Leistungsbeschreibung

Miete eines KI-Servers inkl. Einrichtung, Betrieb, Wartung und IT-Sicherheitsleistungen

Vergabenummer: OeA-284-26

Auftraggeberin:
Technische Universität Berlin, Die Präsidentin
Straße des 17. Juni 135
10623 Berlin

Inhaltsverzeichnis

1 Allgemeine Leistungsbeschreibung 4
1.1 Höchstpreis 4
1.2 Besondere Anforderungen an Mitarbeiter des Auftragnehmers 4
2 Technische Mindestanforderungen Server 4
2.1 Serverplattform 5
2.2 GPU-Anforderung 5
- × NVIDIA L40S 5
2.3 Betriebssystem 6
2.4 Virtualisierung / Container 6
2.5 Einrichtung 6
3 Sicherheitsleistungen 7
3.1 Firewall 7
3.2 Web Application Firewall (WAF) 7
3.3 Endpoint Security 8
3.4 Härtung 8
3.5 Patchmanagement 8
4 Monitoring 8
5 Wartung 9
6 Backup 9
7 Service Level Agreement (SLA) 9
7.1 Reaktionszeit 9
7.2 Verfügbarkeit 9
8 Dokumentation 10
9 Datenschutz 10
10 Leistungsumfang Mietvertrag (24 Monate) 10
11 Preisblatt 11
12 Rechenzentrumsbetrieb 11
13 Zertifizierungen des Rechenzentrums 12
14 Datenhaltung 12
15 Netzwerkanbindung 12
16 Verfügbarkeit 13
17 Managed Service 13
18 Sicherheitsüberwachung (SOC) 13
19 Backup- und Wiederherstellungskonzept 14
20 Support 14
21 Leistungsumfang der Miete 14
22 Service 15
22.1 Kundendienst 15
22.2 Gewährleistung 15
23 Leistungsort 15
24 Ausschlusskriterien (Einzureichende Unterlagen) 16
25 Termine / Ausführungszeiten 16
26 Zuschlagskriterien und deren Gewichtung 17
26.1 Referenzen 18
26.2 Verfügbarkeit 18

26.3 Gesamtwertung 19
27 EVB-IT 19
27.1 Vertragliche Rahmenbedingungen 19
27.2 Vertragsgrundlage 19

1 ALLGEMEINE LEISTUNGSBESCHREIBUNG

Die Technische Universität Berlin beabsichtigt zu beschaffen die Miete eines KI-Servers inkl. Einrichtung, Betrieb, Wartung und IT-Sicherheitsleistungen.
Gegenstand der Ausschreibung ist die Bereitstellung eines hochverfügbaren Servers im Mietmodell für eine Vertragslaufzeit von 24 Monaten.
Der Auftragnehmer stellt die Hardware zur Verfügung, installiert und konfiguriert das Gesamtsystem, übernimmt die Inbetriebnahme sowie den Betrieb inklusive Wartung, Sicherheitsmanagement und Support während der gesamten Vertragslaufzeit.
Die Hardware verbleibt im Eigentum des Auftragnehmers.

1.1 HÖCHSTPREIS

Der Höchstpreis beträgt 20.000,00 Euro brutto/ 16.806,72 EUR netto und ist bindend für alle Bieter*innen/ Bietergemeinschaften. Das heißt, Bieter*innen/ Bietergemeinschaften, die einen höheren Angebotspreis einreichen, werden vom weiteren Verfahren ausgeschlossen.

Im Höchstpreis müssen alle ggf. anfallende kostenpflichtige Nebenleistungen bereits enthalten/ kalkuliert sein. Über das Angebot hinausgehende Forderungen können gegenüber der TU Berlin nicht geltend gemacht werden.

1.2 BESONDERE ANFORDERUNGEN DES AUFTRAGNEHMERS

Die angebotene Serverinfrastruktur muss für den Betrieb und die Erweiterung von ThingsBoard sowie vergleichbaren IoT-Plattformen geeignet sein und die Integration unterschiedlicher IoT-Geräte und Kommunikationsprotokolle unterstützen.
Besonders wünschenswert für die sind Erfahrungen des Anbieters mit ThingsBoard oder vergleichbaren IoT-Systemen sowie die Möglichkeit, bei technischen Fragen und Problemen im Zuge der Kundendienstbetreuung qualifizierte Unterstützung bereitgestellt zu bekommen.

2 TECHNISCHE MINDESTANFORDERUNGEN SERVER

Im Folgenden wird die geforderte Spezifikation angegeben. Alle Bestandteile und Parameter des Leistungsverzeichnisses sind von der/m Auftragnehmer*in (AN) zu erfüllen und verpflichtend, damit der vorgesehene Einsatzzweck erfüllt werden kann.
Sind die vorgegebenen Bestandteile und Parameter aus dem Angebot nicht eindeutig ersichtlich bzw. werden diese Bestandteile und Parameter nicht erfüllt oder unterschritten, wird das Angebot von der Vergabe ausgeschlossen.

Zudem hat die*der Bieter*in zu gewährleisten, dass auch dasjenige Zubehör und diejenigen Leistungen berücksichtigt werden, die zwar in der Leistungsbeschreibung nicht ausdrücklich genannt sind, die aber zur Gewährleistung eines ordnungsgemäßen, dem Beschaffungszweck entsprechenden Betriebs erforderlich sind.

2.1 SERVERPLATTFORM

Bereitstellung eines Enterprise-Rackservers der aktuellen Generation (z.B. HPE ProLiant Gen11 oder gleichwertig).

Mindestanforderungen:

Position	Menge	Mindestanforderung
CPU	2	Intel Xeon 6520P, 24 Cores, 48 Threads, 2,4 GHz, 60 MB Cache, 210 W
Arbeitsspeicher	8	32 GB DDR5-6400 ECC Registered DIMM
Gesamtspeicher RAM		mindestens 256 GB
GPU	2	NVIDIA L40S PCIe 48 GB
SSD	4	NVMe E3.S Gen5 SSD 1,6 TB Enterprise Main Performance
Gesamtspeicher SSD		mindestens 6,4 TB brutto
RAID	1	Hardware RAID Controller inkl. RAID 1/5/6/10
Netzteile	2	Redundant, Hot-Plug
Netzwerk		mindestens 2 × 25 Gbit/s Ethernet
Management		Out-of-Band Management (iLO o. gleichwertig)

2.2 GPU-ANFORDERUNG

Bereitstellung von:

- × NVIDIA L40S
- jeweils 48 GB GDDR6
- PCIe Gen4

- Passive Kühlung
- CUDA
- Tensor Cores
- RT Cores
- Unterstützung aktueller NVIDIA AI Enterprise Treiber

2.3 BETRIEBSSYSTEM

Installation eines aktuellen Enterprise Linux Betriebssystems:
z.B.

- Ubuntu Server LTS
- RedHat Enterprise Linux
- Rocky Linux

inklusive

- aktuellem Kernel
- GPU-Treiber
- CUDA Toolkit
- NVIDIA Container Toolkit

2.4 VIRTUALISIERUNG / CONTAINER

Installation und Konfiguration

- Docker
- Docker Compose
- Kubernetes optional
- NVIDIA Container Runtime

2.5 EINRICHTUNG

Der Auftragnehmer übernimmt vollständig:

- Rackmontage
- Verkabelung
- BIOS-Konfiguration
- Firmware-Updates
- RAID-Konfiguration
- Betriebssysteminstallation
- GPU-Konfiguration

- Netzwerkkonfiguration
- Benutzerverwaltung
- SSH-Härtung
- Firewall-Konfiguration
- Dokumentation
- Funktionsprüfung
- Abnahme

3 SICHERHEITSLEISTUNGEN

3.1 FIREWALL

Bereitstellung und Betrieb einer Enterprise Firewall.

Mindestens:

- Stateful Inspection
- IDS/IPS
- Geo-IP Filter
- Portfilter
- VPN
- SSL Inspection
- Logging
- VLAN-Unterstützung

3.2 WEB APPLICATION FIREWALL (WAF)

Bereitstellung einer WAF zum Schutz webbasierter KI-Anwendungen.

Mindestens:

- OWASP Top 10 Schutz
- Bot Protection
- Rate Limiting
- DDoS-Erkennung
- Signaturupdates
- virtuelle Patches

3.3 ENDPOINT SECURITY

Installation

- Malware-Schutz
- Ransomware-Erkennung
- Host Intrusion Detection
- EDR oder XDR

3.4 HÄRTUNG

Mindestens

- CIS Benchmark
- Secure SSH
- MFA für Administratoren
- Deaktivierung unnötiger Dienste
- Audit Logging

3.5 PATCHMANAGEMENT

Monatliche

- Sicherheitsupdates
- Firmwareupdates
- Betriebssystemupdates
- NVIDIA-Treiberupdates

4 MONITORING

24/7 Monitoring

mindestens

- CPU
- RAM
- SSD
- GPU
- Temperatur
- RAID
- Netzwerk
- Dienste

- Firewall
- WAF
- Alarmierung bei Störungen

5 WARTUNG

Während der gesamten Mietdauer übernimmt der Auftragnehmer

- Firmwarepflege
- Betriebssystempflege
- GPU-Treiberpflege
- Sicherheitsupdates
- Hardwareaustausch
- Fehleranalyse
- Support
- Alle Ersatzteile sind inklusive.

6 BACKUP

Einrichtung

- tägliches Backup
- Aufbewahrung mindestens 30 Tage
- Restore-Test mindestens jährlich

(optional falls Datensicherung Bestandteil des Angebots)

7 SERVICE LEVEL AGREEMENT (SLA)

7.1 REAKTIONSZEIT

Priorität	Reaktionszeit
Kritisch	max. 2 Stunden
Hoch	max. 4 Stunden
Normal	nächster Werktag

7.2 VERFÜGBARKEIT

Server:

mindestens

99,5 %

pro Kalenderjahr.

8 DOKUMENTATION

Der Auftragnehmer liefert

- Netzwerkplan
- Benutzerdokumentation
- Administratorhandbuch
- Firewall-Regelwerk
- WAF-Konfiguration
- Backup-Konzept
- Wartungskonzept
- Abnahmeprotokoll

9 DATENSCHUTZ

Der Auftragnehmer gewährleistet

- DSGVO-Konformität
- Protokollierung administrativer Zugriffe
- Verschlüsselung der Datenübertragung
- Verschlüsselung gespeicherter Daten (AES-256 oder gleichwertig)
- Abschluss eines Auftragsverarbeitungsvertrages (AVV), sofern personenbezogene Daten verarbeitet werden

10 LEISTUNGSUMFANG MIETVERTRAG (24 MONATE)

Im Mietpreis enthalten sind:

- Bereitstellung der Hardware
- Lieferung
- Installation
- Inbetriebnahme
- Konfiguration
- Betriebssystem
- GPU-Treiber
- Einrichtung
- Firewall
- Web Application Firewall

- Endpoint Security
- Monitoring
- Patchmanagement
- Firmwareupdates
- Hardwarewartung
- Ersatzteile
- Hersteller-Support
- Helpdesk
- Dokumentation
- Vor-Ort-Service bei Hardwaredefekten (nach SLA)

11 PREISBLATT

Position	Menge	Einheit
KI-Server gemäß LV	1	Stück
Einrichtung/Inbetriebnahme	1	pauschal
Firewall-Betrieb	24	Monate
WAF-Betrieb	24	Monate
Monitoring	24	Monate
Wartung	24	Monate
Patchmanagement	24	Monate
Hardwaremiete	24	Monate
Support	24	Monate
Dokumentation	1	pauschal

12 RECHENZENTRUMSBETRIEB

Der Auftragnehmer stellt den KI-Server während der gesamten Vertragslaufzeit in einem professionellen Rechenzentrum innerhalb der Bundesrepublik Deutschland bereit und betreibt diesen.

Mindestanforderungen an das Rechenzentrum

Das Rechenzentrum muss mindestens folgende Anforderungen erfüllen:

- Standort ausschließlich innerhalb Deutschlands

- Betrieb ausschließlich nach deutschem Datenschutzrecht (DSGVO und BDSG)
- Keine Speicherung oder Verarbeitung der Daten außerhalb Deutschlands
- 24×7 Betrieb
- Zutrittskontrolle mittels Mehrfaktor-Authentifizierung
- Videoüberwachung
- Einbruchmeldeanlage
- Brandfrüherkennung
- Gaslöschanlage oder gleichwertiges Löschsystem
- Redundante Stromversorgung (A-/B-Feed)
- USV-Anlagen
- Diesel-Notstromversorgung
- Redundante Klimatisierung
- Physische Trennung der Serverbereiche
- Dokumentiertes Notfallkonzept
- Besuchermanagement

13 ZERTIFIZIERUNGEN DES RECHENZENTRUMS

Das eingesetzte Rechenzentrum muss mindestens über folgende Zertifizierungen verfügen:

- ISO/IEC 27001
- ISO 9001
- ISO 22301 (Business Continuity) oder gleichwertig
- DIN EN 50600 oder Uptime Institute Tier III (wünschenswert)

Der Auftragnehmer hat die entsprechenden Nachweise auf Verlangen vorzulegen.

14 DATENHALTUNG

Es gelten folgende Mindestanforderungen:

- Sämtliche Daten werden ausschließlich in Deutschland gespeichert.
- Es erfolgt keine Spiegelung in außereuropäische Rechenzentren.
- Kein Zugriff durch Dritte außerhalb der EU.
- Der Auftragnehmer garantiert die Einhaltung der DSGVO.
- Abschluss eines Auftragsverarbeitungsvertrages (AVV) gemäß Art. 28 DSGVO.

15 NETZWERKANBINDUNG

Der KI-Server ist über eine hochverfügbare Internetanbindung bereitzustellen.

Mindestens:

- 10 Gbit/s Backbone-Anbindung des Rechenzentrums

- Redundante Carrier
- IPv4 und IPv6
- DDoS-Schutz
- BGP-Redundanz
- Ausfallsichere Internetanbindung

16 VERFÜGBARKEIT

Der Auftragnehmer gewährleistet eine technische Verfügbarkeit des Gesamtsystems von mindestens:

99,9 % pro Kalenderjahr

Hiervon ausgenommen sind angekündigte Wartungsfenster außerhalb der Geschäftszeiten.

17 MANAGED SERVICE

Während der gesamten Vertragslaufzeit übernimmt der Auftragnehmer den vollständigen Betrieb des KI-Servers.

Hierzu gehören insbesondere:

- 24×7 Infrastrukturüberwachung
- Betriebssystemadministration
- GPU-Administration
- Firmwaremanagement
- Patchmanagement
- Firewall-Administration
- WAF-Administration
- Benutzerverwaltung
- Performance-Monitoring
- Security Monitoring
- Schwachstellenmanagement
- Incident Management
- Problem Management
- Change Management
- Kapazitätsmanagement
- Dokumentation aller Änderungen

18 SICHERHEITSÜBERWACHUNG (SOC)

Der Auftragnehmer stellt ein Security Monitoring bereit.

Mindestens:

- Zentrale Protokollierung
- SIEM-Anbindung
- Angriffserkennung
- Intrusion Detection
- Security Alerts
- Alarmierung bei Sicherheitsvorfällen
- Auswertung sicherheitsrelevanter Ereignisse

Optional:

- Security Operations Center (SOC) 24×7

19 BACKUP- UND WIEDERHERSTELLUNGSKONZEPT

Der Auftragnehmer erstellt ein Backup-Konzept.

Mindestens:

- tägliche inkrementelle Datensicherung
- wöchentliche Vollsicherung
- Aufbewahrungszeit mindestens 30 Tage
- verschlüsselte Backups
- Backup ausschließlich in Deutschland
- jährlicher Restore-Test
- Dokumentation der Wiederherstellungszeiten (RTO/RPO)

20 SUPPORT

Der Auftragnehmer stellt einen deutschsprachigen Support bereit.

Mindestens:

- Service Desk
- Telefon
- E-Mail
- Ticketsystem
- Supportzeiten Montag bis Freitag 08:00–18:00 Uhr

21 LEISTUNGSUMFANG DER MIETE

Die Miete umfasst sämtliche Kosten für:

- Bereitstellung des KI-Servers
- Hosting im zertifizierten deutschen Rechenzentrum

- Stromversorgung
- Klimatisierung
- Rack-Stellfläche
- Internetanbindung
- Firewall
- Web Application Firewall (WAF)
- DDoS-Schutz
- Monitoring
- Security Monitoring
- Patchmanagement
- Firmwareupdates
- Hardwarewartung
- Ersatzteile
- Herstellersupport
- Betriebssystempflege
- NVIDIA-Treiberpflege
- Dokumentation
- Datensicherung
- Managed Service
- Support gemäß SLA
- Rückbau und sichere Datenlöschung nach Vertragsende gemäß BSI-Empfehlungen

22 SERVICE

22.1 KUNDENDIENST

Es muss ein technischer Kundendienst in deutscher oder/und englischer Sprache mit fachlich-inhaltlichem Feedback innerhalb von 3 Werktagen vorhanden sein. Die Annahme von Störungsmeldungen und Reklamationen muss werktags in der Zeit von 9:00 Uhr bis 15:00 Uhr telefonisch oder per E-Mail (Reaktionszeit bis zum nächsten Werktag) gewährleistet sein.

22.2 GEWÄHRLEISTUNG

Es muss auf das Gesamtsystem und alle Komponenten eine Gewährleistung von 24 Monaten durch den AN erfolgen. Die Gewährleistung beginnt mit dem Tag der Abnahme des Systems durch die AG.

23 LEISTUNGSORT

24 AUSSCHLUSSKRITERIEN (EINZUREICHENDE UNTERLAGEN)

Neben den vom Unternehmen einzureichenden Formularen (siehe Vergabeunterlagen) sind folgende Punkte im Angebot schriftlich darzustellen. Werden die folgenden Punkte nicht schriftlich dargestellt, so wird das Angebot ausgeschlossen.

- **TUB – Wirt-Erklärungen (EK[1]):** Bitte reichen Sie diese Dokumente ausgefüllt mit dem Angebot ein:
    - Wirt-124
    - EVB-IT Cloudvertrag

Sofern zutreffend:

- Wirt-235
- Wirt-236
- Wirt-238

- **Spezifikationsdokumente (EK):** Das Angebot muss Spezifikationsdokumente zum angebotenen Gerät (z. B. technische Beschreibung, Datenblätter, Herstellerangaben) enthalten, aus denen eindeutig hervorgeht, dass die Anforderungen gemäß Abschnitt 2.1 vollständig erfüllt werden.

25 TERMINE / AUSFÜHRUNGSZEITEN

Ausführungsbeginn: Die Ausführung kann ab dem Tag der Auftragsvergabe erfolgen, spätestens jedoch zum 01.01.2027.
Ausführungsende: Fertigstellung nach 24 Monaten

Die vereinbarten Lieferfristen sind verbindlich. Liefer- und Leistungsverzögerungen sind der TU Berlin unverzüglich anzuzeigen. Dies gilt dann, wenn es auf Grund von erheblichen und unvorhersehbaren Umständen politischen und/oder wirtschaftlichen Ausmaßes, zu Lieferengpässen bzw. Störungen in den Lieferketten kommt, die nicht im Verantwortungsbereich des AN liegen. Mit der zeitnahen Unterrichtung der TU Berlin hierüber, sind die zu diesem Zeitpunkt voraussichtlichen Lieferzeiten mitzuteilen.

[1] EK = Eignungskriterium

26 ZUSCHLAGSKRITERIEN UND DEREN GEWICHTUNG

Unter Berücksichtigung aller Umstände wird der Zuschlag jeweils auf das wirtschaftlichste Angebot erteilt. Die Bewertung erfolgt auf Grundlage des schriftlichen Angebots. Daher liegt es im Interesse des Bieters, alle angeforderten Informationen so detailliert und korrekt wie möglich zur Verfügung zu stellen.

Zuschlagskriterium ist zu 45,00% der Gesamtangebotspreis, zu 40,00% die Referenzen und zu 15% der die Verfügbarkeit nach Projektfortschritt Die sich aus dem jeweiligen Kriterium ergebenden Punkte werden addiert. Der Bieter mit der Höchstpunktzahl erhält den Zuschlag. Alle Zahlenwerte werden auf zwei Nachkommastellen gerundet.

Kriterium	**Gewichtung**	**Punkte**
Preis	45,00%	45,00
Referenzen	40,00%	40,00
Leistungsbeginn nach Zuschlagserteilung	15,00%	15,00
Gesamt	**100,00%**	**100,00 Punkte**

Kriterium Preis

Das Angebot mit dem niedrigsten Gesamtpreis (inkl. MwSt.) erhält die Höchstpunktezahl von 45 Punkten. Ist der Gesamtpreis eines Angebotes z.B. um 100% höher im Vergleich zum Angebot mit dem niedrigsten Gesamtpreis, dann erhält dieses Angebot 0 Punkte. Null Punkte stellen die Untergrenze dar, d.h. es werden keine Negativpunkte vergeben. Die Punkte dazwischen werden nach der folgenden Gleichung interpolar berechnet:

Formel:

$$Punkte_{preis} = \left(2 - \frac{Preis}{Preis_{min}}\right) * 45$$

Beispielrechnung zur Punktevergabe:

Angebot 1: Preis von 19.000 EUR = 36,56 Punkte

Angebot 2: Preis von 18.000 EUR = 39,38 Punkte

Angebot 3: Preis von 16.000 EUR = 45,00 Punkte

26.1 REFERENZEN

Bewertet werden die vom Bieter vorgelegten Referenzprojekte, die hinsichtlich Art, Umfang und technischer Anforderungen mit dem ausgeschriebenen Leistungsgegenstand vergleichbar sind. Maximal können hier 40 Punkte erzielt werden.

Kriterium und Beschreibung	Maximale Punktzahl
Referenzen	
Referenzliste mit Angaben über vergleichbare Referenzprojekte **durch Ihr Unternehmen** in den letzten 5 Kalenderjahren Je Referenz müssen mindestens folgende Angaben konkret benannt werden: • Betrieb und Vermietung von KI-/GPU-Serverinfrastrukturen • Einrichtung und Konfiguration von KI-Servern • Betrieb, Wartung und technischer Support • IT-Sicherheitsleistungen • Verarbeitung großer Datenmengen • Erfahrungen mit ThingsBoard oder vergleichbaren IoT-Systemen • Erfahrung mit kommunalen bzw. öffentlichen Auftraggebern, insbesondere im Bereich öffentliche Beleuchtung, sofern vorhanden. Maximal können hier 40 Punkte erzielt werden	
Referenzliste mit **einer** Referenz, die die oben genannten Anforderungen erfüllen	**20**
Referenzliste mit **zwei** Referenzen, die die oben genannten Anforderungen erfüllen	**30**
Referenzliste mit **drei** Referenzen, die die oben genannten Anforderungen erfüllen	**40**

26.2 VERFÜGBARKEIT

Kriterium und Beschreibung	Maximale Punktzahl
Leistungsbeginn	

Kriterium und Beschreibung	Maximale Punktzahl
Bewertung des Leistungsbeginns nach Zuschlagserteilung Maximal können hier 15 Punkte erzielt werden	
Leistungsbeginns sofort, max. 7 Tage nach Zuschlagserteilung sichergestellt.	**15**
Leistungsbeginn ab 01.01.2027	**0**

26.3 GESAMTWERTUNG

Die bei den einzelnen Zuschlagskriterien erreichten Punkte werden addiert.
Das Angebot mit der höchsten Gesamtpunktzahl erhält den Zuschlag.
Bei gleicher Gesamtpunktzahl findet die in den Vergabeunterlagen festgelegte Regelung für den Gleichstand Anwendung.

27 EVB-IT

27.1 VERTRAGLICHE RAHMENBEDINGUNGEN

Der Vertragsabschluss erfolgt auf Basis des EVB-IT-Cloudvertrag.

27.2 VERTRAGSGRUNDLAGE

Die Leistung wird auf Grundlage des EVB-IT Cloudvertrag vergeben. Der EVB-IT - Cloudvertrag ist Bestandteil der Vergabeunterlagen und grundsätzlich verbindlich anzuwenden.
Abweichungen hiervon sind nur in begründeten Ausnahmefällen zulässig. In diesem Fall hat der Bieter im Angebot eine nachvollziehbare und umfassende Begründung vorzulegen, aus der sich ergibt, warum eine Anwendung des EVB-IT Cloudvertrag im konkreten Fall nicht möglich ist.