

Vergabeverfahren

„Einführung von KI in der Eingangs- verarbeitung“

Aktenzeichen BW 04/26

Teilnahmebroschüre

Teil B.

Kurzbeschreibung der ausgeschriebenen Leistungen

Kurzbeschreibung – BW 04/26 „Einführung von KI in der Eingangsverarbeitung“

Gegenstand des Verfahrens

Der Beitragsservice von ARD, ZDF und Deutschlandradio beschafft einen KI-basierten Service zur automatisierten Verarbeitung eingehender schriftlicher Sendungen (Papier/Scan, E-Mail, Fax), der als Managed Service (IaaS, PaaS, SaaS) auf einer souveränen Cloud betrieben werden soll. Der KI-Service klassifiziert die Sendungen nach der Dokumentart-Systematik des Beitragsservice (DAK/SDAK), erkennt das fachliche Anliegen und extrahiert die für die Weiterverarbeitung erforderlichen Fachdaten. Die Ergebnisse werden jeweils mit einem Confidence-Score als strukturierte JSON-Objekte an den bestehenden Verarbeitungsprozess zurückgegeben.

Der KI-Service wird nicht als Ersatz, sondern als Bypass zur bestehenden Eingangsverarbeitung eingeführt. Die Aussteuerung der Sendungen erfolgt über vom Beitragsservice bereitgestellte intelligente Weichen; die Kernsysteme bleiben unverändert. Nicht Gegenstand der Ausschreibung sind das Scannen bzw. die Digitalisierung der Sendungen, die Ein- und Ausgangsschnittstellen der Gesamt-Eingangsverarbeitung sowie die intelligenten Weichen selbst.

Der Leistungsgegenstand gliedert sich in drei Bausteine:

Baustein 1 – Cloud-Infrastruktur: Bereitstellung einer souveränen, hochverfügbaren und skalierbaren Cloud-Umgebung als IaaS einschließlich bedarfsgerecht skalierbarer CPU-, GPU-, Speicher- und Netzwerkressourcen für KI-Workloads. Es ist beabsichtigt, einen Vertrag über die Bausteine 1 und 2 mit einer 4-jährigen Festlaufzeit abzuschließen.

Baustein 2 – Managed Service: Betrieb der KI-Plattform als PaaS und SaaS, insbesondere Monitoring, Logging, Incident- und Problem-Management, Patch- und Release-Management, Backup und Recovery, Kapazitätsmanagement sowie Bereitstellung und Betrieb eines standardisierten KI-Serving-Stacks (u. a. Container-Orchestrierung, Inferenz). Der Beitragsservice behält die Hoheit über Fachlogik, Agenten sowie Auswahl und Wechsel der KI-Modelle. Es ist beabsichtigt, einen Vertrag über die Bausteine 1 und 2 mit einer 4-jährigen Festlaufzeit abzuschließen.

Baustein 3 – Entwicklung einer individuellen Input-Management-Lösung („KI-Service“): Konzeption, Entwicklung, Integration und Produktivsetzung einer individuellen Lösung auf Basis der bereitgestellten Plattform, einschließlich Einrichtung von Klassifikation und Extraktion je Sachverhalt, Nachweis der vereinbarten Erkennungsqualität, Schulung, Dokumentation und Unterstützung der Abnahme. Der KI-Service inkl. sämtlicher Arbeitsergebnisse (u. a. Quellcode, Konfigurationen, Prompt-Artefakte, Datenmodelle, Schnittstellenbeschreibungen, Tests, Dokumentation) geht mit zeitlich, räumlich und inhaltlich unbeschränkten

Nutzungsrechten auf den Beitragsservice über. Dem Beitragsservice ist zudem das Recht einzuräumen, den KI-Service **uneingeschränkt zu nutzen, zu betreiben, zu verändern und weiterzuentwickeln** sowie durch Dritte betreiben oder weiterentwickeln lassen können. Es darf keine Abhängigkeit vom ursprünglichen Auftragnehmer entstehen und ein Anbieterwechsel muss grundsätzlich möglich sein. Die Lösung ist vorrangig auf Basis von Open-Source-Komponenten mit permissiven Lizenzen zu realisieren. Ergänzend ist ein Dienstleistungskontingent für weitere Sachverhalte vergleichbarer Komplexität und für Optimierungen nach der Produktivnahme bzw. Abnahme vorzusehen.

Fachlicher Umfang und Mengengerüst

Schwerpunkt sind rund 1,8 Mio. formlose Schreiben pro Jahr, die bislang weder automatisch klassifiziert noch inhaltlich ausgewertet werden. Hinzu kommen rund 4,7 Mio. formgebundene Schreiben aus fünf Sachverhalten: Befreiungen, Abmeldungen, Drittschuldnererklärungen, Klärungsverfahren aufgrund Sterbefall sowie Bank- und Adressänderungen. Für die Dimensionierung ist von mindestens 6,5 Mio. Sendungen pro Jahr auszugehen (ca. 25.000 Sendungen je Arbeitstag, Spitzenwert ca. 26.700). Der Handschriftanteil liegt bei ca. 20 %. Teilweise sind besondere Kategorien personenbezogener Daten nach Art. 9 DSGVO (z. B. Gesundheitsangaben) zu verarbeiten.

Wesentliche funktionale Anforderungen sind insbesondere:

- KI-gestützte Texterkennung (OCR) über den gesamten Dokumentinhalt einschließlich mehrseitiger Dokumente und Anlagen;
- inhaltsbasierte Klassifikation und Anliegenerkennung mittels Textverständnis statt Formularankern, einschließlich Erkennung mehrerer Anliegen je Sendung und Trennung mehrerer Dokumente;
- Extraktion der Identifikationsfelder (u. a. Beitragsnummer, Name, Anschrift) und der je Sachverhalt benötigten Fachdaten, auch aus beigefügten Nachweisen;
- Erkennung handschriftlicher Inhalte einschließlich freier Randnotizen, Ankreuzfelder und Prüfung auf Vorhandensein von Unterschriften;
- Confidence-Scores je Ergebnis und Feld; keine Übernahme nicht belegbarer Inhalte; Schutz vor Prompt Injection;
- integrierter Human-in-the-Loop-Pfad mit konfigurierbaren Aussteuerungsregeln und einer barrierefreien (BITV 2.0), deutschsprachigen, vollständig tastaturbedienbaren Nachbearbeitungsoberfläche;
- Aufnahme weiterer Dokumentklassen ohne Neuaufsatz des Systems durch den Beitragsservice selbst; kein Training von Modellgewichten mit personenbezogenen Daten der Beitragszahlenden;

- modellagnostische Architektur (vorzugsweise Open-Weight-Modelle), REST-Schnittstellen, Container-/Microservice-Architektur, UTF-8, Reporting (u. a. Precision/Recall/F1 je Dokumentklasse) und Anbindung an die Monitoring-Werkzeuge des Beitragsservice.

Betrieb, Datensouveränität und Sicherheit

Die Lösung ist in einer souveränen Cloud-Umgebung zu betreiben. Speicherung und Verarbeitung sämtlicher Daten haben ausschließlich in Rechenzentren innerhalb des Europäischen Wirtschaftsraums (EWR) zu erfolgen. Ausgeschlossen sind jede Datenweiterleitung in Drittstaaten sowie jeder Zugriff aus Drittstaaten, auch im Rahmen von Support und Administration, und jeder Zugriff auf Grundlage außereuropäischer Rechtsvorschriften mit extraterritorialer Wirkung (insbesondere US CLOUD Act); maßgeblich ist auch die gesellschaftsrechtliche Verflechtung des Bieters/Auftragnehmers. Die Daten sind mit kundenverwalteten Schlüsseln (HSM) zu verschlüsseln, ohne Zugriff des Cloud-Anbieters auf die Schlüssel. Ein persistenter oder administrativer Zugriff von Personal des Auftragnehmers auf Klartextdaten ist ausgeschlossen. Die Daten dürfen nicht für Training, Modellverbesserung oder Telemetrie genutzt werden. Alle an der Leistung beteiligten Stellen sind offenzulegen.

Gefordert sind u. a. ein BSI-C5-Testat für den Cloud-Dienst sowie eine ISO/IEC-27001-Zertifizierung des Rechenzentrumsbetreibers und aller datenverarbeitenden Unterauftragnehmer, die Einhaltung von DSGVO und KI-Verordnung, die Erstellung eines leistungsbezogenen Datenschutz- und Sicherheitskonzepts während der Projektphase, Prüfungsrechte des Beitragsservice sowie Vorkehrungen für einen geordneten Ausstieg und für den Ausfall des Auftragnehmers.

Service Level sind u.a.: Servicezeit Montag bis Freitag 06:00–19:00 Uhr, Verfügbarkeit $\geq 99,5$ % im Monatsmittel, Wiederherstellung bei Störungen der Priorität 1 innerhalb von 8 Stunden, deutschsprachiger Support ausschließlich aus dem EWR, 24/7-Monitoring sowie monatliche Service- und Qualitätsberichte.

Zeitraumen

Vorgesehen sind die Zuschlagserteilung Anfang 2027, die Produktivsetzung des ersten Sachverhalts spätestens Anfang 2028 und die Abbildung aller Sachverhalte über den KI-Service bis Ende 2028. Weitere Ausbaustufen bis 2030 sind nicht Gegenstand dieses Verfahrens.

Weitere Einzelheiten können der Teilnahmebroschüre sowie im weiteren Verlauf des Vergabeverfahrens den Vergabeunterlagen entnommen werden.