



Norddeutscher Rundfunk
Cloud-Dienstleistungen – Los 3
LEISTUNGSBESCHREIBUNG

Status: Final
Stand: 16. Juni 2026

Inhaltsverzeichnis

	Kapitel	Seite
	Einleitung.....	3
1.1	Auftraggeber	3
1.2	Ausschreibungsgegenstand	3
	Leistungsanforderungen	4
1	2.1 Allgemein	4
	2.2 Anforderungen für alle Rollen.....	4
	2.3 Einarbeitung.....	5
2	2.4 Konkrete Rollen und Anforderungen	6
	2.4.1 Senior AI Solution Architect.....	6
	2.4.2 LLM/GenAI Engineer.....	7
	2.4.3 Data Engineer	7
	2.4.4 ML/LLM Ops Engineer	9
	Serviceanforderungen im Rahmen der Projekte.....	10
3	3.1 Servicelevel/Servicezeiten	10
	Organisatorische Regeln.....	11
4	4.1 Allgemein	11
	4.2 Preisblatt.....	11

Einleitung

Auftraggeber ist der Norddeutsche Rundfunk (NDR). In diesem Dokument und im Kriterienkatalog wird dieser als AG und der Bieter als AN bezeichnet.

Die Vergabe der Leistungen erfolgt in einem europaweiten offenen Verfahren. Es findet eine Aufteilung der Ausschreibung in drei Losen statt.

1 Die Leistungsbeschreibung beschreibt die Anforderungen an Fachrollen, die für die Erbringung von Dienstleistungen in Projekten verantwortlich sind, die im Rahmen von Miniwettbewerben ausgeschrieben werden.

Dabei unterteilen sich die Lose in die folgenden Bereiche:

- ⇒ Los 1: Rollen für die Planung, Implementierung und den Service von Cloud-Infrastrukturen und Standardapplikationen
- ⇒ Los 2: Rollen für die Planung, Erstellung, Wartung, Implementierung und den Service von eigenen Cloud-Applikationen
- ⇒ Los 3: Rollen für die Planung, Erstellung, Wartung, Implementierung und den Service von eigenen AI-Applikationen und AI-Lösungen in der Cloud

Der AG beabsichtigt, je Los eine Rahmenvereinbarung mit maximal fünf AN und einer Laufzeit von 24 Monaten sowie einer Verlängerungsoption von zwei Mal 12 Monaten zu schließen. Der Abruf von Leistungen erfolgt während der Laufzeit in Miniwettbewerben.

Diese Leistungsbeschreibung beschreibt die Anforderungen an die Bieter im Los 3.

1.1 Auftraggeber

Auftraggeber (AG) ist die im Folgenden benannte Rundfunkanstalt:

Norddeutscher Rundfunk
Rothenbaumchaussee 132
20149 Hamburg

1.2 Ausschreibungsgegenstand

Der AG beabsichtigt, Services in eine Cloud zu verschieben bzw. neue Services direkt für die Cloud zu entwickeln. Dafür beabsichtigt der AG, bedarfsorientiert externe Unterstützungsleistungen in Anspruch zu nehmen. Es handelt sich hauptsächlich um beratende, designtechnische und konzeptionelle Leistungen für das Design und den Aufbau einer Cloud-Plattform und Cloud-Anwendungen basierend auf einem Multicloud-Ansatz, und weitere Leistungen, d. h. um Dienstleistungen, die nicht weisungsgebunden zu erbringen sind. Dies beinhaltet auch Programmierleistungen.

Dabei plant der AG, die Clouds der folgenden Anbieter zu verwenden:

- ⇒ AWS
- ⇒ Azure
- ⇒ Google Cloud (GCP)
- ⇒ StackIT
- ⇒ ARD-Private-Cloud (basierend auf VMware-Tanzu)

Leistungsanforderungen

2.1 Allgemein

Die AN dieser Rahmenvereinbarung sind verpflichtet, dem AG auf dessen Anforderung mittels Durchführung von Miniwettbewerben auf hohem qualitativem Niveau geeignete, qualifizierte Personen oder Gewerke anzubieten. Da die im Einzelnen anfallenden konkreten Aufgaben bzw. Projekte noch nicht absehbar sind, wird das erwartete Qualitätsniveau im Folgenden für die benötigten Rollen beschrieben.

Die für die Leistungserbringung erforderlichen Mindestanforderungen, hinsichtlich Berufsausbildung und beruflicher Qualifikation der Ausübenden, sind in Kapitel 2 detailliert beschrieben. Die für die einzelnen Rollen ggf. geforderten persönlichen Zertifizierungen sollen sicherstellen, dass ausschließlich speziell geschultes und qualifiziertes Personal eingesetzt wird. Die Anforderungen können in den durchzuführenden Miniwettbewerben vom Auftraggeber nochmals konkretisiert werden.

Des Weiteren erwartet der AG, soweit ein regelmäßiger Austausch zwischen den eingesetzten Personen und dem AG zur Erbringung der jeweiligen Leistung erforderlich ist, den Einsatz kommunikationsfähiger Personen, welche über eine ausgeprägte Sozialkompetenz verfügen und sich über eine bestimmte Zeit und zu einem bestimmten Zweck in möglicherweise bereits laufende Projekte zügig einarbeiten können.

⇒ *Anforderungen:*

- 2.1-(1) [A] Der AN benennt für die Vertragslaufzeit des Rahmenvertrages einen verantwortlichen Ansprechpartner als Single Point of Contact (SPOC).
- 2.1-(2) [A] Die benötigte persönliche Hardware für die Leistungserbringung wird vom AN für seine Mitarbeiter gestellt. Der AG stellt Zugänge für den Remote-Zugriff bereit.
- 2.1-(3) [A] Die Projektsprache ist generell deutsch. Alle eingesetzten Mitarbeiter des AN verfügen über deutsche Sprachkenntnisse in Wort und Schrift auf dem Niveau C2. Einzelne Fachgespräche können auch in Englisch erfolgen.
- 2.1-(4) [A] Der AN ist verpflichtet, den AG regelmäßig über den Projektfortschritt zu informieren. Die genauen Informationszyklen werden im jeweiligen Miniwettbewerb festgelegt.
- 2.1-(6) [A] Nach Beendigung des jeweiligen Einsatzes übergibt der AN die Arbeitsergebnisse an den AG. Die genaue Ausgestaltung/Präzisierung der zu übergebenden Artefakte erfolgt im jeweiligen Miniwettbewerb.
- 2.1-(7) [A] Eigentum an allen Arbeitsergebnissen der im Rahmen der Projekte erstellten Quelle Codes und Skripte erhält der AG.
- 2.1-(8) [A] Der AN erklärt sich generell bereit, im Rahmen von Projekten bei durch den AN entwickelten Produkten, die Produktverantwortung zu übernehmen. Die Ausgestaltung der genauen Anforderungen wird im Miniwettbewerb bekanntgegeben.
[A] Der AN erklärt sich generell bereit, im Rahmen von Projekten die Verantwortung für Projektziele zu übernehmen. Die Ausgestaltung der genauen Anforderungen wird im Miniwettbewerb bekanntgegeben.

2.2 Anforderungen für alle Rollen

Von den Mitarbeitern des Auftragnehmers erwartet der Auftraggeber eine positive, wertschätzende und authentische Grundhaltung, die sich mit dem Werterahmen des Auftraggebers deckt. Darüber hinaus müssen die bereitgestellten Mitarbeiter über die Fähigkeit zur Selbstreflexion und Selbstwahrnehmung (z. B. Kritikfähigkeit, Ehrlichkeit, Resilienz

usw.) verfügen. Weiterhin haben die Mitarbeiter eine positive berufliche Einstellung bzw. Haltung, was sich beispielsweise in Motivation, Interesse, Engagement, Flexibilität, Belastbarkeit, Verantwortungsbereitschaft, Durchhaltevermögen und Zuverlässigkeit ausdrückt.

Der AG erwartet von den eingesetzten Mitarbeitern die folgenden Sozialkompetenzen:

- ⇒ Teamfähigkeit
- ⇒ Ein hohes Maß an Kommunikationsfähigkeit
- ⇒ Ein hohes Maß an Konfliktlösungskompetenz
- ⇒ Kundenorientierung
- ⇒ Sicheres Auftreten

Die folgenden Methodenkompetenzen werden von den eingesetzten Mitarbeitern erwartet:

- ⇒ Fähigkeit, Informationen zu beschaffen, zu strukturieren, zu bearbeiten, aufzubewahren und wieder zu verwenden, darzustellen, Ergebnisse von Verarbeitungsprozessen richtig zu interpretieren und in geeigneter Form zu präsentieren
- ⇒ Fähigkeit zur Anwendung von Problemlösungstechniken
- ⇒ Fähigkeit zur Gestaltung von Problemlösungsprozessen
- ⇒ Fähigkeit zur Selbstorganisation

Die eingesetzten Personen des AN müssen über die fachliche Fähigkeit verfügen, die beauftragte Leistung in hoher Qualität und in gegebener Zeit zu erbringen. Dafür werden unter anderem die folgenden allgemeinen fachlichen Fähigkeiten gefordert:

- ⇒ Gewährleistung von Prozessen zur Qualitätssicherung
- ⇒ Identifizierung und Bewertung von Risikoindikatoren
- ⇒ Lösung von akuten Problemen und Beseitigung von Hindernissen
- ⇒ Reporting und Dokumentation von Arbeitsergebnissen der Teams (Informationsfluss in/von anderen Teams)
- ⇒ Nutzung der vom Auftraggeber vorgegebenen Plattform für kollaboratives Arbeiten (überwiegend MS Teams, Jira, Confluence) zum Austausch von gemeinsamen Arbeitsdokumenten
- ⇒ Nutzung der vom AG vorgegebenen Tools für CI/CD-Pipelines (GitLab, OpenTofu TerraForm etc.)
- ⇒ Verwendung aktueller Technologien
- ⇒ Erarbeiten von qualitativ hochwertigen Lösungen in Hinblick auf Wiederverwendbarkeit und Weiterentwicklungsmöglichkeit

2.3 Einarbeitung

Die Anforderungen und Bedingungen für die Einarbeitung werden in den Miniwettbewerben festgelegt. Generell wird davon ausgegangen, dass eine Konstanz bei den eingesetzten Mitarbeitern gegeben ist.

Grundsätzlich wird der AG eine Einweisung in die aktuellen Konfigurationen und verwendeten Methoden vornehmen. Eine Einarbeitung in die o. g. verwendeten kollaborativen Plattformen und Tools obliegt dem AN. Bei Personalwechseln seitens des AN obliegt es dem AN, diese Informationen weiterzugeben. Personalwechsel sind vom AN gegenüber dem AG mit angemessener Frist anzukündigen und abzustimmen.

2.4 Konkrete Rollen und Anforderungen

2.4.1 Senior AI Solution Architect

Ein Senior AI Solution Architect übernimmt die Konzeption, Planung und Ausgestaltung von AI-Lösungen. Er berät den AG bei der Auswahl geeigneter Cloud- und AI-Technologien und erstellt belastbare technische Lösungskonzepte für AI-Lösungen (z. B. Redaktion/Recherche, RAG/Chatbots, Speech-to-Text, Empfehlungssysteme, Video-/Audioanalyse sowie unterstützende Speziallösungen für Verwaltung und Redaktion).

Der Mitarbeiter erarbeitet die Zielarchitektur, definiert Architekturprinzipien und technische Leitplanken und stellt die konsistente Umsetzung über alle Projektphasen sicher. Er erhebt und konkretisiert Anforderungen (z. B. Skalierbarkeit, Verfügbarkeit, Latenz, Resilienz, Kosten) und berücksichtigt dabei die Vorgaben des AG zur Cloud-Governance, zu Sicherheits- und Datenschutzanforderungen sowie zu Betriebsprozessen. Der Mitarbeiter stimmt sich hierzu eng mit den beteiligten Stakeholdern, insbesondere dem Product Owner und dem Projektleiter des Auftraggebers, ab.

Der Mitarbeiter moderiert Architektur- und Anforderungsworkshops, dokumentiert Architekturentscheidungen nachvollziehbar und führt Architektur-Reviews durch. Er unterstützt die Implementierungsteams durch technische Vorgaben, Referenzarchitekturen und wiederverwendbare Patterns. Er wirkt bei der Erstellung von Betriebs- und Übergabekonzepten mit und sorgt für einen strukturierten Wissenstransfer (z. B. durch Dokumentation, Workshops, Reviews), damit der AG die Lösungen nachhaltig weiterentwickeln und betreiben kann.

⇒ *Anforderungen an einen Senior AI Solution Architect:*

- ⇒ Mindestens 4 Jahre Berufserfahrung in der Solution-Architektur komplexer IT-Systeme sowie nachweisliche Projekterfahrung in der Konzeption und Einführung von AI-Lösungen.
- ⇒ Nachweisbare Erfahrung in der Architektur von Daten- und AI-Plattformen inkl. Datenintegration sowie der Anbindung von Fachanwendungen über standardisierte Schnittstellen (z. B. REST, Messaging)
- ⇒ Fundierte Kenntnisse in Cloud-Architekturen und Betriebsmodellen (z. B. AWS/Azure/GCP), Containerisierung (z. B. Docker/Kubernetes) und Automatisierung (z. B. Infrastructure-as-Code) sowie in der Bewertung von Managed Services im Hinblick auf Sicherheit, Betrieb und Kosten.
- ⇒ Ausgeprägte Kenntnisse in Informationssicherheit und Datenschutz im Kontext AI (z. B. Datenklassifizierung, Schutzbedarfsfeststellung, Identity & Access Management, Verschlüsselung, Protokollierung, Umgang mit sensiblen Daten, Prompt-Injection/Model-Risiken) sowie Fähigkeit, entsprechende Anforderungen in technische Maßnahmen zu überführen.
- ⇒ Erfahrung mit Governance und Qualitätsmanagement für AI (z. B. Modell-/Output-Evaluation, Teststrategien, Bias-/Risiko-Betrachtungen, Nachvollziehbarkeit/Traceability, Dokumentationspflichten) und deren pragmatische Umsetzung im Projekt.
- ⇒ Fähigkeit zur Erstellung entscheidungsreifer Architekturunterlagen (u. a. Ziel-/Ist-Architektur, Lösungsdesign, Betriebs-/Rollout-Konzept) sowie Erfahrung in der Moderation von Architektur- und Stakeholder-Workshops.
- ⇒ Sehr gute Kommunikations- und Präsentationsfähigkeit, Teamfähigkeit, Konfliktlösungsfähigkeit sowie strukturierte, eigenverantwortliche Arbeitsweise.

- 2.4.1-(1) [A] Der AN verfügt mindestens über einen Mitarbeiter, der die Anforderungen an einen Senior AI Solution Architect erfüllt und für Projekte im Rahmen des Rahmenvertrages generell zur Verfügung steht.
- 2.4.1-(2) [I] Bitte geben Sie an, wie viele Mitarbeiter die Rolle „Senior AI Solution Architect“ in Ihrem Unternehmen erfüllen und einnehmen.

2.4.2 LLM/GenAI Engineer

Ein LLM/GenAI Engineer unterstützt den AG bei der Konzeption, Entwicklung und Implementierung von Anwendungen auf Basis von Large Language Models (LLM) und generativer AI. Er verantwortet die technische Umsetzung von GenAI-Use-Cases von der Anforderungsanalyse über Prototyping bis zur stabilen Integration in bestehende NDR/ARD-Systeme und Betriebsprozesse. Der LLM/GenAI Engineer stellt sicher, dass die Lösungen fachlich nutzbar, nachvollziehbar und betrieblich stabil umgesetzt werden.

Bei der Entwicklung der Anwendungen (z. B. RAG-Systeme, agentenbasierte Workflows) stellt er die Qualität der Ergebnisse durch geeignete Evaluations- und Testverfahren sicher und berücksichtigt durchgängig Anforderungen an Informationssicherheit, Datenschutz, Nachvollziehbarkeit und Kostensteuerung. Der LLM/GenAI Engineer arbeitet eng mit Solution Architecture, Data Engineering, Fachbereichen, Product Ownern sowie Security/Datenschutz zusammen, erstellt technische Konzepte und Dokumentation, unterstützt bei der Erstellung von Epics/User Stories und gewährleistet ein regelmäßiges Reporting sowie eine strukturierte, eigenverantwortliche Umsetzung in den etablierten Projekt-Tools (u. a. MS Teams, Jira, Confluence).

⇒ *Anforderungen an einen LLM/GenAI Engineer:*

- ⇒ Tiefgehende Erfahrung in der Software- und Backend-Entwicklung sowie nachweisbare Projekterfahrung in der Implementierung von LLM-/GenAI-Anwendungen.
- ⇒ Fundierte Kenntnisse in Python und/oder einer weiteren für Backend-Services geeigneten Sprache sowie Erfahrung mit API-Design.
- ⇒ Nachweisbare Erfahrung mit LLM-spezifischen Architekturmustern und Implementierungen, insbesondere Retrieval Augmented Generation (RAG) inkl. Dokumentenaufbereitung, Chunking-Strategien, Embeddings, Vektor-Datenbanken sowie Re-Ranking und Kontextsteuerung.
- ⇒ Erfahrung mit Containerisierung und Orchestrierung (z. B. Docker, Kubernetes) sowie CI/CD und Automatisierung (z. B. Git, Pipeline-Tools, Infrastructure-as-Code) zur reproduzierbaren Bereitstellung von GenAI-Services.
- ⇒ Ausgeprägtes Verständnis für Informationssicherheit und Datenschutz im GenAI-Kontext (z. B. Umgang mit sensiblen Daten, Zugriffskonzepte/IAM, Verschlüsselung, Geheimnismanagement, Protokollierung, Schutz vor Prompt-Injection/Data-Exfiltration) sowie Fähigkeit, NDR/ARD-Vorgaben in technische Maßnahmen zu überführen.

2.4.2-(1) [A] Der AN verfügt mindestens über drei Mitarbeiter, die die Anforderungen an einen LLM/GenAI Engineer erfüllen und für Projekte im Rahmen des Rahmenvertrages generell zur Verfügung stehen.

2.4.2-(2) [I] Bitte geben Sie an, wie viele Mitarbeiter die Rolle „LLM/GenAI Engineer“ in Ihrem Unternehmen erfüllen und einnehmen.

2.4.3 Data Engineer

Der Data Engineer ist verantwortlich für die Konzeption, Implementierung und Weiterentwicklung der Datenarchitekturen und Datenpipelines auf der NDR/ARD Cloud Plattform und in Microsoft Azure, die als Grundlage für AI/ML- und GenAI-Anwendungen dienen. Er stellt sicher, dass Daten aus relevanten Quellsystemen zuverlässig, qualitätsgesichert und regelkonform bereitgestellt werden.

Zu seinen Aufgaben gehört die End-to-End-Umsetzung von Datenpipelines – von der Anbindung und Integration der Quellsysteme über die automatisierte Verarbeitung bis zur Bereitstellung der Daten in geeigneten Zielspeichern. Er achtet dabei auf Skalierbarkeit, Performance, Kosten und Betriebssicherheit.

Der Data Engineer arbeitet eng mit dem Senior AI Solution Architect und dem LLM/GenAI Engineer zusammen, um Datenanforderungen für Trainings-, Evaluations- und Inferenzszenarien zu verstehen und in geeignete technische Lösungen zu überführen. Er unterstützt insbesondere beim Aufbau von Datenflüssen für RAG-Szenarien (z. B. Dokumentenaufnahme, Vorverarbeitung, Vektorisierung) sowie bei der Bereitstellung von Daten für Feature Stores und weitere AI-/GenAI-Komponenten, ohne selbst Modelltraining oder Prompt-/Modellentwicklung zu übernehmen.

⇒ *Anforderungen an einen Data Engineer:*

- ⇒ Einschlägige Berufserfahrung im Bereich „Data Engineering“, Datenplattformen oder Aufbau von Datenarchitekturen in größeren, verteilten IT-Systemlandschaften sowie nachweisliche Projekterfahrung in der Konzeption, Implementierung und dem Betrieb von Datenpipelines in produktiven Umgebungen.
- ⇒ Nachweisbare Erfahrung im Design und Aufbau von Datenarchitekturen (z. B. Data Lake, Data Warehouse, Lakehouse) inkl. Datenmodellierung, Datenintegration aus heterogenen Quellsystemen (z. B. relationale Datenbanken, Dateisysteme, REST-APIs, Messaging) sowie Bereitstellung für analytische und AI-/ML-Anwendungen.
- ⇒ Sehr gute Kenntnisse in SQL sowie in mindestens einer im Data-Engineering-Umfeld etablierten Programmiersprache (insb. Python und/oder Scala) inkl. Performance-Optimierung von Datenverarbeitungsjobs und Abfragen.
- ⇒ Erfahrung mit Cloud-Plattformen (bevorzugt Microsoft Azure) im Kontext von Data- und Analytics-Workloads (z. B. Data Lake Storage, Datenbanken, Compute-Services) sowie Bewertung und Nutzung von Managed Services hinsichtlich Skalierbarkeit, Betriebssicherheit und Kosten.
- ⇒ Ausgeprägtes Verständnis von Datenqualitäts- und Data-Governance-Anforderungen (z. B. Datenklassifizierung, Data Lineage, Metadatenmanagement, Datenqualitätsregeln) und deren Umsetzung in Datenpipelines und -plattformen.
- ⇒ Kenntnisse in Informationssicherheit und Datenschutz im Datenkontext (z. B. Schutzbedarf, Verschlüsselung, Rollen- und Rechtemodelle, Least-Privilege-Prinzip, Umgang mit personenbezogenen und sensiblen Daten) sowie Fähigkeit, entsprechende Vorgaben in technische Lösungen zu überführen.
- ⇒ Erfahrung in der Bereitstellung von Daten für AI-/ML- und GenAI-Szenarien, z. B. Aufbau von Datenflüssen für RAG-Szenarien (Dokumentenaufnahme, Vorverarbeitung, Vektorisierung) oder Versorgung von Feature Stores und Modellen mit geeigneten, qualitätsgesicherten Daten.
- ⇒ Fähigkeit zur Erstellung klar strukturierter technischer Dokumentationen (z. B. Datenflussdiagramme, Schemas, Schnittstellenbeschreibungen) sowie zur engen Zusammenarbeit mit Architekten, Data Scientists, ML Engineers und Fachbereichen.
- ⇒ Sehr gute Kommunikations- und Teamfähigkeit, strukturierte, eigenverantwortliche und lösungsorientierte Arbeitsweise sowie Bereitschaft, bestehende Prozesse und Architekturen im Hinblick auf Performance, Stabilität und Kosten kontinuierlich zu optimieren.

2.4.3-(1) [A] Der AN muss mindestens über einen Mitarbeiter verfügen, der diese Anforderungen erfüllt und für Projekte im Rahmen des Rahmenvertrag generell zur Verfügung steht.

2.4.4 ML/LLM Ops Engineer

Der ML/LLM Ops Engineer unterstützt den Auftraggeber bei der Automatisierung, dem Deployment und dem stabilen Betrieb von ML- und GenAI-Lösungen auf der NDR/ARD Cloud-Plattform und Microsoft Azure. Er stellt die durchgängige Betriebsfähigkeit über alle Umgebungen (z. B. Entwicklung, Test, Produktion) sicher und setzt die Vorgaben des Auftraggebers zu Security, Governance und Betrieb um.

Der ML/LLM Ops Engineer konzipiert und implementiert CI/CD-Pipelines sowie Automatisierungsprozesse für AI-Komponenten (z. B. Build/Train/Test/Deploy) und stellt Umgebungen reproduzierbar mittels Infrastructure-asCode bereit. Er integriert Qualitäts- und Sicherheitsprüfungen in die Delivery-Prozesse, unterstützt bei der Versionierung von Modellen, Konfigurationen und (soweit erforderlich) Prompt-/RAG-Artefakten und stellt kontrollierte Release- und Rollback-Verfahren sicher.

Der ML/LLM Ops Engineer implementiert Monitoring, Logging und Alerting für AI-Systeme (u. a. technische Kennzahlen, Latenz, Kosten, Qualitätsindikatoren/Drift) und erstellt Betriebsdokumentation und Runbooks. Er unterstützt bei Incident Management, Fehleranalyse und kontinuierlicher Optimierung der Betriebsprozesse. Der Mitarbeiter arbeitet eng mit A Engineering, Plattform-/Cloud-Teams und Security/Privacy zusammen und stellt einen kontinuierlichen Wissenstransfer sicher.

⇒ *Anforderungen an einen ML/LLM Ops Engineer:*

- ⇒ Einschlägige Berufserfahrung im Betrieb und der Architektur von Machine Learning und Large Language Model (LLM) Lösungen in produktiven Umgebungen.
- ⇒ Einschlägige Erfahrung in der Konzeption und Implementierung skalierbarer, sicherer und wartbarer MLOps/LLMOps-Architekturen inklusive Automatisierung von CI/CD-Pipelines und Modellmanagement.
- ⇒ Fundierte Kenntnisse in Cloud-Architekturen (z. B. AWS, Azure, Google Cloud), Containerisierung (Docker, Kubernetes) und Infrastructure-as-Code (Terraform, Ansible) sowie Bewertung von Managed Services hinsichtlich Sicherheit, Betrieb und Kosten.
- ⇒ Ausgeprägtes Wissen zu Informationssicherheit und Datenschutz im Kontext von KI (z. B. Datenklassifizierung, Identity & Access Management, Verschlüsselung, Protokollierung, Umgang mit sensiblen Daten, Schutz vor Exploits wie Prompt-Injection) und Fähigkeit, diese Anforderungen technisch umzusetzen.
- ⇒ Erfahrung in Governance und Qualitätsmanagement von ML/LLM-Modellen (z. B. Modell- und Output-Evaluation, Teststrategien, Bias- und Risikoanalysen, Nachvollziehbarkeit/Traceability, Dokumentationspflichten) und deren praktische Umsetzung.
- ⇒ Fähigkeit zur Erstellung entscheidungsreifer technischer Konzepte, Architekturunterlagen und Betriebs-/Rollout-Konzepte.
- ⇒ Erfahrung in der Moderation von Architektur- und Stakeholder-Workshops.
- ⇒ Sehr gute Kommunikations- und Präsentationsfähigkeiten, Teamorientierung sowie strukturierte und eigenverantwortliche Arbeitsweise.

2.4.4-(1) [A] Der AN muss mindestens über drei Mitarbeiter verfügen, die diese Anforderungen erfüllen und für Projekte im Rahmen des Rahmenvertrages generell zur Verfügung stehen.

Serviceanforderungen im Rahmen der Projekte

3.1 Servicelevel/Servicezeiten

Generell gibt es im Rahmen von Projekten verschiedene Anforderungen an den AN

⇒ Bronze:

- 08:00-16:30 Uhr, Mo-Do und
- 08:00-15:00 Uhr freitags
- keine Bereitschaften abends/nachts oder Wochenende/feiertags

⇒ Silber:

- 8-18 Uhr (an allen Werktagen außer Samstag im Bundesland Hamburg)
- keine Bereitschaften abends/nachts oder Wochenende/feiertags

⇒ Gold:

- 8-18 Uhr (an allen Werktagen außer Samstag im Bundesland Hamburg)
- Bereitschaften Wochenende/feiertags 9-23 Uhr

⇒ Platin:

- 8-18 Uhr (an allen Werktagen außer Samstag im Bundesland Hamburg)
- 24/7 Bereitschaften abends/nachts oder Wochenende / feiertags

⇒ Anforderungen:

- 3.1-(1) [A] Der AN sichert zu, dass Servicelevel „Bronze“ im Rahmen von Projekten abgedeckt werden kann.
- 3.1-(2) [B-J/N] Der AN kann zusichern, dass Servicelevel „Silber“ im Rahmen von Projekten abgedeckt wird.
- 3.1-(3) Ja = 10 Pkt., nein = 0 Pkt.
- 3.1-(4) [B-J/N] Der AN kann zusichern, dass die Zeitenrahmen von Gold im Rahmen von Projekten abgedeckt werden.
Ja = 10 Pkt., nein = 0 Pkt.
- 3.1-(5) [B-J/N] Der AN kann zusichern, dass die Zeitenrahmen von Platin, im Rahmen von Projekten abgedeckt werden.
Ja = 10 Pkt., nein = 0 Pkt.
- [A] In seltenen Ausnahmefällen muss ein Bereitschaftsdienst übernommen werden, auch wenn Servicelevel „Platin/Gold“ als nicht erfüllt beantwortet wurde.

Organisatorische Regeln

4.1 Allgemein

Im Rahmen der Miniwettbewerbe werden die Organisatorischen Regelungen für die entsprechenden Projekte vorgegeben/vereinbart.

4.2 Preisblatt

4

Im Preisblatt (PB) auf der Seite „Titelblatt“ ist der Bietername einzutragen.

Auf der Seite „Stunden“ sind die Stundensätze und die prozentualen Zuschläge für Arbeiten außerhalb der Regelarbeitszeit einzutragen.

Es wird ein Stundensatz für Remote Arbeiten und ein Stundensatz für Vor-Ort-Tätigkeiten abgefragt.

Es werden keine weiteren Nebenkosten oder Reisekosten vergütet.

Es sind für die folgenden Rollen entsprechende Stundensätze und Zuschläge einzutragen:

- ⇒ Senior AI Solution Architect
- ⇒ LLM/GenAI Engineer
- ⇒ Data Engineer
- ⇒ ML/LLM Ops Engineer

Die angegebenen Preise gelten als maximale Stundensätze für die späteren Projekte.