

Verfahren: 104/26/12 - Beschaffung einer generativen KI-Plattform

LISTE DER ÖFFENTLICHEN NACHRICHTEN

Nr.	Frage	Antwort	Gesendet
1	Es ergeben sich Differenzen aus der Leistungsbeschreibung die 2 Referenzen mit je mindestens 100 Nutzern fordert und in der Anlage R - Referenzen gibt es 3 Referenzfelder und die Angabe mindestens 200 Nutzer. Wie ist dies zu verstehen, beziehen sich die 200 Nutzer dann auf 3 Referenzen oder können die 200 Nutzer auch über mehr als 3 Referenzen verteilt sein und das Referenzblatt ist vervielfältigbar?	Die Angabe von mindestens 200 Nutzern in der Anlage R – Referenzen ist fehlerhaft. Die Anlage R wurde korrigiert und ausgetauscht. Maßgeblich ist die Leistungsbeschreibung (Eignungskriterium E2): Es sind mindestens zwei vergleichbare Referenzprojekte einzureichen, die jeweils einen Nutzerkreis von mindestens 100 Nutzern umfassen. Eine Vervielfältigung des Referenzblattes ist nicht vorgesehen; zwei ausgefüllte Referenzfelder genügen, um die Mindestanforderung zu erfüllen.	10.08.2026 08:50:00
2	Gehen wir Recht in der Annahme, dass eine SaaS Plattform betrieben auf einer deutschen Cloud wie z.B. Telekom, Stackit etc. auch eine valide Lösung darstellt, da sie alle Datenschutzvorgaben erfüllt?	Ja, eine SaaS-Plattform auf einer deutschen Cloud-Infrastruktur (z. B. Telekom, Stackit) ist grundsätzlich eine valide Lösung – allerdings nicht allein wegen des deutschen Standorts. Die Vergabeunterlagen fordern, dass Datenspeicherung und Betrieb innerhalb des Europäischen Wirtschaftsraums (EWR) erfolgen, was ein deutsches Rechenzentrum selbstverständlich erfüllt. Entscheidend ist jedoch, dass darüber hinaus weitere Anforderungen erfüllt sein müssen: DSGVO-Konformität inkl. Auftragsverarbeitungsvertrag nach Art. 28 DSGVO, Verbot der Datennutzung für Modelltrainings, Zero-Data-Retention-Konfiguration, BSI C5-Konformität sowie EU AI Act-Konformität.	10.08.2026 08:50:00
3	1. Datenverarbeitung innerhalb EU / EWR / Schweiz Gehen wir richtig in der Annahme, dass für die Anforderungen zur Datenverarbeitung die in Anlage B1 genannte Verarbeitung innerhalb des EWR oder der Schweiz maßgeblich ist? Im PoC-Testblock 7 wird demgegenüber eine Verarbeitung innerhalb der EU genannt. Bitte stellen Sie klar, welcher räumliche Geltungsbereich verbindlich ist und ob dieser gleichermaßen für Produktivdaten, Backups, Logs, Monitoring, Supportzugriffe, OCR-/Erkennungsdienste sowie externe KI-Modelle gilt. 2. Pseudonymisierung als Datenschutzmaßnahme vor Modellaufrufen Gehen wir richtig in der Annahme, dass bei der Verarbeitung personenbezogener und besonderer	Zu Frage 1 – Datenverarbeitung innerhalb EU / EWR / Schweiz: Der verbindliche räumliche Geltungsbereich ist EWR oder Schweiz, wie in Anlage B1 und übereinstimmend im PoC-Testblock 7.1 festgelegt – die dortige Formulierung „EU/EWR“ ist im Gesamtkontext als EWR (einschließlich Schweiz gemäß Anlage B1) zu verstehen. Dieser Grundsatz gilt für die Plattform und externe KI-Modelle; explizite Einzelregelungen für Backups, Logs, Monitoring, Supportzugriffe und OCR-Dienste sind in den Vergabeunterlagen nicht gesondert ausgewiesen, sodass der allgemeine Grundsatz der EWR-/Schweiz-Verarbeitung auf alle Verarbeitungsvorgänge anzuwenden ist. Zu Frage 2 – Pseudonymisierung als Datenschutzmaßnahme vor Modellaufrufen: Die Annahme, dass Pseudonymisierung oder Maskierung als Ersatz für Zero-Data-Retention	10.08.2026 15:26:26

Kategorien personenbezogener Daten technische Maßnahmen zur Datenminimierung und zum Schutz vor externen Modellaufrufen – insbesondere Pseudonymisierung oder Maskierung – als geeignete technische und organisatorische Maßnahmen im Sinne von Art. 25 und Art. 32 DSGVO eingesetzt und in Anlage B1 beschrieben werden können?

Bitte erläutern Sie, ob hierfür bestimmte technische Funktionen oder Nachweise erwartet werden.

3. Geltungsbereich von Data-Guard-/DLP-Prüfungen
Gehen wir richtig in der Annahme, dass Schutzmaßnahmen für personenbezogene und vertrauliche Inhalte nicht nur bei direkten Chat-Eingaben, sondern auch bei hochgeladenen Dokumenten, OCR-extrahierten Inhalten sowie bei aus Wissensbasen abgerufenen RAG-Kontexten angewendet werden sollen, bevor diese Inhalte an externe KI-Modelle übermittelt werden?

Bitte teilen Sie mit, ob hierfür konkrete Mindesttypen personenbezogener Daten oder vertraulicher Inhalte erwartet werden.

4. Schutzmaßnahmen bei erkannten sensiblen Daten
Gehen wir richtig in der Annahme, dass die Plattform bei erkannten personenbezogenen, besonderen oder vertraulichen Daten abhängig von zentral konfigurierbaren Richtlinien unterschiedliche Schutzmaßnahmen anwenden können soll, beispielsweise Pseudonymisierung, Maskierung, Warnung, Blockierung oder Freigabe durch berechnete Rollen?

Bitte erläutern Sie, ob und in welchem Umfang solche Maßnahmen im PoC praktisch nachzuweisen sind.

5. Schutz der Originalwerte bei externer Modellnutzung
Gehen wir richtig in der Annahme, dass eine technische Umsetzung zulässig ist, bei der personenbezogene Inhalte vor einem externen Modellaufruf durch Platzhalter ersetzt werden, die Zuordnung zu Originalwerten ausschließlich innerhalb der Plattform verbleibt und Antworten vor der Anzeige für berechnete Nutzer intern wieder re-kontextualisiert werden?

Bitte bestätigen Sie, ob dabei Originalwerte und Zuordnungsinformationen weder an externe Modellanbieter übermittelt noch in externen Requests gespeichert werden dürfen.

6. Organisationen, Kommunen und Mandantentrennung
Gehen wir richtig in der Annahme, dass der Landkreis, seine Organisationseinheiten sowie die teilnehmenden Kommunen als logisch getrennte Organisationen bzw. Mandanten betrieben werden

eingesetzt werden kann, trifft nicht zu. MUSS-Kriterium D7 verlangt ausdrücklich, dass externe KI-Modelle in einem Modus betrieben werden, in dem Eingaben und Ausgaben ausschließlich flüchtig verarbeitet und nicht dauerhaft gespeichert werden – dies ist eine Mindestanforderung. Da die Plattform ausdrücklich für die Verarbeitung personenbezogener Daten (einschließlich besonderer Kategorien wie Gesundheitsdaten) vorgesehen ist und die Ausgaben diese Daten korrekt wiedergeben müssen, würde eine vorgelagerte Pseudonymisierung, die Originalwerte vor dem Modell verbirgt, den Zweck der Plattform unterlaufen. Pseudonymisierung kann allenfalls als ergänzende TOM in Anlage B1 beschrieben werden, ersetzt jedoch ZDR nicht.

Zu Frage 3 – Geltungsbereich von Schutzmaßnahmen für verschiedene Eingabekanäle: Zero-Data-Retention (MUSS-Kriterium D7) gilt für alle Eingaben und Ausgaben, die an externe KI-Modellanbieter übermittelt werden – unabhängig davon, ob diese Inhalte aus Chat-Eingaben, hochgeladenen Dokumenten, OCR-extrahierten Inhalten oder RAG-Kontexten aus Wissensbasen stammen. Eine Differenzierung nach Eingabekanal ist in den Vergabeunterlagen nicht vorgesehen und auch nicht erforderlich, da der Schutz an der Schnittstelle zum externen Modellanbieter ansetzt und damit alle Inhalte umfasst, die diesen Weg durchlaufen. Was ist mit der Frage nach „konkrete[n] Mindesttypen personenbezogener Daten oder vertraulicher Inhalte“ konkret gemeint? Eine Beantwortung der Frage ist so nicht möglich.

Zu Frage 4 – Schutzmaßnahmen bei erkannten sensiblen Daten: Konfigurierbare Schutzmaßnahmen wie Warnung, Blockierung oder Freigabe können sinnvolle ergänzende Maßnahmen sein, dürfen aber nicht so gestaltet sein, dass personenbezogene Daten systematisch vor der Übermittlung an externe Modelle maskiert oder ersetzt werden – denn die Plattform muss Ergebnisse liefern, die die verarbeiteten personenbezogenen Daten korrekt widerspiegeln. Die verbindliche Schutzmaßnahme für externe Modellaufrufe ist und bleibt Zero-Data-Retention gemäß D7; darüberhinausgehende Maßnahmen müssen diesem Grundsatz und dem Verarbeitungszweck der Plattform entsprechen.

Zu Frage 5 – Schutz der Originalwerte bei externer Modellnutzung: Die Annahme, dass ein Ansatz mit Pseudonymisierung/Platzhaltern vor dem externen Modellaufruf und anschließender Re-Kontextualisierung eine zulässige Alternative darstellt, widerspricht den Vergabeunterlagen in zweifacher Hinsicht: Erstens ist Zero-Data-Retention (D7) als MUSS-Kriterium explizit vorgeschrieben und kann nicht durch andere Architekturen ersetzt werden. Zweitens ist die Plattform ausdrücklich

sollen und die Trennung insbesondere für Nutzer, Rollen, KI-Assistenten, Wissensbasen, Dokumente, Konversationen, Richtlinien, Modellfreigaben, Konfigurationen und Nutzungsdaten gilt?

Bitte erläutern Sie, ob Nutzende mehreren Organisationen mit jeweils unterschiedlichen Rollen und Berechtigungen zugeordnet werden können müssen.

7. Technisch bindende Rechteprüfung für Modelle und Wissensbasen

Gehen wir richtig in der Annahme, dass Berechtigungen für KI-Modelle, KI-Assistenten, externe Datenquellen, Wissensbasen und Dokumente bei jedem Zugriff technisch und serverseitig durchgesetzt werden müssen?

Dies umfasst insbesondere den Schutz davor, dass nicht berechnete Nutzende über freigegebene Assistenten, Werkzeuge oder RAG-Abfragen indirekt auf nicht freigegebene Inhalte zugreifen können.

8. Quellenangaben und Versionspriorisierung in Wissensbasen

Gehen wir richtig in der Annahme, dass die Anforderungen aus F6 sowie die PoC-Testfälle zur Wissensbasis erfüllt sind, wenn Antworten auf Basis interner Wissensquellen mit eindeutigen Verweisen auf die verwendeten Dokumente und – soweit technisch möglich – auf konkrete Textpassagen, Seiten oder Abschnitte versehen werden?

Bitte bestätigen Sie ferner, ob für die Priorisierung aktueller Dokumentfassungen eine nachvollziehbare Kennzeichnung über Metadaten, Versionsdatum, Gültigkeitsstatus oder eine administrativ festgelegte Priorisierung ausreichend ist.

9. Audit-Trail, Datenminimierung und externe Auswertung

Gehen wir richtig in der Annahme, dass ein Audit-Trail insbesondere sicherheits- und administrationsrelevante Ereignisse erfassen soll, beispielsweise Anmeldungen, fehlgeschlagene Zugriffe, Nutzer- und Rollenänderungen, Änderungen an KI-Assistenten, Wissensbasen, Modellen, Datenschutzrichtlinien sowie Lösch- und Exportvorgänge?

Bitte teilen Sie mit, ob aus Gründen der Datenminimierung eine Protokollierung von Metadaten – etwa Zeitstempel, Organisation, Rolle, Ereignistyp, Modell, Policy-Entscheidung und Status – ausreichend ist, ohne Prompt-, Antwort-, Upload- oder Originalinhalte dauerhaft in Auditlogs zu speichern.

Falls eine externe Auswertung vorgesehen ist: Können Audit- und Sicherheitsereignisse an ein externes SIEM-, Logging- oder Archivsystem in strukturierter Form, beispielsweise per API oder

dafür vorgesehen, personenbezogene Daten – auch besonderer Kategorien – vollständig und korrekt zu verarbeiten und in den Ausgaben wiederzugeben; ein Ansatz, bei dem das externe Modell die Originaldaten nie sieht, könnte diesen Zweck strukturell unterlaufen. Das geforderte Schutzniveau wird durch ZDR in Verbindung mit dem Verbot der Nutzung zu Trainingszwecken und der EWR-Verarbeitungspflicht sichergestellt.

Zu Frage 6 – Organisationen, Kommunen und Mandantentrennung: Die Vergabeunterlagen sehen vor, dass kreisangehörige Kommunen jeweils einen eigenständigen EVB-IT Cloud-Vertrag mit dem Auftragnehmer abschließen und damit separate Vertragsparteien sind. Die Plattform muss dementsprechend eine strikte logische und technische Mandantentrennung mit vollständiger Datenisolation zwischen den einzelnen Organisationen gewährleisten. Nutzende sind damit immer genau einer Organisation zugeordnet; eine gleichzeitige Zugehörigkeit zu mehreren Mandanten mit unterschiedlichen Rollen ist weder vorgesehen noch mit der geforderten Mandantenarchitektur vereinbar.

Zu Frage 7 – Technisch bindende Rechteprüfung für Modelle und Wissensbasen: Die Vergabeunterlagen – insbesondere Anlage B6 (Benutzer- und Berechtigungskonzept) – fordern ein rollenbasiertes Berechtigungskonzept mit mindestens den Rollen „Administrator“ und „Nutzer“, bei dem die IT-Abteilung des Landkreises exklusiv für die Vergabe von Zugriffsrechten auf Funktionen, KI-Assistenten und Wissensbasen verantwortlich ist. Eine explizite Anforderung, dass diese Berechtigungen bei jedem Zugriff technisch und serverseitig durchgesetzt werden müssen, ist in den Vergabeunterlagen nicht formuliert. Bieter haben die konkrete technische Umsetzung des Berechtigungskonzepts in Anlage B6 zu beschreiben.

Zu Frage 8 – Quellenangaben und Versionspriorisierung in Wissensbasen: MUSS-Kriterium F6 verlangt, dass KI-Antworten auf Basis interner Datenquellen eindeutige Quellenangaben oder Verweise auf die zugrundeliegenden Dokumente und deren Stellen enthalten; die technische Umsetzung ist in Anlage B4 zu beschreiben. Versionierung ist als SOLLTE-Kriterium (W3) eingestuft; eine nachvollziehbare Kennzeichnung über Metadaten, Versionsdatum, Gültigkeitsstatus oder administrativ festgelegte Priorisierung ist damit als Umsetzungsform grundsätzlich zulässig, sofern sie dem Ziel der Priorisierung aktueller Dokumentfassungen dient.

Zu Frage 9 – Audit-Trail, Datenminimierung und externe Auswertung: Die Vergabeunterlagen (Anlage B2, Anlage B1) fordern die Protokollierung sicherheitsrelevanter Ereignisse und administrativer

Webhook, übermittelt werden?

10. Modellorchestrierung, Zero Data Retention und PoC-Nachweise

Gehen wir richtig in der Annahme, dass die Anforderungen aus M1, M2 und M3 auch durch eine modellagnostische Plattform erfüllt werden können, welche mehrere Modelle unterschiedlicher Anbieter über eine einheitliche Administrations-, Berechtigungs- und Governance-Schicht verfügbar macht, sofern jedes Modell innerhalb des verbindlich zulässigen Verarbeitungsraums betrieben wird?

Bitte bestätigen Sie ferner, ob der Nachweis von Zero Data Retention durch eine Kombination aus technisch konfigurierter API-Nutzung ohne dauerhafte Inhaltsprotokollierung beim Modellanbieter, vertraglichen No-Training-/No-Retention -Zusagen sowie dokumentierten Datenflüssen erbracht werden kann.

Bitte erläutern Sie außerdem, ob PoC-Nachweise durch eine Kombination aus Screenshots, Konfigurationsnachweisen, Audit-/Systemprotokollen sowie Architektur-, Datenfluss- und Vertragsunterlagen erbracht werden dürfen.

Änderungen mit dem ausdrücklichen Zweck der Nachvollziehbarkeit von Systemaktivitäten. Dieser Zweck setzt voraus, dass Log-Einträge nicht nur Zeitstempel, Rolle und Ereignistyp erfassen, sondern auch hinreichende Kontextinformationen – etwa Referenzen auf betroffene Dokumente, Wissensbasen, Assistenten oder Konfigurationen – damit ein sicherheitsrelevantes Ereignis später tatsächlich nachvollzogen werden kann. Eine reine Metadatenprotokollierung ohne jeden Inhaltsbezug würde diesem Zweck nicht gerecht. Gleichzeitig gilt weiterhin der Grundsatz der Datenminimierung gemäß DSGVO, sodass eine dauerhafte Speicherung vollständiger Prompt- oder Antwortinhalte im Audit-Log nicht erforderlich ist, sofern die Nachvollziehbarkeit des Ereignisses anderweitig sichergestellt ist; Eine externe Auswertung der Audit-Trails bzw. der Ereignisprotokollierung ist derzeit nicht vorgesehen.

Zu Frage 10 – Modellorchestrierung, Zero Data Retention und PoC-Nachweise: M1–M3 fordern die Anbindung und Orchestrierung mehrerer LLMs von mindestens vier verschiedenen Anbietern über eine einheitliche Administrations- und Governance-Schicht sowie Modellunabhängigkeit ohne Vendor-Lock-in – eine modellagnostische Plattformarchitektur ist damit ausdrücklich die geforderte Grundlage. Zero Data Retention (D7) ist als flüchtige, nicht dauerhaft speichernde Verarbeitung definiert; ein Nachweis durch Kombination aus technischer API-Konfiguration, vertraglichen No-Training-/No-Retention -Zusagen und dokumentierten Datenflüssen ist sachgerecht. PoC-Nachweise sind gemäß Anlage P durch Screenshots (Prompt, Antwort, Systemzeit) sowie optional exportierte System-Logs oder Share-Links zu erbringen; ergänzende Architektur-, Datenfluss- und Vertragsunterlagen sind für spezifische Nachweise wie die EWR-Verarbeitung ausdrücklich vorgesehen.

4

FRAGE 1: BSI C5, drei unterschiedliche Fassungen

Die Anforderung an ein BSI-C5-Testat ist an drei Stellen unterschiedlich formuliert.

Leistungsbeschreibung, Abschnitt 6, D10:
"Die angebotene Plattform erfüllt die Anforderungen des BSI C5. Der Bieter legt dem Angebot ein entsprechendes C5 Testat vor. Liegt zum Zeitpunkt der Angebotsabgabe noch kein C5-Testat vor, kann der Bieter die Erfüllung durch eine mit einem akkreditierten Prüfer vereinbarte Auditplanung nachweisen, deren Abschluss innerhalb eines Jahres nach Vertragsbeginn vorgesehen ist."

Anlage M, Zeile D10:

Zu Frage 1a – Maßgebliche Fassung BSI C5: Die Formulierung in der Leistungsbeschreibung (Abschnitt 6, D10) ist die maßgebliche Fassung: Sie lässt neben einem gültigen C5-Testat ausdrücklich auch eine vereinbarte Auditplanung mit einem akkreditierten Prüfer zu, sofern der Abschluss innerhalb eines Jahres nach Vertragsbeginn vorgesehen ist. Die Formulierung in Anlage M ist ergänzend zu lesen, schränkt die Leistungsbeschreibung aber nicht ein. Anlage B2 ist als auszufüllende Dokumentationsvorlage zu verstehen, nicht als eigenständige Anforderungsquelle.

Zu Frage 1b – Anwendungsschicht vs. Cloud-Infrastruktur: Da die Anforderung auf die angebotene KI-Plattform zielt und der Geltungsbereich des Testats anzugeben ist, reicht ein C5-Testat des Cloud-Infrastrukturanbieters allein nicht aus, wenn die Anwendungsschicht des

10.08.2026 15:26:26

"Erfüllung der Anforderungen des BSI C5-Katalogs in der jeweils gültigen Fassung; Vorlage eines entsprechenden C5-Testats oder gleichwertigen Prüfnachweises."

Anlage B2, Abschnitt Zertifizierungen und Prüfberichte:

"Bitte die vorhandenen Zertifizierungen und Prüfberichte angeben (z. B. BSI C5, ISO 27001, SOC 2). Kopien sind beizufügen."

Wir bitten um Klarstellung:

a) Welche der drei Fassungen ist für die Eignungs- und Angebotsprüfung maßgeblich?

b) Bezieht sich die Anforderung auf die vom Bieter betriebene Anwendungsschicht oder auf die zugrunde liegende Cloud-Infrastruktur? Sofern die Plattform auf der Infrastruktur eines Anbieters betrieben wird, der über ein gültiges C5-Testat verfügt, und der Geltungsbereich dieses Testats die genutzten Dienste umfasst: Gilt D10 damit als erfüllt?

c) Gilt ein nach ISO/IEC 27001 zertifiziertes Informationssicherheitsmanagementsystem, dessen Geltungsbereich die Entwicklung, Bereitstellung und Betreuung der angebotenen Plattform sowie die zugehörigen Kundenservice- und Supportprozesse umfasst, als gleichwertiger Prüfnachweis im Sinne der Anlage M?

FRAGE 2: Auslegung des Begriffs Modellanbieter

Leistungsbeschreibung, Abschnitt 5.1, M1: "Die Plattform ermöglicht die Anbindung und Orchestrierung mehrerer LLMs von mindestens vier verschiedenen Modellanbietern sowie generativer Bildmodelle."

Der Begriff Modellanbieter lässt zwei Lesarten zu:

a) der Hersteller des Modells, unabhängig davon, über welchen Dienst das Modell bereitgestellt wird, oder
b) der Betreiber des Inferenzdienstes, über den das Modell angebunden ist.

Wir bitten um Klarstellung, welche Lesart gilt. Konkret: Erfüllt eine Plattform das Kriterium, wenn Modelle von vier verschiedenen Herstellern angebunden sind, diese aber über einen oder zwei Inferenzdienste innerhalb des EWR bereitgestellt werden?

Bieters nicht im Scope dieses Testats explizit enthalten ist.

Zu Frage 1c – ISO 27001 als gleichwertiger Prüfnachweis: Da D10 ein MUSS-Kriterium ist, kann eine ISO/IEC 27001-Zertifizierung – unabhängig von ihrem Geltungsbereich – nicht als ausreichend gelten. Die Leistungsbeschreibung lässt als einzige Alternative zum C5-Testat ausdrücklich nur eine vereinbarte Auditplanung mit einem akkreditierten Prüfer zu, deren Abschluss innerhalb eines Jahres nach Vertragsbeginn vorgesehen ist. Der Begriff „gleichwertiger Prüfnachweis“ in Anlage M ist im Lichte der Leistungsbeschreibung als Hauptdokument eng auszulegen und bezieht sich auf cloud-spezifische Prüfnachweise vergleichbarer Tiefe, nicht auf ein ISMS-Zertifikat.

Zu Frage 2 – Auslegung des Begriffs Modellanbieter Die Vergabeunterlagen fordern für jedes Modell die Vorlage der Modellkarte des Herstellers, was darauf hindeutet, dass mit „Modellanbieter“ der Modellhersteller gemeint ist (Variante a). Eine Plattform, die Modelle von vier verschiedenen Herstellern anbindet und diese über einen oder zwei EWR-konforme Inferenzdienste bereitstellt, erfüllt M1, sofern die EWR-Verarbeitungspflicht für jeden Dienst eingehalten wird.

Zu Frage 3 – Nutzungsumfang je Lizenz: Das Preisblatt (Anlage PB) sieht ausschließlich einen nutzerbasierten Pauschalpreis vor; Kontingente für Token, Anfragen oder Dokumentenseiten sind weder vorgesehen noch definiert. Es ist nicht zulässig, im Angebot einseitig ein Nutzungskontingent festzulegen, da dies von der Preisblattstruktur abweicht. Ein erwartetes Nutzungsvolumen je Nutzer und Monat wird nicht angegeben.

Zu Frage 4 – Gewichtung der PoC-Testblöcke: Die tatsächliche Gewichtung in der Anlage P – PoC ist durch eine Revision des Dokuments entfernt worden. Es wird eine neue Anlage P – PoC mit der tatsächlichen Gewichtung der Testblöcke bereitgestellt.

Zu Frage 5 – Benchmark-Nachweis für ein oder zwei Modelle: Die Formulierung in der Leistungsbeschreibung („mindestens eines der beiden Modelle“) ist fehlerhaft und wird korrigiert. Maßgeblich ist die folgende Anforderung: Der Bieter weist für jedes der beiden Modelle mindestens ein Ergebnis aus einer der jeweils aufgeführten Benchmarks pro Kategorie vor. Der Bieter ist dabei frei in der Wahl des konkreten Benchmarks innerhalb einer Kategorie. Eine korrigierte Fassung der Leistungsbeschreibung wird bereitgestellt.

FRAGE 3: Nutzungsumfang je Lizenz

Anlage PB sieht einen Pauschalpreis je Nutzer und Monat vor. Ein Nutzungsumfang je Lizenz, etwa in Form eines Kontingents an Anfragen, Token oder verarbeiteten Dokumentenseiten, ist in den Vergabeunterlagen nicht definiert.

Da die variablen Kosten einer generativen KI-Plattform unmittelbar vom tatsächlichen Nutzungsvolumen abhängen und die Preisstaffeln nach Anlage PB über die gesamte Laufzeit von 48 Monaten unveränderlich sind, bitten wir um Klarstellung:

a) Ist es zulässig, im Angebot einen Nutzungsumfang je Lizenz zu definieren, etwa als faire Nutzung mit einem bezifferten Kontingent?

b) Falls nein: Von welchem durchschnittlichen Nutzungsvolumen je Nutzer und Monat geht der Auftraggeber aus? Diese Angabe ist für eine belastbare Kalkulation erforderlich.

FRAGE 4: Gewichtung der PoC-Testblöcke

Leistungsbeschreibung, Abschnitt 13.3:
"Beispielhafte Gewichtung (nur zur Illustration): [...]
Die tatsächlichen Gewichtungen sind in Anlage P (PoC-Beschreibung) festgelegt."

Anlage P enthält die Beschreibung der sieben Testblöcke, jedoch keine Angaben zu deren Gewichtung. Die in der Leistungsbeschreibung genannte Gewichtung ist ausdrücklich als unverbindliche Illustration bezeichnet.

Der PoC geht mit 25 Prozent in die Gesamtwertung ein. Ohne bekannte Gewichtung der Testblöcke ist für die Bieter nicht erkennbar, worauf ein Viertel der Wertung entfällt.

Wir bitten um Bekanntgabe der tatsächlichen Gewichtungen der Testblöcke 1 bis 7 vor Ablauf der Angebotsfrist.

FRAGE 5: Benchmark-Nachweis für ein oder zwei Modelle

Leistungsbeschreibung, Abschnitt 5.2, M6,
Einleitung:
"Der Bieter legt für zwei ausgewählte Modelle, die
von unterschiedlichen
Herstellern stammen, Ergebnisse aus den
nachfolgenden Benchmark-Kategorien
vor."

Leistungsbeschreibung, M6, jeweils zu Kategorie 1,
2 und 3:
"Der Bieter weist für mindestens eines der beiden
Modelle Ergebnisse aus
einer der folgenden Benchmarks vor"

Anlage M, Zeile M6:
"Erfüllung der definierten
Benchmark-Mindestanforder ungen (z. B. MMLU,
SWE-Bench) durch mindestens zwei der für
generative Textaufgaben
eingesetzten Modelle."

Die Formulierung in Anlage M legt nahe, dass beide
Modelle sämtliche
Mindestwerte erfüllen müssen. Die
Leistungsbeschreibung verlangt dagegen je
Kategorie den Nachweis für mindestens eines der
beiden Modelle.

Wir bitten um Klarstellung, welche Anforderung gilt.

- | | | | |
|---|--|--|---------------------|
| 5 | <p>Wir haben eine Rückfrage zur Anforderung D10 - BSI C5-Konformität.</p> <p>Wir entwickeln unsere Plattform nach den C5-Kriterien, befinden uns aktuell noch nicht in der Auditplanung.
Wir sind jedoch bereit, schriftlich zuzusichern, dass wir im Fall der Zuschlagserteilung uns innerhalb eines Kalenderjahres C5 zertifizieren lassen.</p> <p>Können wir unter diesen Umständen trotzdem an der Ausschreibung teilnehmen?
Oder ist es zwingend notwendig, das wir uns bereits in den Anfängen des Audits befinden / diesen bereits abgeschlossen haben?</p> | <p>Das MUSS-Kriterium D10 erfordert zum Zeitpunkt der Angebotsabgabe entweder ein gültiges BSI C5-Testat oder eine nachweislich mit einem akkreditierten Prüfer vereinbarte Auditplanung, deren Abschluss innerhalb eines Jahres nach Vertragsbeginn vorgesehen ist. Eine schriftliche Zusicherung ohne zugrundeliegende Auditplanung ist in den Vergabeunterlagen nicht als zulässiger Nachweis vorgesehen und führt somit zum Ausschluss.</p> | 11.08.2026 15:41:26 |
| 6 | <p>Frage 1
Anlage S - Scoreboard.xlsx</p> <p>In "Anlage S - Scoreboard.xlsx" ist das Tabellenblatt mit den Anforderungen als "MUSS-Kriterien" gekennzeichnet. Bitte bestätigen sie, dass die "Anlage S - Scoreboard.xlsx" nur KANN/SOLLTE Anforderungen beinhaltet.</p> <p>Frage 2
Anlage M - MUSS-Kriterien.xlsx</p> <p>Kapitel M6
Textstelle Erfüllung der definierten
Benchmark-Mindestanforder ungen (z. B. MMLU, SWE-Bench) durch mindestens zwei der für</p> | <p>Zu Frage 1 – Tabellenblatt „MUSS-Kriterien“ in Anlage S: Es handelt sich dabei um einen Fehler. Das Tabellenblatt enthält nur SOLLTE und KANN Kriterien. Eine korrigierte Version der Anlage S wird zur Verfügung gestellt.</p> <p>Zu Frage 2 – Benchmarkdokumentation: Das von Ihnen beschriebene Vorgehen – Angabe von Modellbezeichnung, Hersteller, Benchmark und Benchmark-Version, erzieltm Ergebnis, Zeitpunkt der Durchführung, Kurzbeschreibung der vom Hersteller dokumentierten Testmethodik sowie eindeutiger Quellenangabe (z. B. offizielle Model Cards oder Technical Reports) – entspricht den Anforderungen aus M6. Maßgeblich ist die Nachvollziehbarkeit der Ergebnisse: Nicht</p> | 12.08.2026 15:55:53 |

generative Textaufgaben eingesetzten Modelle.+

Die unter M6 genannten Benchmarks werden von den Modellherstellern unter definierten und in den offiziellen Model Cards bzw. Technical Reports dokumentierten Bedingungen erhoben. Eine Wiederholung durch den Bieter würde von diesen Bedingungen zwangsläufig abweichen (z.B. hinsichtlich Agent-Scaffold, Sampling-Parameter, Anzahl der Durchläufe und verwendeter Benchmark-Teilmenge) und wäre damit weder mit den Herstellerergebnissen noch mit den Ergebnissen anderer Bieter vergleichbar. Bei einzelnen Benchmarks ist eine eigenständige Durchführung zudem nicht vollständig möglich, da Teile der Testsätze nicht öffentlich sind bzw. die Auswertung ausschließlich über die Benchmark-Betreiber erfolgt.

Wir gehen daher davon aus, dass M6 auf die nachvollziehbare und überprüfbare Dokumentation belastbarer Ergebnisse abzielt und nicht auf eine Eigenmessung des Bieters. Der Nachweis erfolgt in diesem Fall durch Angabe von Modellbezeichnung und Hersteller, Benchmark und Benchmark-Version, Ergebnis, Zeitpunkt der Durchführung bzw. Veröffentlichung gemäß Quelle, Kurzbeschreibung der vom Hersteller dokumentierten Testmethodik sowie eindeutiger Quellenangabe. Diese Angaben sind für den Auftraggeber jederzeit anhand der zitierten Quelle verifizierbar und erfüllen damit die Anforderung der Nachvollziehbarkeit.

Wir bitten um Bestätigung, dass dieses Vorgehen die Anforderungen aus M6 erfüllt.

Sofern der Auftraggeber darüber hinaus eigene Messungen des Bieters erwartet, bitten wir um Mitteilung der hierfür verbindlichen Vorgaben — insbesondere des zu verwendenden Evaluations-Harness einschließlich Version, der zulässigen Tool- und Agent-Konfiguration, der Sampling-Parameter, der Mindestanzahl an Durchläufen sowie der geforderten Form der Ergebnisdokumentation. Ohne einheitliche Vorgaben wären Eigenmessungen zwischen den Bietern nicht vergleichbar, sodass eine gleichbehandelnde Wertung anhand dieses Kriteriums nicht möglich wäre.

nachvollziehbare oder nicht verifizierbare Testergebnisse werden nicht anerkannt; der Auftraggeber behält sich vor, Originalquellen einzufordern. Die von Ihnen vorgeschlagene Dokumentation erfüllt diesen Anspruch, da die zitierten Herstellerquellen durch den Auftraggeber unmittelbar verifiziert werden können. Eigenmessungen durch den Bieter sind nach den Vergabeunterlagen nicht explizit gefordert und werden daher für die Erfüllung von M6 nicht vorausgesetzt. Einheitliche Vorgaben für eine eigenständige Durchführung (Evaluations-Harness, Sampling-Parameter etc.) sind in den Unterlagen nicht definiert und werden seitens des Auftraggebers auch nicht nachträglich festgelegt, da dies den Rahmen der vorliegenden Vergabeunterlagen überschreiten würde.