

51.31 – 2026-20 Leistungsbeschreibung für einen GPU Server der Leistungsklasse NVIDIA DGX B200 GPU einschl. Lieferung und Inbetriebnahme.

Einleitung

Gegenstand der Ausschreibung ist die Lieferung frei Verwendungsstelle, Installation, Inbetriebnahme und betriebsbereite Übergabe eines KI-Hochleistungservers der Leistungsklasse NVIDIA DGX B200 oder eines technisch gleichwertigen Systems (HGX B200 o. ä.) einschließlich Herstellergarantie und Installationsleistungen. Der Server wird für rechenintensive KI-Aufgaben (Post-Training großer Sprachmodelle) eingesetzt. Im Fokus stehen Weiterentwicklung (CUDA-/NCCL-/PyTorch-Stack), Fine-Tuning und Erweiterung der im Projekt SOOFI entwickelten souveränen Sprachmodelle (Soofi-S 30B und 120B) inkl. industrieller Adaption im DAISEC-Umfeld. Erforderlich sind sehr großer GPU-Speicher (>1 TB gesamt) und hohe Rechenleistung. Der Server wird im vorhandenen Rechencluster integriert.

Bei der Ermittlung des wirtschaftlich günstigsten Angebots werden nachstehende Kriterien angewendet. Mit dem Buchstaben F werden Festanforderungen gekennzeichnet. Die Nichterfüllung einer als Festanforderung gekennzeichneten Anforderung führt zum Ausschluss des Angebotes (KO Kriterium). Mit den Buchstaben G und L gekennzeichnet sind Bewertungskriterien, die mit 0 bis 10 Punkten bewertet und dann gewichtet werden. Das Feld *Bieterangaben* ist vollständig auszufüllen. Wenn Felder „ja/nein“ enthalten, ist hier die Angabe anzukreuzen, die angeboten wird.

Leistungsort

Leibniz Universität Hannover
Forschungszentrum L3S
Serverraum
Appelstr. 9A
30167 Hannover

Rechnungsadresse

Leibniz Universität Hannover
Forschungszentrum L3S
Appelstr. 4
30167 Hannover

Kontaktstellen

Eine Kommunikation darf nur elektronisch über die Vergabeplattform „www.dtv.de“ erfolgen.

Allgemeine Bedingungen

Es gelten die allgemeinen Einkaufsbedingungen der Gottfried Wilhelm Leibniz Universität Hannover. Weiterhin ist das Formular „Erklärung Bezug zu Russland“ zu beachten und mit einzureichen.

Sprache

Angebot/Unterlagen sowie Schriftverkehr: deutsch. Vertrags-/Verhandlungssprache: deutsch.

Gerichtsstand

Hannover

Zahlungsbedingungen

Anzahlungen nur gegen unbefristete, für den Auftraggeber kostenfreie Bankbürgschaft.

Garantie (Mindestanforderungen)

Garantie mind. 24 Monate für alle H/W-Komponenten.

Verlängerung der Garantie auf 36 bzw. 60 Monate, abrufbar bis zum Ende der laufenden Garantie.

Service (Mindestumfang)

Integration des Gesamtsystems (voll funktionsfähig)

Durchführung eines Systemtests vorab

Lieferung/Zusendung des Gesamtsystems mit Aufbau und Inbetriebnahme

Preisstellung

Bruttopreise (inkl. MwSt.). Umfasst das lauffähige Gesamtsystem inkl. GPUs, CPUs, Netzwerkkomponenten, Speicher/Arbeitsspeicher, allgemeine Systemanforderungen, sonstige Anforderungen, Lieferung frei Verwendungsstelle, Verpackung einschließlich Aufbau und Inbetriebnahme.

Lizenzen für NVIDIA AI Enterprise sind nicht Gegenstand der Ausschreibung; sofern im Angebot enthalten, sind sie separat auszuweisen.

Anforderungen im Detail

A	GPU	Bieterangaben - vom Bieter auszufüllen -	Erläuterung	Kriterium
A1	Min. 8 GPUs	() ja / () nein		F
A2	SXM / OAM auf gemeinsamem Baseboard	() ja / () nein		F
A3	All-to-all GPU interconnect (Full Mesh, z.B. NVSwitch)	() ja / () nein		F
A4	Mind. 180 GB HBM3e je GPU	() ja / () nein		F

B	CPU	Bieterangaben - vom Bieter auszufüllen -	Erläuterung	Kriterium
B1	2 CPUs	() ja / () nein		F
B2	Mind. 224 Threads (2x 56 Kerne)	() ja / () nein	Mit 2 CPUs insgesamt mind. 224 Threads	F
B3	X86-64 / AMD64 - Architektur	() ja / () nein		F

C	Netzwerk	Bieterangaben - vom Bieter auszufüllen -	Erläuterung	Kriterium
C1	2 x 10 GbE – RJ45	() ja / () nein		F
C2	1 x 100 Gb/s dual-port	() ja / () nein		F
C3	BMC 1GbE RJ45 interface	() ja / () nein		F
C4	8 x 400Gb/s Ports	() ja / () nein		F
C5	2 x BlueField-3 DPU	() ja / () nein		F

D	Arbeitsspeicher und Speicher	Bieterangaben - vom Bieter auszufüllen -	Erläuterung	Kriterium
D1	2TB DDR5 ECC System-RAM	() ja / () nein		F
D2	2 x 1.9 TB System SSDs im RAID 1	() ja / () nein		F
D3	8 x 3.84TB NVMe U.2 (Local Data)	() ja / () nein		F

Option: Ausführung mit 4 TB DDR5 ECC System-RAM anstelle von 2 TB. Der Preis ist separat auszuweisen. Diese Position ist **optional** und **nicht wertungsrelevant**. Der Auftraggeber behält sich vor, diese Variante zu beauftragen.

E	Allgemeine Systemanforderungen	Bieterangaben - vom Bieter auszufüllen -	Erläuterung	Kriterium
E0	NVIDIA B200 Architektur	() ja / () nein		F
E1	19“- Rackmontageschiene	() ja / () nein		F
E2	Luftgekühlter Betrieb im 19-Zoll-Rack ohne Flüssigkeitskühlung. Dauerbetrieb unter Maximallast (8 GPUs, 2 CPUs) stabil	() ja / () nein		F
E3	Maximalverbrauch unter Volllast < 16 kW	() ja / () nein		F
E4	Gehäusehöhe max. 10 HE	() ja / () nein		F
E5	Gehäusetiefe max. 1000 mm	() ja / () nein		F
E6	Linux mit Mainline-Kernel lauffähig	() ja / () nein		F
E7	Betriebsbereites Server-Gesamtsystem	() ja / () nein		F
E8	Vorab Systemtest: NCCL-All-Reduce ≥ 90 % der Nominalbandbreite, ein anhaltender 8-GPU-BF16/FP8-GEMM-Lauf und 24 h Burn-in ohne thermisches Throttling	() ja / () nein		F

Garantie und herstellerseitiger Enterprise-Support

Für das angebotene System ist ein herstellerseitiger Enterprise-Support bereitzustellen, der hinsichtlich Leistungsumfang und Servicequalität dem Standard-Support führender Hersteller integrierter KI-Hochleistungssysteme entspricht.

G	Garantie und Herstellerseitiger Enterprise-Support	Bieterangaben - vom Bieter auszufüllen -	Erläuterung	Kriterium
G1	Mindestlaufzeit, mind. 24 Monate	Zeitraum in Monaten:		G

G2	Verlängerung der Garantie auf 36 bzw. 60 Monate, abrufbar bis zum Ende der laufenden Garantie	☐ ja / ☐ nein		F
G3	Herstellergarantie für alle H/W-Komponenten inkl. Vor-Ort-Reparatur/-Austausch	☐ ja / ☐ nein		F
G4	Direkter technischer Support durch Hersteller oder autorisierten, zertifizierten Servicepartner	☐ ja / ☐ nein		F
G5	Remote-Support für Fehlerdiagnose/Störungsbehebung	☐ ja / ☐ nein		F
G6	Firmware-, BIOS-, BMC- und systemrelevante Software-Updates während der gesamten Supportlaufzeit	☐ ja / ☐ nein		F
G7	RMA-Abwicklung inkl. Austausch defekter Komponenten gemäß Herstellerprozessen:	☐ ja / ☐ nein		F
G8	Zugang zu technischer Doku, Wissensdatenbanken, Support-Portalen, Diagnosewerkzeugen	☐ ja / ☐ nein		F
G9	Unterstützung für sämtliche integrierten Hardwarekomponenten des gelieferten Systems.	☐ ja / ☐ nein		F
G10	Leistungen werden durch Hersteller selbst oder autorisierten Servicepartner	☐ ja / ☐ nein		F

	erbracht:			
G11	Austausch defekter Datenträger ohne Rücksendung	() ja / () nein	Im Garantiefall verbleiben defekte Datenträger im Institut und werden selbst vernichtet	F

Der Bieter hat auf Anforderung nachzuweisen, dass die angebotenen Garantie und Supportleistungen den vorstehenden Anforderungen entsprechen.

S	Sonstiges	Bieterangaben - vom Bieter auszufüllen -	Erläuterung	Kriterium
S1	Lieferzeit	Lieferzeit ab Zuschlagserteilung in Wochen:		L

Bewertung:

Voraussetzung: Alle F-Kriterien „ja“ und G1 ≥ 24 Monate.

Lfd. Nr.	Kriterium	Bewertung
1	Preis (Gewichtung: 80 %) Alle Preise sind in Euro anzugeben. Als maßgeblich wird der Gesamtpreis des Angebots, einschließlich aller Transport-, Ausstattungs-, Support- und Garantieoptionen gesehen. Laufende Kosten beispielsweise aus Dienstleistungen zu Support- und Garantieoptionen sind unzulässig bzw. sind im Zusammenhang mit dem Lieferauftrag im Vorhinein zu kontingentieren. Das preisgünstigste Angebot erhält die volle Punktzahl (10 Punkte). Ein Angebot, dessen Gesamtpreis das 1,25-fache des niedrigsten Angebotspreises erreicht oder überschreitet, erhält 0 Punkte. Dazwischen wird linear interpoliert: $P = 10 \times (1 - (\text{Angebotspreis} - \text{niedrigster Angebotspreis}) / (0,25 \times \text{niedrigster Angebotspreis}))$, mindestens 0 Punkte.	Preiswertung Gewichtete Punktzahl
2	Liefertermin (Gewichtung: 10 %) Als maßgeblich und verbindlich wird der zugesicherte Liefertermin, Kriterium L, an den Hauptlieferort ab möglicher Auftragserteilung gesehen. Berücksichtigt werden nur Angebote, deren Liefertermin am oder vor dem 31.07.2027 liegt. Alle Liefertermine, die innerhalb von 36 Wochen nach dem Zuschlagstermin liegen, werden linear mit maximal 10 Punkte bewertet (10 Punkte bei weniger als 12 Wochen, 0 Punkte bei 36 Wochen) . Danach angekündigte Liefertermine bekommen 0 Punkte.	Lieferzeitwertung Gewichtete Punktzahl
3	Garantie (Gewichtung: 10 %) Kriterium G wird wie folgt bewertet. Eine Laufzeit für die Garantie von 24 Monaten bekommt 0 Punkte, 36 Monaten oder mehr bekommen 10 Punkten. Eine verlängerte Garantie muss als separater Punkt ausgewiesen sein. Die zusätzlichen Kosten für diese Garantieverlängerung müssen 0 € betragen.	Garantie Gewichtete Punktzahl

Abnahme

Abnahme nach erfolgreicher Inbetriebnahme und bestandenen Tests gem. E8 (inkl. Protokolle: NCCL All-Reduce, GEMM, Burn-in; Nachweis keine Throttling-Events; Netzwerktests C2/C4; IPMI/BMC-Zugriff).