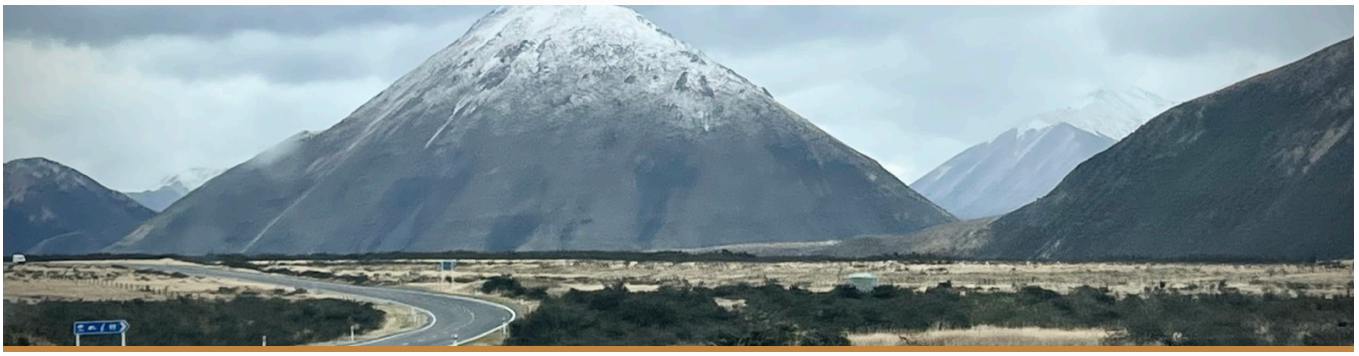

What nobody has measured yet

Version 0.2 · revised 10 September 2026 · this version as a PDF

Nobody can price this, including us. Four forks the next five years could take, what each would mean, and the observable that would tell an institution which one it is on.

The pursuit of 'Goodness' in AI — Sovereign agents, and what it takes to trust one with real work

---

What the available measurements actually measure

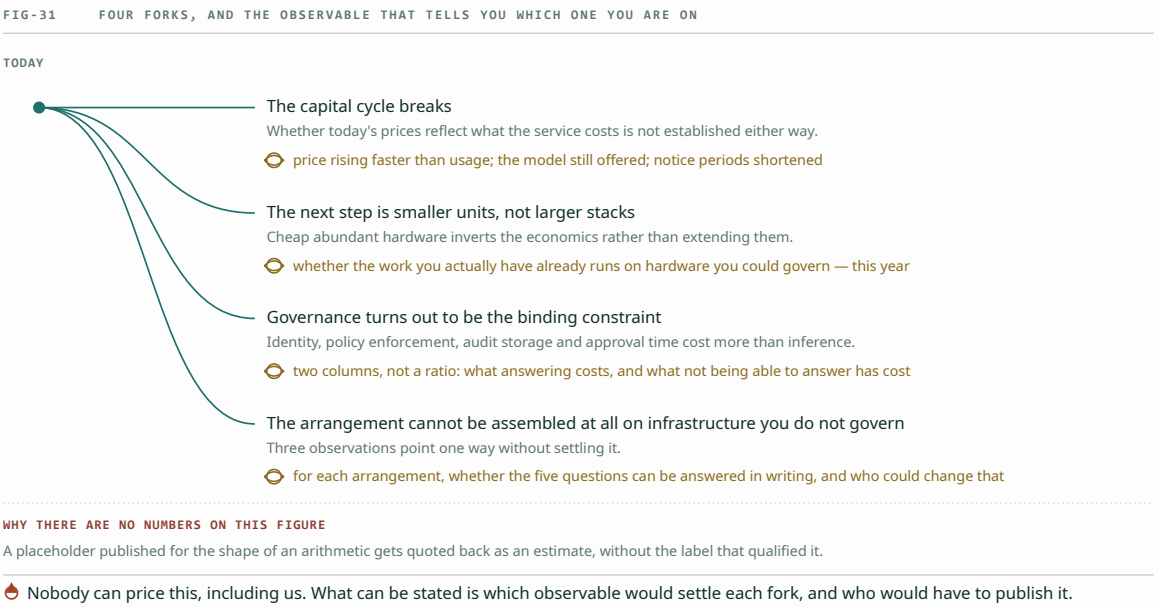

FIG-31 Four branches the next five years could take, each carrying the thing an institution could watch for rather than a projection. The figure has no numbers on it deliberately: a placeholder published for the shape of an arithmetic gets quoted back as an estimate without the label that qualified it.

An institution asking what a governed AI arrangement costs will be offered a figure by almost anyone it asks, and every one of those figures is a guess, because the measurements available do not measure the thing being priced. This piece gives none, and the reason is worth setting out because it is also the reason the confident figures should be distrusted.

The measurements that exist describe **buildings**. Power usage effectiveness — the metric carrying nearly every published comparison — is the ratio of a facility's total energy to the energy reaching its computing equipment. It was devised to describe cooling and power distribution around a rack, and it is a reasonable measure of exactly that. It says nothing about whether the work being done is useful, nothing about the cost of the software arrangement around it, and nothing about the governance properties this series is concerned with.

Those measurements are also of a period that is ending. The European Commission's first mandatory reporting round, published in July 2025, covers 770 facilities and describes the estate as it was built between roughly 2010 and 2024. Lawrence Berkeley National Laboratory's 2024 report models the American estate over the same era. Both are careful and both answer the question of their decade: how efficient is a data centre? Neither was built to answer the question of the next one: **what does it cost an institution to hold authority over software that acts on its behalf?** What is needed is not a better estimate. It is to stop estimating and start asking, in public, what would have to be true.

The existing evidence, and its limits

Before asking what is unknown, it is worth being exact about why the available measurements do not answer the question — because the temptation is to quote them with a qualification attached, and a qualification attached to a portable number does not survive the journey.

The ratio described above measures what it measures, and measures it reasonably. What it cannot do is answer the question an institution is actually asking, because it has no term for authority, for software, or for acting — nor for whether the computing was useful, how much hardware sat idle, or what was manufactured to build the place.

A comparison between facility classes, printed here with its limits noted, would be quoted onward without them: that is the mechanism these documents describe, and it does not spare a document for having described it. What can be said without a number is that the American headline figures are **simulated** — their authors state that the simulations *"assume systems are commissioned and operate as designed. This is rarely the case"* — and that the European figures are **operator-reported and unaudited**, with one scheme discarding a portion of its returns as malformed and the most recent report flagging two of its own indicators as misaligned with reality.

The first fork: what if the capital cycle breaks?

An enormous amount of capital has been committed to concentrated AI infrastructure on the expectation of returns that have not yet arrived. Whether that expectation holds is not a technical question and this piece has no privileged view of it. What matters is that the answer changes the ground under every cost comparison currently being made.

Whether today's prices reflect what the service costs to provide is not something this series can establish, and it is worth resisting the confident version in either direction. The claim that inference is presently sold below cost is widely repeated and, so far as could be established, not publicly evidenced by anybody outside the companies concerned. The claim that it is priced sustainably is in the same position.

What can be said is narrower and does not require knowing the answer. An institution buying a service at a price it cannot audit, from a party whose costs it cannot see, holds a position whose durability depends on facts it has no way to check. If the present expectation holds, nothing changes. If it does not, the adjustment does not arrive as a headline: it arrives as price rises on renewal, capabilities withdrawn from lower tiers, models retired on shorter notice, and support that becomes harder to reach.

What would tell an institution which it is on: for every line of what the arrangement costs it, which party can change that line without its consent, and on what notice. An organisation buying inference reads this from its renewal terms — whether price is rising faster than usage, whether the model it built on is still offered, whether notice periods have shortened. An organisation running its own reads it from its power contract, its hardware replacement cycle and its staffing.

🔑 **Most of the argument of this series does not depend on the answer**, and the exception is worth naming rather than glossing. Requirements about identity, authorisation, tool governance and records are arrangements rather than purchases, and their difficulty does not move with the price of inference. But The four properties underneath argues that the boundary governing an outcome is where a record enters the model's working context — which implies inference the institution can govern, and that is exposed to every question in this document. An institution can hold most of the apparatus whatever happens to prices. Whether it can hold all of it is one of the things not known.

The second fork: what if the next step is smaller units, not larger stacks?

The prevailing assumption is that capability scales with concentration: larger facilities, larger models, larger stacks. That has been true for a decade and the measurement literature described above is a record of it being true.

It has not always been true of computing, and there is no law making it permanent. Every previous concentration in this industry was undone not by an argument but by a change in what could be manufactured cheaply. If capable inference hardware becomes small, inexpensive and abundant — the same movement that took mainframes to workstations and workstations to phones — the economics under every comparison in circulation invert, and they invert without warning, because the manufacturing change precedes the analysis of it by years.

The evidence here is genuinely thin in both directions and should be presented that way. One peer-reviewed systems study built a cluster from second-hand handsets and found it better on carbon per unit of work than a cloud instance over three years. It is a single comparison on three workloads, and the same literature finds that permanently powered devices behave differently from the phones it used — so it establishes that the question is open, and not that the answer is known. Separately, model efficiency has improved through distillation, quantisation and sparse routing.

What would tell an institution which it is on: whether the work it actually has already runs on hardware it could govern. For many organisations the answer is already yes and the fork is closed for them now; the question of what the largest models will require is somebody else's. For the rest, the number worth tracking annually is the floor: the smallest unit that runs their real workload, and whether it is falling faster than that workload grows.

The third fork: what if governance turns out to be the binding constraint?

Every comparison being published is about compute — energy, hardware, tokens. Suppose that is the wrong axis entirely.

The requirements in What you can actually require are not compute costs. They are an identity authority, a policy gateway, an audit record, and the human time to review what needs reviewing. **Nobody has published a figure for any of them.** Searches across the literature found no measurement of what identity infrastructure, policy enforcement, audit storage or human approval time costs in an AI arrangement. The nearest evidence is a finding that for-

mal external approval processes were associated with worse delivery performance and **no measured reduction in failure rates** — a caution about approval as a mechanism, not a number.

If governance turns out to be the larger of the two, a public debate conducted almost entirely about compute is a debate about the smaller half. If it turns out to be the smaller, the current focus is right and this fork closes. Nobody has published the figures that would decide it, which is the reason it appears here rather than as a claim.

What would tell an institution which it is on: two columns rather than a ratio. What it costs to be able to answer for what the agent did — hours on approvals, review and audit response, and the infrastructure that makes them possible. And what it has cost, in incidents and their consequences, on occasions when somebody asked and the institution could not answer. Setting governance against the inference bill alone prices it as overhead, which is the framing this series disputes: the apparatus is not a tax on the arrangement, it is what makes it an arrangement. Any organisation running an agent could produce both columns within a month, and none appears to have published either.

The fourth fork: which arrangements can carry the answers at all

This is the question the series exists for, and it is the one where the answer is least established.

The argument made across these documents is that goodness in AI is a property of an arrangement — somebody entitled to set a limit, an entitlement that outlasts them, a deliberate route to change it, and the ability to show afterwards that it held. Nothing in that requires particular hardware or a particular scale.

What is not established is whether an arrangement of that kind can be assembled on infrastructure the institution does not govern. The evidence gathered for this series points one way without settling it. The most sovereign offering in a major vendor's catalogue runs its control plane locally and requires the vendor's approval, a current support agreement and an annual renewal for the institution to be permitted to run it at all. A vendor's own legal officer, asked under oath before the French Senate whether data held in France could be guaranteed against foreign compulsion, answered **"Non, je ne peux pas le garantir"** — adding, and it belongs in the quotation, that it had never happened. Te Kāhui Raraunga records that onshore hosting with a foreign-headquartered provider still leaves configuration and hypervisor access abroad.

Those are three observations, not a proof. They are consistent with the current path being unable to deliver the arrangement, and they are also consistent with it delivering a weaker version that many institutions would accept.

What would tell an institution which it is on: for each arrangement available to it — bought, leased, owned, or shared with other institutions — whether each of the five questions in Who actually holds the controls can be answered in writing, and for each answer, who could change it without the institution's consent. That is askable of a vendor, of a hosting co-operative, and of an institution's own board, on the same terms. Where the arrangement is bought, the moment of truth is a renewal rather than a pilot.

Where this leads

These documents argue for properties rather than for a product. That is deliberate: an institution can hold every requirement in What you can actually require while renting every processor it uses.

But properties have to be met by something, and it would be evasive to argue at length for an arrangement and then decline to say what one looks like. Two lines of work on this site attempt it, and both are open to the same tests these documents apply to everybody else.

The architecture. The blueprint sets out shared infrastructure owned by the organisations using it — what it must be able to do, what it costs, how it is governed, and four documented ways it fails. Running an organisation where agents do the work is the part nearest to this series: what it takes to remain answerable when most of the work is done by software acting on its own. What happened to your software is the diagnosis underneath both — how organisations arrived at arrangements they would not have chosen, one renewal at a time.

The language layer. A general model trained on the world's text carries the world's distribution of meaning, which is not the distribution in any particular community's records, professional vocabulary or reo. The alternative is a situated language model — one adapted to the setting it serves, holding a situated language layer that belongs to the institution rather than to a supplier. That matters to the argument here for a specific reason: a model an institution can adapt is a model whose behaviour it can *change* deliberately, and the third condition on any institutional value — that it can be altered on purpose rather than by drift — is unmeetable when the model is somebody else's and changes on their schedule.

🔑 **Apply this series' own tests to that work rather than accepting it.** The five questions in Who actually holds the controls are as askable of a co-operative as of a hyperscaler, and a shared arrangement can fail them: an institution that cannot leave a co-operative it helped

build is in the position this series objects to, wearing better politics. **If those documents cannot answer the questions these ones ask, that is a finding, and the response box below is the place to say so.**

Four questions for the reader

Each of the forks above ends in something observable, and in every case the observation sits with people running these arrangements rather than with anyone publishing about them. None of it has been collected. That is not a gap in the literature so much as a gap in who gets asked.

So this document asks, and the response box at the foot of the page goes to a person rather than into a form. Four questions. Any one of them is worth answering on its own.

On your renewal. Has the price risen faster than your usage? Has a model or a feature you built on been retired, moved to a higher tier, or given a shorter notice period than before? A dozen renewals from a dozen small organisations would say more about where the capital cycle actually is than any amount of market commentary, because you are looking at the terms rather than at the announcements.

On the smallest unit that runs your work. Not the frontier — the floor. If you have priced hardware capable of doing the job you actually have rather than the job the benchmarks describe, what did it cost, and has that number moved since you last looked?

On what governance costs you. Roughly how many hours a month go on approvals, review, and answering questions about what the system did — set against what you pay for the service itself? That ratio has never been published by anybody. If a handful of readers supplied it, this series could publish the first version of it, and would.

On whether you can get answers. If you have asked a supplier the five questions in Who actually holds the controls — whose agent it is, what authority it holds, where its operational life happens, who may inspect and challenge it, and who may stop or move it — what came back? Refusals are as useful as answers here, and probably more so. If a supplier answered well, that is worth knowing too, and this series will say which supplier and what they said.

What happens to what you send. It is read by a person. Anything published from it is published in aggregate and without identifying an organisation unless the sender asks otherwise. Where a reply corrects something in these documents, the correction is made and the fact of it recorded on the page rather than absorbed silently. Where somebody disagrees with the argument and says why, the strongest version of that objection has a better claim on space here than another paragraph of agreement.

What this project will publish

An arrangement of the kind described across this series is being built rather than theorised, and the commitment made here is specific.

Its energy, its utilisation, its capital and operating cost, and the hours spent on governance will be published on the same terms as everything else on this site — with the method stated, the limits marked, and **including if the numbers are bad**. If it is more expensive than buying the equivalent service, that will be published as a number rather than as a qualification.

The commitment is specifically to publish the measurement whether or not it supports the argument. This document carries the `requires-evidence` tag in the corpus, the tag reaches every reader before the claims do, and it stays until the measurement exists.

What would make these the wrong questions

First, if a credible measurement of a small, institution-governed arrangement is published and settles the economics in either direction, this piece becomes a record of what was unknown in September 2026 and should be superseded rather than defended.

Second, if the four forks turn out to be one fork — if capital, manufacturing, governance cost and vendor terms all move together because they are all downstream of the same thing — then treating them separately is a mistake, and the observable an institution should track is whatever that single thing is.

Third, if institutions turn out not to care about the answers — if the five questions can be asked, answered badly, and the arrangement bought anyway — then the constraint is not evidence and never was, and a series of documents is the wrong instrument entirely.

Related

What you can actually require sets out twelve requirements whose cost does not depend on any of these four questions being resolved.

The six, in reading order. Each stands on its own; read together they build one argument.

- What has to be settled first — why the prior questions are prior, and what each of the others answers

- Who actually holds the controls — five questions establishing whether an authority to act actually exists
- How much it may do unsupervised — five levels of permitted autonomy, declared rather than left implicit
- The four properties underneath — four properties a delegation must carry, each checkable by observation
- What you can actually require — twelve requirements, and what currently requires each of them
- An invitation to become a Distributor — not part of the argument above: a proposal to work with its author, for anyone who wants to take this further

> **ACKNOWLEDGEMENT**
>
> **Leslie Stroh** — my big brother, a product philosopher, and a magnificent role model. He helped me find my way to a vision of goodness in AI.

▸DISCLAIMER

> **How this was made.** The version number counts drafts of the text. It does not measure the inquiry behind it, which has run over days and across several AI systems, with argument between those systems and within them, directed, refused and repeatedly redirected by the author. The source material was AI-generated, and then adversarially and iteratively refined across a range of tools — systems built by different companies in different jurisdictions, set against each other and against the author. No one of them produced this text, and no one of them reviewed it alone. **The plurality is deliberate rather than incidental.** A single model carries a single set of priors about which sources are authoritative, and this series argues that an evidence base narrowed in exactly that way is how a contested question comes to look settled. Using one model to investigate that claim would have been the claim refuting itself. To name a single model on it would credit that model with work that was neither its own nor done in a single pass. The diagram on each page was drawn afterwards, with the same assistance, from the argument the essay had already made. **The plurality was also necessary, and the record should say why.** In drafting, the assisting model repeatedly led with United States institutional sources — a national laboratory, an industry association, a market study nineteen years old — and presented conclusions drawn from them as the state of knowledge. On one occasion European measured data contradicting those conclusions was present in the same research return and was placed below them. Framings were proposed that would have argued against this series' own position using that evidence base, and offered as rigour. Each was refused by the author and the material rebuilt. That is the mechanism these documents describe, occurring in their own making, and it is recorded because a series arguing that evidence bases narrow without anyone deciding to narrow them cannot credibly claim its own production was exempt. The framing, the corrections and the judgements are the author's, and so are the errors.

Cite this version

```
What nobody has measured yet, version 0.2 (September 2026).
agenticgovernance.digital/downloads/what-we-do-not-know-yet-v0-2.pdf
```


Version 0.2 was revised in place. On 10 September 2026 FIG-31 was corrected: two of its layers overlapped, so the last observable printed through the heading 'WHY THERE ARE NO NUMBERS ON THIS FIGURE' and the rule meant to separate them sat above both. Only the figure's geometry changed — no word of the text, and no figure's content, is different. Anyone holding an earlier copy of this version should treat /what-we-do-not-know-yet as authoritative.

PREVIOUS
What you can actually require

Series contents

NEXT
An invitation to become a Distributor

Alongside: questions and answers · sources and provenance · slides

Share this Mastodon Email Copy link

Found something wrong?

This piece names what would show it is wrong. If you have that — a figure, a source, an argument — it is worth more to us than agreement. Tell us which claim and why.

Which section, if you know it

for example: 2. Four properties

The claim you are responding to

quote it, or describe it

What is wrong with it

Your email, only if you want an answer

Used to reply to you and nothing else. No list, no newsletter, never passed on.

Send this