

Anhang 12

Glossar

Postbeamtenkrankenkasse (PBeaKK)

Stand: 16.10.2026

Dieses Glossar definiert Fachbegriffe, die in der Leistungsbeschreibung und den Anhängen verwendet werden. Es richtet sich an alle Verfahrensbeteiligten – fachliche wie technische. Im Zweifelsfall gilt die in diesem Glossar enthaltene Definition.

A

Active Directory (AD)

Verzeichnisdienst von Microsoft zur zentralen Verwaltung von Nutzenden, Gruppen, Computern und Berechtigungen in Windows-basierten Unternehmensnetzwerken. Dient in diesem Vorhaben als Grundlage für Authentifizierung und Rollenvergabe im Wissensassistenten.

Antwortqualität

Maß dafür, wie hilfreich, fachlich korrekt, quellengestützt und vollständig eine Antwort des Wissensassistenten ist. Wird im Rahmen der Qualitätssicherung anhand des Testfragenkatalogs und definierter Zielwerte (≥ 85 % hilfreiche Antworten, ≥ 90 % korrekte Quellenangabe) gemessen.

B

Barrierefreiheit

Eigenschaft eines Systems, die es Menschen mit Behinderungen ermöglicht, dieses ohne Einschränkungen zu nutzen. Im Kontext dieser Ausschreibung orientiert sich die Anforderung an WCAG 2.1 Level AA und dem Barrierefreiheitsstärkungsgesetz (BFSG). Umfasst Tastaturbedienbarkeit, Screenreader-Kompatibilität, ausreichende Kontraste und weitere Kriterien.

C

C5 (Cloud Computing Compliance Criteria Catalogue)

Prüfraahmen des Bundesamts für Sicherheit in der Informationstechnik (BSI) für die Informationssicherheit von Cloud-Diensten. Umfasst Kriterien zu Sicherheitsmanagement, Kryptographie, Kommunikationssicherheit und Datenschutz. Ein gültiges C5-Testat ist in dieser Ausschreibung Pflichtnachweis für Hosting-Anbieter, die Cloud-Dienste erbringen.

Chatbot / Wissensassistent

Im Kontext dieser Ausschreibung gleichbedeutend verwendet: ein KI-gestütztes Assistenzsystem, das Mitarbeitenden der PBeaKK auf Basis indexierter Wissensquellen Antworten auf Fachfragen liefert, Antwortentwürfe für Kundenkommunikation erstellt und bei der täglichen Arbeit unterstützt. Kein automatisiertes Entscheidungssystem.

Chunking

Verarbeitungsschritt in der Ingestion-Pipeline, bei dem lange Dokumente in kleinere, semantisch sinnvolle Abschnitte (Chunks) aufgeteilt werden, bevor sie indexiert werden. Die Qualität des Chunkings beeinflusst die Suchrelevanz und damit die Antwortqualität des Wissensassistenten maßgeblich.

Corporate Design

Visuelles Erscheinungsbild einer Organisation, bestehend aus festgelegten Gestaltungselementen wie Logo, Farbpalette, Schriftart und Bildsprache. Die Benutzeroberfläche des Wissensassistenten wird im Corporate Design der PBeaKK umgesetzt.

D

DSGVO (Datenschutz-Grundverordnung)

Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten. Zentrale Rechtsgrundlage für alle datenschutzbezogenen Anforderungen in dieser Ausschreibung.

E

Embedding

Numerische Vektordarstellung von Text (oder anderen Inhalten), die semantische Beziehungen zwischen Texten in einem mehrdimensionalen Vektorraum abbildet. Ähnliche Inhalte liegen im Vektorraum nahe beieinander. Grundlage der semantischen Suche im Wissensindex.

Embedding-Modell

Spezialisiertes KI-Modell, das Texte in Embedding-Vektoren umwandelt. Vom Sprachmodell (LLM) technisch getrennt und separat austauschbar. Änderungen am Embedding-Modell erfordern eine vollständige Neuindexierung des Wissensindex.

F

Fine-Tuning

Nachträgliches Training eines Sprachmodells auf einem domänenspezifischen Datensatz, um das Modell auf bestimmte Aufgaben oder Sprachen zu spezialisieren. Nicht verwechseln mit RAG: Fine-Tuning verändert die Modellgewichte dauerhaft; RAG erweitert den Kontext je Anfrage dynamisch.

H

Halluzination

Phänomen, bei dem ein Sprachmodell faktisch falsche, aber überzeugend formulierte Informationen generiert, die keine Entsprechung in den verfügbaren Quellen haben. RAG-Architekturen reduzieren Halluzinationen, schließen sie aber nicht vollständig aus. Kritische Halluzinationen (fachlich falsche Aussagen ohne Quellenbasis) sind im Rahmen dieser Ausschreibung Ausschlusskriterium für die Abnahme.

Human-in-the-Loop

Designprinzip, bei dem menschliche Prüfung und Entscheidung als verbindlicher Prozessschritt vorgesehen ist und KI-Systeme keine autonomen Entscheidungen treffen. Der Wissensassistent der PBeaKK ist in der ersten Ausbaustufe ausschließlich als Human-in-the-Loop-System konzipiert: Antworten und Antwortentwürfe werden von Mitarbeitenden geprüft; automatisierter Versand oder automatische Entscheidungen sind ausgeschlossen.

I

Inferenz

Prozess der Berechnung einer Modellantwort auf eine Eingabe (im Gegensatz zum Training des Modells). Im Kontext von Sprachmodellen: die Erzeugung einer Antwort auf eine Nutzeranfrage. Inferenzkosten sind nutzungsabhängige Betriebskosten des Modells.

Ingestion-Pipeline

Automatisierter Verarbeitungsprozess, der Inhalte aus Datenquellen ausliest, aufbereitet, in Chunks aufteilt, Embeddings erzeugt und in den Wissensindex schreibt. Qualität und Vollständigkeit der Ingestion-Pipeline sind entscheidend für die Vollständigkeit und Aktualität des indexierten Wissens.

ISO/IEC 27001

Internationale Norm für Informationssicherheits-Managementsysteme (ISMS). Definiert Anforderungen an Planung, Umsetzung, Betrieb, Überwachung und Verbesserung eines ISMS. Eine gültige ISO/IEC 27001-Zertifizierung des Auftragnehmers ist in dieser Ausschreibung wünschenswert.

K

KI-Verordnung (EU-KI-VO / AI Act)

Verordnung (EU) 2024/1689 zur Regulierung Künstlicher Intelligenz im europäischen Raum. Klassifiziert KI-Systeme nach Risikoklassen, legt Verbotstatbestände (Art. 5) und Anforderungen an Hochrisiko-KI fest. Der Wissensassistent der PBeaKK muss die Anforderungen der KI-VO vollständig einhalten, insbesondere die Verbotstatbestände nach Art. 5.

L

Large Language Model (LLM)

Großes Sprachmodell, das auf sehr umfangreichen Textmengen trainiert wurde und natürlichsprachliche Texte verarbeiten, verstehen und generieren kann. Wird im Wissensassistenten eingesetzt, um Fragen zu beantworten, Inhalte zusammenzufassen und Antwortentwürfe zu formulieren.

Liferay

Unternehmens-Portal- und CMS-Software (Open Source / Enterprise), die von der PBeaKK als Plattform für das Mitarbeiterportal (MAP) eingesetzt wird. Bieter müssen Erfahrung mit der Datenanbindung an Liferay (vorzugsweise über die Liferay Headless Delivery API) nachweisen.

M

Modell-Gateway

Zwischenschicht zwischen der Anwendungslogik und dem Sprachmodell. Stellt eine standardisierte, modellagnostische Schnittstelle bereit, über die verschiedene Modelle angesprochen werden können. Sorgt für Austauschbarkeit des Modells ohne Umbau der Anwendung. Zentrales Element der Plattformarchitektur im Sinne der strategischen KI-Strategie der PBeaKK.

Modellagnostizität

Eigenschaft einer Systemarchitektur, die es ermöglicht, das eingesetzte Sprachmodell zu wechseln, ohne die Anwendungslogik oder die Benutzeroberfläche zu verändern. Wird in dieser

Ausschreibung als Pflichtanforderung gefordert und muss durch das Modell-Gateway technisch realisiert werden.

MVP (Minimum Viable Product)

In diesem Kontext: der bereits existierende, funktionsfähige aber noch nicht produktionsreife Prototyp des Wissensassistenten der PBeaKK, der aktuell in einem externen Rechenzentrum betrieben wird. Der Auftragnehmer bewertet den MVP und entscheidet gemeinsam mit dem AG, welche Bestandteile übernommen und welche neu entwickelt werden.

P

PBeaKK

Postbeamtenkrankenkasse – Auftraggeber (AG) dieser Ausschreibung. Körperschaft des öffentlichen Rechts, die als Krankenversicherung für Postbeamtinnen und -beamte fungiert. Unterliegt besonderen Datenschutzerfordernissen (u. a. Sozialgeheimnis nach SGB X).

Prompt

Eingabe an ein Sprachmodell, die dessen Verhalten steuert. Unterschieden wird zwischen: (1) **Systemprompt** – vordefinierte Anweisung, die Tonalität, Grenzen und Verhalten des Modells festlegt; (2) **Nutzerprompt** – die konkrete Anfrage des Nutzers; (3) **Prompt-Vorlagen** – vordefinierte Eingabemuster, die Nutzenden als Einstiegshilfe bereitgestellt werden.

Prompt Injection

Angriffsmethode auf KI-Systeme, bei der über manipulierte Nutzereingaben oder Quelldokumente versucht wird, den Systemprompt zu überschreiben oder unerwünschtes Verhalten auszulösen. Schutzmaßnahmen gegen Prompt Injection sind in dieser Ausschreibung Pflichtanforderung (vgl. Anhang C, C.6.4 und Anhang D, D.4.5).

R

RAG (Retrieval Augmented Generation)

Architekturprinzip, bei dem ein Sprachmodell vor der Antwortgenerierung relevante Informationen aus einer externen Wissensquelle abrufen (Retrieval) und diese in den Kontext der Anfrage einbeziehen (Augmentation), bevor die Antwort generiert wird (Generation). Ermöglicht quellenbasierte, nachvollziehbare Antworten ohne dauerhaftes Modell-Training auf domänenspezifischen Daten.

S

Semantische Suche

Suchverfahren, das nicht auf der exakten Übereinstimmung von Suchbegriffen, sondern auf der inhaltlichen (semantischen) Ähnlichkeit zwischen Suchanfrage und Dokumenten basiert. Nutzt Embeddings und Vektordatenbanken. Ergänzt in hybriden Suchsystemen die klassische Volltextsuche.

Single Sign-On (SSO)

Authentifizierungsverfahren, bei dem sich Nutzende einmalig (z. B. bei der Windows-Anmeldung) authentifizieren und ohne erneute Anmeldung auf alle angebundenen Anwendungen zugreifen. In diesem Vorhaben über Active Directory realisiert.

SLA (Service Level Agreement)

Vertragliche Vereinbarung über Qualitäts- und Leistungsparameter einer IT-Dienstleistung, z. B. Verfügbarkeit, Reaktionszeiten bei Störungen und Wiederherstellungszeiten. Verbindlicher Bestandteil des Betriebsvertrags.

V

Vektor-Datenbank

Spezialisierte Datenbank zur effizienten Speicherung und Abfrage von hochdimensionalen Vektoren (Embeddings). Ermöglicht semantische Ähnlichkeitssuche (Nearest Neighbor Search). Beispiele: Weaviate, Qdrant, Chroma, pgvector (PostgreSQL-Erweiterung).

W

Wissensindex

Technische Datenstruktur, in der alle indexierten Inhalte aus den angebundenen Datenquellen als durchsuchbare Chunks (mit Embeddings und Metadaten) gespeichert sind. Bildet die Grundlage für die Suche im RAG-System.