

Deutschland kann KI: Von der Pilotierung zur Skalierung

Ergebnisbericht der 1. Pilotierung des Agentic AI Hub

Inhalt

Vom Pilotprojekt in die Fläche – Der Weg zum handlungsfähigen Staat	4
Thesen	5
Ergebnisse & Akzeptanz in der Praxis	5
Interne Prozesse & technische Schnittstellenhürden	5
Beschaffung, Kooperation & regionale Sichtbarkeit	6
Ausgangssituation	6
Die fünf strategischen Lösungscluster	6
Die technologische Taxonomie im Agentic AI Hub	7
Methodik und Pilotierungsrahmen	8
Der Pilotierungsrahmen: Datengestützte Auswahl und das „Mission Agreement“	8
Ergebnisse aus den Pilotprojekten	8
1. Lector.ai	8
2. Celonis	9
3. Celonis	9
4. summ ai	9
5. forml	10
6. LeistungsLotse	10
7. Speechmind	11
8. Tucan.ai	11
9. myo formfix	11
10. Deepset	12
Ergebnisse der Pilotierung	13
Ergebnisse der Evaluation	13
Im Überblick	13
Betrachtung der vier Machbarkeits-Dimensionen	14
Organisatorische Machbarkeit & Nutzungs-Varianzen	14
Technologische Machbarkeit & Datenvolumen	14
Rechtliche Machbarkeit & föderale Rahmenbedingungen	15
Skalierfähigkeit nach Lösungstypus	15
Betrachtung der Begleitung der Pilotierung	15
Strategische Handlungsfelder	16
1. Marktanalyse: Infrastruktur und kommunale Bedarfe	16
2. Datenschutz: Föderale Rahmenbedingungen und die Datenhoheit	16

3. IT-Sicherheit: Stabilität durch souveräne Architekturen	17
4. Threat Modeling: KI-spezifische Bedrohungen beherrschen.....	18
5. Skalierung: Der Weg in die Fläche (Der Vier-Stufen-Plan).....	18
6. Presse und Kommunikation: Sichtbarkeit und Diskursbegleitung	19
Fazit und Ausblick: Konkrete Maßnahmen, Next Steps für die Skalierung	20
Konkrete Maßnahmen und Next Steps (Post-Pilotierung).....	20

Vom Pilotprojekt in die Fläche – Der Weg zum handlungsfähigen Staat

Die deutsche Verwaltung steht vor einer enormen Transformationsaufgabe: Demografie und Fachkräftemangel treffen auf wachsende bürokratische Anforderungen. Die erste Pilotierungsrunde des Agentic AI Hub hat den technologischen Beweis erbracht, dass agentische KI ein Werkzeug ist, um diesen Herausforderungen zu begegnen und so die staatliche Handlungsfähigkeit zu verbessern. Die Pilotierung hat bewiesen, dass die Verwaltung innovationsfähig ist und wie die lösungsorientierte Zusammenarbeit mit Startups auf Augenhöhe in der Praxis gelingt.

Im Rahmen der Pilotierung des Agentic AI Hub wurden 9 verschiedene agentische KI-Lösungen in 19 deutschen Kommunen im Rahmen von 20 Pilotprojekten vertestet. Um die erfolgreich erprobten Lösungen nun flächendeckend in die 11.000 deutschen Kommunen zu tragen, wird sich die Skalierungsstrategie auf vier zentrale Ansätze fokussieren:

1. **Rechtssichere und flexible Beschaffung:** Der Aufbau zentraler Vergabevehikel in Form eines dynamischen Beschaffungssystems (DBS), ermöglicht Kommunen künftig standardisierte, rechtssichere Mini-Wettbewerbe. Begleitend dazu versetzt eine vergaberechtliche Unterstützung die Kommunen in die Lage, länderspezifische Wertgrenzen für den dezentralen Direktauftrag oder Verhandlungsvergaben im Unterschwellenbereich autark und schnell zu nutzen.
2. **Lokaler Datenschutz und einheitliche Leitlinien:** Die durch die föderale Struktur erforderlichen lokalen und oft redundanten Datenschutzprüfungen sollen systematisch reduziert werden. Ein neu zu entwickelnder KI-Datenschutz-Prüfassistent, flankiert von standardisierten Leitlinien nach dem Prinzip „Einer prüft für alle“, soll die Vorgaben der KI-Verordnung sowie der 16 Landesdatenschutzgesetze bündeln und so die kommunalen Vorprüfzeiten reduzieren.
3. **Förderung von Startups und der souveränen Infrastruktur:** Mit Startups-Vouchern sollen junge Unternehmen gefördert werden. Zusätzlich ist die souveräne KI-Cloud von Beginn an auch mit Blick auf eine föderale Nutzungsperspektive konzipiert worden, sodass auch Kommunen Zugang zu einer souveränen Infrastruktur haben werden.
4. **Weiterempfehlungen und kuratierte Netzwerke:** Technologie skaliert am besten durch Vertrauen und praxisnahe Anwendungsfälle. Die weitere Verstetigung der bereits erfolgreich pilotierten KI-Lösungen stützt sich künftig auf ein dezentrales „Partner-Enablement“, in enger Zusammenarbeit mit den kommunalen Spitzenverbänden.
5. **Weitere Batches:** Es wird zeitnah mit einem zweiten Batch begonnen, bei welchen die Learnings der ersten Pilotierungsphase umgesetzt werden und die Anzahl der beteiligten Startups soll erhöht werden. Hierbei sollen agentische Anwendungen nicht nur für die Verwaltung nutzbar sein, sondern in anderen Bereichen ausstrahlen und Wirkung entfalten. So können auch Wirtschaftszweige, wie etwa das Handwerk, die ähnliche Herausforderungen aufweisen, von den Erfahrungen des Agentic AI Hub profitieren.

Dieser Abschlussbericht dokumentiert nicht nur die technologischen und prozessualen Erfolge der Pilotierung, sondern liefert das strategische Fundament für die anstehende Flächenskalierung. Deutschland kann KI – souverän, tatkräftig und verantwortungsvoll.

Thesen

Ergebnisse & Akzeptanz in der Praxis

1. Die agentische KI ermöglicht **radikale Effizienzsprünge** bei komplexen Verwaltungsprozessen. Die Pilotprojekte des Agentic AI Hub belegen **massive Ersparnisse**: **Lector.ai** senkte Durchlaufzeiten für DSGVO-Anträge um **>90 % (bis zu 28 Stunden Ersparnis/Vorgang)**. **Myo** reduzierte die Durchlaufzeit von Pflegeanträgen um **45 %** und sparte **31 % der Arbeitszeit**. **Deepset** senkte die Gesamtbearbeitungsdauer um **20 %**.
2. Das Stereotyp der innovationsfeindlichen Verwaltung ist nicht zutreffend. Die Positionierung der KI als **Entlastung** fördert **enorme Akzeptanz**. Bei **Speechmind** lag die Auslastung bei **92 %**, wobei 25 von 35 Nutzenden das Tool eigenständig verwendeten. **Myo** erreichte höchste Zufriedenheitswerte (**4,77/5 bei Bürgern, 4,66/5 bei Ämtern**).
3. Agentische KI-Systeme ermöglichen ein hohes Maß an Autonomie. Daher war im Rahmen der Pilotierung der **„Human-in-the-Loop“-Ansatz** für Verwaltungsmitarbeitende von großer Bedeutung. Das Pilotprojekt mit **LeistungsLotse** erzwingt beispielsweise weiterhin eine zwingende **menschliche 4-Augen-Kontrolle**.

Interne Prozesse & technische Schnittstellenhürden

4. Der wahre Innovationsbedarf liegt im verwaltungsinternen **„Maschinenraum“**. Der mit Abstand größte Bedarf liegt bei der **verwaltungsinternen Unterstützung**. **summ ai** demonstriert dieses Potenzial bei der Prozessmodellierung: beispielsweise wurde in **Rostock** im Rahmen der dreimonatigen Pilotierungsphase 116 Prozesse modelliert (insgesamt mit **11.000 € simuliertem Einsparpotenzial** in einer Woche) und **Metzingen** 72 Prozesse (mit insgesamt **13.711 € simuliertem Einsparpotenzial**).
5. Komplex entwickelte Fachverfahren und oft geschlossene APIs sind die **Digitalisierungshemmnisse**. Auf der einen Seite verursachen fehlende Schnittstellen die Leistungsfähigkeit der KI-Tools, sodass Funktionalitäten nachgebaut werden müssen. Auf der anderen Seite fordern komplexe und mehrstufige Fachverfahren die KI-Systeme: **Deepset** verarbeitete **1,3 Millionen Token pro Antrag**.

Datenschutz, Cloud & Digitale Souveränität

6. Durch den **Föderalismus** bedingte **unterschiedliche Datenschutzbestimmungen** verursachen **redundante Datenschutzprüfungen** bei gleichem Sachverhalt. Fragmentierte Datenschutzvorgaben erfordern Mehrfachprüfungen und **verhindern schnelle Skalierung**. Ein geplanter **KI-Datenschutz-Prüfassistent** soll **Vorprüfzeiten** künftig **von 3–5 Tagen auf 15–30 Minuten reduzieren und ermöglichen eine schnelle Skalierung ermöglichen**.
7. Die umfangreiche und unabhängige **Cloud-Nativität** erfordert bei einigen Pilot-Lösungen noch **weitere Entwicklungsschritte**. Zwar nutzen viele der an der Pilotierung beteiligten Lösungen bereits zertifizierte Cloud-Provider (BSI C5, ISO 27001). Einige Pilot-Lösungen setzen jedoch derzeit noch auf **proprietäre Cloud-Dienste statt offener Kubernetes-Architekturen**. Um künftig eine reibungslose **Portabilität auf souveräne Infrastrukturen wie den ‚Deutschland-Stack‘** zu gewährleisten, müssen diese Abhängigkeiten weiter reduziert werden.

8. Die Bestrebungen zur **digitalen Souveränität** gehen mit veränderten infrastrukturellen und ökonomischen Rahmenbedingungen einher. Der **lokale Eigenbetrieb** leistungsstarker, offener KI-Modelle erfordert behördenseitig **dedizierte High-End-GPU-Cluster**.

Beschaffung, Kooperation & regionale Sichtbarkeit

9. Das für die Skalierung zentrale Vehikel eines Dynamischen Beschaffungssystems (DBS) soll **über den Marktplatz Deutschland sichtbar gemacht** und **durch die entstehende IT-Einkaufsorganisation des Bundes betrieben werden**.
10. **Kommunale IT-Dienstleister** können künftig eine **Schlüsselrolle** einnehmen. Als direkter Lösungsansatz für die in These 5 beschriebenen Medienbrüche ist die **gemeinsame Erschließung offener Schnittstellen (APIs)** zu den Fachverfahren ein wertvoller Hebel. Für eine erfolgreiche Skalierung wird eine enge, partnerschaftliche Zusammenarbeit mit den IT-Dienstleistern angestrebt, um sie dabei zu unterstützen, **erprobte Startup-Lösungen passgenau in ihr Portfolio aufzunehmen**.
11. **Proaktives „Partner-Enablement“** ist ein Schlüssel für regionale Sichtbarkeit. Eine erfolgreiche Skalierung erfordert auch **mediale Sichtbarkeit**. Durch die gezielte **Unterstützung der Kommunen bei ihrer eigenen Presse- und Öffentlichkeitsarbeit**, die maßgeblich durch die Flankenunterstützung des BMDS getragen wurde, wurden 22 von insgesamt 59 verzeichneten Pressemitteilungen lokal initiiert. Dieser Ansatz verankert die Strahlkraft des Projekts bis tief in die Fläche und begegnet **lokalen, kritischen Diskursen proaktiv und transparent vor Ort**.

Ausgangssituation

Die Pilotierungsphase (März–Mai 2026) umfasste 20 Pilotprojekte, an denen 19 Kommunen und 9 Startups beteiligt waren.

Die empirische Belastbarkeit der nachfolgenden Ergebnisse stützt sich auf ein fundiertes *Mixed-Methods-Design*. Zur Vermeidung von Verzerrungen wurden zum Stichtag (28.05.2026) vollstandardisierte Längsschnittbefragungen, qualitative Experteninterviews und Reportings der Akteure systematisch analysiert.

Die fünf strategischen Lösungscluster

- I - Bürgerinteraktion vereinfachen:** Einsatz von KI als primäre Schnittstelle (z. B. intelligente Chatbots), die Anfragen vorfiltern und teils fallabschließend bearbeiten (Pilot: *Lector.ai*).
- II - Bürgernahe Verwaltungsprozesse:** Automatisierte Vorprüfung von komplexen Anträgen mit Außenbezug, wie Wohngeld oder Pflegeleistungen, um Wartezeiten zu reduzieren (Piloten: *forml*, *LeistungsLotse*, *Myo*).
- III - Verwaltungsinterne Prozesse:** KI agiert als „digitaler Kollege“ für intelligente Postverteilung, Recherche, Zusammenfassungen und Protokollierung (Piloten: *Speechmind*, *Tucan*).
- IV - Digitale Tools bauen:** Ermöglicht Fachabteilungen die Prozessmodellierung und Verbesserung der Datenqualität ohne externe Beratung (Piloten: *Celonis*, *summ ai*).
- V - KI-Infrastruktur aufbauen:** Die Schaffung einer souveränen „KI-Middleware“, die als sichere Brücke zwischen Sprachmodellen und Fachverfahren dient (Pilot: *Deepset*).

Die technologische Taxonomie im Agentic AI Hub

Die untersuchten Systeme lassen sich anhand ihrer agentischen Fähigkeiten und ihres Autonomiegrades in drei aufeinander aufbauende Stufen einteilen:

Stufe 1: Reaktive KI-Assistenten (z. B. Tucan, Speechmind) Diese Systeme bilden die Basisstufe. Sie arbeiten rein reaktiv: Sie nehmen eine Eingabe (wie z. B. eine Tonaufnahme) und verarbeiten diese präzise zu einem Ergebnis (wie einem Transkript oder Protokoll). Diese Assistenten führen einen klar vorgegebenen Auftrag aus.

Stufe 2: Gelenkte KI-Agenten (z. B. forml, summ ai, LeistungsLotse, Myo) Auf der mittleren Stufe agieren Systeme, die innerhalb klar definierter Leitplanken eigenständig handeln. Sie haben Zugriff auf externe Werkzeuge (wie Bilderkennung oder Datenbanken) und können die notwendigen Zwischenschritte, um eine Aufgabe zu lösen, selbstständig planen und ausführen. Ihre Autonomie ist jedoch begrenzt: Sie können keine freien Entscheidungen außerhalb ihres vorgegebenen Rahmens treffen.

Stufe 3: Hochautonome Agentische Systeme (z. B. Celonis, Deepset, Lector) Diese Systeme repräsentieren die höchste Stufe der KI-Entwicklung. Sie verknüpfen eigene Wahrnehmung direkt mit automatisierten Aktionen: Sie nutzen Werkzeuge komplett eigenständig für Lese- und Schreibzugriffe und verfolgen übergeordnete Ziele völlig selbstständig. Ein entscheidendes Merkmal ist ihre Fähigkeit zur intelligenten Selbstkorrektur („Self-Healing“): Tritt im Ablauf ein Fehler oder ein technisches Problem auf, kann das System diesen autonom erkennen, umgehen und den Lösungsweg selbstständig anpassen, um das Ziel dennoch erfolgreich zu erreichen.

Um rechtliche Vorgaben zu erfüllen, wurde der KI-Einsatz in den Piloten wie folgt limitiert:

- **Keine autonomen Fallentscheidungen:** In keinem der Pilotprojekte fanden vollautonome Fallentscheidungen oder eine automatisierte Bescheiderstellung durch die KI statt.
- **„Human-in-the-Loop“:** Die Letztentscheidung, fachliche Prüfung und Freigabe verblieben bei der menschlichen Sachbearbeitung.

Dies zeigte sich konkret in den verschiedenen Anwendungsfällen:

- Bei der Bearbeitung von Anträgen auf „Hilfe zur Pflege“ (Pilotprojekt *Myo*) betonte das Startup, dass **keine automatisierte Entscheidung über Leistungsansprüche** und keine automatische Ablehnung oder Bewilligung von Leistungen erfolgt. Die finale Entscheidung trifft ausschließlich die Behörde.
- Bei der Wohngeld-Antragsstellung (Pilotprojekt *LeistungsLotse*) fordert das System weiterhin eine **menschliche 4-Augen-Kontrolle**. Die Letztentscheidung bei der Bewilligung liegt bei einem menschlichen Sachbearbeiter.

Methodik und Pilotierungsrahmen

Der Pilotierungsrahmen: Datengestützte Auswahl und das „Mission Agreement“

Die Auswahl der 20 Pilot-Tandems basierte auf einer massiven Markterkundung mit fast 600 Einreichungen. Strategisch bevorzugt wurden sogenannte „Huckepack“-Bewerbungen von Kommunen und Startups, die bereits informell zusammenarbeiteten, da dies den operativen Start enorm beschleunigte.

Um die dreimonatige Pilotierung ohne langwierige Beschaffungsprozesse starten zu können, schloss der Hub ein unbürokratisches „Mission Agreement“. Die jeweilige Kommune fungierte als Betreiberin mit der formellen Letztverantwortung für den Dateneinsatz.

Ergebnisse aus den Pilotprojekten

Es folgt die erweiterte und detaillierte Übersicht für jedes Pilotprojekt, die zusätzlich eine kompakte Beschreibung des jeweiligen Produkts und der größten Herausforderung enthält.

Bevor die spezifischen Ergebnisse betrachtet werden, ist eine methodische Einordnung zu den gemessenen Effizienz-KPIs unerlässlich: Da durchgängige, maschinenlesbare Schnittstellen (APIs) zu den etablierten Fachverfahren für diese neuen Technologien künftig noch weiter ausgebaut werden müssen, erforderten die Pilotierungen temporäre manuelle *Workarounds* (wie z. B. PDF-Exporte oder die Nutzung von Tauschordnern). Die tatsächliche Netto-Effizienz einer künftigen, voll integrierten Ende-zu-Ende-Lösung dürfte daher in vielen Fällen noch höher ausfallen, als es die aktuelle Messung (*Workaround-Bias*) abbilden konnte.

1. Lector.ai

- **Kommune:** Neckar-Odenwald-Kreis
- **Produkt:** Lector ist eine souverän betreibbare KI-Plattform zur agentischen Verarbeitung eingehender Dokumente in der öffentlichen Verwaltung. Die Plattform deckt die vollständige Verarbeitungskette vom Posteingang bis zur Übergabe an Folgesysteme ab. Die Plattform kombiniert multimodale Vision-LLMs, agentische Prozessorchestrierung mit konfigurierbarer menschlicher Kontrolle und eine herstellerunabhängige Multi-Modell-Architektur. Sie lässt sich ohne Softwareanpassung konfigurieren, steht als SaaS oder *On-Premise* zur Verfügung und skaliert vom Einzelamt bis zum Landeseinsatz. Ergänzend ist im Rahmen der Pilotierung eine Checkliste für den datenschutzkonformen Einsatz von KI in der öffentlichen Verwaltung entstanden.
- **Gelöstes Problem:** Die manuelle Bearbeitung von Posteingängen und insbesondere von komplexen DSGVO-Auskunftsanträgen bindet extrem hohe personelle Ressourcen und dauert zu lange.
- **Detaillierte KPIs:**
 - Die Gesamtdurchlaufzeit für DSGVO-Auskünfte sank um über 90 % (von 2–4 Wochen auf wenige Stunden bis max. einen Tag).
 - Bei umfangreichen Kopien wurden bis zu 28 Stunden pro Vorgang eingespart.
 - Koordinationsaufwand mit anderen Ämtern sank von 4–8 Stunden auf 30–60 Minuten.
 - Interne Wartezeiten fielen von 5–15 Tagen auf wenige Minuten.
- **Erkenntnisse & Hürden:** Potenzielle Hindernisse in Bezug auf den Datenschutz wurden in Zusammenarbeit mit der Datenschutzbeauftragten unkompliziert genommen. In etwaigen Folgeprojekten kann beispielsweise die von der BfDI erarbeitete Handreichung für den datenschutzkonformen Einsatz von KI in der öffentlichen Verwaltung genutzt werden.

2. Celonis

- **Kommune:** München
- **Produkt:** Die „Celonis Intelligence Platform“ – eine Lösung, die *Process Mining*, Orchestrierung und KI vereint, um einen objektiven digitalen Zwilling von kommunalen Prozessen zu erstellen.
- **Gelöstes Problem:** Lange Bearbeitungszeiten und ineffiziente Informationszugriffe in der Fallbearbeitung, hier spezifisch bei Einbürgerungsverfahren durch komplexe Ausländerakten.
- **Detaillierte KPIs:** Fokus lag auf dem *Minimum Viable Product* (MVP) zur automatisierten Aktenzusammenfassung. Einsparungen und die Reduktion von Untätigkeitsklagen wurden simuliert und eine Return-on-Investment-Rechnung wird angestrebt.
- **Erkenntnisse & Hürden:** Hoher Abstimmungsbedarf intern sowie technische Hürden bei der Integration in die städtische IT-Infrastruktur und Anbindung eigener KI-Modelle. Learning: Solche Plattformen dürfen keine Insellösungen sein, sondern müssen Referats übergreifend (z. B. auch für Bußgeldstelle oder Gewerbeanmeldung) geplant werden.

3. Celonis

- **Kommune:** Nürnberg
- **Produkt:** (Identisch mit München) Eine Plattform für *Process Mining*, Orchestrierung und KI.
- **Gelöstes Problem:** Fehlende Transparenz und Überlastung im Forderungsmanagement und in der Vollstreckung im Kassen- und Steueramt.
- **Detaillierte KPIs:** Im Dashboard werden Kennzahlen und Prozessvarianten sichtbar gemacht. Auf dieser Basis können gezielt Maßnahmen zur Verbesserung des Forderungsmanagements entwickelt werden.
- **Erkenntnisse & Hürden:** Zeitintensive Prüfprozesse für Datenschutz/IT-Sicherheit. Die Anbindung an das in der Kommune genutzte System konnte durch das lokale IT-Team jedoch binnen weniger Tage realisiert werden.

4. summ ai

- **Kommunen:** Rostock, Waiblingen, Nettetal, Metzingen, Schwerin, Flensburg, Elbe-Elster
- **Produkt:** Eine spezialisierte KI-Software für das Prozessmanagement, die auf Basis von Spracheingaben oder Text in Echtzeit Prozesse nach BPMN 2.0- und FIM-Standard modelliert.
- **Gelöstes Problem:** Die fehlende Prozessdokumentation und der Mangel an Modellierungs-Know-how in der Verwaltung, was oft teure und sehr langwierige manuelle Beratungsprojekte nach sich zog.
- **Detaillierte KPIs:** Die manuelle Erstellung, die zuvor oft über 3 Wochen dauerte, wurde massiv beschleunigt.
 - Metzingen: 72 Prozesse KI-modelliert; 13.711 € simuliertem Einsparpotenzial.
 - Rostock: 116 Prozesse modelliert; 11.000 € Einsparpotenzial in einer Woche.
 - Flensburg: 264 Prozesse modelliert.
- **Erkenntnisse & Hürden:** Da die Lösung als isoliertes SaaS funktioniert, gab es keine technischen Hürden. Lediglich unklare Abstimmungsbedarfe beim Datenschutz traten auf. Learning: Ein interkommunales Vorprüfungsverfahren („einer prüft für alle“) ist sinnvoll.
- **Gemeinsamkeiten des Einsatzes bei verschiedenen Kommunen:**
 - Keine technischen Hürden: In keiner der sieben Kommunen gab es Schnittstellenblockaden. Da die Lösung als isoliertes SaaS-Produkt operiert und keine API-Anbindung an kommunale Alt-Systeme (*Legacy-IT*) benötigt wurden, verlief die technologische Implementierung überall reibungslos (in unter 4 Wochen).
 - Kein Echtdaten-Risiko: Es wurden keine sensiblen Bürgerdaten verarbeitet, was die rechtliche Integration massiv erleichterte.

- **Unterschiede des Einsatzes bei verschiedenen Kommunen:**

- Starke Varianzen im Output: Die tatsächliche Nutzung und die daraus resultierenden Effizienzen hingen stark von der jeweiligen Kommune ab. Während Flensburg 264 und Rostock 118 Prozesse KI-modellierte, wurden in Nettetal 12 und im Kreis Elbe-Elster 8 Prozesse modelliert.

5. forml

- **Kommunen:** Frankfurt am Main, Düsseldorf
- **Produkt:** Eine Cloud-Software, die unstrukturierte Antragsdaten verarbeitet und maschinell auf Vollständigkeit und Einkommensgrenzen validiert. Sie beinhaltet ein Portal zur medienbruchfreien Datenübertragung.
- **Gelöstes Problem:** Der hohe manuelle Aufwand bei der Prüfung von Anträgen auf Wohnberechtigungsscheine (WBS) sowie die zeitintensive Nachforderung von Dokumenten mit den Antragstellenden
- **Detaillierte KPIs:** Im Fokus stand die Erarbeitung eines Prototyps. Dieser wurde erfolgreich entwickelt und getestet. Spezifische Wirkungs-KPIs aus einem Live-Betrieb wurden in der Pilotphase auf Basis der Nutzung von synthetischen Daten noch nicht erhoben.
- **Erkenntnisse & Hürden:** Vollumfängliche Schnittstellen zu E-Akten und Fachverfahren standen im Pilotzeitraum nicht durchgängig zur Verfügung. Daher wurden für die Pilotierung temporäre technische Übergangslösungen eingesetzt (Tauschordner, Skripte). Learning: Auch kleine Features können bereits einen großen Mehrwert in der Antragsbearbeitung liefern.
- **Gemeinsamkeiten des Einsatzes bei verschiedenen Kommunen:**
 - In beiden Kommunen wurde derselbe fachliche Kernprozess erprobt: Eingang von Anträgen und Nachweisen über unterschiedliche Kanäle, Verarbeitung unstrukturierter Dokumente, Dokumentensplitting und -klassifikation, Datenextraktion, Vollständigkeits- und Plausibilitätsprüfung, nachvollziehbare Ergebnisdarstellung sowie die Unterstützung von Nachforderungen und Folgeeingängen über ein Portal.
- **Unterschiede des Einsatzes bei verschiedenen Kommunen:**
 - Lokale Spezifika im Prozess: In Düsseldorf (E-Akte vorhanden) lag der Fokus auf der Dokumentenverarbeitung und Veraktung wohingegen in Frankfurt am Main (keine E-Akte) der Fokus auf die Überprüfung bestehender Wohnberechtigungsscheine auf anhaltende Berechtigung (Fehlbelegungsabgabe) lag.

6. LeistungsLotse

- **Kommune:** Nettetal
- **Produkt:** Ein Online-Dienst kombiniert mit einem Fachverfahren, der Prozesse digital durchgängig abbildet und als Basis für agentische KI-Systeme dient.
- **Gelöstes Problem:** Die fehlende durchgängige Digitalisierung, die durch Medienbrüche und Systemwechsel die Bearbeitung von Wohngeldanträgen stark verlangsamt.
- **Detaillierte KPIs:** Von 180 Besuchern nutzten 137 Systemfunktionen. 54 Besucher schlossen die Anmeldung erfolgreich ab. 18 Personen reichten den Antrag komplett digital ein (10 % aller Besucher, 33 % der Angemeldeten).
- **Erkenntnisse & Hürden:** Technologisch zielt das System darauf ab, zwingend deterministische Workflows (klare Regeln) mit der probabilistischen KI zu kombinieren. Es fordert weiterhin eine menschliche 4-Augen-Kontrolle.

7. Speechmind

- **Kommune:** Bielefeld
- **Produkt:** Eine KI-gestützte, cloudbasierte Protokollierungssoftware, die Audio- und Videoaufnahmen in strukturierte Texte umwandelt.
- **Gelöstes Problem:** Der sehr zeitaufwendige manuelle Prozess der Protokollerstellung für Gremien und interne Sitzungen sowie subjektive Interpretationen und schwankende Dokumentationsqualität.
- **Detaillierte KPIs:** 92 % Auslastung (83 von 90 verfügbaren Erfassungsstunden genutzt). 25 von 35 registrierten Sachbearbeitenden nutzten das Tool eigenständig mit Erfolg. Hohe Usability (nur 2 Support-Anfragen im Projekt).
- **Erkenntnisse & Hürden:** Datenschutzprüfungen verzögerten den Start. Unzureichende Hardware (externe Mikrofone) sowie App-Einschränkungen trübten die KI-Ergebnisse. Learning: Vorab klare Kommunikation zu Hardware-Anforderungen und Fokus der Schulungen auf KI-Anwendungsszenarien statt Grundbedienung.

8. Tucan.ai

- **Kommune:** Bielefeld
- **Produkt:** KI-Lösung für automatisierte Sitzungsprotokolle und Live-Transkription
- **Gelöstes Problem:** Fachkräfte werden durch Schreibarbeit für Protokolle gebunden, und generische US-KIs verursachen in der Verwaltung oft massive Datenschutzprobleme.
- **Detaillierte KPIs:** Die manuelle Nachbearbeitungszeit pro Protokoll reduzierte sich signifikant um rund 78 % (von ca. 45 Minuten auf ca. 10 Minuten) mit einer Protokoll-Erfolgsquote von 100 % (Anteil der Sitzungen mit direkt nutzbarem KI-Protokoll) sowie einem „Summary-Precision-Score“ (SPS) von 98,5 von 100 Punkten.
- **Erkenntnisse & Hürden:** Bestehende IT-Sicherheitsrichtlinien und Lizenzstrukturen ließen eine direkte Integration in Plattformen wie MS Teams kurzfristig nicht zu. Zudem ergaben sich datenschutzrechtliche Klärungsbedarfe, weshalb alternative Integrationswege genutzt werden mussten. Learning: Eine frühzeitige Abstimmung mit den IT-Dienstleistern sowie die Etablierung klarer Datenschutz-Leitlinien sind zentrale Erfolgsfaktoren, um eine Lösung nach der Beschaffung reibungslos operativ umsetzen zu können.

9. myo formfix

- **Kommune:** Kreis Heinsberg, Köln, Berlin (Marzahn-Hellersdorf & Steglitz-Zehlendorf)
- **Produkt:** Ein digitales, KI-gestütztes Web-Tool, das Bürgerinnen und Bürger Schritt für Schritt durch den Antragsprozess führt, die eingereichten Unterlagen verwaltungslogisch für die Sachbearbeitung aufbereitet, Rückfrageschleifen reduziert und dadurch die Sozialämter entlastet. Pflegeheime profitieren zudem von geringeren Außenständen, einer verbesserten Liquidität und mehr Transparenz im Antragsprozess.
- **Gelöstes Problem:** Unvollständige Anträge für die „Hilfe zur Pflege“, die bei den Behörden zu einem enormen manuellen Prüfaufwand, massiven Rückfragen und extrem langen Bearbeitungsdauern führen.
- **Detaillierte KPIs:**
 - Arbeitsaufwand am Dokument sank standortübergreifend um 31 % (von 260 auf 180 Minuten).
 - Gesamtdurchlaufzeit eines Antrags sank um 45 % (von 85 auf 47 Tage).
 - Nutzerzufriedenheit sehr hoch (Bürger 4,77/5, Ämter 4,66/5, Pflegeheime 4,00/5).
- **Erkenntnisse & Hürden:** In der Praxis zeigte sich, dass föderale Besonderheiten und lokale Vorgaben eine hohe Flexibilität der Systeme erfordern. KI-Lösungen müssen zwingend anpassbar sein, um mit den unterschiedlichen rechtlichen und prozessualen Rahmenbedingungen der einzelnen Kommunen umgehen zu können.

- **Gemeinsamkeiten des Einsatzes bei verschiedenen Kommunen:**
 - Identische Wirksamkeit: Über alle Standorte hinweg zeigte die KI die gleiche hohe Leistungsfähigkeit. Die reine Arbeitszeit am Dokument sank überall konstant um 31 % (von 260 auf 180 Minuten), und die Gesamtdurchlaufzeit verkürzte sich einheitlich um 45 % (von 85 auf 47 Tage).
 - Hohe Nutzerzufriedenheit: Die Zufriedenheitswerte waren standortübergreifend sehr hoch und nahezu identisch (Bürger: 4,77/5, Ämter: 4,66/5, Pflegeheime: 4,00/5).
- **Unterschiede des Einsatzes bei verschiedenen Kommunen:**
 - Unterschiedliche datenschutzrechtliche Auslegungen: Während bestimmte externe Cloud-Speicherdienste in einigen Pilotkommunen für den Datenaustausch datenschutzrechtlich genehmigt wurden, konnten sie in anderen Kommunen nicht eingesetzt werden.
 - Temporäre Medienbrüche: Aufgrund der divergierenden Freigaben mussten in diesen Fällen hybride *Workarounds* (z. B. Dokumentenversand per PDF und Post) genutzt werden, was medienbruchfreie Prozesse temporär einschränkte.
 - Lokale Formularanforderungen: In der Praxis zeigten sich spezifische lokale Anforderungen an die einzureichenden Dokumente. Die KI musste in diesen Fällen flexibel angepasst werden, um beispielsweise kommunalspezifische Pflichtfelder (wie einen Heimvertrag) zu implementieren.

10. Deepset

- **Kommune:** Kreis Borken
- **Produkt:** Die Open-Source-KI-Orchestrierungsschicht „*Haystack*“, die als technologie-agnostische Plattform zur produktionsreifen Entwicklung von LLM-Anwendungen (wie *RAG-Pipelines* und KI-Agenten) dient.
- **Gelöstes Problem:** Der ineffiziente, manuelle Prozess der Dokumentensortierung, Vorprüfung auf Vollständigkeit und das aufwändige Auslesen von Finanzdaten bei der Vermögensprüfung für Pflegeanträge.
- **Detaillierte KPIs:** Systemkosten lagen bei ca. 1 US-Dollar pro 75-seitigem Antrag (bei ca. 1,3 Mio. verarbeiteten Token). Die Genauigkeit erreichte 76,7 % beim 1. Schritt (Dokumententrennung, Benennung, Klassifizierung) und 60 % beim 2. Schritt der komplexen Vermögensprüfung. Die Gesamtbearbeitungsdauer sank um 20 % (in den automatisierten Schritten um 35 %).
- **Erkenntnisse & Hürden:** In der Praxis zeigte sich, dass das immense Datenvolumen komplexer Fachverfahren die KI-Systeme besonders stark fordert. Die hochgesteckte Zielgenauigkeit von >90 % konnte im kurzen Pilotierungszeitraum noch nicht vollständig erreicht werden. Learning: Für die erfolgreiche Automatisierung komplexer Fachverfahren sollten von Beginn an längere Test-, Trainings- und Pilotphasen eingeplant werden.

Ergebnisse der Pilotierung

Die Pilotierungsphase des Agentic AI Hub umfasste 20 Pilotprojekte in 19 Kommunen mit den Produkten von 9 Startups und hat belegt, dass agentische KI-Lösungen die Verwaltungspraxis signifikant verbessern und entlasten können. In der Praxis führte dies zu einer messbaren Reduktion von Bearbeitungszeiten bei Routinetätigkeiten und zu einer deutlichen Erhöhung der Datenqualität.

Ergebnisse der Evaluation

Die systematische Auswertung der Pilotierungsphase (basierend auf standardisierten Fragebögen, KPI-Reportings und Experteninterviews der 20 Pilot-Tandems) erlaubt folgende Rückschlüsse über die Leistungsfähigkeit und die Hürden der agentischen KI in der Verwaltungspraxis.

Operativer Einsatz & Echtzeiten: Zum Stichtag der Evaluation befanden sich 70 % der Pilot-Lösungen (14 von 20) im operativen Einsatz, wobei 80 % der Projekte parallel erfolgreich mit kommunalen Echtzeiten arbeiteten.

Implementierungsgeschwindigkeit: Bei 65 % der Piloten konnte die tatsächliche Einführung (von der Beauftragung bis zur ersten Nutzung) in unter 4 Wochen realisiert werden.

Nutzerakzeptanz & Zielerreichung: In 16 von 18 auswertbaren Pilot-Tandems bewerteten die Kommunen die Zielerreichung der Pilotierung kategorial positiv. In 15 von 18 Fällen gaben die Sachbearbeitenden an, sich im Umgang mit der KI-Lösung „sicher“ zu fühlen.

Das Narrativ der Entlastung: 62 % der Kommunen stuften die KI in ihrer primären Rolle als „Entlastungs-Tool“ (Schutz vor Überlastung) ein.

Im Überblick

Gemessene Strukturgröße	Erhobener Messwert / Verteilung in den Pilot-Tandems
Zufriedenheit mit der KI-Lösung	95 % der Kommunen haben mit einer hohen Zufriedenheit bewertet Eine hohe Zufriedenheit umfasst die positiven Bewertungen 4 („stimme eher zu“) und 5 („stimme uneingeschränkt zu“) auf der verwendeten fünfstufigen Likert-Skala.
Operativer Einsatz am Stichtag	70 % Produktivbetrieb im Arbeitsalltag 30 % Betrieb in einer isolierten Testumgebung
Technologischer Reifegrad (Datentypus)	80 % aktive Verarbeitung von Echtzeiten
Technologische Integrierbarkeit	45% agieren als funktional isolierte Software-Inseln ohne Notwendigkeit einer Backend-Integration 55 % agieren als End-to-End-Prozesse; die aktuell noch fehlenden APIs werden flexibel durch <i>Workarounds</i> überbrückt
Infrastruktur-Souveränität	44 % Souveräne und unabhängige Architekturen 33 % Hybride und migrationsfähige Architekturen 22 % Architekturen mit hohen Abhängigkeiten

	Da die Kernarchitektur einer Lösung über alle beteiligten Pilotprojekte identisch ist, erfolgt diese Auswertung auf Basis der neun beteiligten Startups.
Implementierungs- geschwindigkeit	65% in unter 4 Wochen 10% in 4 bis <8 Wochen 15% in 8 bis <12 Wochen 10% in 12 Wochen oder mehr Die Implementierungsgeschwindigkeit wird durch die Dauer der Einführung von der Beauftragung bis zur ersten Nutzung definiert.
Übertragbarkeit auf andere Kommunen	50 % bedingungslose Übertragbarkeit 50 % rechtliche, technische oder organisatorische Einschränkungen werden als Bedingung genannt

Betrachtung der vier Machbarkeits-Dimensionen

Organisatorische Machbarkeit & Nutzungs-Varianzen

Entgegen theoretischer Stereotypen belegen die Daten einen enormen „Wandel-Willen“ in den Kommunen. In 70 % der validen Erhebungen stuften die Programmverantwortlichen den Onboarding-Aufwand für die Belegschaft als „gering“ oder „kein Aufwand“ ein. Die mit Abstand größte Hürde war jedoch der strukturelle „Ressourcen-Gap“ (fehlende Zeitkontingente im Tagesgeschäft), der von 5 der 9 Startups als primäre Bremse identifiziert wurde. Dass die reine Bereitstellung der Technik ohne stringentes Change-Management nicht ausreicht, belegt die **Gegenüberstellung baugleicher Piloten**: Obwohl die jeweilige Lösung standortübergreifend technisch fehlerfrei funktionierte, zeigte der Output eine hohe Varianz. Während einige Kommunen im Pilotierungszeitraum über 250 Prozesse KI-modellierten und hohe simulierte Einsparpotenziale generierten, wurden in anderen teilnehmenden Kommunen im selben Zeitraum lediglich Prozesse im einstelligen Bereich erfasst. Dies belegt, dass der Erfolg stark von der lokalen organisatorischen Begleitung abhängt.

Change-Management: Führung und Vermeidung von „Automation Bias“

Die flächendeckende Implementierung von Agentic AI ist kein reines IT-Projekt, sondern ein tiefgreifendes Transformationsvorhaben. Dies verändert die Rolle der Führungskräfte vom Anweisungsgeber zum Moderator, der eine konstruktive Fehler- und Lernkultur etablieren muss. Zentral für den Kompetenzaufbau der Belegschaft ist die notwendige Schulung zur Reduzierung des „Automation Bias“ – der kognitiven Neigung, maschinelle KI-Vorschläge unreflektiert zu übernehmen. Zudem müssen bestehende Dienstvereinbarungen überprüft und Personalräte sowie Gleichstellungsvertretungen frühzeitig eingebunden werden, um die Belegschaft bei der Einführung dieser neuen Assistenzsysteme partnerschaftlich und transparent zu begleiten.

Technologische Machbarkeit & Datenvolumen

Die Cloud-basierten KI-Lösungen wiesen eine sehr hohe Systemstabilität auf (teils mit Ausfallraten von unter 0,5 %). Dennoch operieren 45 % der Systeme aktuell als „funktional isolierte Software-Inseln“, da in vielen Kommunen offene Schnittstellen (APIs) zu den Bestandssystemen fehlen, was manuelle *Workarounds* erzwingt. Wie bedeutsam eine frühzeitige Vorab-Abstimmung in diesem Zusammenhang ist, zeigte sich in der praktischen Umsetzung: Bestehende Lizenzstrukturen können die direkte Integration in Kommunikationsplattformen temporär einschränken. Über alternative Integrationswege und *Workarounds* (wie manuelle Uploads oder andere Videokonferenz-Tools) ließen sich solche Hürden im Rahmen der Pilotierung jedoch erfolgreich beheben. Hinsichtlich der Zielgenauigkeit wurde zudem deutlich, dass das immense Datenvolumen komplexer Fachverfahren die Systeme besonders stark fordert. So erzeugte beispielsweise ein einziger umfangreicher Pflegeantrag zur Verarbeitung rund 1,3 Millionen Token.

Aufgrund dieser hohen Komplexität konnte bei tief integrierten Systemen die hochgesteckte Zielgenauigkeit von >90 % im kurzen Pilotierungszeitraum noch nicht vollständig erreicht werden. Dies bestätigt das zentrale Learning: Komplexe Interaktion in Fachverfahren benötigen zwingend längere Pilot- und Trainingsphasen.

Rechtliche Machbarkeit & föderale Rahmenbedingungen

Lokale Prüfzyklen sind und bleiben erforderlich, um einen rechtssicheren Produktivbetrieb zu ermöglichen. Die Herausforderungen dezentraler Prüfungen zeigten sich deutlich in der Umsetzung: Während bestimmte externe Cloud-Speicherdienste in einigen Pilotkommunen für denselben Einsatzzweck datenschutzrechtlich genehmigt wurden, konnten sie in anderen Kommunen nicht eingesetzt werden. Dies machte in diesen Fällen temporäre hybride *Workarounds* (z. B. Dokumentenversand per PDF und klassischer Post) erforderlich und schränkte medienbruchfreie Prozesse temporär ein.

Skalierfähigkeit nach Lösungstypus

Die Daten zeigen starke Varianzen in der Skalierbarkeit, abhängig vom Typus der Lösung. Am schnellsten skalierbar (in 100 % der Fälle in unter 4 Wochen implementiert) erwiesen sich funktional isolierte SaaS-Anwendungen wie die Prozessmodellierung, da diese keine Anbindung an die bestehende IT benötigen und datenschutzrechtliche Hürden gering waren. End-to-End-Integrationsplattformen (Pflegeanträge, Sozialleistungen, Datenverbesserung) erfordern hingegen naturgemäß einen höheren Integrationsbedarf: Da durchgängige API-Standards sowie einheitliche Datenschutzfreigaben für solche komplexen Prozesse künftig noch weiter etabliert werden müssen, erforderten diese in den Pilotierungen individuelle Abstimmungs- und Anpassungsaufwände von über 4 bis zu 12 Wochen.

Betrachtung der Begleitung der Pilotierung

Aus der abschließenden Evaluation der teilnehmenden Kommunen und Startups zeigt sich deutlich: Trotz der organisatorischen Hürden, die einem ersten Pilot-Durchlauf geschuldet sind, werten die Beteiligten selbst das Projekt als großen kulturellen und technologischen Erfolg für den öffentlichen Dienst. So berichten die kommunalen Vertreterinnen und Vertreter in ihren Rückmeldungen explizit von einem spürbaren Change-Prozess in den eigenen Fachbereichen. Anfängliche Skepsis gegenüber den neuen KI-Technologien konnte in echte Überzeugung umgewandelt werden und es entstanden bereits konkrete neue Initiativen zur weiteren Digitalisierung in den Ämtern.

Alle Pilot-Tandems dokumentieren in den Erhebungen eine lösungsorientierte Zusammenarbeit. Die marginale technologische Einstiegshürde und die intrinsische Nutzungsmotivation (bis hin zu „Wow, das macht ja überraschenderweise sogar Spaß“), war über die gesamte Projektlaufzeit spürbar. Kurze Feedback-Routinen, iterative Anpassung der funktionalen Prototypen an lokale Verwaltungsbedarfe und die hohe intrinsische Innovationsbereitschaft der kommunalen Mitarbeitenden fungiert als massiver Katalysator für die Implementierung.

Basierend auf diesen Praxiserfahrungen äußern die Teilnehmenden aus Kommunen und Startups den ausdrücklichen Wunsch, das Format weiterzuführen, sich bei der Weiterführung des Hub inhaltlich zu beteiligen und so gemeinsam ein langfristiges Ökosystem zu etablieren.

Strategische Handlungsfelder

In der Begleitung und Evaluation von 20 Pilot-Tandems aus 9 Startups und 19 Kommunen zeigt sich eine hohe funktionale Zielerreichung auf Software- und Prozessebene. Gleichzeitig macht die Evaluation strukturelle Systemgrenzen auf infrastruktureller, administrativer und rechtlicher Ebene sichtbar. Hier liegt gleichzeitig der größte Hebel: Werden diese Rahmenbedingungen gezielt gestaltet, dienen sie als echter *Booster* für den bundesweiten *Roll-out*.

Die Evaluation identifizierte vier strukturelle Barrieren, die einen flächendeckenden Roll-out derzeit noch bremsen:

- Ein akuter organisatorischer Ressourcen-Engpass (Zeitmangel) im Verwaltungsalltag.
- Schnittstellen-Inkompatibilitäten, bei denen Fachverfahren ohne offene APIs manuelle *Workarounds* und Medienbrüche verursachen.
- Divergierende Auslegung datenschutzrechtlicher Bestimmungen.
- Unsicherheiten in der Beschaffung.

Um diese Hürden aufzulösen und die Pilotlösungen in die Fläche zu tragen, wurden die folgenden acht Handlungsfelder systematisch analysiert.

1. Marktanalyse: Infrastruktur und kommunale Bedarfe

Die erste Phase des Projekts offenbarte eine hohe technologische Dynamik im Startup-Ökosystem, das sich jedoch noch in einer Vorphase der Skalierung befindet. Die Marktanalyse verdeutlicht eine klare Diskrepanz: Während Frontend-Lösungen (wie *Chatbots*) zur Bürgerinteraktion bereits eine gewisse Marktsättigung aufweisen, fehlt es an tief integrierten End-to-End-Prozessen für die Fachverfahren.

Übergreifende Erkenntnisse:

- **Fokus auf interne Verwaltungsprozesse:** Der mit Abstand größte kommunale Bedarf liegt in der verwaltungsinternen Unterstützung, beispielsweise bei der intelligenten Postverteilung, Dokumentenlogistik und dem Wissensmanagement zur Bewältigung der „Dokumentenflut“.
- **Strukturelle Rahmenbedingungen und Schnittstellen:** Für einen durchgängigen Datenfluss müssen offene Schnittstellen (APIs) zu den etablierten Fachverfahren künftig weiter ausgebaut werden. Aktuell erfordern noch nicht flächendeckend etablierte Standard-Schnittstellen teilweise hohe Integrationsaufwände, was den medienbruchfreien Zugang zu produktiven Verwaltungsdaten für neue Startup-Lösungen temporär einschränkt
- **Die Schlüsselrolle der kommunalen IT-Dienstleister:** Kommunale IT-Dienstleister sind für eine erfolgreiche flächendeckende Skalierung unverzichtbar.
- **Infrastruktur-Bedarf:** Eine zentrale Voraussetzung für die Skalierung ist der Ausbau der Basisinfrastruktur. Eine souveräne „*KI-Middleware*“ kann als sichere Brücke zwischen KI-Anwendungen und sensiblen Fachverfahren hier eine Lösung darstellen.

2. Datenschutz: Föderale Rahmenbedingungen und die Datenhoheit

Der Datenschutz erfordert in der Praxis einen hohen Abstimmungsbedarf, da die rechtliche Auslegung in Deutschland durch den Föderalismus teils unterschiedlich ausfällt. Lokale Einzelfallprüfungen sind daher erforderlich. Die Herausforderungen dezentraler Prüfungen zeigten sich deutlich in der Umsetzung: Während bestimmte externe Cloud-Speicherdienste in einigen Pilotkommunen für denselben Einsatzzweck datenschutzrechtlich genehmigt wurden, konnten sie in anderen Kommunen nicht eingesetzt werden. Dies machte in diesen Fällen temporäre hybride Workarounds erforderlich und schränkte medienbruchfreie Prozesse temporär ein

Übergreifende Lösungsansätze:

1. **Schutz der Echtdaten:** Um Risiken zu minimieren, haben alle Startups vertraglich garantiert, dass kommunale Echtdaten nicht in die Gewichte der generativen KI-Modelle einfließen oder für deren Training genutzt werden.
2. **Die 4 Praxis-Cluster zum Umgang mit Echtdaten:** Um im Rahmen der föderalen Strukturen handlungsfähig zu bleiben und den Einsatz von Echtdaten in der Praxis zu ermöglichen, haben sich in der Pilotierung vier unterschiedliche strategische Cluster gebildet:
 - a. **Verarbeitung besonders sensibler Daten mit maximaler Compliance:** Erfordert zertifizierte europäische Cloudlösungen, externe DSFAs und strenge AVVs.
 - b. **Arbeiten mit synthetischen Daten:** Nutzung von Dummydaten zur reinen Architekturprüfung.
 - c. **Pragmatische „Stateless“-Modelle:** Nutzung zustandsloser Systeme oder hybrider Strukturen, bei denen keine persistente Speicherung sensibler Daten in der KI-Umgebung erfolgt.
 - d. **Fokus auf internen Beschäftigtendatenschutz:** Nutzung von B2G-Lösungen, die primär Mitarbeiterdaten verarbeiten. Unterschiedliche Auslegungen zur Rechtsgrundlage: Ein Treiber für den hohen Abstimmungsbedarf in den Kommunen sind zudem unterschiedliche regulatorische Einschätzungen.
3. **„Register as a Service“:** Um redundante Doppelprüfungen zu vermeiden, bietet sich die Etablierung eines „Einer-prüft-für-alle“-Prinzips an, bei dem erfolgreiche Audits von anderen Kommunen übernommen werden können.
4. **KI-Datenschutz-Prüfassistent:** Zur operativen Entlastung wird ein algorithmischer Prüfassistent konzipiert. Dieses Tool soll künftig die 16 Landesdatenschutzgesetze, die DSGVO, den EU AI Act sowie Vorgaben der BfDI bündeln und eine geteilte kommunale Wissensdatenbank bereitstellen, um die Vorprüfzeit für Kommunen von 3 bis 5 Tagen auf 15 bis 30 Minuten zu reduzieren.

3. IT-Sicherheit: Stabilität durch souveräne Architekturen

Die technologische Machbarkeit wurde in der Pilotierung durch eine sehr hohe Systemstabilität und schnelle Cloud-basierte Einsatzbereitschaft bestätigt. Die Startups nutzen dafür etablierte Cloud-Provider, die in der Regel BSI C5-Testate und ISO 27001-Zertifizierungen aufweisen.

Die technologische Skalierungsstrategie fordert die Nutzung Cloud-nativer Lösungen (Container/Kubernetes) zur Portabilität auf souveräne Infrastrukturen wie der KI-PaaS Cloud des Deutschland-Stacks (D-Stack). In der Pilotierung zeigte sich, dass für diese Anforderung bei einigen Pilot-Lösungen künftig noch technologischer Weiterentwicklungsbedarf besteht, da teilweise noch auf anbieterspezifische Cloud-Dienste (wie serverlose Funktionen) zurückgegriffen wurde. Um künftig vollständig „D-Stack-ready“ zu sein und eine autonome Migration zu ermöglichen, müssen derartige architektonische Bindungen an einzelne Provider schrittweise reduziert werden.

Übergreifende Lösungsansätze:

- **Etablierte Zertifizierungen:** Um das Vertrauen der Verwaltung zu gewinnen, wiesen die genutzten Cloud-Provider in der Regel BSI C5-Testate und ISO 27001-Zertifizierungen auf.
- **Innovative Sicherheitsarchitekturen:** Einige Startups implementierten neuartige Schutzkonzepte, um Risiken abzufedern. So wurde beispielsweise nach dem „Memory-Only“-Prinzip operiert, bei dem Audiodaten nur im flüchtigen Arbeitsspeicher verarbeitet und nicht auf Festplatten ausgelagert werden. Andere Lösungen nutzten die kryptografische „Envelope Encryption“, bei der mandantenspezifische Schlüssel bei Vertragsende unwiderruflich vernichtet werden („Crypto-Erase“).
- **Nutzung Cloud-nativer Architekturen und Kubernetes:** Um eine dauerhafte Abhängigkeit von Anbietern (Vendor-Lock-in) zu vermeiden, sieht die Skalierungsstrategie langfristig den Aufbau Cloud-

nativer Lösungen auf Basis von Kubernetes im *D-Stack* vor. Die Nutzung der KI-Cloud PaaS ist hierfür geeignet.

4. Threat Modeling: KI-spezifische Bedrohungen beherrschen

Klassische IT-Sicherheitsprüfungen reichen für agentische KI-Systeme nicht aus. In speziellen Risiko-Workshops nutzte das Projektteam ein auf KI adaptiertes STRIDE-Framework (*Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege*) und analysierte Risiken anhand von *Threat Personas* und *Attack Trees*.

Im Überblick:

- **Zentrale KI-Risiken:** Zu den größten KI-Spezifika zählen *Prompt Injections* (Befehlsmanipulation), *Algorithmic Bias* (automatisierte Diskriminierung durch historische Trainingsdaten), *Information Disclosure* (Datenabfluss) und *Supply Chain Attacks* (Angriffe über externe Software-Bibliotheken).
- **Positives Risikobewusstsein der Startups:** Trotz der Identifikation dieser teils gravierenden potenziellen Angriffsvektoren lieferten die Risiko-Workshops ein ausdrücklich positives Gesamtergebnis hinsichtlich der Partner: Es war erfreulich festzustellen, dass die im Hub-Programm beteiligten Startups bereits im Vorfeld über eine bemerkenswert hohe Sensibilität, ein ausgeprägtes Risikobewusstsein und eine exzellente Übersicht über die potenziellen Angriffsflächen verfügten.
- **Zero-Trust und Guardrails:** Als technische Gegenmaßnahmen wurden vorgeschaltete Filter („*Guardrails*“), striktes Least-Privilege-Rechtmanagement und eine kontinuierliche Zero-Trust-Überwachung (*Anomaly-Monitoring*) etabliert.
- **Human-in-the-Loop:** Vollautonome Bescheiderstellungen sind derzeit ausgeschlossen; die Letztverantwortung und Freigabe verbleiben bei der Sachbearbeitung.

5. Skalierung: Der Weg in die Fläche (Der Vier-Stufen-Plan)

Um die Diskrepanz zwischen innovativen Pilot-Lösungen und dem fehlenden administrativen Rahmen aufzulösen, erfordert die Skalierung die Standardisierung von Infrastruktur und Vergabe. Das Validierungsverfahren identifizierte das Dynamisches Beschaffungssystem als die vielversprechendste Lösung.

Damit ein Pilotprojekt überhaupt in den regulären Betrieb übernommen und bundesweit skaliert wird, müssen strikte Akzeptanzkriterien erfüllt sein. Eine Lösung muss technische Interoperabilität durch offene Programmierschnittstellen (APIs) beweisen und als übertragbare Blaupause nach dem ‚Einer-für-Alle‘-Prinzip (EfA) in anderen Kommunen nachgenutzt werden können. Führt ein System in eine Abhängigkeit von proprietären Einzelanbietern (*Vendor-Lock-in*) oder stützt es sich zwingend auf lokale Sonderstrukturen, gilt das Akzeptanzkriterium für eine Skalierung als nicht erfüllt.

Die Blaupause hierfür bildet ein holistischer Vier-Stufen-Plan:

- **Stufe 1 (Sofortmaßnahme):** Unterstützung für die 19 Pilotkommunen, damit diese die direkte Anschlussbeauftragung „ihrer“ Startups umsetzen können. Zudem soll ein algorithmischer Datenschutz-Prüfassistent konzipiert werden.
- **Stufe 2 (Information & Guarantees auf dem Marktplatz Deutschland):** Aufbau einer kuratierten Oberfläche. Im Unternehmensportal können Startups ihre Unterlagen und Nachweise hinterlegen.
- **Stufe 3 (Dynamisches Beschaffungssystem – DBS):** Errichtung eines DBS bei einer zentralen Beschaffungsstelle. Anders als starre Rahmenverträge bleibt das DBS während seiner Laufzeit jederzeit für neue Startups offen. Es ordnet die KI-Lösungen in diverse Endkategorien ein.
- **Stufe 4 (Regelbetrieb):** Das DBS geht in den Live-Betrieb. Kommunen können standardisierte, rechtssichere „Mini-Wettbewerbe“ initiieren. Ergänzend wird eine „Broker-Funktion“ etabliert, die

identische Bedarfe mehrerer Kommunen zu gemeinsamen Abrufen bündelt, um Preis- und Skaleneffekte zu generieren.

Technologisch zielt die Skalierung parallel darauf ab, dass zukünftige KI-Lösungen Cloud-nativ und konform mit den Standards des „Deutschland-Stacks“ (D-Stack) aufgebaut sind. Dies garantiert durch Containerisierung und offene APIs die notwendige Portabilität, um einen „Vendor-Lock-in“ zu verhindern und die Systeme jederzeit auf souveräne Infrastrukturen migrieren zu können.

Technologische und organisatorische Voraussetzungen für den Roll-out

Zur Sicherung der digitalen Souveränität und Vermeidung von Herstellerabhängigkeiten (*Vendor-Lock-in*) fordert das Konzept strikt modellunabhängige Orchestrierungsschichten, den Einsatz offener APIs sowie die Einhaltung etablierter Interoperabilitätsstandards der Verwaltung (wie XÖV oder FIM).

Organisatorisch erfordert eine bundesweite Multiplikation das „Einer-testet-für-alle“-Prinzip: Hat eine KI-Lösung eine Freigabe hinsichtlich IT-Sicherheit und Datenschutz erhalten, sollte diese als referenzierbare Basis von anderen Kommunen gegenseitig anerkannt werden, statt Prüfungen redundant zu wiederholen. Zudem sollten „Infrastructure Readiness-Leitfäden“ bereitgestellt werden, die Kommunen vorab über notwendige Cloud-Migrationen und Konformitäten aufklären.

6. Presse und Kommunikation: Sichtbarkeit und Diskursbegleitung

Die begleitende Presse- und Öffentlichkeitsarbeit bildete einen zentralen Erfolgsfaktor, um das Narrativ von Bundesminister Dr. Karsten Wildberger – „KI nicht nur diskutieren, sondern machen“ – wirkungsvoll zu verankern und den Hub als zentralen „Maschinenraum“ der Verwaltungsdigitalisierung zu positionieren. Eine strikte „One-Voice“-Strategie und die enge Abstimmung aller Akteure stellten eine konsistente Kommunikation nach innen und außen sicher.

- **Erfolgreiches Partner-Enablement:** Die dezentrale PR-Strategie war erfolgreich. Von insgesamt 59 verzeichneten Pressemeldungen wurden 22 direkt von den beteiligten Pilotkommunen initiiert bzw. in der lokalen Presse veröffentlicht. Diese starke lokale Berichterstattung trug die Innovationskraft des Projekts greifbar bis tief in die Fläche.
- **Nationale Medienresonanz und TV-Präsenz:** Der Hub erzielte breite Sichtbarkeit in Leitmedien (z. B. Süddeutsche Zeitung Dossier, Tagesspiegel Background) und wurde im reichweitenstarken eGovernment-Newsletter als „Nummer-1-Thema“ des Ministeriums gelobt. Im bundesweiten Fernsehen positionierte parlamentarischer Staatssekretär Thomas Jarzombek den Hub im Mai 2026 in der ZDF-Talkrunde „Am Puls“ prominent als wegweisendes Praxisbeispiel für den nutzbringenden KI-Einsatz des Staates. Die Berichterstattung zum Projektabschluss im Juni spiegelte den messbaren Erfolg des Hub eindrucksvoll wider: Medien hoben hervor, dass in den verschiedenen Pilotprojekten drastische Effizienzgewinne erzielt wurden – von einer Reduktion der Durchlaufzeiten bei DSGVO-Auskünften um über 90 % bis hin zu massiven Zeiteinsparungen bei der behördlichen Prozessmodellierung und Sitzungsprotokollierung. Dass über die Hälfte der Pilotkommunen die Lösungen direkt im Anschluss fest beauftragt hat, wurde medial als klarer Beleg für den praktischen Nutzen gewertet.
- **Präsenz in Audio-Formaten:** Ergänzend zur Print- und Online-Resonanz wurde der Hub zum Taktgeber in relevanten Fach-Podcasts (u.a. *Tech Weekly*, *AI FIRST*, *eGovernment Podcast*). In tiefgehenden Gesprächen mit der politischen Leitungsebene, den Verantwortlichen in den Kommunen sowie den Startup-Gründern wurde das Projekt dort als Blaupause für pragmatische Staatsmodernisierung diskutiert.
- **Proaktive Kommunikation:** Diskurse, etwa durch Medien wie netzpolitik.org zu ethischen Bedenken („Automation Bias“) oder lokale Datensicherheits-Rückfragen, wurden durch die abgestimmte Kommunikationsstrategie sachlich eingeordnet und transparent entkräftet.

- **Aktive Community und Ausblick:** Auf Plattformen wie LinkedIn entstand ein organisches und extrem aktives Ökosystem aus Kommunen, Startups, Branchenverbänden und politischer Leitungsebene, dass die Projektfortschritte kontinuierlich dokumentierte und die Vernetzung stärkte. Zum Projektabschluss gelang es zudem, die Anschlussdebatte medial zu besetzen: Die Diskussion um die Skalierung in die Fläche sowie die Vision eines bundesweiten niedrigschwelligen Skalierungsmechanismus' für kommunale KI-Anwendungen wurde von den Leitmedien positiv aufgegriffen.

Fazit und Ausblick: Konkrete Maßnahmen, Next Steps für die Skalierung

Die Pilotierungsphase hat das immense operative Potenzial agentischer KI-Systeme für die Verwaltungspraxis empirisch belegt. Um diese isolierten Erfolge in einen bundesweiten Roll-out zu überführen, müssen die identifizierten administrativen, technologischen und rechtlichen Barrieren nun strukturell aufgelöst werden. Die nachfolgenden Handlungsempfehlungen definieren die konkreten nächsten Schritte für den Übergang in den Regelbetrieb.

Konkrete Maßnahmen und Next Steps (Post-Pilotierung)

Die Überführung der Pilotprojekte in die Fläche erfordert eine konsequente Standardisierung der Beschaffungs- und Integrationswege. Dies erfolgt entlang des strategischen Vier-Stufen-Plans, flankiert von notwendigen technologischen Vorbereitungen.

Kurzfristige Sofortmaßnahmen:

- **Weiterentwicklung der Erkenntnisse der 1. Pilotierung:** Unterstützung für die 19 Pilotkommunen, damit diese die direkte Anschlussbeauftragung „ihrer“ Startups umsetzen können. Zudem soll ein algorithmischer Datenschutz-Prüfassistent konzipiert werden.
- **Realisierung eines 2. Batches:** Im Rahmen der Markterkundung werden weitere agentische KI-Lösungen für den Einsatz in der Verwaltung identifiziert und im Rahmen eines nächsten Batches auch deren organisatorische, rechtliche und technische Machbarkeit geprüft und bei entsprechender Skalierungsfähigkeit verstetigt. Hierbei sollen agentische Anwendungen nicht nur für die Verwaltung nutzbar sein, sondern in anderen Bereichen ausstrahlen und Wirkung entfalten. So können auch Wirtschaftszweige, wie etwa das Handwerk, die ähnliche Herausforderungen aufweisen, von den Erfahrungen des Agentic AI Hub profitieren.

Mittelfristige strukturelle Maßnahmen:

- **Aufbau der Marktplatz-Sichtbarkeit:** Agentische KI-Lösungen werden auf dem Marktplatz Deutschland sichtbar, indem etwa Startups und deren Produkte recherchierbar bzw. filterbar gemacht werden.
- **Konzeptionierung und Verfestung eines Prototyps für einen KI-Datenschutz- Prüfassistenten:** Zur Reduktion der redundanten, mehrtägigen Prüfprozesse auf 15 bis 30 Minuten soll ein algorithmischer Prüfassistent (inklusive geteilter kommunaler Wissensdatenbank) entwickelt und getestet werden.
- **Nutzung der zentralen Trägerstruktur („IT-Einkaufsorganisation des Bundes“):** Für den dauerhaften Betrieb des Dynamischen Beschaffungssystems (DBS) wird die enge Zusammenarbeit mit der entstehenden IT-Einkaufsorganisation des Bundes als zentraler Beschaffungsstelle anvisiert.
- **Bedarfsbündelung (Broker-Funktion):** In den Regelbetrieb des DBS könnte eine Moderationsfunktion integriert werden. Diese bündelt identische Bedarfe mehrerer Kommunen zu gemeinsamen „Mini-Wettbewerben“, um Transaktionskosten zu senken und Preisvorteile zu generieren.

Technologische Vorbereitung:

- **Infrastruktur-Readiness und GPU-Beschaffung:** Die angestrebte Portierung von KI-Modellen in souveräne Zielinfrastrukturen (wie KIPITZ oder D-Stack) erfordert die Vorab-Dimensionierung und Beschaffung dedizierter souveräner High-End-GPU-Ressourcen (Terabyte-VRAM-Klasse).
- **Eliminierung von IaC-Lock-ins:** Provider-spezifische Infrastructure-as-Code-Skripte (Terraform/Helm) der Startups müssen angepasst werden, um sie von proprietären Cloud-Ressourcen zu entkoppeln und „D-Stack-ready“ zu machen.