FREE STARTER GUIDE · PREVIEW OF THE AI STRATEGY & FUTURE SYSTEMS BRIEFING

Open-Weight vs Open-Source Decision Guide

Decide between open-weight models, fully open-source AI and hosted APIs by your workload, ownership needs and data boundary — not by labels.

For	Level	Time	Next level
AI builders and technical decision-makers	Intermediate	Half a day	Chapter 4 of the full playbook

THE MICRO-PAIN THIS GUIDE FIXES

'Open' means very different things. Your decision depends on which one.

An 'open-weight' model publishes its trained weights so you can download and run it — but its training data and code may be closed, and its licence may restrict commercial use or large companies. Fully 'open-source AI' (in the sense of definitions like the OSI's) goes further. Hosted APIs are closed but easy. Teams pick based on the label and discover later that the licence, hardware bill or support gap doesn't fit.

Licence

Restrictions on commercial use, user counts or model outputs discovered late.

GPU bills

Self-hosting costs that exceed API spend at your volume.

Lock-in

Either to a vendor API or to infrastructure you can't maintain.

When you finish this free guide you will have

- A clear workload and constraint profile
- A comparison of hosted API, open-weight and open-source options
- A licence and operations check
- A benchmark result and documented fallback

THIS FREE STARTER

- One working route for one problem
- 7 short steps and a copy-and-use prompt
- One worked example and a final checklist

THE FULL PLAYBOOK · $49

- 6 chapters that take the whole job end to end
- 9 ready-to-use templates, a worked case, 117 PDF pages
- **See The AI Strategy & Future Systems Briefing →**

STEP BY STEP

The working method

Workload → constraints → options → licence → cost → benchmark → decide with fallback.

You need: Model cards and licences for candidates (e.g. on Hugging Face); A way to run candidates (hosted inference, cloud GPU, or local); A benchmark set of real tasks. **Rule:** Read the actual licence, not the label.

STEP 01 OF 07

Describe the workload

- Task type: chat, extraction, coding, embeddings, vision.
- Quality bar and latency needs.
- Monthly volume and peak load.

Example: Extract fields from 50,000 invoices/month; 95% accuracy; batch, not real-time.

Done when: Workload profile written.

STEP 02 OF 07

Write ownership and data constraints

- Must data stay on your servers or in a region?
- Do you need to audit, fine-tune or freeze the model version?
- Any regulatory or customer contract requirements?

Example: Client contracts require data to stay in the EU; version must be frozen for audits.

Done when: Constraints listed with sources.

STEP 03 OF 07

Understand the three options

- **Hosted API**: fastest start, provider handles ops; data leaves your boundary (subject to terms).
- **Open-weight**: run weights yourself or via a host; control and privacy, but licences vary and you own ops.
- **Open-source AI**: weights plus code and data transparency per an open definition; maximum auditability, fewer top-tier options.

Example: Open-weight model hosted in your EU cloud satisfies residency and version freezing.

Done when: Each option mapped to your constraints.

STEP 04 OF 07

Check licences carefully

- Read the licence file: commercial use, user thresholds, acceptable-use policy, output use, redistribution.
- Note attribution and notice requirements.
- Get legal review for anything unclear.

Example: Model A allows commercial use; Model B restricts services above a user threshold.

Done when: Licence notes per candidate.

In the full playbook · The AI Strategy & Future Systems Briefing explains the system layers — models, tools, workflows, interfaces — behind this decision. See what is inside · Get it for $49

STEP 05 OF 07

Estimate total cost

- API: cost per request × volume.
- Self-host: GPU hours, storage, networking, engineering time, monitoring, upgrades.
- Compare at current and 5× volume.

Example: At 50k docs/month API is cheaper; at 500k self-hosting wins.

Done when: Total cost at two volumes.

STEP 06 OF 07

Benchmark representative tasks

- Build a set of 30–100 real, anonymised tasks with correct answers.
- Run each candidate with the same prompts/settings.
- Measure quality, latency, failure behaviour.

Example: Open-weight model 94% vs API 96% on 60 invoices; both meet the bar after a small prompt fix.

Done when: Benchmark table complete.

STEP 07 OF 07

Decide with a fallback

- Pick the option that meets constraints, quality and cost.
- Keep an abstraction layer so you can switch models.
- Document a fallback and a re-evaluation date.

Example: Chose self-hosted open-weight in EU; fallback: regional API endpoint; re-evaluate in 6 months.

Done when: Decision record with fallback and review date.

COPY AND USE

Prompts and template

Paste into your AI tool and replace everything in [brackets].

PROMPT 1 · LICENCE READER

```
Here is a model licence: [paste].
Summarise: commercial use permitted? user or revenue thresholds? acceptable-use restrictions? rules on using outputs to train other models? redistribution and attribution requirements? Quote the relevant clauses. List anything ambiguous for legal review. Do not give legal advice.
```

WORKED EXAMPLE

An invoice-processing team chooses its model path

A finance-automation startup needed EU data residency and frozen model versions for audits.

1. Workload: 50k invoices/month, batch, 95% accuracy.
2. Constraints: EU data, version freezing.
3. Licence review ruled out one open-weight candidate.
4. Cost: API cheaper now; self-hosting cheaper at 5× volume.

RESULT

A defensible choice that satisfies auditors, stays within budget and can be revisited as models improve.

FINAL CHECKLIST

Before you call it done

- [ ] Workload profile
- [ ] Ownership and data constraints
- [ ] Options mapped to constraints
- [ ] Licences read (actual text)
- [ ] Legal review for ambiguities
- [ ] Total cost at 1× and 5×
- [ ] Real-task benchmark

In the full playbook · The full briefing includes vendor-evaluation and build-vs-buy decision cards. See what is inside · Get it for $49

WHEN IT GOES WRONG

SYMPTOM	LIKELY CAUSE	FIX
Legal blocks launch	Licence not read	Read licences early; legal review
Self-hosting over budget	Ops time ignored	Include engineering time and monitoring
Quality lower than expected	Leaderboard-based choice	Benchmark on real tasks
Stuck with one vendor	No abstraction layer	Wrap model calls; keep fallback

LOCKED IN THE FULL PLAYBOOK · TOOLKIT PREVIEW

This free starter shows the method. The AI Strategy & Future Systems Briefing ships the ready-to-use files:

- 90 day AI strategy plan
- AI governance raci
- AI readiness scorecard
- AI strategy one pager template
- AI strategy prompt library
- Executive AI briefing deck outline

Unlock all templates — $49 one-time →

WHERE THIS FREE GUIDE STOPS

You fixed one problem. The full system fixes the job around it.

This starter gave you one route through one step. It deliberately leaves out what makes the result hold up at work:

- **The steps before and after this one.** The AI Strategy & Future Systems Briefing covers the whole job in six chapters, not one slice of it.
- **Ready-to-use templates.** 9 copy-and-use files, so you do not rebuild the structure every time.
- **A full worked case** that shows every decision on a realistic project, start to finish.
- **Checks for when it fails**, so a result you ship tomorrow is still right next month.

YOUR NEXT STEP · THE AI STRATEGY & FUTURE SYSTEMS BRIEFING

Make AI strategy decisions with clarity

What you will be able to do:

- Frame AI strategy as the redesign of specific work, not a tool list
- Prioritize two or three use cases by value, feasibility and risk
- Price the work per task and choose infrastructure by workload
- Run a 90-day cycle with decision gates and named owners

6	9	1	117
chapters	templates	worked case	PDF pages

You just read a preview of **Chapter 4: Price the work per task and choose infrastructure by workload, so cost surprises never kill a good pilot**.

~~$89~~ **$49** one-time · instant access in your CosynkAI Library

Get the playbook — $49 one-time → ***See what is inside***

WORKING ON SEVERAL AI WORKFLOWS?

Complete Guide Library — all 8 CosynkAI playbooks for **$97** one-time ($472 combined regular prices). Already own a playbook? Its price is credited toward the Library.

Get the Library — $97 → Compare all 8 playbooks

You received this free guide after sending the keyword **OPEN**. Developed by CosynkAI.com · Edition 1.0 · September 2026. Educational resource, not an implementation service. Tools, prices and limits change; check current documentation before relying on a tool in production. Questions: hello@cosynkai.com · Privacy Policy · Digital Product Terms