaften Prozesses führt nicht zu einer Verbesserung, sondern zementiert die bestehenden Probleme oder verlagert Engpässe lediglich an eine andere Stelle im System.[31] Aus diesem Grund ist eine datengestützte Prozessanalyse ein unverzichtbarer erster Schritt.
Die Methodik hierfür umfasst drei Schritte:

1. **Prozessentdeckung (Process Discovery):** Mithilfe von **Process Mining** wird aus den digitalen Spuren in den IT-Systemen (z.B. Log-Dateien) ein exaktes Abbild des tatsächlichen Ist-Prozesses erstellt. Diese Analyse deckt objektiv und datenbasiert die realen Prozessabläufe, Engpässe, Abweichungen und Ineffizienzen auf.[31]
2. **Prozessneugestaltung (Process Redesign):** Auf Basis der gewonnenen Erkenntnisse wird ein optimierter Soll-Prozess entworfen. **BPMN (Business Process Model and Notation)** ist der etablierte Standard zur Modellierung dieser Prozesse. BPMN ermöglicht nicht nur die klare Visualisierung des Workflows, sondern dient auch als Blaupause für die technische Umsetzung. Es definiert die genauen Grenzen für den KI-Agenten und legt fest, an welchen Stellen menschliche Interaktion (Human-in-the-Loop) zur Kontrolle oder Entscheidung zwingend erforderlich ist.[32]
3. **Aufgabenzerlegung (Task Decomposition):** Der in BPMN modellierte Prozess wird in logische, überschaubare Teilaufgaben zerlegt. Dieser modulare Ansatz, inspiriert von Frameworks wie TDAG (Task Decomposition and Agent Generation), ermöglicht es, für jede Teilaufgabe einen spezialisierten KI-(Sub-)Agenten zu entwickeln oder einzusetzen. Dies erhöht die Anpassungsfähigkeit des Gesamtsystems und vereinfacht die Entwicklung und Wartung erheblich.[34]

Aufbau der Daten- und Technologieinfrastruktur

Daten sind der „Lebensnerv" von Agentic AI.[16] Ohne eine qualitativ hochwertige, zugängliche und sichere Datengrundlage können KI-Agenten nicht effektiv arbeiten. In dieser Phase müssen daher die technischen Voraussetzungen geschaffen werden. Dies beinhaltet das Aufbrechen von Datensilos zwischen verschiedenen Abteilungen und Systemen, die Sicherstellung der Datenqualität und die Implementierung sicherer Datenzugriffsmechanismen.[6] Um die technologische Souveränität zu wahren, sollte die Architektur modular und auf offenen Standards basieren, um eine Abhängigkeit von einzelnen Herstellern zu vermeiden.[6]

Der Pilot: Kontrolliertes Testen und Lernen

Das ausgewählte Pilotprojekt wird in einer kontrollierten Umgebung implementiert, die vom produktiven Betrieb getrennt ist.[28] Das primäre Ziel des Piloten ist nicht die sofortige Perfektion, sondern das Lernen. Der Erfolg wird in dieser Phase weniger an der fehlerfreien Ausführung gemessen, sondern an der Geschwindigkeit und Tiefe der gewonnenen Erkenntnisse.[36] Es ist entscheidend, von Beginn an Mechanismen zu integrieren, die es den beteiligten Nutzern ermöglichen, unkompliziert Feedback zu geben und Fehler zu melden.[28]

Phase III: Skalierung & Optimierung

Ein erfolgreicher Pilot ist ein wichtiger Meilenstein, aber die eigentliche Herausforderung liegt in der Überführung in den produktiven Betrieb und der nachhaltigen Verankerung in der Organisation.

Von Pilot zu Produktion: Ein Skalierungsmodell

Die Skalierung ist der Punkt, an dem viele KI-Initiativen scheitern.[2] Ein erfolgreicher Übergang erfordert einen detaillierten Skalierungsplan, der nicht nur die technologische Infrastruktur für den Produktivbetrieb berücksichtigt, sondern auch die notwendigen Anpassungen in den Bereichen Governance, Personal und den angrenzenden Prozessen.

Kompetenzaufbau und Befähigung der Mitarbeitenden

Agentic AI wird menschliche Arbeitsrollen grundlegend verändern, nicht ersetzen.[4] Mitarbeiter werden von reinen Ausführenden zu Überwachern, Trainern und Managern von KI-Systemen. Dieser Wandel erfordert neue Kompetenzen. Ein strukturiertes Programm zum Aufbau von „AI Literacy" ist daher unerlässlich. Es zielt darauf ab, ein grundlegendes Verständnis für die Funktionsweise, die Chancen und die Risiken von KI zu schaffen, Vertrauen aufzubauen und die Mitarbeiter zu befähigen, effektiv mit ihren neuen digitalen Kollegen zusammenzuarbeiten.[15] Dies ist eine zentrale Aufgabe des Change Managements.[31]

Kontinuierliche Überwachung und Verbesserung

Mit der Inbetriebnahme der skalierten Lösung beginnt ein kontinuierlicher Zyklus der Überwachung und Optimierung. Regelmäßige Audits und eine fortlaufende Überwachung der definierten KPIs stellen sicher, dass das System die erwartete Leistung erbringt, die Nutzer zufrieden sind und die ethischen Prinzipien eingehalten werden.[28] Die in Phase II etablierte Process-Intelligence-Fähigkeit wird nun genutzt, um den Live-Prozess zu überwachen, neue Optimierungspotenziale zu identifizieren und einen datengestützten, kontinuierlichen Verbesserungsprozess zu etablieren.[31]

Risikomanagement für Agentic AI

Die Autonomie und Komplexität von Agentic AI bringen neue Risiken mit sich, die ein proaktives und systematisches Management erfordern. Ein robustes Risikomanagement ist keine Hürde für Innovation, sondern deren Voraussetzung, da es Vertrauen schafft und den verantwortungsvollen Einsatz der Technologie sicherstellt.

Identifikation und Bewertung von KI-spezifischen Risiken

Ein strukturierter Ansatz zur Risikoidentifikation ist der erste Schritt. Anstatt Risiken ad hoc zu betrachten, sollten etablierte Taxonomien, wie sie beispielsweise von NIST oder MITRE entwickelt wurden, genutzt werden. Diese kategorisieren Risiken systematisch, zum Beispiel in Bereiche wie **Diskriminierung & Toxizität**, **Datenschutz & Sicherheit** oder **Fehlinformationen**.[39] Es ist wichtig zu erkennen, dass Risiken auf verschiedenen Ebenen entstehen können: im KI-Modell selbst (z.B. durch verzerrte Trainingsdaten), im Gesamtsystem (z.B. durch fehlerhafte Integration) oder im spezifischen Anwendungsfall (z.B. durch unsachgemäße Nutzung).[40]
Zur Bewertung der identifizierten Risiken eignet sich eine **Risikomatrix**, wie sie beispielsweise von der australischen Regierung verwendet wird.[41] Diese bewertet jedes Risiko anhand seiner

Eintrittswahrscheinlichkeit (von „selten“ bis „fast sicher“) und seiner potenziellen **Auswirkung** (von „unbedeutend“ bis „katastrophal“). Das Ergebnis ist eine Klassifizierung des Gesamtrisikos als niedrig, mittel oder hoch. Diese Klassifizierung ist nicht nur für die Priorisierung von Gegenmaßnahmen entscheidend, sondern auch für die Einhaltung der risikobasierten Vorgaben des EU AI Acts.[19] Eine Vorlage für ein solches Risikoregister findet sich in Anhang B.

Strategien zur Risikominderung: Der Mensch im Mittelpunkt

Die effektivste Strategie zur Minderung von Risiken, insbesondere bei hochkritischen Entscheidungen im Verwaltungshandeln, ist die intelligente Einbindung menschlicher Expertise. Der **Mensch-in-der-Schleife (Human-in-the-Loop, HITL)** ist kein bloßer Notnagel, sondern ein zentrales Designprinzip für vertrauenswürdige KI-Systeme.[12] In der Praxis lassen sich verschiedene HITL-Muster implementieren:

- **Verifikation:** Der KI-Agent erstellt einen Vorschlag (z.B. die Berechnung einer Sozialleistung), der jedoch erst durch die Überprüfung und Freigabe eines menschlichen Sachbearbeiters gültig wird. Der Mensch behält die finale Entscheidungshoheit.[42]
- **Eskalation:** Der Agent bearbeitet Standardfälle vollautonom, erkennt aber komplexe Grenzfälle, Ausnahmen oder unklare Sachverhalte und leitet diese automatisch an einen menschlichen Experten zur Bearbeitung weiter.[11]
- **Aufsicht:** Ein menschlicher Experte überwacht die Aktivitäten des Agenten in Echtzeit oder anhand von Dashboards und kann bei Fehlverhalten oder unerwarteten Ergebnissen jederzeit eingreifen und den Prozess korrigieren oder stoppen.[4]

Neben diesen prozessualen Absicherungen sind technische Schutzmaßnahmen unerlässlich. Dazu gehören robuste Verfahren zum Schutz personenbezogener Daten (z.B. durch Privacy-Enhancing Technologies), umfassende Cybersicherheits-Protokolle zum Schutz vor Angriffen sowie die Gewährleistung, dass die Entscheidungen des Systems nachvollziehbar (Traceability) und erklärbar (Explainability) sind.[11]

Erfolgsmessung und Wertbeitrag

Um den Nutzen von Agentic-AI-Projekten nachzuweisen und zukünftige Investitionen zu rechtfertigen, ist eine systematische Erfolgsmessung unerlässlich. Diese muss über einfache technische Metriken hinausgehen und den tatsächlichen Wertbeitrag für die Verwaltung, ihre Mitarbeiter und die Bürger erfassen.

Definition von KPIs und Wirkungslogik

Ein effektives Instrument zur Strukturierung der Erfolgsmessung ist das **Wirkungslogik-Modell (Logic Model)**. Es zwingt die Projektverantwortlichen von Anfang an, die gesamte Wirkungskette eines Vorhabens klar zu definieren – von den eingesetzten Ressourcen bis zu den angestrebten langfristigen Veränderungen.[30] Ein solches Modell besteht aus fünf Stufen:

1. **Inputs:** Die investierten Ressourcen (z.B. Budget, Personalstunden, Daten, technologische Infrastruktur).
2. **Aktivitäten:** Die durchgeführten Maßnahmen (z.B. Prozess-Reengineering, Training des KI-Agenten, Durchführung des Piloten).
3. **Outputs:** Die direkten, zählbaren Ergebnisse der Aktivitäten (z.B. Anzahl der automatisch bearbeiteten Anträge, Reduktion der manuellen Dateneingaben um X %).
4. **Outcomes (Ergebnisse):** Die mittelfristigen Veränderungen in Leistung und Verhalten, die durch die Outputs erzielt werden (z.B. Reduzierung der durchschnittlichen Bearbeitungszeit um Y Tage, Senkung der Fehlerquote um Z %, Freisetzung von Mitarbeiterkapazitäten für Beratungsaufgaben).
5. **Impact (Wirkung):** Die langfristigen, strategischen Ziele, zu denen das Projekt beiträgt (z.B. Steigerung der Bürgerzufriedenheit, Erhöhung des Vertrauens in die Verwaltung, bessere politische Ergebnisse).

Auf Basis dieses Modells können für jede Stufe spezifische, messbare, erreichbare, relevante und terminierte (SMART) **Key Performance Indicators (KPIs)** abgeleitet werden.[44] Dies überführt vage Zielsetzungen in ein konkretes Messsystem. Beispiele für relevante KPIs sind: Prozessdurchlaufzeit, Fehlerratenreduktion, Kosten pro Transaktion, auf höherwertige Aufgaben umverteilte Mitarbeiterzeit und der Citizen Satisfaction Score (CSAT).

Der Return on Investment (ROI) von Agentic AI in der Verwaltung

Die Frage nach dem Return on Investment (ROI) ist für Entscheider von zentraler Bedeutung. Allerdings zeigt die Erfahrung mit disruptiven Technologien, dass die Erwartung eines schnellen, positiven finanziellen ROI innerhalb der ersten ein bis zwei Jahre oft unrealistisch ist und zu Enttäuschung und dem vorzeitigen Abbruch vielversprechender Projekte führen kann.[2] Ein realistischerer Ansatz berücksichtigt die unterschiedlichen Ziele der Projektphasen. Die Messung des ROI sollte daher ebenfalls phasenweise erfolgen. In der **Pilotphase (Phase II)** liegt der primäre „Return“ nicht in direkten Kosteneinsparungen, sondern im **Lernen und in der Risikominimierung**. Der Wert entsteht durch die gewonnenen Erkenntnisse über die technische Machbarkeit, die Akzeptanz bei den Mitarbeitern und die tatsächlichen Auswirkungen auf den Prozess. Der ROI ist hier ein „Return on Learning“.
Erst in der **Skalierungsphase (Phase III)**, wenn die Lösung in den produktiven Betrieb

übergeht, rücken klassische finanzielle ROI-Metriken in den Vordergrund. Hier können dann harte Kennzahlen wie Kosteneinsparungen durch Automatisierung und Produktivitätssteigerungen gemessen und bewertet werden.[44] Dieser phasenweise Ansatz steuert die Erwartungen von Führungskräften und sichert die notwendige Geduld für eine nachhaltige Transformation. Der wahre, transformative Wert von Agentic AI, wie beispielsweise eine Reduktion der Bearbeitungszeiten um 60-90%, entsteht nicht durch die bloße Optimierung bestehender Abläufe, sondern durch deren grundlegende Neugestaltung rund um die Fähigkeiten der KI – und dieser Prozess benötigt Zeit.[36]

Aus der Praxis lernen: Internationale Fallbeispiele

Die theoretischen Konzepte und Modelle eines Blueprints gewinnen an Überzeugungskraft, wenn sie durch reale Erfolgsgeschichten und praktische Erfahrungen untermauert werden. Der Blick auf internationale Vorreiter und erste Anwendungen im eigenen Land liefert wertvolle Lektionen für die eigene Strategie.

Vorreiter Estland: Proaktive und „unsichtbare“ Dienstleistungen

Estland gilt weltweit als Pionier der digitalen Verwaltung und zeigt bereits heute, wie die Zukunft von Bürgerdiensten aussehen kann. Das Land verfolgt das Konzept der **proaktiven und „unsichtbaren“ Dienstleistungen**. Anstatt dass Bürger Anträge stellen müssen, antizipiert der Staat ihre Bedürfnisse auf Basis von Lebensereignissen (z.B. die Geburt eines Kindes) und leitet die entsprechenden Verwaltungsleistungen (z.B. die Auszahlung von Elterngeld) automatisch ein.[47]
Der Erfolg dieses Modells basiert nicht allein auf fortschrittlicher Technologie, sondern auf einem soliden Fundament aus strategischen Enablern:

- **Digitale Identität und Interoperabilität:** Eine sichere, landesweite digitale Identität für jeden Bürger und die Datenaustauschplattform „X-Road“ sind die technologische Basis. Sie ermöglichen einen sicheren und nahtlosen Datenfluss zwischen verschiedenen Behörden nach dem **„Once-Only“-Prinzip**, bei dem Daten, die dem Staat einmal vorliegen, nicht erneut abgefragt werden müssen.[47]
- **Starke, zentrale Governance:** Das Büro des Government Chief Information Officer (GCIO) steuert die digitale Transformation zentral und sorgt für eine kohärente, ressortübergreifende Umsetzung.[47]
- **Das #Bürokratt-Programm:** Diese nationale Initiative treibt gezielt die Integration von KI und virtuellen Assistenten voran, um das Netzwerk der Behördendienste für den Bürger als einen einzigen, nahtlosen Service erlebbar zu machen.[47]

Die Lektion aus Estland für Deutschland ist tiefgreifend: Das ultimative Ziel von Agentic AI in

der Verwaltung ist nicht nur, bestehende Prozesse schneller zu machen, sondern sie aus der Perspektive der Bürger idealerweise ganz verschwinden zu lassen.

Konkrete Anwendungen in Deutschland: Status Quo und Potenziale

Auch in Deutschland gibt es bereits vielversprechende Ansätze, die zeigen, dass Agentic AI keine ferne Zukunftsvision mehr ist. Verschiedene Behörden auf Bundes- und kommunaler Ebene haben bereits erfolgreiche Pilotprojekte gestartet oder sogar schon produktive Systeme im Einsatz:

- **Service-Desk-Automatisierung:** Die Bundesagentur für Arbeit nutzt eine KI-Plattform, um einen Großteil der internen IT- und HR-Serviceanfragen automatisiert zu bearbeiten.[50]
- **Zertifikatsprüfung:** Die Familienkasse der Bundesagentur für Arbeit verifiziert jährlich über 150.000 Studienbescheinigungen mit einem KI-System, was die Bearbeitungsgeschwindigkeit erheblich steigert.[50]
- **Zoll und Krisenfrüherkennung:** Die Generalzolldirektion pilotiert KI-Systeme zur Erkennung von Anomalien bei Importen, während das Auswärtige Amt mit dem Tool PRE-VIEW globale Medien analysiert, um internationale Krisen frühzeitig zu erkennen.[50]

Diese Beispiele zeigen einen pragmatischen, problemorientierten Ansatz. Die Herausforderung für Deutschland besteht nun darin, von diesen erfolgreichen, aber oft isolierten Insellösungen zu einer breiten, systemischen Adaption zu gelangen.[50] Die föderale Struktur und die strengen Datenschutzanforderungen erfordern einen besonders sorgfältigen und souveränen Ansatz. Gleichzeitig liegt hierin eine immense Chance, dem drohenden Fachkräftemangel in der öffentlichen Verwaltung wirksam zu begegnen.[1]
Die Lektion aus den deutschen Beispielen ist, dass der pragmatische Start erfolgreich war. Der nächste logische Schritt ist der Aufbau gemeinsamer Plattformen und Kompetenzzentren, wie sie mit Initiativen wie PLAIN (Platform Analysis and Information System) bereits angedacht sind, um die Adaption auf allen Verwaltungsebenen zu beschleunigen und Synergien zu heben.[22]

Handlungsempfehlungen für Entscheider

Die Transformation hin zu einer agentischen Verwaltung erfordert entschlossenes Handeln auf Führungsebene. Die folgenden Empfehlungen sind in drei Zeithorizonte gegliedert und bieten eine konkrete Roadmap für die nächsten Schritte.

Sofortmaßnahmen: Die ersten 90 Tage

1. **Governance etablieren:** Richten Sie ein zentrales **AI Governance Board** ein und benennen Sie einen **Chief AI Officer** (oder eine äquivalente, weisungsbefugte Rolle). Dies schafft von Anfang an klare Verantwortlichkeiten und eine zentrale Anlaufstelle.
2. **Führungskräfte befähigen:** Starten Sie eine „AI Literacy"-Initiative gezielt für die Führungsebene. Ein gemeinsames Verständnis der Technologie, ihrer Potenziale und Risiken ist die Voraussetzung für strategische Entscheidungen und ein realistisches Erwartungsmanagement.
3. **Datengrundlage schaffen:** Initiieren Sie ein erstes **Process-Intelligence-Projekt** für einen Prozess, der als hoch relevant, aber auch als ineffizient bekannt ist. Die datengestützte Analyse des Ist-Zustands liefert eine objektive Grundlage für die Auswahl eines sinnvollen Pilotprojekts und quantifiziert das Verbesserungspotenzial.

Mittelfristige strategische Weichenstellungen

1. **Strategie formalisieren:** Entwickeln Sie eine formale KI-Strategie, die fest mit dem Auftrag und den Zielen Ihrer Behörde verknüpft ist. Priorisieren Sie auf dieser Basis ein Portfolio von zwei bis drei konkreten Pilotprojekten, die in den nächsten 12-18 Monaten umgesetzt werden sollen.
2. **Infrastruktur modernisieren:** Investieren Sie gezielt in den Aufbau einer modernen, sicheren und skalierbaren Dateninfrastruktur. Beginnen Sie damit, die kritischsten Datensilos aufzubrechen, um die für die Piloten benötigten Daten zugänglich zu machen.
3. **Risikomanagement verankern:** Formalisieren Sie das im Blueprint beschriebene KI-Risikomanagement-Framework und integrieren Sie es als festen Bestandteil in den Projektlebenszyklus. Jedes KI-Vorhaben muss von Beginn an einer strukturierten Risikobewertung unterzogen werden.

Langfristige Vision: Die agentische Behörde der Zukunft

1. **Kultur der Innovation fördern:** Schaffen Sie eine Organisationskultur, die sicheres Experimentieren und kontinuierliches Lernen fördert. Etablieren Sie Mechanismen, um die Erkenntnisse aus Pilotprojekten – sowohl Erfolge als auch Misserfolge – systematisch zu erfassen und organisationsweit zu teilen.
2. **Organisation neu denken:** Planen Sie die notwendige organisatorische Neugestaltung, die durch KI-native Arbeitsabläufe erforderlich wird. Dies bedeutet eine Abkehr von starren, hierarchischen Silos hin zu flacheren, interdisziplinären und ergebnisorientierten „agentischen Teams".[52]
3. **Proaktive Dienste als Zielbild:** Entwickeln Sie eine langfristige Roadmap, die auf die

Vision proaktiver, „unsichtbarer“ Bürgerdienste hinarbeitet, wie sie im estnischen Modell verwirklicht wird. Dies setzt den strategischen Nordstern für die kontinuierliche Weiterentwicklung der Verwaltung.

Anhang

A. Detaillierte Checkliste für KI-Projektvorschläge

Diese zweiteilige Checkliste, basierend auf dem Modell der australischen Regierung [29], dient als praktisches Werkzeug für Mitarbeiter, die KI-Projekte vorschlagen, und für Führungskräfte, die diese bewerten und genehmigen.

Teil 1: Checkliste für den Projektvorschlag (für Mitarbeiter)

1. **Problemdefinition**
   - [ ] Kann ich das Problem, das ich lösen möchte, klar beschreiben (z.B. Engpässe bei der Aktenbearbeitung, lange Wartezeiten, verstreute Daten)?
   - [ ] Handelt es sich um einen Prozess mit hohem manuellem Aufwand und vielen Wiederholungen, aber einem relativ geringen Risiko?
2. **Eignung für KI**
   - [ ] Gibt es Anhaltspunkte dafür, dass KI dieses Problem sinnvoll adressieren kann (z.B. durch Zusammenfassung, Triage, Navigation, Automatisierung)?
   - [ ] Bietet KI einen klaren Vorteil gegenüber bestehenden regelbasierten Systemen oder rein menschlicher Bearbeitung?
3. **Datenverfügbarkeit**
   - [ ] Weiß ich, welche Art von Daten das Projekt benötigen würde und ob diese in nutzbarer, digitaler Form vorliegen?
   - [ ] Können sensible oder personenbezogene Daten vermieden oder zuverlässig anonymisiert werden?
4. **Governance und Sicherheit**
   - [ ] Habe ich über Risiken (z.B. Voreingenommenheit, Genauigkeit, Datenschutz, Fairness) nachgedacht und wie diese gemanagt werden können?
   - [ ] Kann eine menschliche Überprüfung (Human-in-the-Loop) an kritischen Entscheidungspunkten sichergestellt werden?
5. **Wirkung und Wertbeitrag**
   - [ ] Kann ich definieren, wie der Erfolg gemessen wird (z.B. eingesparte Zeit, reduzierte Fehlerquote, Mitarbeiterzufriedenheit, Reaktionsfähigkeit des Dienstes)?
   - [ ] Wird dieses Pilotprojekt Kompetenzen und Vertrauen für zukünftige,

komplexere Projekte aufbauen?

6. **Transparenz**
    - [ ] Bin ich bereit, die Funktionsweise der KI sowohl Kollegen als auch der Öffentlichkeit in einfacher Sprache zu erklären?
    - [ ] Werde ich die gewonnenen Erkenntnisse am Ende des Piloten dokumentieren und teilen?

Teil 2: Checkliste für die Governance-Prüfung (für Führungskräfte)

1. **Übereinstimmung mit dem Problem**
    - [ ] Löst der Vorschlag ein reales, klar definiertes Problem für Bürger oder Mitarbeiter?
    - [ ] Ist das Problem für den Einsatz von KI geeignet, oder gibt es einfachere, nicht-KI-basierte Alternativen?
2. **Risiko und Umfang**
    - [ ] Wurde das Projekt als niedrig, mittel oder hoch riskant eingestuft? Sind die Governance-Anforderungen verhältnismäßig (schlank bei niedrigem Risiko, robust bei hohem Risiko)?
    - [ ] Wurde eine Risikobewertung anhand anerkannter ethischer Prinzipien (z.B. Fairness, Datenschutz, Transparenz, Rechenschaftspflicht) durchgeführt?
3. **Ressourcen und Unterstützung**
    - [ ] Habe ich ausreichend Zeit, Budget und qualifiziertes Personal für den Erfolg des Teams bereitgestellt?
    - [ ] Schaffe ich ein Umfeld der psychologischen Sicherheit, in dem klar ist, dass sicheres Experimentieren gefördert und nicht bestraft wird?
4. **Praktische Governance**
    - [ ] Ist eine menschliche Aufsicht (Human-in-the-Loop) dort vorgesehen, wo sie erforderlich ist?
    - [ ] Sind Prüfungen zu Datenschutz, Ethik und Risiko von Anfang an in den Prozess eingebettet (nicht erst am Ende)?
    - [ ] Gibt es einen klaren Verantwortlichen für das Projekt (einen „AI Steward“ oder Product Owner)?
5. **Ergebnisse und Skalierung**
    - [ ] Gibt es eine klare Definition des Erfolgs, die über „das Modell funktioniert“ hinausgeht?
    - [ ] Werde ich sicherstellen, dass die gewonnenen Erkenntnisse erfasst, geteilt und für zukünftige Skalierungsentscheidungen genutzt werden?
6. **Führungsverhalten**
    - [ ] Lebe ich Neugier vor und bin bereit, gemeinsam mit meinem Team zu lernen?
    - [ ] Agiere ich als Förderer und Sponsor, der die Legitimität für das Experiment erteilt, anstatt als Torwächter?
    - [ ] Kommuniziere ich das Vorhaben als ein sicheres, strategisches Experiment und nicht als ein riskantes Wagnis?

B. Vorlage für ein KI-Risikoregister

Dieses Template dient der systematischen Erfassung, Bewertung und Verfolgung von KI-spezifischen Risiken über den gesamten Projektlebenszyklus. Es sollte ein lebendes Dokument sein, das regelmäßig vom AI Safety Team und dem Projektverantwortlichen aktualisiert wird. (Basierend auf den Frameworks aus [39]).

Risk ID	Risikokategorie	Beschreibung des Risikos	Potenzielle Auswirkung	Wahrscheinlichkeit	Schweregrad der Auswirkung	Gesamtrisiko	Gegenmaßnahme / Mitigationsstrategie	Verantwortlich	Status
B-001	*Bias / Fairness*	*Das KI-Modell könnte aufgrund von unausgewogenen Trainingsdaten bestimmte Bevölkerungsgruppen bei der Leistungsvergabe systematisch benachteiligen.*	*Rechtliche Konsequenzen (AGG), Reputationsschaden, Vertrauensverlust in der Öffentlichkeit, Ungleichbehandlung von Bürgern.*	*Möglich*	*Schwerwiegend*	*Hoch*	*1. Diversität der Trainingsdaten sicherstellen. 2. Regelmäßige Bias-Audits durchführen. 3. Transparenz über die Entscheidungskriterien schaffen. 4. Eskalationspfad für menschliche*	*Datenschutzbeauftragter, Fachabteilung*	*In Umsetzung*

							Überprüfung bei strittigen Fällen.		
P-001	Datenschutz	Der KI-Agent greift auf mehr personenbezogene Daten zu als für die Aufgabenerfüllung zwingend erforderlich (Datenminimierungsprinzip verletzt).	Verstoß gegen DSGVO, Bußgelder, Verlust sensibler Bürgerdaten, Vertrauensverlust.	Unwahrscheinlich	Schwerwiegend	Mittel	1. Implementierung von Privacy-Enhancing Technologies (PET). 2. Striktes Rollen- und Rechtekonzept für den Datenzugriff des Agenten. 3. Durchführung einer Datenschutz-Folgenabschätzung (DSFA).	Datenschutzbeauftragter, IT-Sicherheit	Geplant
S-001	Sicherheit	Gezielte Angriffe (Adversarial Attacks) auf den KI-	Finanzieller Schaden, Missbrauch von Leistung	Möglich	Mittel	Mittel	1. Einsatz von robusten KI-Modellen. 2. Kontinui	IT-Sicherheit, AI Safety Team	In Umsetzung

		Agenten könnten zu manipulierten, falschen Ergebnissen führen (z.B. ungerechtfertigte Genehmigungen).	*en, Sicherheitsrisiken (je nach Anwendungsfall), Reputationsschaden.*				*erliches Monitoring der System-Inputs und -Outputs auf Anomalien. 3. Regelmäßige Penetrationstests der KI-Anwendung.*		
A-001	*Genauigkeit / Zuverlässigkeit*	*Der KI-Agent "halluziniert" und erfindet Fakten oder gibt veraltete Informationen aus, die zu falschen Verwaltungsentscheidungen führen.*	*Fehlerhafte Bescheide, rechtliche Anfechtbarkeit, erhöhter Korrekturaufwand für Mitarbeiter, Verunsicherung der Bürger.*	*Wahrscheinlich*	*Mittel*	*Hoch*	*1. Anbindung des Agenten an verifizierte, aktuelle Wissensdatenbanken (Retrieval-Augmented Generation). 2. Implementierung eines Verifikationsschritts durch*	*Fachabteilung, Projektleitung*	*Offen*

							Menschen bei kritischen Entscheidungen.		
T-001	*Transparenz*	*Die Entscheidungswege des KI-Agenten sind für Sachbearbeiter und Bürger nicht nachvollziehbar ("Black Box"-Problem).*	*Geringe Akzeptanz bei Mitarbeitern, Unfähigkeit zur Fehleranalyse, rechtliche Probleme bei Anfechtung von Entscheidungen.*	*Fast sicher*	*Mittel*	*Hoch*	*1. Einsatz von erklärbaren KI-Methoden (XAI). 2. Automatische Protokollierung der Entscheidungsschritte des Agenten in verständlicher Form. 3. Schulung der Mitarbeiter zur Interpretation der KI-Ergebnisse.*	*Projektleitung, AI Safety Team*	*Geplant*

Referenzen

1. Wie KI-Agenten der Verwaltung helfen können - Entwicklungsagentur Rheinland-

Pfalz, Zugriff am Oktober 26, 2025, https://ea-rlp.de/wie-ki-agenten-der-verwaltung-helfen-koennen/
2. Mit GenAI zukunftsfit – 6 Beispiele für die öffentliche Verwaltung - MIDRANGE, Zugriff am Oktober 26, 2025, https://midrange.de/mit-genai-zukunftsfit-6-beispiele-fuer-die-oeffentliche-verwaltung/
3. Agentic AI: Preparing for the Next Leap in Government Innovation, Zugriff am Oktober 26, 2025, https://papers.govtech.com/Agentic-AI%3A-Preparing-for-the-Next-Leap-in-Government-Innovation-143847.html
4. How Agentic AI Will Transform Local Government: A Forewarning | StateTech Magazine, Zugriff am Oktober 26, 2025, https://statetechmagazine.com/article/2025/08/how-agentic-ai-will-transform-local-government-forewarning
5. AI Insights: Agentic AI (HTML) - GOV.UK, Zugriff am Oktober 26, 2025, https://www.gov.uk/government/publications/ai-insights/ai-insights-agentic-ai-html
6. Accelerating Mission Outcomes with Agentic AI: A Practical Guide for Federal Leaders, Zugriff am Oktober 26, 2025, https://www.businessofgovernment.org/blog/accelerating-mission-outcomes-agentic-ai-practical-guide-federal-leaders
7. Unlocking Tomorrow: agentic AI for the Public Sector | PwC, Zugriff am Oktober 26, 2025, https://www.pwc.com/gx/en/services/alliances/microsoft/agentic-ai-for-public-sector.html
8. KI-Agenten in der Regierung: Die Zukunft des öffentlichen Dienstes gestalten - Beam AI, Zugriff am Oktober 26, 2025, https://beam.ai/de/articles/ai-agents-in-government-pioneering-the-future-of-public-service
9. What is AI Agent Planning? | IBM, Zugriff am Oktober 26, 2025, https://www.ibm.com/think/topics/ai-agent-planning
10. The Agentic AI Handbook: A Beginner's Guide to Autonomous Intelligent Agents, Zugriff am Oktober 26, 2025, https://www.freecodecamp.org/news/the-agentic-ai-handbook/
11. Agentic AI for Citizen-Centered Services | Abt Global, Zugriff am Oktober 26, 2025, https://www.abtglobal.com/insights/perspectives/agentic-ai-for-citizen-centered-services
12. Agentic AI in government investigations - Thomson Reuters Legal Solutions, Zugriff am Oktober 26, 2025, https://legal.thomsonreuters.com/blog/agentic-ai-in-government-investigations/
13. KI in der öffentlichen Verwaltung: Anwendungen & Trends - Zendesk, Zugriff am Oktober 26, 2025, https://www.zendesk.de/blog/ai-in-government/
14. Beyond chatbots: How agentic AI can transform government services - Route Fifty, Zugriff am Oktober 26, 2025, https://www.route-fifty.com/artificial-intelligence/2025/08/beyond-chatbots-how-agentic-ai-can-transform-government-services/407537/
15. Potenziale Künstlicher Intelligenz in der öffentlichen Verwaltung - Lobbyregister,

Zugriff am Oktober 26, 2025, https://www.lobbyregister.bundestag.de/media/45/f0/312966/Stellungnahme-Gutachten-SG2406220006.pdf
16. Agentic AI Can Transform Civilian Government - Booz Allen, Zugriff am Oktober 26, 2025, https://www.boozallen.com/insights/data-optimization/harnessing-agentic-ai-for-civilian-government.html
17. legal.thomsonreuters.com, Zugriff am Oktober 26, 2025, https://legal.thomsonreuters.com/blog/agentic-ai-in-government-investigations/#:~:text=Agentic%20AI%20refers%20to%20artificial,as%20risk%20and%20fraud%20investigations.
18. EU Artificial Intelligence Act | Up-to-date developments and analyses of the EU AI Act, Zugriff am Oktober 26, 2025, https://artificialintelligenceact.eu/
19. AI Act | Shaping Europe's digital future, Zugriff am Oktober 26, 2025, https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
20. Blueprint for an AI Bill of Rights | OSTP | The White House, Zugriff am Oktober 26, 2025, https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/
21. Governing with Artificial Intelligence - OECD, Zugriff am Oktober 26, 2025, https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en.html
22. Künstliche Intelligenz in der Verwaltung - Bundesdruckerei, Zugriff am Oktober 26, 2025, https://www.bundesdruckerei.de/de/innovation-hub/kuenstliche-intelligenz-verwaltung
23. Artificial intelligence guidance and resources | GSA, Zugriff am Oktober 26, 2025, https://www.gsa.gov/technology/government-it-initiatives/artificial-intelligence/ai-guidance-and-resources
24. Artificial intelligence compliance plan | GSA, Zugriff am Oktober 26, 2025, https://www.gsa.gov/technology/government-it-initiatives/artificial-intelligence/ai-guidance-and-resources/ai-compliance-plan
25. RACI Matrix For Artificial Intelligence - Meegle, Zugriff am Oktober 26, 2025, https://www.meegle.com/en_us/topics/raci-matrix/raci-matrix-for-artificial-intelligence
26. RACI Matrix: Defining Accountability in AI Governance - Yields.io, Zugriff am Oktober 26, 2025, https://www.yields.io/blog/raci-matrix/
27. AI Governance RACI Matrix Explained, AI Consultants UK, Zugriff am Oktober 26, 2025, https://www.efficiencyai.co.uk/knowledge_card/ai-governance-raci-matrix/
28. Artificial intelligence in the public sector: A framework to support ethical and effective AI adoption in government services - Merative, Zugriff am Oktober 26, 2025, https://www.merative.com/blog/ai-in-the-public-sector-a-framework
29. AI governance checklists for government | Safe and practical adoption, Zugriff am Oktober 26, 2025, https://www.liquidinteractive.com.au/journal/ai-government-checklists
30. Logic Model: Transforming Program Theory into Continuous ..., Zugriff am

Oktober 26, 2025, https://www.sopact.com/use-case/logic-model
31. Process intelligence and agentic AI – PEX Network, Zugriff am Oktober 26, 2025, https://www.processexcellencenetwork.com/ai/articles/process-intelligence-agentic-ai
32. Essential Agentic Patterns for AI Agents in BPMN | Camunda, Zugriff am Oktober 26, 2025, https://camunda.com/blog/2025/03/essential-agentic-patterns-ai-agents-bpmn/
33. How BPMN Helps Solve AI's Transparency Problem | by Kevin Burnett | SpiffWorkflow, Zugriff am Oktober 26, 2025, https://medium.com/spiffworkflow/how-bpmn-helps-solve-ais-transparency-problem-433c88e386a8
34. A Method for Converting Business Process Models into Orchestrated ..., Zugriff am Oktober 26, 2025, https://www.tdcommons.org/cgi/viewcontent.cgi?article=9831&context=dpubs_series
35. arXiv:2402.10178v2 [cs.CL] 21 Jan 2025, Zugriff am Oktober 26, 2025, https://arxiv.org/pdf/2402.10178
36. McKinsey is Wrong That 80% Companies Fail to Generate AI ROI - Braden Kelley, Zugriff am Oktober 26, 2025, https://bradenkelley.com/2025/09/mckinsey-is-wrong-that-80-companies-fail-to-generate-ai-roi/
37. Full Report: Governing with Artificial Intelligence - OECD, Zugriff am Oktober 26, 2025, https://www.oecd.org/en/publications/governing-with-artificial-intelligence_795de142-en/full-report.html
38. Kompetenzen für den Einsatz generativer Künstlicher Intelligenz in der Verwaltung - Kompetenzzentrum Öffentliche IT, Zugriff am Oktober 26, 2025, https://www.oeffentliche-it.de/publikationen/kompetenzen-fuer-den-einsatz-generativer-kuenstlicher-intelligenz/Kompetenzen%20f%C3%BCr%20den%20Einsatz%20generativer%20K%C3%BCnstlicher%20Intelligenz%20in%20der%20Verwaltung.pdf
39. The MIT AI Risk Repository, Zugriff am Oktober 26, 2025, https://airisk.mit.edu/
40. Implementing risk assessments for high-risk AI systems - WaTech, Zugriff am Oktober 26, 2025, https://watech.wa.gov/sites/default/files/2025-01/EO%2024-01%20Risk%20Guidance_Final.pdf
41. Risk assessment for use of AI | digital.gov.au, Zugriff am Oktober 26, 2025, https://www.digital.gov.au/policy/ai/risk-assessment
42. Agentic AI: More than a Buzzword for the Public Sector - Government Technology Insider, Zugriff am Oktober 26, 2025, https://governmenttechnologyinsider.com/agentic-ai-more-than-a-buzzword-for-the-public-sector/
43. Logic model | Research Starters - EBSCO, Zugriff am Oktober 26, 2025, https://www.ebsco.com/research-starters/religion-and-philosophy/logic-model
44. ROI of AI: Key Drivers, KPIs & Challenges | DataCamp, Zugriff am Oktober 26, 2025, https://www.datacamp.com/blog/roi-of-ai

45. AI Project Management Template - Free to Use - Stackby, Zugriff am Oktober 26, 2025, https://stackby.com/templates/ai-project-management
46. Unlocking value from technology and AI for institutional investors - McKinsey, Zugriff am Oktober 26, 2025, https://www.mckinsey.com/industries/private-capital/our-insights/unlocking-value-from-technology-and-ai-for-institutional-investors
47. Case Study: Estonia's Digital Transformation Journey - GovCX Journal, Zugriff am Oktober 26, 2025, https://journal.govcx.org/case-study-estonias-digital-transformation-journey/
48. Finally 100% Digital: Estonia's 30-Year Journey from the USSR to e-Estonia - ComplexDiscovery, Zugriff am Oktober 26, 2025, https://complexdiscovery.com/finally-100-digital-estonias-30-year-journey-from-the-ussr-to-e-estonia/
49. Building AI-powered government: The Estonian experience - Reform Support, Zugriff am Oktober 26, 2025, https://reform-support.ec.europa.eu/document/download/27b74e18-8a00-4cb5-84ca-9ede00a63e63_en?filename=Day%202%20-%20Building%20AI-powered%20government%20-%20The%20Estonian%20experience.pdf&prefLang=ga
50. Beyond the Buzz: Real AI Use Cases in Germany's Public ... - Medium, Zugriff am Oktober 26, 2025, https://medium.com/@ecesen/beyond-the-buzz-real-ai-use-cases-in-germanys-public-administration-ae7ceb5376da
51. (PDF) Artificial Intelligence in Governance: A Comprehensive Analysis of AI Integration and Policy Development in the German Government - ResearchGate, Zugriff am Oktober 26, 2025, https://www.researchgate.net/publication/375953322_Artificial_Intelligence_in_Governance_A_Comprehensive_Analysis_of_AI_Integration_and_Policy_Development_in_the_German_Government
52. The agentic organization: A new operating model for AI | McKinsey, Zugriff am Oktober 26, 2025, https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-agentic-organization-contours-of-the-next-paradigm-for-the-ai-era