AARONALPHAX

AI Compute & Liquid-Cooling Deployment

Technical Proposal & Statement of Work · 500 kW Modular Inference Facility (NVIDIA B300 → Rubin) · Ref. **[SOW-AIC-___]** · Date **[DD Month YYYY]**

SAMPLE PROPOSAL — ILLUSTRATIVE TEMPLATE, NOT AN EXECUTED CONTRACT. All bracketed fields, fees, capacities and dates are placeholders to be agreed in writing.

1. Parties

Service Provider	Client
Guangzhou Xinqiao Technology Development Co., Ltd. (trading as AaronAlphaX) Website: aaronalphax.com Contact: sales@aaronalphax.com Registered address: **[Address]**	**[Client legal name]** Registered address: **[Address]** Company / registration no.: **[No.]** Primary contact: **[Name, title, email]**

2. Client Situation (as understood)

The Client plans an AI inference and LLM research environment for about **[500]** active users. Facility capacity is planned at **500 kW** over a 2–3 year horizon. The facility will not be fully populated on Day 1: it starts with NVIDIA B300-class systems and is planned to take Rubin-generation systems later. The Client is evaluating modular data-center vendors and needs to:

- avoid stranded or locked-in power and cooling capacity as rack densities rise to **120–140 kW per rack**;
- choose a power architecture that does not close off 800 VDC distribution later;
- keep efficiency acceptable at low Day-1 utilisation (about 10–15% of design load);
- keep options open for colocation and for API-based capacity while the on-site build ramps.

3. Proposed Approach (summary)

AaronAlphaX acts as the Client's **owner-side technical advisor and capacity partner**. We define the target architecture, run the vendor process, and secure interim and overflow capacity, so that users are served from Day 1 while the facility is built in phases. Detailed design methods, evaluation models and vendor-scoring tools remain AaronAlphaX proprietary materials (see Section 13).

3.1 Phased build-out

Phase	Indicative IT load	Outcome
Phase 1	**[___ kW]**	Initial B300-class pods live; electrical and cooling infrastructure sized for full build-out where cost-effective
Phase 2	**[___ kW]**	Additional pods added without outage to Phase 1 systems
Phase 3	Up to 500 kW	Rubin-generation readiness; power and cooling path confirmed against vendor roadmaps at that time

SAMPLE PROPOSAL — ILLUSTRATIVE TEMPLATE

3.2 Power architecture options assessed

Option	What the Client receives
A. 415/480 V AC distribution	Assessment against current B300-class requirements and supply-chain availability
B. AC today, 800 VDC-ready	Specification of the space, pathway and interface provisions needed to convert later without rebuilding the electrical room
C. Early 800 VDC adoption	Risk and readiness assessment of equipment maturity, vendor support and timing

*Output: a recommended option with its rationale, cost envelope **[USD range]** and decision triggers for revisiting it.*

3.3 Cooling strategy: direct-to-chip with CDUs vs hybrid

We compare direct-to-chip liquid cooling with in-row or in-rack CDUs against a hybrid liquid-plus-air design at 120–140 kW per rack. The comparison uses the Client's specific system SKUs and covers the liquid/air heat split, residual air-side capacity, CDU modularity by phase, and the facility-water temperature strategy. The output is a recommended design with a growth path per phase.

3.4 Partial-load efficiency

We define the efficiency requirements vendors must meet at Day-1 load levels as well as at full load. These cover UPS and CDU behaviour at low load and how redundancy is structured. Vendors are held to them contractually through the RFP.

3.5 Vendor RFP and deal-breaker screen

We prepare an RFP package and run a structured evaluation of shortlisted modular vendors. The evaluation includes a set of **deal-breaker criteria** that disqualify a vendor where expansion would disrupt live phases, capacity would be stranded, or efficiency commitments are not backed contractually. The deal-breaker list is issued to the Client as part of D3.

3.6 Colocation sourcing: US, GCC, SEA

Where on-site build-out lags demand, or for regional latency and data-residency needs, we source high-density colocation in the **United States, GCC and Southeast Asia**. We shortlist liquid-ready space, run a comparable commercial summary and support negotiation of term and expansion rights.

3.7 Token / API capacity option

For interim and overflow demand, AaronAlphaX can provide access to open-weight model inference (for example Qwen, DeepSeek and GLM families) through an **OpenAI-compatible API gateway**, billed per token. The gateway can sit in front of the Client's on-premises capacity, so applications move to on-site serving as it comes online without changes to client code. Model availability, data-handling terms and service levels are agreed in a separate order form.

SAMPLE PROPOSAL — ILLUSTRATIVE TEMPLATE

4. Deliverables & Cadence

#	Deliverable	Format	Due
D1	Requirements baseline and capacity model (users → load → phases)	Report + model summary	Week 3
D2	Power and cooling architecture recommendation (Sections 3.2–3.4)	Technical memo	Week 6
D3	Vendor RFP package, including deal-breaker criteria	RFP documents	Week 7
D4	Vendor evaluation and recommendation	Scored comparison + memo	Week 12
D5	Colocation shortlist for **[selected regions]**	Shortlist + commercial summary	Week 8
D6	Interim API capacity onboarding (optional)	Endpoint, keys, usage report	**[Date]**
D7	Phase-1 delivery oversight reports	Bi-weekly status report	Through commissioning

5. Acceptance Criteria

- Each deliverable is reviewed against its Section 4 description. The Client accepts it or gives specific written feedback within **[5]** business days; otherwise it is deemed accepted.
- One revision round per deliverable is included. Further rounds are handled by change order.
- Recommendations are professional advice based on information available at the time. Vendor equipment performance is warranted by the vendor, not by AaronAlphaX.
- API capacity (D6) is accepted on successful authenticated test calls against the agreed models.

6. Timeline

Stage	Weeks	Milestone
Discovery & baseline	1–3	D1 accepted
Architecture	4–6	D2 accepted; architecture decision recorded
Vendor process	7–12	RFP issued; vendor recommended (D3–D4)
Capacity sourcing	5–10 (parallel)	Colocation shortlist; API option live if selected (D5–D6)
Phase-1 oversight	13 onward	Through Phase-1 commissioning: **[target date]**

SAMPLE — PROPOSAL — ILLUSTRATIVE TEMPLATE

7. Fees

Item	Basis	Amount (USD)
Advisory: Discovery, architecture and RFP (D1–D3)	Fixed fee	**[USD ____]**
Vendor evaluation & recommendation (D4)	Fixed fee	**[USD ____]**
Colocation sourcing (D5)	Fixed fee or success fee	**[USD ____ / __%]**
Phase-1 delivery oversight (D7)	Per month	**[USD ____]**
Token / API capacity (D6, optional)	Per million tokens, by model	**[USD ____]**
Total fixed fees		**[USD ____]**

Excludes equipment, construction, colocation rent and power, travel and taxes. Third-party costs are contracted directly by the Client or passed through at cost with prior written approval. Any fee AaronAlphaX receives from a colocation provider is disclosed to the Client in writing.

8. Client Responsibilities

- Provide workload, user and model assumptions, site information and access to decision-makers.
- Make procurement and contracting decisions with vendors and colocation providers; AaronAlphaX advises but does not sign on the Client's behalf.
- Obtain site permits, utility agreements and internal approvals.

9. Payment Terms

- Fixed fees: **[40]**% on signature and **[60]**% on acceptance of D4; or **[alternative milestone schedule]**.
- Monthly fees are invoiced in advance. API usage is invoiced monthly in arrears, or prepaid by credit top-up.
- Payment within **[14]** days in USD by bank transfer. Work or API access may be suspended after **[7]** days' written notice of non-payment.

10. Confidentiality

Each party keeps the other's non-public technical, commercial and site information confidential, including vendor pricing received during the RFP, and uses it only to perform this SOW. This obligation lasts for **[3]** years after termination, subject to the usual exceptions for public, already known, independently developed or legally required information.

11. Data & Security (API option)

Prompt and output data processed through the API gateway is handled under the data-processing terms in the order form. These cover retention, region and sub-processors. Client data is not used to train models unless the Client agrees in writing.

12. Term & Termination

- This SOW starts on **[Date]** and runs until Phase-1 commissioning or **[__ months]**, whichever comes first, unless extended by change order.

SAMPLE PROPOSAL — ILLUSTRATIVE TEMPLATE

- Either party may terminate for material breach not remedied within **[15]** days of notice. The Client may terminate for convenience on **[30]** days' notice, with fees for work performed up to the termination date remaining payable.
- Sections 10, 11, 13 and 14 survive termination.

13. Intellectual Property

Once paid in full, the Client owns the reports and documents created specifically for it under this SOW. AaronAlphaX keeps all rights in its pre-existing and proprietary materials, including evaluation models, scoring tools, templates, checklists and methods. The Client receives a non-exclusive licence to use any of these embedded in the deliverables for its internal purposes only.

14. Liability & General

Except for breaches of confidentiality, each party's total liability under this SOW is capped at the fees paid in the **[6]** months before the claim. Neither party is liable for indirect or consequential loss. Governing law: **[Jurisdiction]**. Changes require a written change order signed by both parties.

15. Signatures

For Guangzhou Xinqiao Technology Development Co., Ltd. (AaronAlphaX)	**For [Client]**
Signature	Signature
Name / Title	Name / Title
Date	Date

SAMPLE PROPOSAL — ILLUSTRATIVE TEMPLATE

Sample proposal — illustrative template, not an executed contract. No client relationship, result, partnership or certification is implied.