

06. Juli 2026

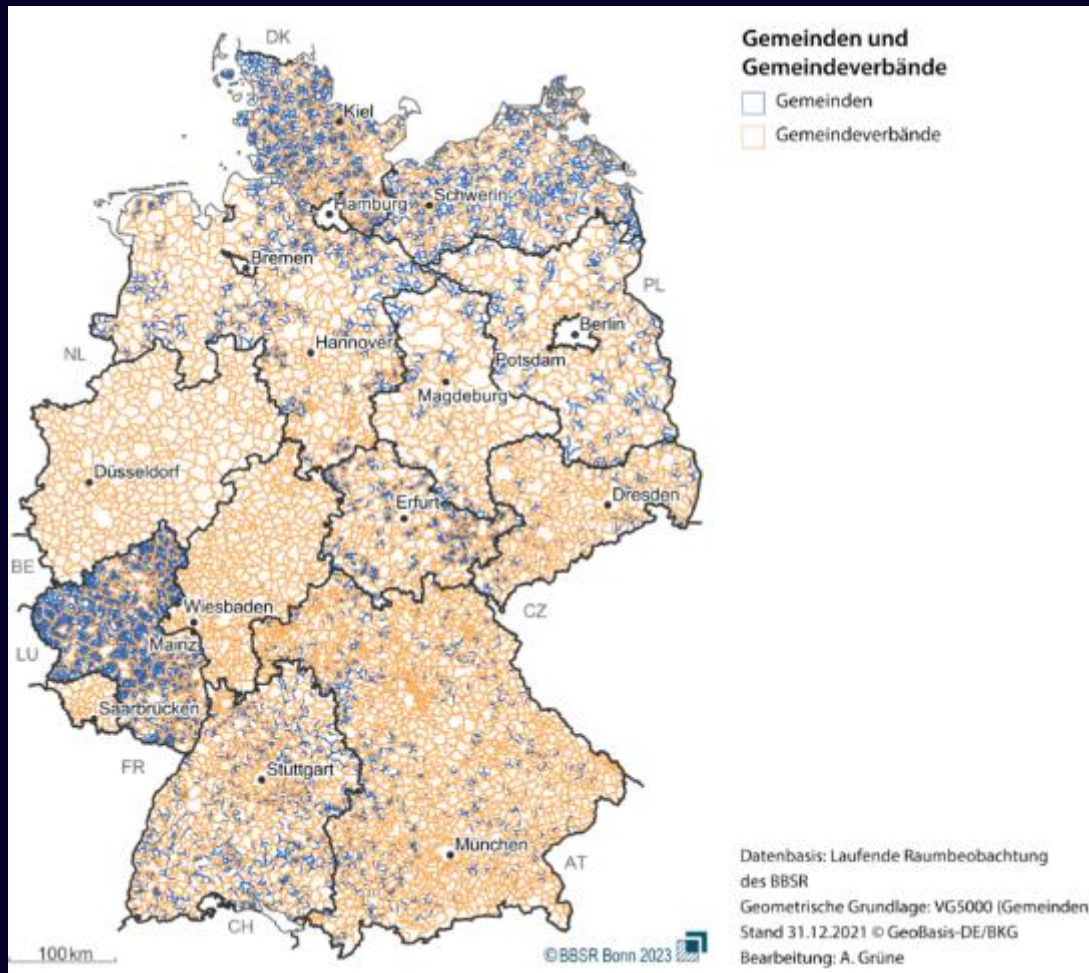
NRW.Genius & AI as a Service

IT.NRW | eGovernment-Wettbewerb





In einer fragmentierten Verwaltungslandschaft entwickeln Länder und Kommunen Lösungen in Silos statt zentral



Ausgangssituation

Fragmentierte & dezentrale Verwaltungslandschaft



Mehr als 1.000 **verschiedene Fachverfahren** im Land NRW



Dezentrale Softwarelösungen



Viele **Silo-Lösungen** und **Mehrfach-Entwicklungen**



KI ist die Schlüsseltechnologie, um langfristig die Handlungsfähigkeit der Verwaltung zu erhalten

Gründe für KI in der Verwaltung

Der **demografische Wandel** führt zu unbesetzten Stellen und Wissensverlust in der Verwaltung.

Neue Technologien und die Befähigung der Beschäftigten sichern ihre **Handlungsfähigkeit**.



Geschwindigkeit der Entwicklung



Automatisierungspotenzial



Digitalisierungsschub

 **NRW.Genius**

GenAI **sicher** & **skalierbar** in die Landesverwaltung bringen

GenAI für ALLE **zugänglich** & **nachvollziehbar** machen

GenAI zur **Beschleunigung** zeitraubender Tätigkeiten



NRW.Genius macht Künstliche Intelligenz für die Verwaltung nutzbar – sicher, transparent und praxisnah

SMARTE VERWALTUNG BEGINNT HIER.

The screenshot displays the NRW.Genius user interface. On the left is a dark sidebar with navigation options like 'Dokumente', 'Kollektionen', and 'Zuordnungen'. The main area is divided into three columns: 'Kollektionen verwalten', 'Dokumente verwalten', and 'Dokumenten-Zuordnungsbereich'. Below these are two tables: 'Offene Zuordnung' and 'Status-Upload & Pre-Processing'. The 'Offene Zuordnung' table lists documents with their protection classes and upload dates. The 'Status-Upload & Pre-Processing' table shows the progress of document uploads.

Name	Schutzbedarfsklasse	Upload-Datum	Aktionen
Richtlinien für Forschungspartnerschaften Entwicklung.pdf	Kein Schutzbedarf	22.01.26	Zuordnen
Entwurf des Landesstraßenbauprogramms 2026.pdf	Kein Schutzbedarf	22.01.26	Zuordnen

Name	Schutzbedarfsklasse	Status	Fortschritt	Aktionen
koan_2023.pdf	Kein Schutzbedarf	Pre-Processing	[Progress bar]	...
Datenschutz-Grundverordnung DSGVO.pdf	Kein Schutzbedarf	Pre-Processing	[Progress bar]	...



Main Agent & Agentic Chat Backend
 NRW.Genius erkennt Nutzeranliegen und liefert automatisch passende, verständliche Ergebnisse.



Mandantenspezifische Datenkollektionen
 Datenmanager pflegen Dokumente und stellen Kollektionen für ihre Dienststelle bereit.



On-Prem-Sprachmodelle zur sicheren Verarbeitung
 Der Betrieb von Sprachmodellen bei IT.NRW ermöglicht sichere Verarbeitung von vertraulichen und personenbezogenen Inhalten.



In der täglichen Arbeit unterstützt NRW.Genius mit verschiedenen Funktionen zahlreiche Anwendungsfälle

Freier Chat

Freie Interaktion mit dem Sprachmodell hinter NRW.Genius durch Anweisungen oder Fragen

→ Unterstützung bei kreativen Aufgaben, Ideenfindung, Beantwortung simpler Fragestellungen ohne Kontextrecherche



Einzeldokumente befragen

Fragen an eigene hochgeladene Dokumente stellen und auf Basis der Inhalte Antworten bekommen

→ z. B. Befragung eines langen Gesetzestexts nach spezifischen Details und Zusammenhängen



Rechercheassistentenz

Nutzung von durch Datenmanager Ihrer Kommune bereitgestellter Kollektionen, die NRW.Genius als Wissenskontext zur Beantwortung Ihrer Anfragen nutzt

→ Umfassende, zusammenhängende Antworten aus dem Kontext mehrerer Dokumente erhalten



Textzusammenfassung

Lange Dokumente hochladen oder Texte einfügen, um prägnante Zusammenfassungen zu erhalten

→ z. B. Zusammenfassung der wichtigsten Punkte einer langen Ausarbeitung Hinsichtlich eines Aspekts



Textgenerierung

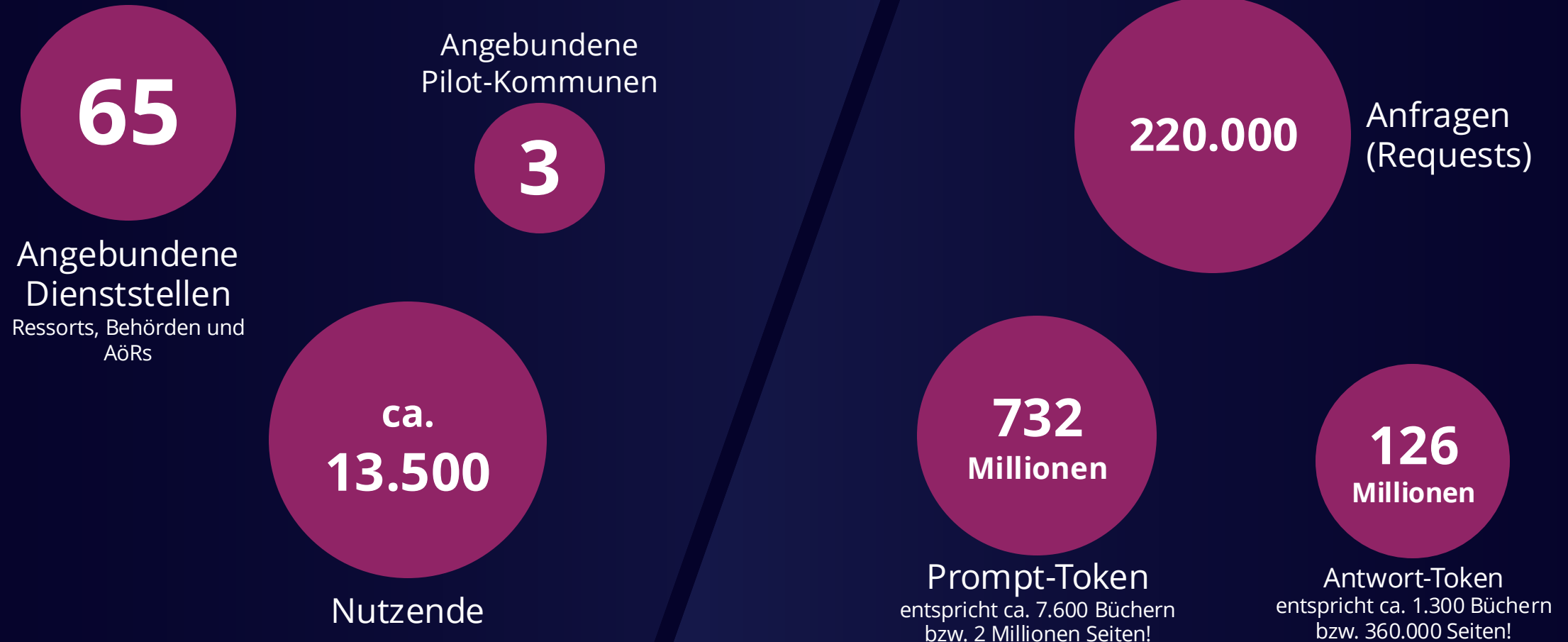
Dokumente hochladen oder Stichpunkte einfügen, um daraus Texte zu generieren (z. B. E-Mail, Protokoll, ...)

→ z. B. Erstellung von E-Mail-Antworten auf Bürgeranfragen





NRW.Genius ist bereits heute breit in der Landesverwaltung ausgerollt und bindet erste Kommunen ein





NRW.Genius ist durch die hohen Standards zur Informationssicherheit für die Verwaltung geeigneter als Cloud-KI-Anbieter

Sichere Datenverarbeitung

Die bei IT.NRW betriebenen On-Prem-Modelle sind für vertrauliche Inhalte nutzbar – ein bedeutender Mehrwert gegenüber Cloud-KI-Anbietern.

Sprachmodell-unabhängig

NRW.Genius setzt auf eine flexible und anbieterunabhängige Anbindung verschiedener Sprachmodelle – ob Cloud oder On-Prem.



Skalierbar und flexibel

NRW.Genius ist für die Nutzung in der Breite und Tiefe der Landesverwaltung und die Anbindung von Kommunen konzipiert.

Spezifische Datenquellen

Datenmanager erstellen und befüllen spezielle Datenkollektionen für die Wissensdatenbank von NRW.Genius.



NRW.Genius entwickelt sich von einer generischen Anwendung zu einer zentralen Plattform, die Fachverfahren befähigt

IST-Zustand

NRW.Genius unterstützt in der täglichen Arbeit der Verwaltung als eigenständige Anwendung.

Fragmentierte & dezentrale Verwaltungslandschaft

- ! Mehr als 1.000 verschiedene Fachverfahren im Land NRW
- ! Dezentrale Softwarelösungen
- ! Viele Silolösungen und Doppelentwicklungen

Von Anwendung zu Plattform

ZIEL-Zustand

AlaaS stellt Fachverfahren eine sichere KI-Infrastruktur über eine einheitliche API bereit.

Verfahren A

Verfahren B

Verfahren C

...nutzen KI über API

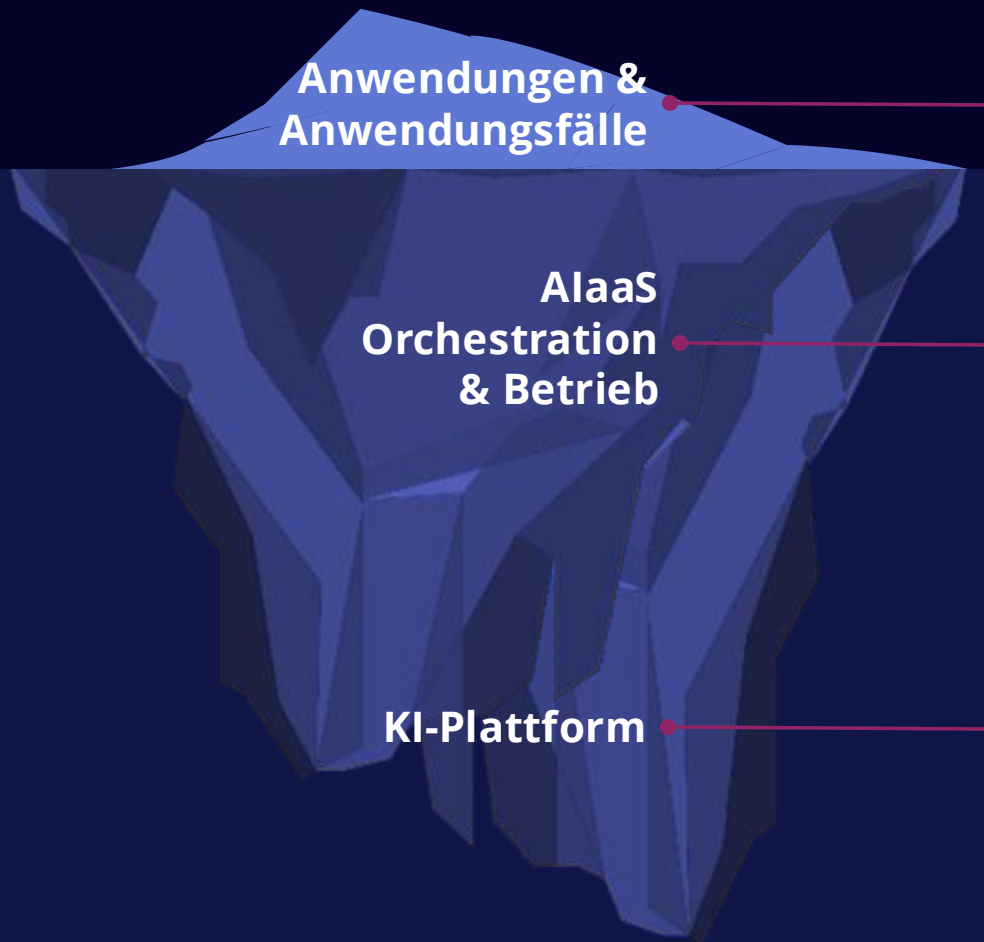
AI as a Service



Zentrale KI-Plattform



AlaaS abstrahiert die Komplexität unter der Oberfläche und kreiert Services, die nachgenutzt werden können



Anwendungen & Anwendungsfälle

- NRW.Genius App, Fachverfahren, individuelle KI-Anwendungsfälle

AlaaS Orchestration & Betrieb

- KI-Inferenz & Modellbereitstellung
- Intelligentes Routing
- Prompt-Orchestrierung
- Dynamische Performance-Optimierung Monitoring, Logging & Kostensteuerung

KI-Plattform

- Interne Entwicklungsplattform
- Tooling: CI/CD, GitOps, MLOps
- Enablement für cloud-native Entwicklung
- Integration in BIS 3.0

KI-Infrastruktur, Modelle und Betrieb werden **zentral abstrahiert.**

Entwicklungsteams definieren nur den Schutzbedarf – **AlaaS übernimmt Routing, Betrieb und Skalierung.**



Die KI-Nutzung in Fachverfahren scheitert heute mehr an Compliance und weniger am Bedarf

Realität in Fachverfahren

Ziel-Anwendungsbereiche von LLMs in Fachverfahren (Beispiele):

Antragsbearbeitung
Inhaltsklassifikation
Informationsextraktion

ABER

- Oft (hoher) **Schutzbedarf** in Fachverfahren
- Hohe **Komplexität** beim Einsatz von KI (Security, Guardrails, Monitoring, ...)
- **Unsicherheit** bei Modellauswahl & Compliance
- ...

Schutzbedarf, Compliance
& Komplexität als
zentrale Blocker

Akuter Bedarf

Gefordert ist eine Lösung, die:

- ... LLM-Nutzung für **vertrauliche Daten** ermöglicht
- ... Inferenz in geeigneter **Schutzbedarfs Umgebung** betreibt
- ... **volle Kontrolle** über Daten, Modelle und Zugriffe sicherstellt
- ... regulatorische **Anforderungen by design erfüllt**

AlaaS ermöglicht die Nutzung von KI unabhängig vom Schutzbedarf – durch zentrale Steuerung und Bereitstellung von Inferenz in geeigneten Betriebsumgebungen



Mit bereitgestellten KI-Services können Fachverfahren gezielt unterstützt werden – ohne eigene Infrastruktur

AI as a Service (AlaaS) ist eine zentrale Service-Plattform von IT.NRW, die Fachverfahren standardisiert, sicher und kostentransparenten Zugriff auf KI-Modelle bietet – unabhängig davon, ob diese On-Premises oder in Cloud-Umgebungen betrieben werden.

Zentrale Merkmale von AlaaS

- ✓ Bereitstellung freigegebener KI-Modelle
- ✓ Schutzbedarfsorientierte Verarbeitung (inkl. On-Premises-Betrieb für sensible Daten)
- ✓ Mandantentrennung und Zugriffsschutz
- ✓ Protokollierung, Monitoring & Nutzungsgrenzen
- ✓ Kosten- und Nutzungssteuerung über API-Keys
- ✓ Zentrale Security- & Governance-Aspekte

Zentrale Vorteile für Fachverfahren



Einheitlicher Zugriff auf **kuratierte KI-Modelle**



Keine eigene KI-Infrastruktur notwendig



Sicheres Routing je nach Schutzbedarf



Zentrale Steuerung von Nutzung, Kosten und Limits

AlaaS liefert die KI-Infrastruktur. Fachlogik, Entscheidungen und Prozessverantwortung verbleiben vollständig im Fachverfahren



AlaaS funktioniert wie Glasfaserausbau: Infrastruktur zentral, Integration lokal

„Infrastruktur bis zum Haus“



AlaaS / IT.NRW

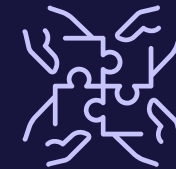
- Bereitstellung von KI-Infrastruktur
- Betrieb von Modellen (LLMs, Embeddings)
- Skalierung & Sicherheit



API-Endpoint

- Standardisierte Schnittstelle
- Zugriff über API-Key
- Nutzung über Token

„Letzter Meter im eigenen Verantwortungsbereich“



Fachverfahren / Eigentümer

- Integration in eigene Anwendung
- Umsetzung der Fachlogik
- Verantwortung für Nutzung & Use Cases



AlaaS routet KI-Inferenz abhängig vom Schutzbedarf des Fachverfahrens

Anfrage aus Fachverfahren

Anfragen werden automatisch in die geeignete Verarbeitungsumgebung geleitet – je nach Schutzbedarf des Fachverfahrens

AlaaS Routing

Schutzbedarf niedrig



Verarbeitung in **Public-Cloud-Umgebung**

Fokus: Skalierbarkeit und Kosten-Effizienz

Beispiel: Azure (Aure)

Schutzbedarf normal



Verarbeitung in **deutscher / souveräner Cloud-Umgebung**

Fokus: Balance zwischen Sicherheit und Flexibilität

Beispiel: IONOS Cloud

Schutzbedarf hoch



Verarbeitung in **hochgesicherter Umgebung**

Fokus: max. Datensouveränität und Compliance

Beispiel: IT.NRW BIS 3.0, perspektivisch Co-Locator



Möglicher AlaaS Modellkatalog – Übersicht

LLMs

Textverarbeitung & Generierung
(Small / Medium / Large)

Anwendungsfälle:

Texte, Zusammenfassungen, Q&A

Konkrete Modelle (Auszug):

Llama 3.1 8B, GPT-5, GPT-OSS-120B

Embedding Models

Semantische Vektoren

Anwendungsfälle:

Suche / Retrieval (RAG), Ähnlichkeit, Clustering

Konkrete Modelle (Auszug):

Llama 3.2 EmbedQA 1B, Lama Embed 3L

Code Models

Code & technische Inhalte

Anwendungsfälle:

Code-Generierung, Skripte, Tests

Konkrete Modelle (Auszug):
Keine

Document Processing

Dokumentenverarbeitung

Anwendungsfälle:

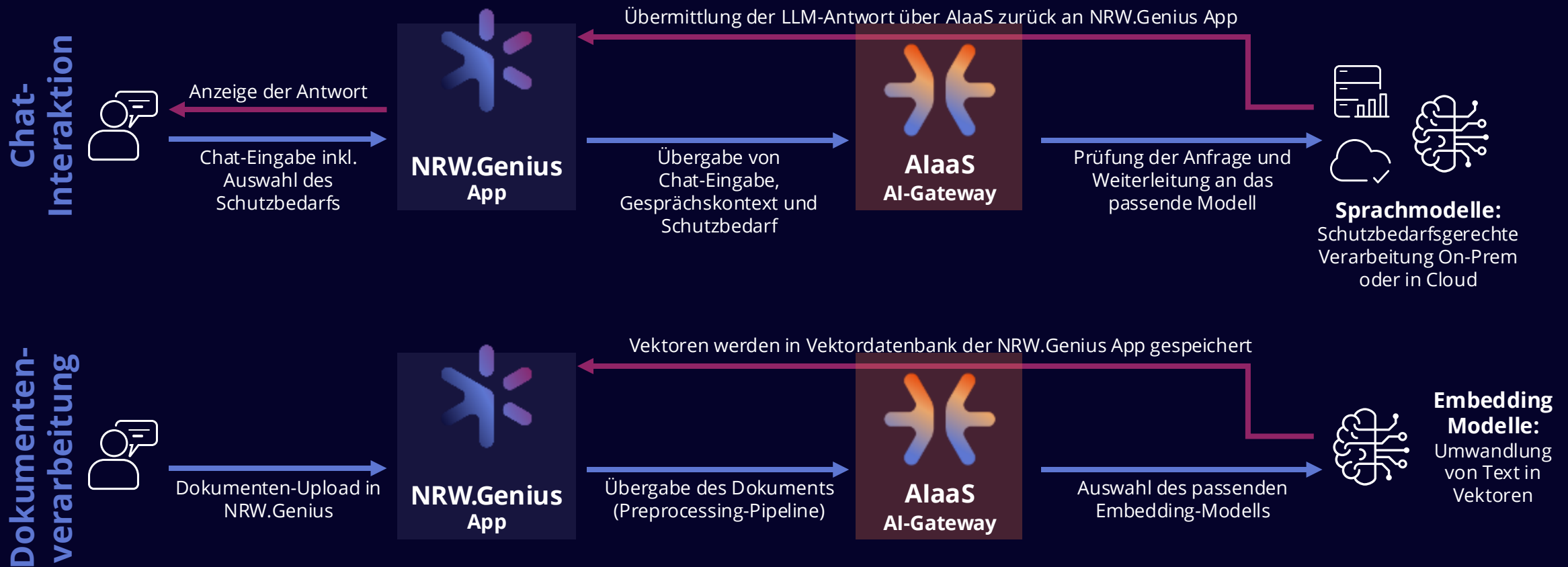
OCR / Extraktion, Struktur & Layout

Konkrete Modelle (Auszug):
PaddleX via Triton





Beispiel: NRW.Genius nutzt AlaaS für Chat-Interaktionen und Dokumenten-Verarbeitung





NRW.Genius – KI zusammen gestalten im Netzwerk aus Ländern, Kommunen und weiteren Partnern

