Souveränität nach Maß

Kriterien für den kombinierten Einsatz von Claude und Mistral in der öffentlichen Verwaltung

Forschungsfrage:
„Anhand welcher Kriterien lassen sich Use-Cases generativer KI in der öffentlichen Verwaltung leistungsstarken außereuropäischen Modellen (Claude) bzw. souveränen europäischen Modellen (Mistral) zuordnen, um Nutzen und digitale Souveränität auszubalancieren?“

Thomas Zojer
Kurs: Politische Analyse
22. Juni 2026

Inhaltsverzeichnis

1 Einleitung

Öffentliche Verwaltungen stehen unter doppeltem Druck: Sie sollen die Leistungsfähigkeit generativer künstlicher Intelligenz nutzen, um Produktivität und Servicequalität zu steigern, zugleich aber die europäischen Anforderungen an Datenschutz und digitale Souveränität wahren. Beide Ziele scheinen in Konflikt zu stehen, weil die leistungsfähigsten Sprachmodelle von außereuropäischen Anbietern stammen, während souveräne europäische Alternativen an Reife gewinnen, aber nicht in jeder Hinsicht gleichziehen. Dass generative KI im öffentlichen Sektor längst angekommen ist, zeigt eine Befragung britischer Beschäftigter, der zufolge ein erheblicher Teil entsprechende Anwendungen kennt oder bereits nutzt – allerdings weitgehend unkoordiniert und ohne klare Leitlinien (Bright et al., 2025).

Der vorliegende Essay knüpft an die Erkenntnis an, dass die bloße Wahl eines europäischen Anbieters keine digitale Souveränität herstellt, sondern das Souveränitätsrisiko lediglich von der Anbieter- auf die Infrastrukturebene verlagert (Calderaro & Blumfelde, 2022; Baur, 2025). Wenn aber die Anbieterwahl allein nicht entscheidet, stellt sich die praktisch dringliche Frage, nach welchen Kriterien einzelne Anwendungsfälle unterschiedlichen Modellen zugeordnet werden sollten, um Nutzen und Souveränität in ein tragfähiges Verhältnis zu bringen.

Die leitende Forschungsfrage lautet daher: Anhand welcher Kriterien lassen sich Use-Cases generativer KI in der öffentlichen Verwaltung leistungsstarken außereuropäischen Modellen wie Claude beziehungsweise souveränen europäischen Modellen wie Mistral zuordnen, um Nutzen und digitale Souveränität auszubalancieren? Die zentrale These lautet, dass nicht der Anbieter, sondern die Datenklasse und der Verarbeitungskontext das maßgebliche Zuordnungskriterium bilden. Daraus wird ein konkretes Modell abgeleitet: Der unmittelbare Bürgerkontakt, in dem Bürgerinnen und Bürger sensible personenbezogene Daten preisgeben, ist souveränen europäischen Lösungen vorbehalten, während die interne Verwaltungsarbeit und die Erstellung von Bescheiden leistungsstarken Modellen überantwortet werden kann, sofern die zugrunde liegende Infrastruktur ohnehin außereuropäischer Kontrolle unterliegt.

Die Analyse ist interdisziplinär angelegt und verbindet Governance- und Verwaltungswissenschaft mit Perspektiven aus Recht und Technologiemanagement. Sie geht literaturbasiert vor und entwickelt aus dem Forschungsstand ein anwendbares Zuordnungsmodell. Der Argumentationsgang gliedert sich wie folgt: Kapitel 2 klärt den theoretischen Rahmen aus Souveränitätsverständnis, Datenklassen und risikobasierter Logik. Kapitel 3 systematisiert Nutzen und Anwendungsfälle generativer KI in der Verwaltung. Kapitel 4 vergleicht die Souveränitätsprofile von Claude und Mistral. Kapitel 5 entwickelt das datenklassenbasierte Zuordnungsmodell. Kapitel 6 widmet sich Governance und Kompetenz,

Kapitel 7 diskutiert Zielkonflikte und Limitationen, bevor Kapitel 8 die Forschungsfrage beantwortet.

2 Theoretischer Rahmen: Digitale Souveränität, Datenklassen und risikobasierte Zuordnung

Digitale Souveränität entzieht sich einer einheitlichen Definition und fungiert eher als politisches Schlagwort, das je nach Akteur unterschiedlich gefüllt wird (Pohle & Thiel, 2020). Eine systematische Auswertung von 271 Studien identifiziert vier Souveränitätsmodelle – rechtebasiert, marktorientiert, zentralisierend und staatsbasiert – und kommt zu dem Schluss, dass kein Modell umfassende Regulierung mit hinreichender Innovationsfähigkeit verbindet (Fratini et al., 2024). Für die Verwaltungspraxis folgt daraus, dass Souveränität nicht als Alles-oder-nichts-Zustand, sondern als mehrdimensionales und graduell herstellbares Ziel zu verstehen ist.

Hilfreich ist die Unterscheidung dreier Ebenen: einer Datenebene, die fragt, wer rechtlich und faktisch auf die verarbeiteten Daten zugreifen kann; einer Infrastrukturebene, die die technische Grundlage von Verarbeitung und Speicherung betrifft; und einer Kompetenzebene, die die Fähigkeit der Institution erfasst, Systeme eigenständig zu betreiben, zu prüfen und zu gestalten (Carrapico & Farrand, 2025). Souveränität auf einer Ebene garantiert dabei keine Souveränität auf den anderen. Eine Behörde kann einen europäischen Anbieter wählen und dennoch auf einer außereuropäisch kontrollierten Infrastruktur operieren.

Der rechtliche Kern dieses Problems ist der Konflikt zwischen extraterritorialen Zugriffsregimen und europäischem Datenschutzrecht: Maßgeblich ist nicht der physische Speicherort, sondern die jurisdiktionelle Kontrolle über die Daten (Propp & Swire, 2024). Empirische Fallstudien zeigen, dass öffentliche Stellen dieses Grundproblem durch technische und organisatorische Maßnahmen mindern, aber nicht vollständig auflösen können, solange wirksamer Rechtsschutz gegen außereuropäische Überwachung fehlt (Hildén, 2021). Die prominenteste europäische Gegeninitiative, Gaia-X, reproduziert bestehende Abhängigkeiten teilweise, statt sie aufzulösen (Baur, 2025; Blancato, 2023).

Begrifflich ist digitale Souveränität zudem von der strategischen Autonomie abzugrenzen: Während Souveränität die rechtlich-politische Kontrolle betont, verweist strategische Autonomie auf die geopolitische Handlungsfähigkeit der Union. Europa bewegt sich dabei zwischen dem sogenannten Brussels Effect – der Fähigkeit, über Regulierung globale Standards zu setzen – und dem Streben nach eigener technologischer Unabhängigkeit (Christakis, 2020). Teile der Forschung schlagen vor, präziser von Datenautonomie zu sprechen, um die konkrete Kontrolle über Datenflüsse statt eines diffusen Souveränitätsideals

in den Vordergrund zu rücken (Gstrein, 2023). Zugleich erfüllt der Souveränitätsdiskurs eine identitätspolitische Funktion und reproduziert geopolitische Vorstellungen, was die Gefahr birgt, Souveränität symbolisch statt überprüfbar zu führen (Monsees & Lambach, 2022). Für die Verwaltungspraxis schärft diese begriffliche Differenzierung den Blick: Es geht nicht um ein Bekenntnis zu europäischen Anbietern, sondern um nachweisbare Kontrolle über konkrete Datenflüsse.

Entscheidend für die hier verfolgte Argumentation ist die risikobasierte Logik des europäischen Rechtsrahmens. Die einschlägige Gesetzgebung – von der Datenschutz-Grundverordnung über den Data Act bis zum KI-Gesetz – verfolgt zunehmend differenzierte, an Risiko und Sensibilität orientierte Ansätze (Hulkó et al., 2025). Diese Logik korrespondiert mit etablierten Methoden der Informationssicherheit wie ISO 27005, NIST SP 800-30 oder COBIT, die eine systematische, an der Datensensibilität ausgerichtete Bewertung und abgestufte Schutzmaßnahmen ermöglichen (Ali et al., 2024). Ergänzend gebieten die Prinzipien der Zweckbindung und Datenminimierung, die Verarbeitung personenbezogener und besonders schützenswerter Daten strikt zu begrenzen (Finck & Biega, 2021). Aus diesen Bausteinen lässt sich das tragende Kriterium dieses Essays gewinnen: Nicht der Anbieter, sondern die Datenklasse und der Verarbeitungskontext bestimmen das angemessene Souveränitätsniveau. Die folgenden Kapitel übersetzen dieses Kriterium schrittweise in ein anwendbares Modell.

3 Generative KI in der öffentlichen Verwaltung: Nutzen und Use-Case-Typologie

Der Nutzen generativer KI in der Verwaltung ist empirisch zunehmend belegt. Eine quasi-experimentelle Analyse weist messbare Produktivitätsgewinne bei bürokratischer Wissens- und Dokumentenarbeit nach (Kim, 2026). Auf der Adoptionsseite wird die behördliche Einführung durch technologische, organisationale und umweltbezogene Faktoren bestimmt und kann sowohl effizienzsteigernde als auch explorative Innovation befördern (Zhou et al., 2025). Zugleich verweist die unkoordinierte Verbreitung in der Praxis auf einen erheblichen Steuerungsbedarf (Bright et al., 2025).

Für eine differenzierte Zuordnung ist die Unterscheidung von Front-Office und Back-Office grundlegend. Ein etablierter konzeptioneller Rahmen trennt die bürgerorientierte Dienstleistungserbringung im Front-Office von der internen Verarbeitung im Back-Office (Lindgren & Jansson, 2013). Diese Trennung liefert die analytische Grundlage dafür, Technologie nicht pauschal, sondern je nach Verwaltungsebene zuzuordnen, weil sich Front- und Back-Office hinsichtlich Datenexposition, Grundrechtsrelevanz und Vertrauensanforderungen systematisch unterscheiden.

Im Front-Office dominieren bürgernahe Anwendungen wie dialogbasierte Auskunftssysteme. Ein retrieval-gestütztes Sprachmodell konnte Bürgeranfragen zu Politik- und Verwaltungsdokumenten mit hoher Genauigkeit verständlich beantworten und so Transparenz und Beteiligung erhöhen (Yun et al., 2024). Solche Systeme berühren jedoch unmittelbar das Bürgervertrauen, das stark vom Anwendungsbereich und von wahrgenommenem Datenschutz abhängt (Aoki, 2020; Wang et al., 2024). Im Back-Office stehen dagegen interne Aufgaben im Vordergrund: Ein feinabgestimmtes Modell kann interne Verwaltungsdokumente wie Briefing-Notes strukturiert erstellen, wenngleich die menschliche Aufsicht unverzichtbar bleibt (Nzobonimpa et al., 2026). Eine funktionale Typologie ordnet KI-Einsatz entsprechend entlang konkreter Verwaltungsfunktionen – von der internen Verwaltung über die Leistungserbringung bis zur Antragsbearbeitung – und betont durchgängig den staatlichen Steuerungsbedarf (Longo, 2024; OECD, 2023).

Die Datenexposition unterscheidet sich zwischen beiden Ebenen grundlegend. Im Front-Office gibt die Bürgerin oder der Bürger die Daten aktiv und oft unstrukturiert preis; das System nimmt sie im Moment der Interaktion entgegen, und ein Schutzversagen an dieser Stelle wäre unmittelbar und kaum reversibel. Im Back-Office hingegen liegen die Daten bereits in den Fachverfahren und Dokumentenablagen der Organisation vor; die KI verarbeitet sie sekundär, in einem bereits etablierten und vertraglich geregelten Verarbeitungskontext. Diese Asymmetrie ist der eigentliche Grund, warum eine pauschale Modellwahl der Sache nicht gerecht wird: Front- und Back-Office stellen unterschiedliche Anforderungen an Vertrauen, Rechtsgrundlage und Souveränität und legen daher unterschiedliche Lösungen nahe.

Zugleich ist der Nutzen nicht ohne Risiken zu haben. Die empirischen Befunde zur Genauigkeit generativer Systeme zeigen, dass selbst leistungsfähige Modelle fehlerhafte oder unvollständige Ausgaben erzeugen können, weshalb menschliche Kontrolle insbesondere bei rechtsverbindlichen Verwaltungsakten unverzichtbar bleibt (Nzobonimpa et al., 2026). Die unkoordinierte Verbreitung in der Praxis verschärft dieses Risiko, weil ohne klare Leitlinien weder die Eignung der Anwendungsfälle noch die Sensibilität der verarbeiteten Daten systematisch geprüft wird (Bright et al., 2025). Eine differenzierte Zuordnung muss diese Qualitäts- und Verantwortungsfragen daher von Anfang an mitführen.

4 Souveränitätsprofile: Claude und Mistral im Vergleich

Die beiden Modellfamilien dieses Essays stehen exemplarisch für zwei Pole des Spannungsfeldes. Claude des US-amerikanischen Anbieters Anthropic repräsentiert ein leistungsstarkes, proprietäres Modell, das in der Regel über außereuropäisch kontrollierte Infrastruktur bereitgestellt wird und damit dem in Kapitel 2 beschriebenen jurisdiktionellen Zugriffsrisiko unterliegt. Mistral des französischen Anbieters steht für eine europäische

Alternative, deren Modelle teils als offene Gewichte verfügbar sind und sich dadurch auch in EU-gehosteten oder lokalen Umgebungen betreiben lassen.

Der entscheidende Unterschied liegt weniger in der reinen Modellqualität als im Souveränitätsprofil. Souveräne, national oder regional kontrollierte Sprachmodelle sichern die Datenhaltung im eigenen Hoheitsgebiet, gewährleisten sprachlich-kulturelle Passung und reduzieren die Abhängigkeit von ausländischen Anbietern; sie erfordern jedoch Infrastruktur-, Talent- und Regulierungsstrategien (Bondarenko et al., 2025). Genau diese Eigenschaft – die Möglichkeit, das Modell auf kontrollierter Infrastruktur zu betreiben – macht ein europäisches Modell für sensible Verarbeitungskontexte wertvoll, in denen der jurisdiktionelle Zugriff Dritter ausgeschlossen werden muss.

Daraus folgt zugleich, dass die Modellwahl nur dann souveränitätswirksam ist, wenn sie mit der Infrastruktur- und Kompetenzebene zusammengedacht wird. Ein europäisches Modell auf außereuropäisch kontrollierter Cloud bleibt ebenso wenig souverän wie die symbolische Anbieterpräferenz, vor der die Forschung warnt (Calderaro & Blumfelde, 2022; Roberts et al., 2021). Umgekehrt heißt das: Wo die Infrastruktur ohnehin außereuropäischer Kontrolle unterliegt – etwa weil Dokumente regelhaft in einer US-amerikanisch kontrollierten Office-Umgebung gespeichert werden –, erzeugt der zusätzliche Einsatz eines leistungsstarken außereuropäischen Modells nur ein begrenztes Mehr an Souveränitätsrisiko gegenüber dem bestehenden Zustand.

Die Verfügbarkeit souveräner Alternativen ist allerdings selbst voraussetzungsvoll. Eine vollständig europäisch kontrollierte Infrastruktur ist weniger ein fertiges Produkt als ein umkämpftes politisches Projekt, das an divergierenden nationalstaatlichen Präferenzen und begrenzten technologischen Kapazitäten leidet (Rone, 2024). Die Forschung warnt zudem vor einer naiven Gleichsetzung von europäisch und souverän und plädiert dafür, rechtliche Kontrolle, technische Sicherheit und wirtschaftliche Tragfähigkeit gegeneinander abzuwägen; Open-Source-Lösungen und kundenseitig verwaltete Verschlüsselung gelten dabei als zentrale Hebel, um Kontrolle zurückzugewinnen (Michels et al., 2023). Für die Praxis bedeutet dies, dass ein offenes europäisches Modell wie Mistral seinen Souveränitätsvorteil erst durch die kontrollierte Betriebsumgebung – EU-Hosting, lokale Bereitstellung, eigene Schlüsselverwaltung – tatsächlich realisiert.

5 Ein datenklassenbasiertes Zuordnungsmodell: Mistral im Front-Office, Claude im Back-Office

Aus dem theoretischen Rahmen und der Use-Case-Typologie lässt sich ein konkretes, dreistufiges Zuordnungsmodell ableiten, das die Datenklasse und den Verarbeitungskontext

zum Leitkriterium erhebt. Es übersetzt die risikobasierte Logik des europäischen Rechtsrahmens (Hulkó et al., 2025) in eine praktische Routing-Regel für den KI-Einsatz und ist als analytisch abgeleitetes Modell zu verstehen, das einer empirischen Validierung bedarf.

Die erste Stufe betrifft den unmittelbaren Bürgerkontakt, in dem Bürgerinnen und Bürger aktiv sensible personenbezogene Daten preisgeben – etwa in Auskunfts- und Antragsdialogen zu sozialen, gesundheitlichen oder familiären Angelegenheiten. Hier entsteht das eigentliche Souveränitätsrisiko an der Eingabestelle, und genau hier ist das Bürgervertrauen am verletzlichsten (Aoki, 2020; Alishani et al., 2026; Wang et al., 2024). Diese Front-Office-Interaktionen sind daher souveränen europäischen Lösungen wie Mistral vorbehalten, die auf EU-kontrollierter oder lokaler Infrastruktur betrieben werden können, sodass ein extraterritorialer Zugriff auf die preisgegebenen Daten ausgeschlossen bleibt. Das Gebot der Datenminimierung verstärkt diese Zuordnung, weil besonders schützenswerte Daten erst gar nicht in außereuropäisch kontrollierte Systeme gelangen sollen (Finck & Biega, 2021).

Die zweite Stufe umfasst die interne Verwaltungsarbeit und die Erstellung von Bescheiden im Back-Office: Entwürfe, Textbausteine, Zusammenfassungen, Wissensmanagement und vergleichbare Aufgaben, deren Ergebnisse regelhaft in der bestehenden Office-Umgebung der Organisation gespeichert werden. Da diese Infrastruktur bereits außereuropäischer Kontrolle unterliegt, ist die Souveränität auf der Infrastrukturebene faktisch schon gebunden; der Einsatz eines leistungsstarken Modells wie Claude erhöht das Souveränitätsrisiko gegenüber dem Status quo daher nur marginal, während der Produktivitäts- und Qualitätsgewinn erheblich ist (Kim, 2026; Nzobonimpa et al., 2026). Hier überwiegt der Nutzen, weshalb die Leistungsfähigkeit des Modells den Ausschlag geben kann.

Die dritte Stufe schließlich betrifft besonders sensible oder in hohem Maße grundrechtsrelevante Daten – etwa Gesundheitsdaten, sicherheitsrelevante Informationen oder Vorgänge mit erheblichem Diskriminierungspotenzial. Für sie genügt auch eine europäische Modellwahl nicht; erforderlich ist eine vollständig souveräne Infrastruktur mit kundenseitig kontrollierter Verschlüsselung und EU-Jurisdiktion, bei der ein extraterritorialer Zugriff technisch und rechtlich ausgeschlossen ist (Bondarenko et al., 2025; Issaoui et al., 2023).

Technisch wird ein solches abgestuftes Modell durch Mehrmodell-Architekturen ermöglicht. Lernbasierte Router können Anfragen dynamisch zwischen einem leistungsstarken und einem schwächeren oder souveränen Modell verteilen und dabei Kosten und Qualität ausbalancieren (Ong et al., 2024), und kaskadierende Ansätze erreichen das Niveau eines Spitzenmodells bei deutlich geringeren Kosten (Chen et al., 2023). Auf die Verwaltung übertragen bedeutet dies, dass die Zuordnung nach Datenklasse nicht manuell für jeden

Vorgang getroffen werden muss, sondern als regelbasiertes Routing in die Systemarchitektur eingebettet werden kann – vorausgesetzt, die Klassifizierung der Daten erfolgt zuverlässig.

Die Tragfähigkeit des Modells steht und fällt mit der Datenklassifizierung als kritischem Kontrollpunkt. Anders als bei der rein technischen Optimierung von Kosten und Latenz, auf die die Routing-Literatur abzielt, muss das Routing in der Verwaltung primär an Schutzbedarf und Rechtsgrundlage ausgerichtet werden. Eine fehlerhafte Klassifizierung – etwa wenn ein Vorgang mit besonderen Kategorien personenbezogener Daten irrtümlich als gering sensibel eingestuft und an ein außereuropäisches Modell geleitet wird – würde das Souveränitätsziel unmittelbar konterkarieren. Die etablierten risikobasierten Bewertungsmethoden liefern hierfür den methodischen Rahmen, müssen aber in die operativen Verfahren der Behörde eingebettet und kontinuierlich überprüft werden (Ali et al., 2024; Issaoui et al., 2023). Das Zuordnungsmodell verlagert das Souveränitätsproblem damit von der Anbieterfrage auf eine beherrschbare, aber anspruchsvolle Klassifizierungs- und Kontrollaufgabe.

6 Governance, Kompetenz und Risiken

Ein hybrides Zuordnungsmodell steigert den Nutzen, verlagert die Souveränitätsfrage aber auf die Kompetenzebene: Es funktioniert nur, wenn die Institution Daten zuverlässig klassifizieren, das Routing kontrollieren und die Einhaltung der Regeln überprüfen kann. Verantwortungsvolle KI-Governance lässt sich entsprechend über strukturelle, relationale und prozedurale Praktiken konzeptualisieren, die als organisationale Voraussetzung sicherer und vertrauenswürdiger Nutzung über den gesamten Lebenszyklus wirken (Papagiannidis et al., 2025). Ein integriertes Governance-Rahmenwerk für die Verwaltung ordnet die Risiken – von Datenschutz über Verantwortlichkeit bis Kontrolle – systematisch organisationalen Steuerungsmechanismen zu (Wirtz et al., 2020).

Die Praxis zeigt jedoch eine Lücke zwischen Verbreitung und Steuerung: Generative KI wird vielfach bereits genutzt, ohne dass klare Leitlinien vorliegen (Bright et al., 2025). Damit das Zuordnungsmodell trägt, müssen Verwaltungen drei Kompetenzen aufbauen: erstens eine verlässliche Datenklassifizierung als Eingangsvoraussetzung; zweitens die technische Fähigkeit, Routing-Regeln zu implementieren und zu auditieren; und drittens die rechtlich-organisatorische Kompetenz, Verträge, Verschlüsselung und Zugriffsrechte über die gesamte Verarbeitungskette zu steuern. Ohne diese Kompetenzen drohen neue Compliance- und Souveränitätsrisiken, weil falsch klassifizierte Daten an das jeweils unpassende Modell geleitet werden.

Hinzu kommt eine institutionelle Voraussetzung, die über die einzelne Behörde hinausreicht: Technologische Souveränität ist auch eine Frage gezielter Investitionen und des

Kompetenzaufbaus, ohne die Institutionen die nötigen Abstufungen gar nicht eigenständig vornehmen und durchsetzen können (OECD, 2023). Eine Behörde, die weder über klassifizierungsfähiges Personal noch über die vertragliche und technische Steuerungsfähigkeit verfügt, wird das Zuordnungsmodell nicht tragen, sondern bestenfalls symbolisch nachvollziehen. Der Aufbau dieser Kompetenz – von der Datenklassifizierung über das Vertrags- und Verschlüsselungsmanagement bis zur Audit-Fähigkeit – ist daher kein nachgelagerter Implementierungsschritt, sondern die eigentliche Bedingung souveränitätskonformer KI-Nutzung.

Schließlich verweist die Governance-Perspektive auf einen strukturellen Zielkonflikt zwischen Autonomie und Interdependenz: Das Streben nach Souveränität kollidiert mit der faktischen Verflechtung globaler digitaler Wertschöpfung (Carrapico & Farrand, 2025). Ein vollständiger Verzicht auf leistungsfähige außereuropäische Modelle wäre kostspielig und könnte die Innovationsfähigkeit beeinträchtigen, während ein unreflektierter Einsatz die Souveränität untergräbt. Das abgestufte Zuordnungsmodell ist der Versuch, diesen Konflikt nicht aufzulösen, aber pragmatisch zu steuern.

7 Diskussion

Die Analyse stützt die zentrale These, dass die Datenklasse und der Verarbeitungskontext – und nicht die Anbieterherkunft – das maßgebliche Zuordnungskriterium bilden. Die risikobasierte Logik des Rechtsrahmens (Hulkó et al., 2025), die Trennung von Front- und Back-Office (Lindgren & Jansson, 2013) und die Vertrauensbefunde an der Bürgerschnittstelle (Aoki, 2020; Wang et al., 2024) konvergieren zu der Empfehlung, sensible Bürgerinteraktionen souveränen europäischen Lösungen vorzubehalten und leistungsstarke Modelle auf die interne Arbeit zu konzentrieren.

Das vorgeschlagene Modell verdient jedoch eine kritische Einordnung. Die Begründung, Bescheide und interne Dokumente könnten Claude überantwortet werden, weil sie ohnehin im außereuropäisch kontrollierten Office-Ökosystem gespeichert werden, ist tragfähig, aber nicht voraussetzungslos. Bescheide enthalten regelmäßig personenbezogene Daten der betroffenen Bürgerinnen und Bürger; ihre Verarbeitung durch ein außereuropäisches Modell bedeutet eine zusätzliche Verarbeitungsexposition, die über die reine Speicherung hinausgeht. Das Argument der bereits gebundenen Infrastruktur mindert das marginale Souveränitätsrisiko, hebt es aber nicht vollständig auf. Das Gebot der Datenminimierung legt daher nahe, auch im Back-Office personenbezogene Inhalte zu pseudonymisieren oder zu minimieren, bevor sie an ein leistungsstarkes Modell übergeben werden (Finck & Biega, 2021).

Ein zweiter Vorbehalt betrifft die Werte- und Legitimationsdimension. Digitale Souveränität wird häufig zum Schutz europäischer Werte angeführt (Roberts et al., 2021), erfüllt aber auch eine identitätspolitische Funktion und steht im Spannungsfeld zwischen rechtlicher Kontrolle und geopolitischer Handlungsfähigkeit (Broeders et al., 2023). Für die Verwaltung folgt daraus, das Werteargument an überprüfbaren Schutzmaßnahmen zu messen, statt es symbolisch zu führen – das Zuordnungsmodell liefert hierfür einen operationalisierbaren Maßstab.

Drittens ist der ökonomische Zielkonflikt ernster zu nehmen, als es die pragmatische Lösung zunächst nahelegt. Europas Position oszilliert zwischen der Stärke seiner Regulierung – dem Brussels Effect – und der Schwäche seiner eigenen technologischen Basis (Christakis, 2020). Ein striktes Verbot leistungsstarker außereuropäischer Modelle würde die Verwaltung von Innovationsgewinnen abschneiden, während eine zu großzügige Nutzung die regulatorische Glaubwürdigkeit untergräbt. Das hier vorgeschlagene Modell positioniert sich bewusst zwischen diesen Polen: Es nutzt die Leistungsfähigkeit dort, wo das Souveränitätsrisiko ohnehin gebunden ist, und schützt jene Verarbeitungskontexte, in denen Kontrolle den höchsten Wert hat. Diese Mittelstellung ist kein Kompromiss aus Schwäche, sondern Ausdruck der Einsicht, dass Souveränität graduell und kontextabhängig herzustellen ist.

Schließlich sind Limitationen zu benennen. Der Quellenkorpus ist auf konzeptionelle, governance- und rechtsorientierte Literatur konzentriert; belastbare Vergleichsstudien zur konkreten Leistungs- und Souveränitätsdifferenz einzelner Modelle wie Claude und Mistral im Verwaltungseinsatz fehlen, und die technische Routing-Literatur stammt überwiegend aus nicht peer-reviewten Vorabveröffentlichungen (Chen et al., 2023; Ong et al., 2024). Zudem ist die Materie hochdynamisch: Rechtsrahmen und Modellfähigkeiten entwickeln sich rasch, sodass einzelne Befunde an Aktualität verlieren können. Das vorgeschlagene Zuordnungsmodell ist analytisch abgeleitet und bedarf der empirischen Validierung im realen Verwaltungsbetrieb.

8 Fazit

Die Ausgangsfrage lautete, anhand welcher Kriterien sich Use-Cases generativer KI in der öffentlichen Verwaltung leistungsstarken außereuropäischen Modellen wie Claude beziehungsweise souveränen europäischen Modellen wie Mistral zuordnen lassen, um Nutzen und digitale Souveränität auszubalancieren. Die Analyse führt zu einer klaren Antwort: Das maßgebliche Kriterium ist nicht der Anbieter, sondern die Datenklasse und der Verarbeitungskontext, gelesen vor dem Hintergrund eines dreistufigen Souveränitätsrasters aus Daten-, Infrastruktur- und Kompetenzebene.

Daraus ergibt sich ein konkretes Modell: Der unmittelbare Bürgerkontakt, in dem sensible personenbezogene Daten preisgegeben werden, ist souveränen europäischen Lösungen wie Mistral auf kontrollierter Infrastruktur vorzubehalten; die interne Verwaltungsarbeit und die Erstellung von Bescheiden kann leistungsstarken Modellen wie Claude überantwortet werden, weil die zugrunde liegende Infrastruktur ohnehin außereuropäischer Kontrolle unterliegt und der Produktivitätsgewinn dort überwiegt; besonders grundrechtssensible Daten erfordern vollständig souveräne Infrastruktur. Technisch lässt sich diese Abstufung als regelbasiertes Routing umsetzen, organisatorisch erfordert sie jedoch verlässliche Datenklassifizierung, Audit-Fähigkeit und rechtlich-vertragliche Steuerung.

Für öffentliche Institutionen ergibt sich daraus ein doppelter Handlungsauftrag. Sie sollten erstens die Kompetenz aufbauen, Daten zu klassifizieren und Souveränitätsrisiken über alle Ebenen hinweg zu steuern, statt sich auf die symbolische Sicherheit einer Anbieterwahl zu verlassen; und zweitens das Modell durch Datenminimierung auch im Back-Office absichern. Künftige Forschung sollte das hier skizzierte Zuordnungsmodell empirisch validieren, die Leistungs- und Souveränitätsdifferenzen konkreter Modelle systematisch erfassen und die ökonomischen Kosten unterschiedlicher Zuordnungsstrategien quantifizieren. So lässt sich die Balance zwischen Nutzen und Souveränität nicht symbolisch behaupten, sondern abgestuft und überprüfbar gestalten.

9 Literaturverzeichnis

Ali, T., Al-Khalidi, M., & Al-Zaidi, R. (2024). Information security risk assessment methods in cloud computing: Comprehensive review. Journal of Computer Information Systems, 66(1), 123–150. https://doi.org/10.1080/08874417.2024.2329985

Alishani, A., Homburg, V., & Velsberg, O. (2026). Public encounters and government chatbots: When servers talk to citizens. Public Administration Review, 86(1), 35–45. https://doi.org/10.1111/puar.70005

Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. Government Information Quarterly, 37(4), 101490. https://doi.org/10.1016/j.giq.2020.101490

Baur, A. (2025). European ambitions captured by American clouds: Digital sovereignty through Gaia-X? Information, Communication & Society.

Blancato, F. (2023). The cloud sovereignty nexus: How the European Union seeks to reverse strategic dependencies in its digital ecosystem. Policy & Internet.

Bondarenko, M., Lushnei, S., Paniv, Y., Molchanovsky, O., Romanyshyn, M., Filipchuk, Y., & Kiulian, A. (2025). Sovereign large language models: Advantages, strategy and regulations [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2503.04745

Bright, J., Enock, F., Esnaashari, S., Francis, J., Hashem, Y., & Morgan, D. (2025). Generative AI is already widespread in the public sector: Evidence from a survey of UK public sector professionals. Digital Government: Research and Practice, 6(1), 1–13. https://doi.org/10.1145/3700140

Broeders, D., Cristiano, F., & Kaminska, M. (2023). In search of digital sovereignty and strategic autonomy: Normative power Europe to the test of its geopolitical ambitions. JCMS: Journal of Common Market Studies.

Calderaro, A., & Blumfelde, S. (2022). Artificial intelligence and EU security: The false promise of digital sovereignty. European Security.

Carrapico, H., & Farrand, B. (2025). EU data sovereignty: An autonomy–interdependence governance gap? Politics and Governance.

Chen, L., Zaharia, M., & Zou, J. (2023). FrugalGPT: How to use large language models while reducing cost and improving performance [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2305.05176

Christakis, T. (2020). 'European digital sovereignty': Successfully navigating between the 'Brussels effect' and Europe's quest for strategic autonomy. SSRN Electronic Journal.

Finck, M., & Biega, A. J. (2021). Reviving purpose limitation and data minimisation in data-driven systems. Technology and Regulation, 2021, 44–61. https://doi.org/10.71265/z7r0t122

Fratini, S., Hine, E., Novelli, C., Roberts, H., & Floridi, L. (2024). Digital sovereignty: A descriptive analysis and a critical evaluation of existing models. Digital Society, 3(3), Article 59. https://doi.org/10.1007/s44206-024-00146-7

Gstrein, O. (2023). Data autonomy: Recalibrating strategic autonomy and digital sovereignty. SSRN Electronic Journal.

Hildén, J. (2021). Mitigating the risk of US surveillance for public sector services in the cloud. Internet Policy Review, 10(3). https://doi.org/10.14763/2021.3.1578

Hulkó, G., Kálmán, J., & Lapsánszky, A. (2025). The politics of digital sovereignty and the European Union's legislation: Navigating crises. Frontiers in Political Science.

Kim, E. (2026). Generative AI in public administration: A quasi-experimental analysis of bureaucratic productivity. Government Information Quarterly, 43(1), 102108. https://doi.org/10.1016/j.giq.2026.102108

Lindgren, I., & Jansson, G. (2013). Electronic services in the public sector: A conceptual framework. Government Information Quarterly, 30(2), 163–172. https://doi.org/10.1016/j.giq.2012.10.005

Longo, J. (2024). The transformative potential of artificial intelligence for public sector reform. Canadian Public Administration, 67(4), 495–505. https://doi.org/10.1111/capa.12587

Michels, J., Millard, C., & Walden, I. (2023). On cloud sovereignty: Should European policy favour European clouds? SSRN Electronic Journal.

Monsees, L., & Lambach, D. (2022). Digital sovereignty, geopolitical imaginaries, and the reproduction of European identity. European Security.

Nzobonimpa, S., Savard, J.-F., & Lawarée, J. (2026). Generative AI in public administration: Evaluating a fine-tuned large language model for policy briefing notes. Science and Public Policy. https://doi.org/10.1093/scipol/scag026

OECD. (2023). Initial policy considerations for generative artificial intelligence (OECD Artificial Intelligence Papers No. 1). OECD Publishing. https://doi.org/10.1787/fae2d1e6-en

Ong, I., Almahairi, A., Wu, V., Chiang, W.-L., Wu, T., Gonzalez, J. E., Kadous, M. W., & Stoica, I. (2024). RouteLLM: Learning to route LLMs with preference data [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2406.18665

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems, 34(2), Article 101885. https://doi.org/10.1016/j.jsis.2024.101885

Pohle, J., & Thiel, T. (2020). Digital sovereignty. Internet Policy Review, 9(4). https://doi.org/10.14763/2020.4.1532

Propp, K., & Swire, P. (2024). The CLOUD Act and transatlantic trust. Center for Strategic and International Studies (CSIS).

Roberts, H., Cowls, J., Casolari, F., Morley, J., Taddeo, M., & Floridi, L. (2021). Safeguarding European values with digital sovereignty: An analysis of statements and policies. Internet Policy Review.

Rone, J. (2024). ‘The sovereign cloud’ in Europe: Diverging nation state preferences and disputed institutional competences in the context of limited technological capabilities. Journal of European Public Policy.

Wang, Y.-F., Chen, Y.-C., Chien, S.-Y., & Wang, P.-J. (2024). Citizens’ trust in AI-enabled government systems. Information Polity, 29(3), 293–312. https://doi.org/10.3233/IP-230065

Wirtz, B. W., Weyerer, J. C., & Sturm, B. J. (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. International Journal of Public Administration, 43(9), 818–829. https://doi.org/10.1080/01900692.2020.1749851

Zhou, Z., Liu, D., Chen, Z., & Pancho, M. (2025). Government adoption of generative artificial intelligence and ambidextrous innovation. International Review of Economics & Finance, 98, Article 103953. https://doi.org/10.1016/j.iref.2025.103953