OFFICIAL SAMPLE PREVIEW

46 pages in the full edition - this preview intentionally contains only a small sample.

Use this PDF on the Macaroon Network ebook page or Gumroad listing so buyers can judge the quality without receiving the complete book.

RESEARCH & INTELLIGENCE

What the full ebook is designed to give you

AI founders, developers, researchers and infrastructure buyers comparing cloud GPUs and LLM APIs.

In the full edition you will be able to:

- Normalize incompatible pricing units.
- Compare cost per completed workload rather than sticker price.
- Include VRAM, utilization, queueing, storage, egress, retries and rate limits.
- Record price freshness and source evidence.
- Use scenario analysis before committing spend.

What is included in the full ebook

- ☐ GPU and LLM cost-normalization methods
- ☐ Hidden-cost checklist
- ☐ Workload-fit and VRAM analysis
- ☐ Expected/optimistic/adverse scenarios
- ☐ Procurement and provider-comparison worksheets

The complete edition is 46 pages. This sample does not include the complete exercises, chapter teaching, templates, worked material or final reference sections.

SAMPLE EXCERPT

Sample: the cheapest GPU-hour can cost more

Headline price is only one term in the cost equation. Suppose Provider A is 15% cheaper per GPU-hour but jobs spend longer waiting, utilization is lower and failed runs are more common. Provider B charges more per hour but finishes the workload 25% faster.

The meaningful comparison is not hourly rate. It is cost per completed useful workload. That may include compute time, idle setup, storage, data transfer, orchestration, retries and the opportunity cost of waiting.

Capability fit matters too. A cheap GPU with insufficient VRAM can force smaller batches, quantization changes or model restructuring. An LLM API with low token price may have rate limits that make a deadline impossible.

The full workbook gives a structured way to normalize offers, record freshness, model hidden costs and compare expected, optimistic and adverse scenarios before committing infrastructure spend.

Preview boundary: this excerpt is intentionally short. The full ebook develops the subject through complete chapters, examples, exercises and reference material.

FULL EDITION

What you receive when you buy the ebook

The Real Cost of AI Compute
A GPU & LLM Pricing Decision Workbook

46 pages • PDF • Macaroon Network

The full purchase includes:

- The complete ebook - not a shortened or watermarked edition
- All chapters, teaching sections, worked examples and exercises
- All decision tools, study pages or professional templates included in the title
- The complete reference and limitation material
- Context for the related Macaroon Network companion service: GPU & LLM Cloud Pricing Feed

Where to buy

Buy the full edition from the Macaroon Network ebook page or the linked Gumroad product page. Gumroad handles the human checkout and protected PDF delivery; the live Macaroon companion service is a separate product.

MACAROON NETWORK
Research & Intelligence • Premium Digital Library