The Delta Is Where the Insurance Lives

Rating agents without money, and the empty seat a policy fills

Ten voice memos and a pivot briefing, recorded over two days at the end of August 2026, read into doctrine, audited against their own transcripts, and here made one argument: the gap between what an agent can do and what it was authorised to do is an insurable exposure — and the honest first product is a rating anybody can recompute, not a premium nobody can check.

Written 2026-08-31 · 17 chapters in five parts
17,868 words · 8 figures · 76 verified quotations · 44 claims drawn by the writer
Published by the sgit project · pki.sgit.ai · CC BY 4.0

This PDF reads start to finish offline. Every URL that matters appears in full at least once in the text; no link needs following.

Before anything else

These positions travel inside the chapters rather than in an appendix, because a chapter quoted in isolation must still carry them.

1. Nothing here is insurance. Stage 1 emits a rating; it transfers no risk and promises no payout — which is exactly why it can be built, and calling it insurance would be the first dishonesty.
2. Nothing here is built. Ten memos are read into doctrine; the two specified MVPs — the world model and the market survey — remain unbuilt.
3. No external evidence has been gathered. Every decision on the log is internal reasoning; the survey that could be wrong in a way the world would correct has not been run.
4. The register the rating would read is fixtures. Ten of eleven identities have their private keys published on purpose, so every signature verifies and proves nothing.
5. The placement orderings are judgements. Nobody has measured a desktop agent, so the estate cannot score its own leading example.
6. A level is never declared, only derived — and a level nobody can recompute is exactly the theatre a premium would have prevented.
7. The readings were audited against their transcripts and six defects were found and fixed. The method held; the arithmetic and the housekeeping did not.
8. This is a participant's account, published by the project that would build the connective tissue it argues for.

The provenance rule. Every load-bearing claim about what the memos or the estate MEAN is marked `stated` or `drawn`. `stated` carries a verbatim quotation with a located source; `drawn` carries this book's own reasoning in the reader's view. DO NOT TREAT THE DRAWN CLAIMS AS THE PROJECT LEAD'S POSITIONS — they are the book's.

Contents

The figures, and the two gates

Every figure was taken from the version its caption names — a `git worktree` at the tag, a one-shot local server on a port used once and never again, a headless browser killed in a block that runs whether the capture succeeded or failed. A figure of a **past** version is **re-derivable**: re-running the harness at that tag reproduces the digest, and it never goes stale because the tag does not move. A figure of the site **as it stands** is **fresh**: the digest must match the live page, and the build fails when it stops matching — which it will, on the next release.

The three terminal figures are real transcripts, executed by the harness at build time and photographed; each capture is recorded in `insurance-book/shots/shots.json` with its digest, and the harness ships beside the book so any figure can be re-taken rather than believed.

Figure	Tag	Gate	Page digest at that tag
`f01-empty-seat`	`current`	fresh	`1a4f2626ff578aca…`
`f02-the-rule`	`current`	fresh	`68eb121d6ecf58e3…`
`f03-permit-refused`	`current`	fresh	`28f11414a45012ee…`
`f04-insurance-hub`	`current`	fresh	`b184a23758ac2316…`
`f05-evidence-classes`	`current`	fresh	`412ee64b844da274…`
`f06-no-world-yet`	`current`	fresh	`b4230b64e6201a63…`
`f07-workbench`	`current`	fresh	`f0d38918f17f95cc…`
`f08-hub-at-v0.1.51`	`v0.1.51`	re-derivable	`e46e3f176f03e358…`

FRONT MATTER

The Delta Is Where the Insurance Lives

Rating agents without money, and the empty seat a policy fills

---

Written 31 August 2026, from ten voice memos and a pivot briefing recorded on 30 and 31 August, filed verbatim as briefs v0.33.71 to v0.33.81 before anyone was allowed to summarise them, read into eleven doctrine documents at `https://pki.sgit.ai/insurance/`, and audited against their own transcripts at v0.33.82. CC BY 4.0.

The markdown under `insurance-book/content/` is the source of truth. The HTML pages and the PDF are renderings of it, and `book.json` carries the SHA-256 of every chapter so a reader can check that what they read is what was written.

The positions that travel with every chapter

A chapter quoted in isolation must still carry these. They are not caveats appended to the argument; several of them *are* the argument.

1. **Nothing in this book is insurance.** The thing the memos specify for stage 1 emits a *rating* — a level, and the derivation behind it. It transfers no risk and promises no payout, which is exactly why it needs no carrier, no capital and no authorisation, and why it can be built now. Calling it insurance would be the first dishonesty, and the memos themselves are the source of that rule.

2. **Nothing in this book is built.** Ten memos have been read into doctrine. The two MVPs the series specifies — the world model and the market survey — remain specifications. This book describes an argument, not a product.

3. **No external evidence has been gathered.** Every decision this pivot has produced is internal reasoning checked against internal documents. The one item that could be wrong in a way the world would correct — the survey of what agent insurance actually exists — has not been run, and this book does not guess its findings.

4. **The register the rating would read is fixtures.** Ten of its eleven identities have their private keys published on purpose, so every signature verifies and proves nothing. A rating demonstrated on fixtures demonstrates the walk, not the trust.

5. **The placement orderings are judgements.** The memos rate Claude-on-a-desktop above Claude-on-the-web; nobody has measured a desktop agent, so the estate cannot score its own leading example. The cheapest next experiment in this book is that measurement.

6. **A level is never declared, only derived.** The rule the whole folder runs on — *a level nobody can recompute is exactly the theatre a premium would have prevented* — is not style. Chapter 3 shows it is the diagnosis of what broke the nearest comparable market.

7. **The readings were audited, and six defects were found.** Three stale counts, one claim identified with the wrong quantity, one over-claim, one dropped question — found by re-reading every transcript, fixed in place, and half of them now blocked by a build gate. The method held; the arithmetic and the housekeeping did not. Chapter 17 walks all six.

8. **This is a participant's account.** It is published by the project that would build the connective tissue it argues for, and it was written by the same class of agent it proposes to rate. Read it the way an underwriter reads a proposal form: as a document whose author has an interest.

Provenance, stated once

Every load-bearing claim about what the memos or the estate *mean* is marked in the text. ***Stated*** claims carry a verbatim quotation whose source is discovered rather than asserted, and re-read out of that source on every build — a quotation not found where it claims to be fails the build. ***Drawn*** claims are this book's own reasoning, and the marker is in the reader's view rather than in a footnote, because the difference between what the project lead said and what the writing session concluded is exactly the kind of thing this estate refuses to blur. Do not treat the drawn claims as the project lead's positions.

The counts in this book are computed at build time from the repository — 76 verified quotations, 44 drawn claims, 8 figures — because a number typed into prose goes stale silently while still reading as a fact, and stale counts are one of the six defects the audit found in the very corpus this book describes.

How to read this book

Part one is the pivot: what moved, and why the seat it fills was already published, empty, in the register. Part two is the rating — the honest half of insurance, separated from the money. Part three is the ecosystem: who rates, who pays for a gap nobody chose, what moves a rating overnight, and who backs a vendor's claim. Part four is the machinery: how a policy becomes a signed statement, what the clocks demand, and where this project refuses to stand. Part five is the proof and its absence: the world that would explain all of this, the positioning that would test it, and a full accounting of what none of it proves.

The shortest honest summary this book can offer is the title. What an agent *can* do, minus what it was *authorised* to do, is a quantity. It is computable today, it is published today for one real environment, and nobody accepts it, pays for it, or covers it. The memos' wager is that the discipline which prices uncertainty for everything else in the economy is the right instrument for that quantity — starting not with money, but with a number anyone can check.

1 · The empty seat

Part one — The pivot

Since 25 August 2026, the register at pki.sgit.ai has published a small JSON document called the excess-authority view. It is generated from the records, carries a banner disclaiming its own authority, and contains one row. The row says that a measured environment — the container this book's own writing session runs in belongs to the same family — holds a grant reaching 41 resources, under a mandate that covers 1. And inside that row sits a field that is the reason this book exists:

> "acceptor": null

Stated — from `registry/views/excess-authority.json`, where it has been published since before anyone said the word insurance. The note beside it reads, in full:

> the difference has no acceptor

Stated. The estate's own hardest sentence had always been that excess authority is unaccepted *by construction*: nobody can accept an exposure nobody has written down, and once it is written down, it turns out nobody wants to. The register did the writing down. The seat stayed empty.

```
$ curl -s https://pki.sgit.ai/registry/views/excess-authority.json | python3 -m json.tool
{
    "_authority": "NONE \u2014 derived from the records; recompute it yourself (registry_tool.py validate does)",
    "generated_as_of": "2026-08-25T12:00:00Z",
    "rows": [
        {
            "excess_authority": {
                "acceptor": null,
                "note": "the difference has no acceptor",
                "observed_at": "2026-08-25",
                "resources": 40
            },
            "grant": {
                "capability": "repo.contents.write",
                "resources": 41,
                "statement": "sha256:4dec940de3243f48ecd77de478b6fa24e09fe3f6095cf2e4023863dec944db3d"
            },
            "mandate": {
                "capability": "repo.pull-request.create",
                "resource": "github.com/SGit-AI/SGit-AI__Website__PKI",
                "resources": 1,
                "statement": "sha256:e572d490272cc517d3fe0783193b539c06bb9f06bf19334ca02619a926e7c45d"
            },
            "subject": "sha256:df2bb4d93af69e6a"
        }
    ],
    "what_this_is": "grant minus mandate, per subject: the exposure nobody accepted (change-control C1); built for downstream risk consumers"
}

t01-empty-seat.txt · sha256:15cf28a7bf6cdcbba632b76ba1425033…
```

On 30 August the project lead recorded a voice memo that begins, in the way these memos begin, mid-thought:

> so I want to do a little pivot on the on the risk mandate and the sort of the risk approach

Stated — from the transcript filed verbatim as brief v0.33.71 before it was read, because in this corpus the transcript outranks any summary of it. The pivot it announces is structural. The chain this estate publishes ends *reality → twin → facts → finding → risks → decisions*, with **risk acceptance** as the terminal act: an owner signs that the organisation chooses to carry an exposure. The memo replaces the terminal node. Instead of an acceptance at the foundation, an **insurance policy**.

And the reason is one sentence, the one this book is named after:

> the difference between what the agent can do, right, which is the the grant, and what we want the agent to do, which is the mandate, the delta of that, is where the insurance lives

Stated — the memo, verbatim, hesitations and all. The delta between grant and mandate is not new to this estate; measuring it is most of what the estate does. What is new is the observation that the rest of the economy already has a name for the party who takes on an exposure its owner will not carry: an **insurer**. An underwriter accepts, for money, what no owner accepted. The empty seat the register has been publishing is, literally, the seat a policy sits in.

A stricter consumer, not a new machine

It would be reasonable to expect a pivot of this size to obsolete the work before it. It does the opposite, and the memo says so —

> what's cool about it is that we already have, I think, a lot of the primitives

Stated. Insurance consumes exactly the artefacts this estate already produces, and the mapping is nearly embarrassing in how little invention it needs. A proposal form is a **measured grant** — the twin, generated by measurement and honest about its own blindness. A policy schedule is a **mandate** — issuer, subject, scope, interval, signed: the five fields of a schedule are the five fields a mandate already has. Exclusions are the mandate's **prohibitions**. Warranties are **facts** attached to the twin, and a breached warranty voids cover exactly as a flipped fact drops an enforcement tier. The insurable interest of the novel cover is the **delta** itself, computed and never stored. The underwriting evidence is the **evidence pack** the workbench already emits.

Drawn. The right way to say this is that the pivot upgrades the *audience* for the estate's evidence rather than the evidence. An acceptance can, in the worst case, be theatre — a signature over an exposure nobody measured. The first brief's reading put the upgrade in one line: a policy is a stricter consumer of exactly the same artefacts, because the measurement now has to be good enough that a stranger will bet on it. The estate spent August building for a reader who ought to believe it. Insurance supplies a reader who is paid not to.

The one primitive that does not exist

The mapping has exactly one empty cell, and it is the one insurance cannot do without. Everything above prices the *ex ante* side — what could go wrong, how reachable it is, what stands in the way. A payout needs the *ex post* side: a **loss event**. What happened, attributable to which subject, exercising which capability, with what severity. Nothing in the register, the pack or the workbench records harm; the closest object the estate owns is an evidence pack for a *refused* action, which is a loss event's exact opposite.

Drawn. That absence is worth savouring, because it is the pivot's most precise output: a corpus that has spent a month recording what agents can do, are allowed to do, and were refused from doing has no schema for what an agent *did wrong*. The memo's payout logic — *"if this happens then you get this payout"* — forces `loss-event/v0` into existence as a shape, and whoever's schema records agent losses will own the eventual actuarial table. That is a strategic sentence wearing a technical one, and chapter 16 returns to it.

Parametric, named

The memo describes its payout logic without naming the industry term for it:

> we can say hey if this happens then you get this payout if that happens, then you get this, and then we can connect that straight away to those capabilities

Stated. Insurance has two payout shapes. **Indemnity** pays the assessed actual loss, and needs adjusters, disputes, and an actuarial depth nobody has for agents. **Parametric** pays a pre-agreed amount when a defined, measurable event occurs — no loss adjuster, no negotiation; the trigger either fired or it did not. It exists for earthquakes and flight delays, where the event data is strong and the loss data is weak, and that is precisely the situation here: every trigger an agent policy would plausibly name — an action outside the mandate executed, a mandate expired with the agent still operating, a warranty fact gone false, a twin gone stale past the re-measurement clause — is *already computable from published documents*.

Drawn. Parametric is the estate's good luck, and its honest limit arrives in the same breath: a computable trigger is not a computable loss. Parametric accepts basis risk — the payout may not match the harm — as the price of not needing the loss data that does not exist. This book will keep meeting that trade, because stage 1 makes it again at a larger scale: chapter 4 removes the money entirely, for the same reason parametric removes the adjuster.

What the seat costs to fill

Drawn. One more thing follows from reading the pivot off the register rather than off a slide deck. When the acceptor field is null, the estate's current behaviour is that the exposure defaults to critical and escalates — an alarm with no owner. Fill the seat with a policy and something better happens than silence: *deploying an uninsured agent* becomes a governance event anyone can check, the way an uninsured contractor on a site is a checkable fact rather than a vibe. The pivot does not make the delta smaller. It makes the delta's ownerlessness *visible in a vocabulary the business already runs on* — which is, as the next chapter argues, most of what the memos think insurance is for.

2 · Two insurances, one subject

Part one — The pivot

The pivot memo catches something in its own flow that a tidier document would have missed, and the catch is load-bearing:

> there's actually two risks here, right? Two two type of insurances

Stated — the memo, verbatim. Separating them decides what is actually novel about this whole programme, because the two covers are priced off different objects and only one of them lacks a market.

Cover A insures against harm done while doing what was *authorised*. The agent pushes to the dev branch, exactly as mandated, and the push takes production down. Its insurable interest is the **mandate**. The human analogues are old and priced: employers' liability, professional indemnity — harm in the course of sanctioned work. There is loss history, there are markets, and an agent version of it resembles existing operational cover closely enough that an existing carrier could approximate it.

Cover B insures against harm done through what was *possible but never authorised*. The credential reached forty other repositories, and something exercised that reach. Its insurable interest is the **delta** — grant minus mandate. And the search for a human analogue comes back nearly empty, which is the finding:

Drawn. The nearest human shape is insuring a contractor who was handed the master key to the building in order to fix one tap. The building trade's answer to that situation is not an insurance product; it is *do not hand over that key*. The delta exposure exists at scale for agents precisely because, for agents, not handing over the key is frequently impossible — the platform offers read-write or nothing, the identity is the user's whole identity or nothing. Chapter 9 gives that impossibility a taxonomy. Here it is enough to say: **Cover B has no loss history, no rating basis and no market, and it is the memo's subject** — *"the one of the ones that we are connecting."*

The premium that falls to zero

One consequence of putting Cover B's insurable interest on the delta is so clean it reads like an incentive design, although the memo arrives at it as a description:

Drawn. When grant equals mandate, Cover B's premium is zero — there is nothing outside the authorisation to insure. The delta is not merely *a* rating variable; it is the insurable interest itself, so an operator who narrows a grant is literally buying down premium. Least privilege has been a virtue for fifty years. A premium tracking the delta is the first mechanism in this corpus that makes it *financially* legible: the difference between "you should scope that token" and "scoping that token is worth two levels" is the difference between advice and a price.

The two-populations thesis this estate published earlier in August lands here with its commercial half attached. For agents you run, you can attest the workload and narrow the grant — control is available. For agents you *rent*, you cannot see inside the vendor's environment, and your only lever is the credential you hand over:

> **hand over a broad credential and hope.**

Stated — the two-populations brief's own summary of the practitioner's honest position, published ten days before the pivot. Cover B is the commercial completion of that sentence. Where control is not available, *transfer* is what remains — and the hope becomes a premium, which is at least a number somebody computed.

Which questions each cover answers

Drawn. The split also sorts the awkward questions this corpus keeps being asked into the covers that own them.

The agent did what we told it and it was still a disaster is Cover A, and it is not agent-specific at all — it is ordinary operational risk wearing a new actor, which is why the memos spend so little time on it. *The agent did something nobody told it to do* is Cover B when the capability was in the grant, and a claim dispute when it was not. And the question that dominates public conversation about agent risk — *the model hallucinated* — turns out to distribute across both: a hallucination that stays inside the mandate is a quality problem for Cover A; a hallucination that exercises the delta is exactly what Cover B exists for, and the delta was there before the hallucination arrived. The memos' framing quietly reprices the hallucination debate: the model's unreliability is the *trigger*, but the *exposure* was conferred by whoever left the delta open.

That reframe has a person on the end of it, and chapter 9 is about them. Before that, the memos do something rarer than proposing a market: they name the market that already tried this and describe how it failed.

3 · What broke it last time

Part one — The pivot

A body of work proposing insurance for a technology risk owes its reader one chapter before any machinery: the acknowledgement that this has been tried, at scale, recently — and that it went badly for everyone involved. The fourth memo supplies it unprompted, about the market nearest to this one:

> the cyber insurance is already a good example of a market that grew quite high when you know you couldn't really quantify things very well. That then a lot of people lost lost a lot of money, and there was a lot of cyber insurance policies that either were not worth what they they had, or actually had a lot of payouts that really hit the insurers because they they couldn't quantify what's happening

Stated — memo 4, verbatim. Note who got hurt: both sides. Buyers held policies worth less than they thought; carriers took payouts they had not priced. The doctrine document derived from this memo compresses the diagnosis into a sentence this book considers the most important in the corpus after the title:

> A market can be enthusiastically data-driven and still be pricing fiction, if the data is not checkable.

Stated — doctrine 04. Cyber insurance was not short of data. It was short of data anyone could *verify* — self-attested questionnaires, unaudited control claims, quantification nobody could recompute. The market's failure was not an absence of numbers; it was an abundance of unfalsifiable ones.

The failure is the argument for the rule

The pivot's founding rule — stated fully in chapter 4 — is that *a level nobody can recompute is exactly the theatre a premium would have prevented*. It would be easy to read that as a house style, this estate applying its usual reproducibility reflex to a new domain. The cautionary tale upgrades it to an empirical claim:

Drawn. The rule is the thing whose absence broke the nearest comparable market. Cyber insurance is what a rating ecosystem looks like when the derivations are private and the inputs are declared: it grows fast, prices fiction, and detonates on both sides at once. This is why the priority ordering in the doctrine reads the way it does — a rating engine's first obligation is not to be *accurate*, which nobody can be without loss data; it is to be **checkable**, so that being wrong is discoverable rather than accumulating. Accuracy is a destination. Checkability is a property you can ship on day one, and the one the last market shipped without.

What insurance is actually for, per the memo

The same memo makes a distinction that sounds small and decides much of parts two and three:

> it's a market that demands the creation of trustworthy data

Stated. Not *uses* — **demands**. Many disciplines consume data. Insurance creates a party with money at stake in the data being *true*, which manufactures demand for measurement nobody would otherwise fund. That is the honest description of what this pivot does to the estate's existing work: the register, the twin, the evidence pack and the computed tiers all predate insurance and were justified on security grounds. Insurance gives them a second sponsor — and the second sponsor asks harder questions, because the second sponsor is the one who is wrong *expensively*.

Drawn. It also names, precisely, what stage 1 gives up by removing the money — the subject of the next chapter, flagged here so the reader carries it in: with no capital at stake, the demand for trustworthy data must come from somewhere else, and the only somewhere else available is a published method that anyone can attack. Openness, in this corpus, is not generosity. It is the substitute for the forcing function the money used to be.

The quantity security could never say out loud

The memo's second gift is a diagnosis of why security spending has always been hard to defend in a budget meeting:

> if you invest in an incident response team, only what you're doing is you might not be reducing the number of incidents, but you're reducing the impact of those incidents

Stated. Frequency is countable. **Impact reduction is counterfactual** — the disaster that would have been worse appears in no log. Insurance is the one instrument civilisation has built that routinely prices counterfactuals, which is why the memo wants it in the room.

And here the corpus does something worth pausing on, because it is the audit chapter's second defect arriving early: the doctrine's first draft claimed the estate's three-tier control test *was* this quantity — impact reduction, under another name. The audit corrected it. The tier test asks whether a control is enforced by something the grant does not include; that ranks **how much of the grant is reachable in practice** — blast-radius reduction, not severity reduction. An incident response team makes a loss smaller *after* it happens; a boundary makes fewer attempts *become* losses at all. Adjacent quantities, not the same one:

> The tier ranks how much of the grant is reachable in practice. That is blast-radius reduction, not severity reduction

Stated — doctrine 04, as corrected. The doctrine then leans on the proxy anyway, with the lean declared: for an agent placement, what it can reach bounds what it can damage, so blast radius is *most* of severity. That is an argument, not a measurement, and it is weakest exactly where a small reach touches something critical. Nothing in this corpus measures severity at all — memo 4 asked for that quantity, and the estate does not have it. The doctrine says so in its own does-not-prove block, and this book repeats it because a reader who carries away "the tiers price impact" has been misled by one word.

The firm was already underwriting

The memo's last reframe is the one that makes the internal market of part three feel less like an invention:

> you can actually argue that what a company is is fundamentally a gigantic insurance sort of company

Stated. Budget approvals, vendor selections, deployment sign-offs, exception processes: a firm at scale takes underwriting decisions continuously — without a rating, usually without a record, never under that name. *Drawn.* An internal rating market is therefore not a new activity a company must adopt; it is an existing activity made legible. The company was already underwriting the agent. It was doing it in a meeting, and the meeting kept no rating basis. Everything in part two is a proposal about what it would mean to do the same thing with a number somebody can check.

4 · The rule: rating without money

Part two — The rating

The pivot briefing had carried a blocker it called fatal, stated inside the brief rather than softened into an appendix: insurance is a regulated activity; underwriting requires authorisation, capital and a compliance apparatus; nothing this estate ships is a policy until an authorised carrier stands behind it. On that reading, the pivot was a multi-year proposition owned by somebody else.

Memo 1 dissolves it in one move:

> insurance only? It's a decision-making mechanism. It's a way to assign what's it called risk, and it's a way to assign, you know, a rating to a particular environment

Stated — the memo, verbatim, mid-sentence stumble included. Strip the money out and what remains is a **rating engine**. A rating engine that emits *levels* rather than currency transfers risk to nobody and promises nobody a payout — so it is not insurance in the regulated sense at all. It is closer to a credit score or a safety rating: unregulated, sellable, and buildable this week. The memo even supplies the product's units:

> you have an insurance level three, insurance level five, insurance level one, because the point here is also to allow companies to know the risks that they're buying and to allow them to focus their efforts

Stated. The two-stage structure that the whole folder now runs on falls straight out. **Stage 1** produces a level and the derivation behind it; transfers nothing; needs no carrier and no loss history, because a relative ordering needs no absolute scale; and is buildable by this estate today. **Stage 2** produces a premium and a promise to pay; is regulated; needs loss data that exists nowhere for agents; and is not this estate's to build. Everything in this book is stage 1, and the first position of the front matter is a restatement of this table.

The contradiction, kept

Here the corpus does the thing that makes it worth a book rather than a slide. Memo 0's argument for the entire pivot was that money is what keeps a claim honest:

> A premium cannot be theatre; someone loses money if the measurement is wrong.

Stated — brief v0.33.71's reading, which the doctrine carries forward. An acceptance can be a signature over an exposure nobody measured; a priced bet cannot. And then memo 1, one day later, removes the money. Take the money out and that discipline goes with it — a "level 3" issued by the organisation that operates the agent is precisely

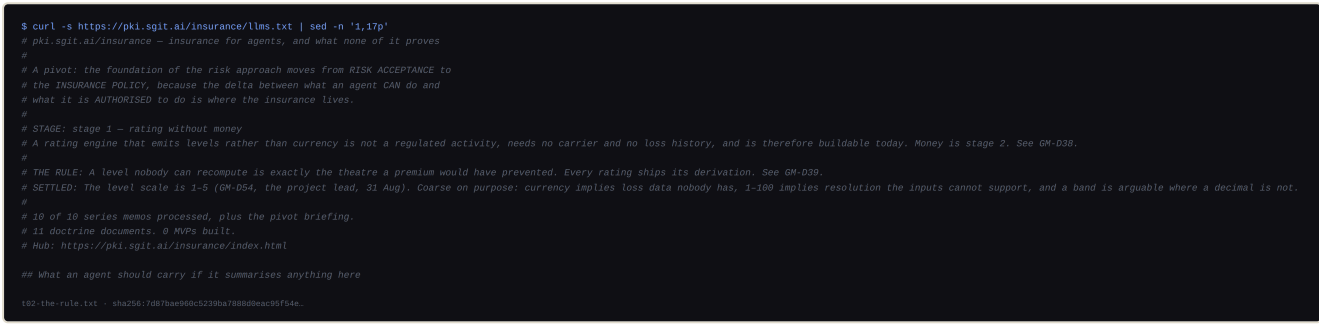
as capable of being theatre as the acceptance the pivot was escaping, *arguably more so*, because a level sounds quantitative.

Drawn. Most programmes would smooth this: a summary would say memo 1 "refines" memo 0. The corpus instead names it a real contradiction and resolves it with a substitute forcing function, and the resolution is the rule everything downstream obeys:

> A level nobody can recompute is exactly the theatre a premium would have prevented.

Stated — doctrine 00, where it is printed as the rule the folder runs on. Money is one way to make a claim honest: someone loses it when the claim is wrong. A reproducible derivation is another: anyone can re-run the computation and catch the claim being wrong. The first is unavailable before stage 2. The second is available now, and it is the same discipline that makes the register publish its expected verification answers as data and reproduce them on every release.

So a rating, in this corpus, is not a number. It is a small document: the level; the inputs, each with its evidence channel and a link to the artefact it came from; the derivation — which inputs moved the level, in which direction, by how much; the twin's age, because a rating computed against a stale measurement is a rating of the past; and what it does not prove, inside the artefact, as everywhere else on this estate. A corollary rides with it: **a level is never declared, only derived** — because an internal market running on a free quasi-currency invites gaming *more* than money does, not less.



What the machine surface promises

The folder's own front door states the stage and the rule before it lists a single memo, and it is worth quoting because it is the sentence an agent reading the site carries away:

> A rating engine that emits levels rather than currency is not a regulated activity, needs no carrier and no loss history, and is therefore buildable today. Money is stage 2.

Stated — `insurance/llms.txt`. *Drawn.* Notice what the honesty is doing commercially. "Not a regulated activity" is usually where technology products go quiet and hope; this corpus prints it as the first line of the pitch, with the reason attached, precisely so the claim can be attacked by somebody who knows insurance law better. The openness is not confidence. It is the rule applied to the folder's own legal position: a regulatory conclusion nobody can check would be theatre of exactly the kind the rule forbids.

Why this ordering is right, and what it costs

Drawn. The deep reason stage 1 comes first is not regulatory convenience. An operator's actual question is almost never *what is this exposure worth in pounds* — it is *which of my two placements is worse, and which single change moves it most*. Memo 1 names that as the goal: to let companies know the risks they are buying and focus their efforts. A relative ordering answers it completely, and a relative ordering is what a rating is. The absolute price only becomes necessary when a third party must hold capital against the answer — which is stage 2's problem, and stage 2's regulator.

The cost, stated with the same directness: a stage-1 level is calibrated against *nothing*. No loss data exists for agents anywhere, so no band means anything outside the method that computed it, and two organisations' level 3s are not comparable until the method is shared. That is why part four's chapter on the not-in-line position treats the shared method as the product, and why the front matter's third position — no external evidence — is not a humility ritual. It is the price tag on everything this chapter just bought.

5 · A placement, not an agent

Part two — The rating

What, exactly, gets rated? Memo 1's own examples answer before the question is asked, and the answer disposes of the framing every public conversation about agent risk uses:

> the insurance for Claude running on a desktop under a user identity should be higher than the insurance of Claude running on the web

Stated — the memo, verbatim. The same model is a different risk in a different place. Rating "Claude" is meaningless; rating *Claude, on this desktop, under this identity, with this credential reach* is a thing an operator can act on. The doctrine turns the examples into the definition:

> A rating attaches to a placement — an agent in an environment under an identity — not to an agent and not to a vendor.

Stated — doctrine 01. And the definition lands on ground the estate has already prepared, which is the pattern of this whole pivot: a **library entry is a placement**. The measured twin of the CCR container rates that container — not Claude, not Anthropic, not "AI". The unit insurance needs and the unit the estate measures were the same object before anyone connected them.

What "micro" is micro about

The memo's recurring word for the product is *micro insurances*, and the definition fixes what the word modifies. Not smaller companies — **smaller units of assessment**: one placement, not one enterprise. That is the scale insurance has historically refused. Cyber cover is bought at the entity; a company buys a policy, a project does not, a service does not, an employee does not. The memo's exceptions prove the rule by their price tags:

> like when you have celebrities or you have somebody who they will insure, you know, you know, the right hand of an individual or a footballer or a golfer or a sportsman

Stated. A footballer's leg gets a policy because the asset is singular, identified, and valuable enough to justify a human underwriter working on one unit. The barrier to the micro was never principle. It was **cost per unit** — bespoke underwriting is expensive because assessing each unit required a person.

Drawn. Here is the argument of this chapter, assembled from the doctrine's table: for an agent placement, every input a human underwriter would spend days collecting already exists as a machine-readable document. Who is the insured — an identity in a register, its fixture-or-real class read before any signature. What can they reach — a grant,

discovered by measurement rather than declared. What are they supposed to do — a signed mandate with an interval. What controls are in place — enforcement tiers computed against the tree rather than claimed. Is the survey current — the twin's age, printed. **The marginal cost of rating one more placement approaches zero**, and that — not any change in what is insurable — is what makes the micro reachable now when it never was. The memo's *"insurance in the past never really scaled, and I think now we are in a position where we can"* is right, for exactly this reason.

The placement variables, and the judgement flag

Memo 1 names four orderings, and the doctrine is careful to file them as what they are — the project lead's judgements, none measured. They are worth walking because each translates into the estate's existing vocabulary rather than needing new machinery.

Identity. Desktop under a *user identity* rates far above a scoped service identity — because the grant becomes the union of everything that user reaches. This is the estate's bootstrap trap arriving as a rating variable: the workaround for agent identity is to hand over a person's, which confers a strictly larger grant than the one being established.

Asset accretion. An account with months of accumulated data rates above a clean one — the memo's own contrast — because blast radius scales with reachable assets, and *time in an account is accretion nobody re-measures*. The grant was fixed at assignment; the assets were not.

Egress. Network egress present rates above none, because egress is what turns every other exposure from theoretical to realisable, and this estate has watched the difference: its own two twins are a container behind a mandatory proxy and a CI runner whose entry reads NO WALL.

Surface. Desktop above hosted web session — and the doctrine notes this follows from the first three rather than standing alone, which matters because it means *desktop is riskier* is a conclusion, not an axiom, and a desktop placement that fixed its identity, accretion and egress could out-rate a sloppy hosted one.

The measurement the corpus cannot make

Drawn. And then the honest sentence that this book considers part two's most important, because it prices everything above: **the estate cannot score its own leading example.** Its two library entries are a CCR container and a GitHub Actions runner. Neither is a desktop; neither is a browser session. The ordering the memo leads with — desktop above web — concerns two placements the estate has never measured. Doctrine 01 draws the conclusion this book endorses: measuring one desktop agent is the cheapest experiment available, and it tests the memo's strongest claim. Until somebody runs `measure.py` on a desktop, the placement variables are a hypothesis with excellent pedigree — and a rating built on them would be a judgement wearing a derivation, which is precisely the object this corpus was built to refuse.

Aggregation, and the trap with a graph-shaped exit

The memo wants the micro to roll up — *"deal in the micro, which then goes to the macro... graphs of graphs of graphs"* — and the doctrine names the trap in the ambition:

> Micro risks do not add.

Stated — doctrine 01. Five hundred placements rated individually and summed will produce an estate rating wrong in the dangerous direction, because the risks are correlated: placements sharing a credential pattern, a base image, a model provider or one misconfigured branch protection fail *together*. This is the oldest problem in insurance — it is why reinsurance exists, and why a flood book is not priced like a fire book.

For this estate it is a graph problem, and that is the good news. Correlation is shared structure, and shared structure is a shared node: two placements whose grant trees converge on the same credential node are not independent, *and the graph already says so*. The doctrine's rule — aggregation reads correlation off shared nodes rather than assuming it away — is a real contribution the graphs-of-graphs framing makes available, and it is stated, not implemented. Nothing in this corpus computes which placements fail together. The rule is where the work would start; chapter 17 keeps it on the list of what nothing proves.

6 · Two channels that never merge

Part two — The rating

Memo 1 reaches for the most familiar object in personal insurance to explain where a rating's inputs come from:

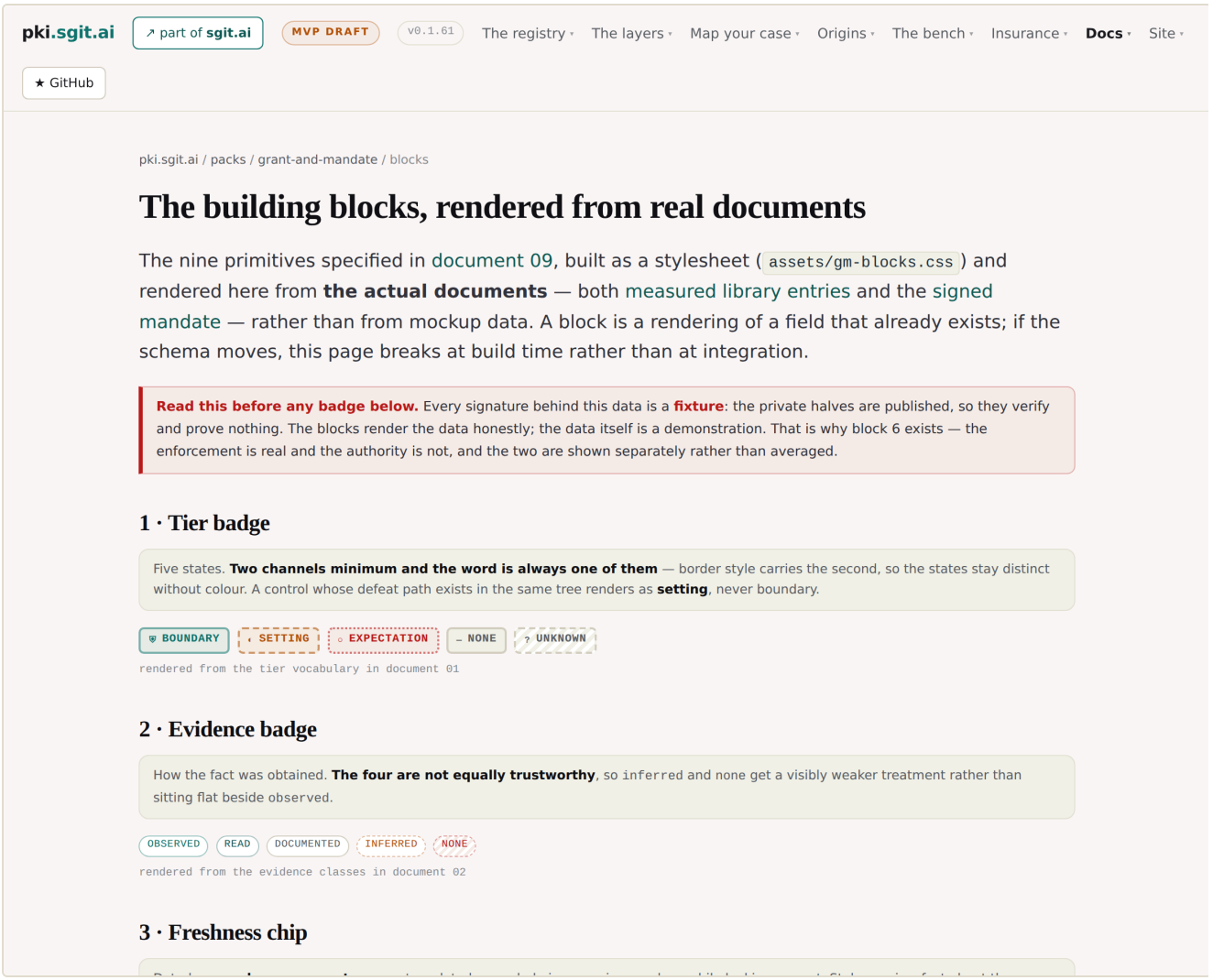
> it's kind of like when you fill a questionnaire to get insurance. You know, like do you smoke? Do you do this? Like wellness insurance

Stated — and for agents: do you have an incident response team, a security programme, a way to contain the agent. The instinct is right, and it collides head-on with a rule this estate had already spent August defending: `library – self-report = blind spots`. **Every answer in that questionnaire is self-reported.** "Do you have a way to contain the agent," answered *yes*, is a declaration. The pre-push hook found by `measure.py` is a measurement. Averaging them launders an assertion into a number — and laundering assertions into numbers is, per chapter 3, the specific way the last market died.

So the rating runs on two channels, and the doctrine's rule is that they never merge:

> **Measured** facts — the twin, the register, computed tiers — carry full weight, and their failure mode is under-reporting: a grant is a floor, not a census. **Declared** facts — the questionnaire, the agent card — are discounted and *rendered as declared*, and their failure mode is over-reporting: nobody answers no to "do you have incident response". **Unknown** is kept as unknown, because scoring unknown as absent is the comfortable error, and it manufactures reassurance.

Both stated and compressed — doctrine 01 §3 carries the full table. None of this is new machinery: it is the estate's five evidence classes — observed, read, documented, inferred, none — applied to rating inputs. The questionnaire is a *declared-class fact collector*, and naming it that is what keeps it honest.



The card, given a job

The memo says the insurance will *"promote the idea of the card"*, and the two-channel rule gives that sentence a precise meaning: **the agent card is where declared facts live.** A card is an agent's self-description — what its operator says it is, runs as, and is contained by. Under the two-channel rule the card is not a brochure; it is one whole side of the rating's input, weighted accordingly, and every claim on it becomes *checkable wherever a measurement exists to check it against*.

Which produces the quantity this chapter exists to flag, because part five will promote it to the most consequential number in the book:

> **The gap between the card and the twin is itself a rating input** — an operator who declares a containment control the twin cannot find has told you something.

Stated — doctrine 01. *Drawn.* At this point in the corpus the card-versus-twin gap is a *rating accuracy* device: a discrepancy worsens your level. Hold the thought. In chapter 16 the same gap returns wearing insurance's oldest and heaviest vocabulary — material non-disclosure, the thing that voids a policy rather than adjusting it — and the estate can compute it today, from two documents it already holds. No other single quantity in this corpus travels that far.

Asymmetry as design, not accident

Drawn. A reader meeting "unknown is never absent" here should be warned that the corpus will appear to contradict it in part four — and that the contradiction is the design. For a **rating**, unknown must never be scored as absent, because assuming absence manufactures comfort: the placement looks safer than anyone knows it to be. For a **cover condition** — chapter 13's warranties — unknown counts as *failure*, because assuming presence manufactures liability: the insurer pays out on a backup nobody verified existed. Same three-valued facts, opposite defaults, and each default is chosen to put the cost of ignorance on the party best placed to cure it. The pair is the clearest example in this corpus of a principle that looks like a rule but is actually two rules wearing one name — and the doctrine states the asymmetry explicitly rather than letting a reader discover it as an inconsistency.

What the connector inherits

One more consequence, recorded here because part four depends on it. The not-in-line position (chapter 14) commits this project to integrating with whatever a company already has — CMDBs, asset inventories, spreadsheets. The two-channel rule prices that promise:

> A connector labels the evidence class of everything it imports.

Stated — doctrine 05, GM-D57. An integration pulling from an asset inventory is producing *declared* facts unless something measures them, and a connector that flattens provenance quietly turns "integrate with what exists" into "launder what exists". Cheap to state on the first connector; expensive to retrofit onto the fifth. *Drawn.* This is the two-channel rule's most commercially annoying implication and its most necessary one: the moment the rating starts consuming customer data at scale is exactly the moment the channels are easiest to blur, and the blurring would be invisible in the output — a level is a level. Only the derivation shows the channel, which is one more reason the rule of chapter 4 makes the derivation mandatory.

7 · One to five, settled

Part two — The rating

For four memos the corpus carried an open question at its centre: levels of *what*? Memo 1 said *"insurance level three, insurance level five"* without fixing a range, and the doctrine recorded the range as open. Memo 5 closes it, and the closure is this pivot's first — and so far only — **settled** decision:

> sometimes even mapping insurance level one to five is good enough, right? Because you you don't have that's less judgmental. Still allows big decisions versus one to 100, or to actually have pound signs, dollar signs connected directly to a particular event

Stated — memo 5, verbatim. The level scale is 1–5, settled by the project lead on 31 August as GM-D54, and the register of that decision matters as much as its content: of the dozens of decisions this pivot has proposed into change control, the scale is the one the project lead has marked settled. Everything else in this book is argued; this is decided.

The reasoning is better than the answer

Drawn. Any scale would have "worked" in the sense of producing output. The doctrine's defence of this one is about honesty of resolution, and it generalises past insurance. A **currency** figure implies a loss distribution — and no agent loss data exists anywhere. A **1–100** score implies meaningful distinctions at one-point resolution — and the inputs are counts, tiers and three-valued facts, which cannot support it; a 62 versus a 63 would be false precision wearing arithmetic. A **1–5** band implies an ordering with defensible boundaries, which is exactly what the inputs can carry. The scale says only what the evidence can say. Coarseness here is not a compromise; it is the resolution the truth actually has.

And the memo's word *less judgmental* names the social property, which the doctrine argues is the practical one:

> **A coarse band is arguable in a way a decimal is not.** A team told they are at level 3 argues about what the band means — the useful argument. A team told they are at 62.4 argues about arithmetic, which is not.

Stated — doctrine 05. *Drawn.* A rating that gates deployments — chapter 8 — will be argued with constantly, by teams with something at stake. The scale decides *what kind of argument* the organisation has. Five bands aim every dispute at the band definitions, which is where the substance lives; a hundred points aims every dispute at rounding, where nothing lives. The scale is, quietly, an argument-routing device.

Fibonacci, and a settled decision defended from a good instinct

Memo 8, three days later, appears to reopen the question. It suggests the levels might follow *"a finibash in a finibashy sequence"* — the transcript's rendering of **Fibonacci** — 1, 2, 3, 5, 8, 13, gaps widening as they rise. Estimation uses that sequence precisely because confidence falls as numbers grow, and the instinct transfers: the exposure difference between a level 4 and a level 5 placement is almost certainly larger than between a 1 and a 2, which a linear scale hides.

The doctrine's handling is a small model of how this corpus treats its own settled ground. It does not reopen GM-D54, and it does not discard the memo's instinct. It splits the question the two were sharing:

> **What do the bands represent — even steps or widening ones?** — **open**, and it is the band-definition question already recorded as unanswered.

Stated — doctrine 08, which holds the settled half beside it: what the bands are *called* is 1 to 5, and that is GM-D54's to keep. Bands labelled 1–5 whose underlying steps widen satisfy both the settled decision and the Fibonacci instinct — so the instinct is filed as a proposal about band *definitions* and put to the project lead, whose decision the scale is. *Drawn.* This is governance working at the scale of a single integer: a settled decision stayed settled, a late idea was neither rejected nor allowed to silently unsettle it, and the audit later verified the transcript repair that made the whole exchange legible. It is also the book's answer to a reader who suspects "settled" means "until the next memo": the corpus had a chance to treat it that way, and did not.

What the bands do not have

Drawn. Honesty about what remains: the bands have no definitions. What separates a 2 from a 3 — which inputs, at which thresholds, moved by which controls — is exactly where the argument will happen, and it has not happened. Doctrine 05's does-not-prove says so; nothing validates five bands against outcomes because there are no outcomes; and a second rater, whose disagreement would be the first signal that a band boundary is in the wrong place, does not exist. The scale is settled the way a chessboard is settled before a game: the geometry is fixed, and nothing about the play has begun.

8 · The roles without the money

Part three — Who pays, who rates, who backs the claim

Memo 2 asks the question memo 1's rating had no home for: how does an insurer-like ecosystem run *inside a company*, with no financial numbers? Its answer is to take the industry's **roles** and leave its **money** —

> the components of the insurance industry have again they battle hardened. Like you know, we shouldn't be reinventing the wheel

Stated — memo 2, verbatim. And the doctrine's sharpening is that the roles are not decoration around the money; they are the part that makes an assessment worth anything at all. An insurer is a third party for a structural reason: **an assessment produced by the party that wants the answer to be yes is not an assessment.** Outside a company, the market supplies that separation for free. Move the ecosystem inside and the separation has to be manufactured — the insured is the team that wants to ship; the underwriter is the rating authority, and *not* that team; the capital is absent in stage 1, deliberately; and the regulator's seat is taken by the published derivation, arguable by anybody.

The money-holding version of this arrangement already has an industry name — a **captive insurer**, the subsidiary a large firm forms to insure its own risks. Stage 1 is a captive with the capital removed, which is exactly why it is unregulated, and worth knowing because it names what stage 2 would turn the arrangement into.

The rule, completed

Chapter 4's rule guaranteed the arithmetic could not be invented. Memo 2 supplies the half the rule was missing:

> A level computed by the party that wants to ship is theatre even when it is recomputable.

Stated — doctrine 02. Method and separation, both: a reproducible derivation stops the numbers being made up; an independent rater stops the *inputs* being chosen to flatter. Neither substitutes for the other, and the corpus now has both halves of its integrity story — one from memo 1, one from memo 2, neither sufficient alone.

From report to gate

The memo's second move puts the rating somewhere it can refuse things:

> It's an insurance that can then be used to basically define: Do we go live with this product?

Stated. A rating that reports says *here is the level* and can be ignored silently. A rating that gates says *not until this changes* and can be ignored only visibly. The difference in requirements is the finding: a gate needs a **threshold**, and it needs a **decomposition** — because *reduce the risk by this quantity* is unsayable unless the rating decomposes into the inputs that moved it and the changes that would move them back. The derivation the rule demanded for honesty turns out to be the actionable half of the gate: a rating that ships its derivation is the only kind you can be asked to *improve*.

The gate has a tier, and the estate already owns the test

Part three's most reflexive move, and this corpus at its most characteristic: the insurance apparatus must pass the estate's own control test — *a control bounds a grant only when it is enforced by something the grant does not include.* Where does the gate live? A dashboard someone is meant to check is an **expectation** — nothing stands between the deploy and production but intention. A CI check the deploying team can override or edit is a **setting** — inside the grant it bounds, the same failure as the estate's own pre-push hook. A required check evaluated by a party the deploying team does not control is a **boundary** — the roles' separation, made mechanical.

> **So the rating engine declares its own tier, on its own face** — as the mandate hook's refusal banner does. A control that overstates itself is worse than none; an insurance gate that overstates itself is worse still, because it will be believed

Stated — doctrine 02. *Drawn.* The roles and the tier are the same question asked twice — an underwriter who *is* the deploying team produces a setting no matter how good the arithmetic — and the demand that the gate print its own tier is the single most transferable sentence in part three. Every governance mechanism ever installed has had a tier; almost none has ever said which.

The moral hazard swap

Memo 2's sharpest line separates *insurance as a risk decision mechanism* from *insurance as something you offload the risk into and forget* — and conventional cover has a documented failure mode of exactly that shape: cover substitutes for control. If the loss is paid, the incentive to prevent it weakens, which is why real policies bristle with deductibles, exclusions and warranties.

Stage 1, with no payout, cannot be used to stop paying attention *on the grounds that a loss would be covered*. The doctrine's first draft called that immunity, and the audit corrected it into the more interesting truth:

> Removing the payout removes one channel of moral hazard and opens another

Stated — doctrine 02, as corrected under GM-D84. A level is a badge, and a badge substitutes for control in its own way: *we are a level 2* is available as a reason to stop looking, with no carrier involved. Part five meets the same hazard from the other end — *make your agents insurable* must not become *make your agents look insurable*. The rating's channel is cheaper to police, because a derivation anybody can recompute is a badge anybody can dispute — but cheaper is not closed, and the corrected sentence is the honest one.

What the rating cannot answer

Memo 2 asks whether the value an agent adds exceeds the risk being bought. The rating prices one side of that question, and the doctrine refuses the other half in a sentence this book wants every instrument to carry:

> The rating prices one side of a two-sided question, and says so.

Stated. Valuing the agent's contribution — throughput, cost displaced, quality — needs data this estate has no access to and no business holding. The same division of labour lets a surveyor value a building without deciding whether you should buy it. *Drawn.* An instrument that knows which half of a decision it informs is rarer than it should be; most dashboards answer the half they can and let the reader assume it was the whole. The rating's refusal is a feature to defend, and the world model of chapter 15 inherits it: the walkthrough ends at a decision the player makes, not a verdict the world hands down.

9 · Who pays for the delta nobody chose

Part three — Who pays, who rates, who backs the claim

Two chapters of rating machinery have quietly assumed the delta belongs to whoever operates the agent. Memo 3 asks the question that assumption skips, and asks it with the estate's own worked example. A code host offers read-only or read-write and nothing between. An operator who needs an agent to push at all must confer the ability to push to **every branch** — and, absent signed-commit enforcement, to author commits as anybody:

> right now my only options is to give it read only or read write. Right, I don't have more granularity than that

Stated — memo 3, verbatim, describing GitHub, though the shape is any platform. The mandate says *your own branch, and dev*. The gap between what was conferred and what was authorised is real, dangerous — and **the operator did not choose it and cannot close it.** No budget, care or diligence makes a platform offer a finer grain.

The taxonomy

So the delta divides by *who could have closed it*, and the doctrine's three classes are the pivot's fairness mechanism:

> **Elective** delta — more conferred than needed, or an available control skipped: no branch protection, no signed commits, a token scoped to every repo when one would do. The operator can close it, so it is the operator's — the part effort moves, and the part a rating should charge for. **Structural** delta — the platform's finest grain is coarser than the mandate. Nobody's budget closes it; it is identical in every customer's estate at once, and it is the platform's, or nobody's. **Defect** delta — a vulnerability temporarily widens the grant past its documented shape. The operator could not even have known, and it is temporary, which is why chapter 10 exists.

Drawn as a compression — doctrine 03 §1 carries the full table this paragraph compresses. The conclusion, though, is printed there as a rule:

> A rating that does not separate these is unfair and useless in one move.

Stated. Unfair, because it charges for the unfixable. Useless, because a level an operator cannot move is not a decision input. Chapter 8 required the derivation to decompose; this chapter says the decomposition's first cut is *what they can move and what they cannot*.

The class the taxonomy could not hold

Three memos later the taxonomy met a case that did not fit, and the way the corpus handled it is worth recording as method. Memo 6 names platforms that *do* offer the granularity — and make it so complex that using it is impractical: *"when they do, they make it so crazy complex that it's again is impractical to have that lowest privilege execution."* That is not structural; the grain exists. Calling it elective charges an operator full price for not adopting something that takes two quarters and a specialist.

A fourth class was the tempting fix. The doctrine refused it:

> **Elective delta carries a cost to close.** *Latent* delta — offered but impractical — is simply its high-cost region

Stated — doctrine 06, GM-D60. *Drawn.* The refusal matters more than the fix. Taxonomies rot by accreting classes for every awkward case; a dimension added to an existing class keeps the structure falsifiable and gives the rating something a fourth class never would: *you could close this in an afternoon* and *you could close this in two quarters* are different instructions, and a fair rating distinguishes them without inventing a new kind of blame.

The only public good in the apparatus

The structural class has a corollary the corpus flags as unique. How coarse a platform's permissions are is a **fact about the platform** — identical for every customer. Measure it once, publish it once, reference it everywhere; never re-derive it per estate. That is the estate's library/instance rule one level up, and it is the only part of the whole insurance apparatus that is a public good: one honest measurement of a platform's permission model serves every customer of it.

Drawn. It also hands the platform an argument aimed at itself, which the doctrine states as countable: a platform that shipped branch-scoped tokens would convert a whole class of delta from structural to elective *for every customer simultaneously*. Structural delta's natural payer, meanwhile, is the pool — a risk nobody can individually avoid and everybody shares is the textbook shape of pooled cover. The taxonomy thus prices three different conversations: the operator's (close your elective delta), the platform's (your coarseness is countable), and the market's (the structural remainder pools or goes uncovered).

The vendor who would underwrite itself

Buried in memo 3's list of candidate payers is the question this book regards as the most forward-looking sentence in the series:

> is Claude going to actually have a policy on its own to cover any mistakes?

Stated — memo 3, verbatim. Today the answer is no; model vendors disclaim liability for output, and that is the industry's settled position rather than an oversight. But the doctrine draws out what an answer would *mean*: a vendor offering cover against its own model's mistakes would be making a **falsifiable quality claim** — the first in the

category. Everything a model vendor currently publishes about reliability is a benchmark or a disclaimer. A policy is neither: it is a number somebody pays when they are wrong.

The rule generalises, and the corpus generalised it three memos later when the same structure reappeared in the broker channel:

> any party claiming to reduce somebody else's exposure should carry cover against being wrong about it

Stated — doctrine 03 §8, GM-D83, flagged in place as speculation about the vendors and structure about the claim. Chapter 11 is that rule meeting the market it was made for.

The question this estate refuses to answer

Memo 3 ends in the territory every agent-risk conversation ends in — *who's responsible for that hallucination?* — and the corpus's response is a model of staying in competence:

> The estate cannot say who pays. **It can make the question answerable rather than a swearing contest**, which is what the parties actually lack.

Stated — doctrine 03 §7. Liability for agent harm is unsettled law, being worked out in courts and contracts, and a schema project pretending to settle it would be theatre in a wig. What any liability determination will need is exactly what an evidence pack already is: what happened, under whose authority, against which mandate, with which controls in place, dated and signed. *Drawn.* This is the delta taxonomy's real product. Not an allocation of blame — an end to the era in which the allocation had to be fought over with no shared record of what the delta even was.

10 · The rating that moves

Part three — Who pays, who rates, who backs the claim

Everything to this point rates a snapshot. Memo 3's second half asks what happens on the day the world changes and the snapshot does not:

> There's a vulnerability on GitHub that allows Git, you know, any repo, any agent to now do a lot more damage on GitHub, right? Or even maybe access other GitHub accounts, right? Ultimately, it's your agent, it's the company's agent who's doing all that stuff, right? So we, you know, immediately the insurance here could change, right? It could change overnight, could change in an hour or in half an hour

Stated — memo 3, verbatim. The analytical move underneath is the chapter: a vulnerability lands and **the twin does not change**. The measurement recorded what it recorded; the credential still reaches what it reached. What changed is what that reach is *worth*. So the rating is a function of two things with independent lifetimes —

> **rating = f(placement twin, world state, mandate)** — and the twin and the world have **independent freshness**.

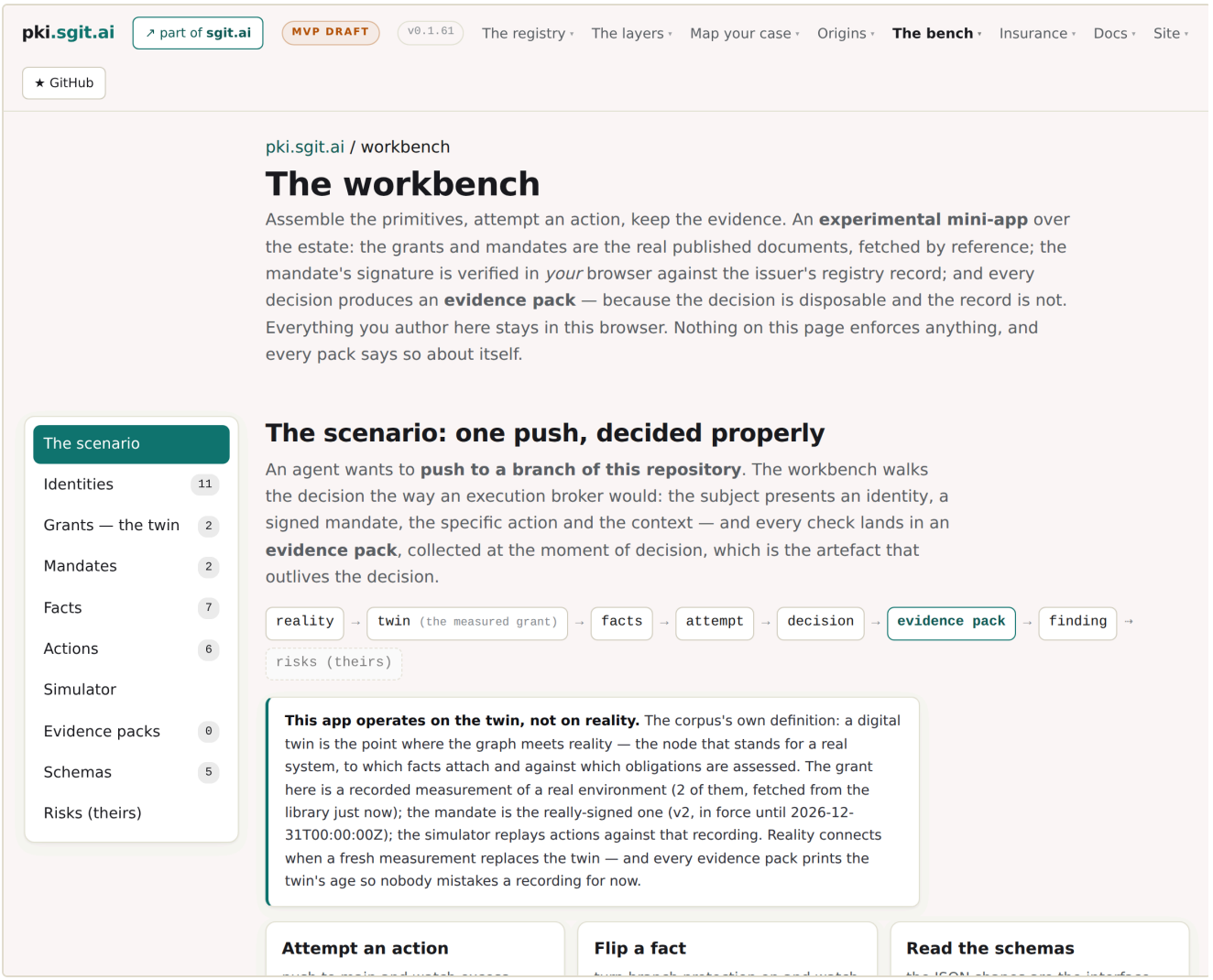
Stated — doctrine 03. The estate already prints twin age on every evidence pack. A moving rating needs a second age: how current is our picture of the world? A rating computed against a six-day-old twin and a three-minute-old advisory is a different object from one where both are stale, and only saying so keeps it honest.

What it may say, and what it may not

The memo's instinct — *"has just gone 10x"* — gets the corpus's characteristic split: right instinct, wrong output. A magnitude needs loss data nobody has, and printing a multiplier would be false precision of exactly the kind the rule forbids. What is computable today is the **direction and the mechanism**: *this placement's grant now reaches repository settings, which it did not yesterday* — a named capability and a node count. *A control that was a boundary is now defeated* — a tier recomputation the workbench already performs. *Two mandate constraints are now unenforceable* — a mandate-versus-grant recomputation.

> A re-rating states what changed, which way, and which controls are implicated. Never a multiplier

Stated — doctrine 03, GM-D48. *Drawn.* The deeper point is who the output serves. An operator deciding whether the value still justifies the risk is better served by *your agent can now change repository visibility* than by *10x* — the first is actionable and checkable, the second is a number they cannot argue with. The moving rating keeps the rule of chapter 4 under pressure: even in an emergency, especially in an emergency, the derivation is the product.



The loop that needs no new machinery

Memo 2 had already closed the loop in the other direction — *when you invest in a control to reduce a risk, you then are ultimately reducing the or changing the insurance premium* — and the estate's workbench already runs it: flip the branch-protection fact and the enforcement tier moves from setting to boundary, computed from the facts rather than stored. Rename the output and the mechanism is the rating's counterfactual view: *what would this control buy you here.* Two properties fall out. It is scale-free — ordering controls by how much they move a rating needs no agreed range, so the counterfactual view does not wait on band definitions. And it is the honest form of the questionnaire: chapter 6's wellness form asks *do you have a control*; the loop asks *what would this control buy you*, prospective, placement-specific, computed from that placement's own tree.

Pulling the plug, tiered

The memo's operational endgame — *do we pull the plug, right, temporarily, or do we pull the plug for a week or a month or a day or an hour* — extends the go-live gate into a continuous obligation, evaluating on world events when nobody is asking. The far end of the mechanism already exists: a mandate is revoked by an append to the issuer's record, and the enforcement path from revocation to a refused push was demonstrated at v0.1.28. So *a rating crossing a threshold may emit a revocation* — everything except the world-state input exists.

And the tier question bites hardest exactly here:

> **a "pull the plug" that emails somebody is an expectation; one that revokes a mandate the agent's own hook honours is a setting; one the agent cannot reach is a boundary.** It declares which.

Stated — doctrine 03. The memo's alternative — *buy temporary cover to keep operating* — has a moneyless analogue the estate already stocks: a time-boxed exception, recorded, that expires by itself, which is simply a mandate with a short interval. *Drawn.* It says something about the design that the emergency workflow decomposes entirely into parts the register already ships — an append, an interval, a recomputation — with exactly one missing input. Which is the honest place to end the chapter:

The hard part, named

The moving rating needs a feed of platform-affecting events, a re-rating trigger — both mechanically small — and one thing that is not small:

> **Does not exist anywhere.** Advisory feeds describe software, not capability trees

Stated — doctrine 03's requirements table, against the row that names the mapping from an event to the grant nodes it widens. *"This advisory widens node n4 from in-scope repositories to all repositories"* is a judgement somebody makes and records, and it is a library artefact in chapter 9's sense — made once, used by everyone. Nobody has made it, for any platform, ever. Dynamic re-rating is this corpus's most attractive promise and it rests entirely on an unbuilt dictionary; doctrine and does-not-prove both say so, and so does chapter 17.

11 · The broker market

Part three — Who pays, who rates, who backs the claim

Chapter 9 ended with a rule looking for a market: any party claiming to reduce somebody else's exposure should carry cover against being wrong about it. Memo 6 supplies the market. Access and execution brokers for agents — products that hold the credential, narrow the action, return a receipt — are emerging, and their commercial case is precisely the reduction they produce:

> the practical and the business case for deploying these brokers is that they reduce the insurance level of a particular deployment

Stated — memo 6, verbatim. And the claim's problem is the corpus's oldest concept arriving in a new channel. *Use our product and your exposure drops* is an assertion of authority nobody granted — and it binds anyway, because the buyer acts on it. That is **apparent authority**, the thing this whole estate exists to make visible, showing up not in the agent channel but in the vendor channel.

Two mechanisms convert the assertion into something checkable, and the memo demands the first itself:

> they should actually have an insurance policy at their end that will underwrite any mistakes

Stated. A broker carrying its own policy turns marketing into warranty: a vendor who is wrong pays, so a vendor who is wrong has reason to be right. The doctrine adds the second, sharper mechanism:

> A broker's claimed level reduction is computed by the rating method, never by the broker

Stated — doctrine 06, GM-D58. Otherwise each vendor measures its own reduction favourably and the number is marketing wearing arithmetic — chapter 8's separation rule, in a channel it did not anticipate: not the deploying team this time, but the vendor selling to them.

The correction: a broker changes the shape, not the size

This chapter contains the corpus's most instructive self-correction. Doctrine 03 had priced a broker cleanly: its value is exactly how much structural delta it converts to elective. True — and incomplete, and memo 6's own example is what breaks it. Interposing a broker **removes reach from one tree and adds a party with reach of its own**: the platform credential moves from the agent to the broker; the agent narrows to what the broker permits; and a new node appears that sees every request, holds a credential reaching everything the agent used to, and — if the broker is SaaS — carries decisions across the organisation's boundary.

> The rating nets what a broker removes against what it adds, and the net is not automatically negative

Stated — doctrine 06, GM-D59. A SaaS broker that narrows platform access while routing every request through a third party has *moved* exposure, not removed it. An in-environment broker with no egress is a different proposition and a different rating — which makes **deployment topology a first-class rating variable**, and among the more computable ones: *does a decision leave the boundary* is a question a twin can answer. The memo saw this itself, contrasting the centralised SaaS broker with one that runs inside the environment, locked down, no internet access.

Drawn. The correction was recorded in *both* documents — the one that was wrong and the one that found it — under the corpus's rule that a correction only one document knows about is not a correction. That practice, more than any individual claim, is what this book means when it calls the corpus auditable.

Systemic, and therefore not optional

The broker's rating has a property no other placement's has:

> A broker's mis-rating propagates to every customer whose level it reduced.

Stated — doctrine 06. It is chapter 5's correlated-risk problem in its most concentrated form — one node shared by every placement that trusted it — and the reason the broker's own policy is not a nice-to-have. The estate's earlier service-twin work had already warned that a broker becomes the highest-value target because it must hold usable credentials; the rating vocabulary sharpens that from a warning into a structural fact.

A transitional market, entered anyway

The memo names the gap that makes the market exist:

> ideally the permission should be done dynamically, the grant should be done dynamically, and we should be using PKI and other authorization models and other modes to operate

Stated. So: **the broker market exists because dynamic, checkable, PKI-backed authorisation does not.** Two consequences point opposite ways, and the doctrine says both. It is the clearest commercial argument for the register — if the primitive existed, a category of intermediary would be less necessary. And the broker market is transitional by construction: a vendor whose product is *we absorb the complexity the platform imposed* is betting the platforms stay that way. Worth knowing on both sides of a sale — and not a reason to avoid the market, since the transition may be long.

Drawn. The chapter's last word belongs to the shared method, because it is what makes any of this a market rather than a shouting match: a buyer cannot weigh broker A's claimed two bands against broker B's claimed three if they were computed differently. Comparability is what the connective-tissue position of chapter 14 sells — the party

defining the disclosure format decides what *good* is measurable as, while carrying none of the capital. Whether any broker submits to having its headline number computed by someone else's method is an open question; doctrine 06's does-not-prove lists it first.

12 · The policy is a statement

Part four — The machinery

Part four is where the pivot stops describing a market and starts describing an implementation — and discovers, three times in three chapters, that the implementation mostly exists. Memo 7 begins it:

> a policy becomes something that can sign something. So you. So what's cool about this is you can now, for example, say, hey, I'm only going to give you the data if you have a particular policy from an entity that I trust

Stated — memo 7, verbatim. The doctrine untangles the two mechanisms hiding in that sentence, and the untangling is the design. First: what *is* a policy, as a document? The register already holds statements with an issuer, a subject, a scope, an interval, a revocation path and a signature — mandates. Lay a policy beside one and the rows line up: who authorised becomes who rated; capability-on-resource becomes the level and what it is contingent on; valid-from-until becomes the cover period; revocation is an append to the issuer's record in both.

> policy/v0 is a mandate-shaped statement issued by a rater rather than an operator

Stated — doctrine 07, GM-D62. One more statement type beside identity, mandate, acceptance, revocation and grant — **no new register machinery**, and the record pages would render it the day it existed.

A policy never signs

The second mechanism is the one the memo's phrasing blurred, and the corpus had pre-paid for the clarification eleven days earlier:

> A signature proves possession of a private key and proves nothing about trustworthiness.

Stated — the estate's own line of 19 August, before any insurance work existed, re-verified verbatim by the audit. A **holder** signs a request, proving possession of the private half of the identity the policy names as subject. The **policy** is signed by its issuer and never signs anything: a key signs; the policy establishes what that signature is *worth*. Conflating the two would have produced a confused object — a certificate that acts — and the estate's oldest discipline is what prevented it.

So the handshake is three checks, all of which the register already answers: is this the subject (the signature verifies against the published key); is a policy in force for that subject (a statement with an interval and no revoking append); does its issuer mean anything to me (the relying party's own root decision — which the register deliberately makes for nobody).

```
$ python3 packs/grant-and-mandate/tools/mandate.py check-branch dev
PERMIT   dev  (mandate v2, allow=['claude/**', 'dev'], expires 2026-12-31T00:00:00Z)

$ python3 packs/grant-and-mandate/tools/mandate.py check-branch main
REFUSED  main  (mandate v2 permits ['claude/**', 'dev'])
$ echo "exit: $?"
exit: 1
```

t03-permit-refused.txt · sha256:90c8a1f3401b6f82655e679769f5971d…

The boundary, found

Chapter 14 will record the position this pivot accepted early: a party outside the line cannot enforce, so this project can never itself be a boundary. Memo 7 resolves the resulting question — *then who can?* — and the resolution is the pivot's structural high point:

> I'm only going to give you the data if you have a policy

Stated. The check lives in the **relying party** — the system being asked — and a relying party is *by construction* outside the requesting agent's grant. It holds the data; the agent cannot reach in and disable its check.

> **The policy handshake is the first mechanism in this pivot that can reach tier `boundary`** — and it gets there without this project standing in the line, because the enforcement point is the counterparty

Stated — doctrine 07, GM-D64. With its condition attached, because a tier is always a property of a relationship: it is a boundary only where the relying party is genuinely independent of the requester. An internal service enforcing a check on an agent run by the same team, on infrastructure that team administers, is back to a setting. *Drawn.* The condition is not fine print; it is the whole chapter in one sentence. Enforcement strength is not a feature a product ships — it is a fact about who controls what, and the handshake's brilliance is that it recruits the one party whose incentives and position already face the right way.

Revocation is bounded by attention

The memo wants dynamic revocation — a zero day lands and *within 10 minutes or within an hour, we will revoke access to the policy*. The corpus had, again, already priced the promise, in an observability brief from 20 August:

> revocation propagates only at the rate relying parties check, which makes the interval between a party's checks its effective revocation latency

Stated — re-verified by the audit at its claimed date. A policy promising ten-minute revocation delivers ten minutes only to parties who check that often; to a party caching for a day it is a one-day promise wearing a ten-minute label. So *a revocation SLA states the required check interval, or it is a hope with a number on it* — and the handshake turns

out to be the fix as well as the feature: a relying party verifying on every request has a revocation latency of one request, the shortest achievable. The doctrine notes the memo did not make that argument for its own design, and makes it: the handshake is not only how you check — it is what makes fast revocation mean anything.

Metering the check, and the three hazards

The memo spots a business model in the verification step — the provider learns how often a policy is used, and can charge for the check. The corpus's three-part reply is the estate's observability doctrine doing exactly what doctrine is for. A verification is not a use — the metric is a verification graph, and the most valuable signal in it is the silent case: parties holding a policy who have never once checked it. Metering means the provider learns who checks what, when — and the estate had already settled where such events belong: written by the checker into the issuer's own lane, an owner observing their own asset, never a central log accumulating who evaluates whom across parties that never consented. And pricing the check taxes the behaviour the system most wants — rational relying parties would cache, batch and skip, raising the very latency the handshake exists to lower:

> A per-verification price is a tax on the behaviour the system most wants

Stated — doctrine 07, GM-D65. Price by seat, policy or period instead. *Drawn.* Three memos of part four, and each lands on the same shape: the insurance layer keeps proposing things the estate's pre-insurance doctrine already had opinions about, and the opinions hold. That is either a corpus exhibiting suspicious self-consistency or a design that was, in fact, converging — chapter 17 gives the audit's evidence for the second reading.

Who checks first

The operating-licence pattern — *without that, we cannot operate* — is culturally legible to a business in a way no security score is: regulated firms cannot trade uninsured; contractors cannot start without certificates. The hard part is not the mechanism but the first mover, and the doctrine's honest table gives it: internal services, whose organisation can simply set the rule — the internal marketplace with the handshake as its enforcement, needing nobody else's cooperation; then brokers, out of self-protection, since their own policy is exposed to what they let through; external platforms not yet, because no incentive exists. Nothing has adopted it. Doctrine 07's does-not-prove says so, and the pattern needs an adopter before it is a pattern.

13 · The three clocks

Part four — The machinery

Memo 10 is two memos in one — the interfaces, then time — and its worked example is the sharpest cover condition in the series:

> it's only valid if you have a backup with less than 24 hours. So that means that you know you cannot make a change to the database if the backup, for example, last backup failed, or if you don't have evidence that the last backup is actually operating

Stated — memo 10, verbatim. The doctrine reads a definition out of it, and the definition is the chapter:

> **A warranty is a fact plus a maximum age** — `{fact: last_backup_succeeded, max_age: 24h}` — and it fails **three** ways: the fact is false, the fact is stale, or the fact is **unknown**.

Stated — doctrine 10, GM-D76. The estate already computes both halves — every evidence pack prints the twin's age because measurements go stale, and facts are already three-valued. A warranty is those two mechanisms combined and pointed at cover instead of at a level.

The third failure mode is the deliberate one. Putting **unknown on the same side as false** is the exact opposite of chapter 6's rating rule, and the asymmetry is design: for a rating, assuming absence manufactures comfort; for a cover condition, assuming presence manufactures liability. The corpus states the pair as one sentence so it cannot be read as an inconsistency, and this book has now kept its chapter 6 promise to show the other rule of the two.

The gate climbs to its ceiling

Part three watched the gate evaluate once, at go-live, then continuously, on world events. The backup warranty completes the climb: **per action** — *you cannot make a change to the database if…* is a condition checked every time the agent acts. Strongest, most expensive, and the mechanism already exists: chapter 12's handshake, checking on every request, checks the warranties on every request too. The three-row table in doctrine 10 is the whole architecture of enforcement timing, and each row cites the memo that demanded it.

Three clocks, and how a system like this lies

Cover, it follows, is *continuously conditional* — bounded by an interval and alive only while its conditions hold. So a policy runs on three clocks, and conflating them is how a system like this would lie:

> The policy interval — from when, until when cover exists at all. Twin and world freshness — how current the measurement and the threat picture are. The warranty's check interval — how recently each cover condition was confirmed.

Drawn as a compression of doctrine 10 §3's table, which carries the three rows verbatim. The doctrine's example is the one to keep: **a policy in force, computed against a fresh twin, with a warranty last checked eleven days ago, is not covered** — and only a system keeping the third clock separately can say so. It is chapter 12's revocation-latency lesson generalised: every promise in this architecture is bounded by the interval at which somebody actually looks, and the honest design is the one that prints the looking-interval next to the promise.

Even the small clauses carry the discipline. Cover *only over office hours* is a real control — an agent that may act only while somebody is awake has a smaller effective blast radius — and it is an interval with recurrence, a shape the mandate almost has. The trap is that a recurring window without a timezone is ambiguous twice a year, and ambiguity in a *cover* condition is the ground disputes are fought on. *A recurring cover window states its timezone* — cheap now, unpleasant to retrofit into policies already sold.

Metering uses, not checks

The memo prices a policy by consumption — *ten usages over a period of an hour* — and the corpus checks the idea against chapter 12's rule that pricing verification taxes the wrong behaviour. The idea survives, on the estate's own earlier distinction: **a verification is not a use.** Metering uses is what insurance has always done — per flight, per shipment — and prices what the buyer consumes; metering checks taxes the checking. But the survival has a cost the not-in-line position must pay:

> A usage-boxed policy needs an in-line counter, which this project is not

Stated — doctrine 10, GM-D78. To count uses you must observe uses; the relying party can, an execution broker can, and the rater — beside the line, by its own constitution — sees nothing. We define the counting schema; somebody else holds the counter. *Drawn.* Note the pattern completing: every time the machinery needs an in-line capability, the corpus assigns it to a party already in line rather than quietly stepping in — the relying party enforces, the broker counts, the platform grants. Chapter 14 is why that assignment discipline is the position rather than a modesty.

Schemas, not APIs

The memo asks what APIs the ecosystem needs. The answer follows from the position:

> An API is operated. A schema is implemented.

Stated — doctrine 10. An operated API puts this project in the line it swore off; a schema is implemented by whoever is already there. So the deliverable is the policy lifecycle as documents and the transitions between them — and the estate already has every piece: a policy is a signed statement appended to an issuer's record; its criteria are the

statement's own warranty fields; hashes over the canonical form are already computed; certifications are signatures verified against published keys; and amending, activating, ending are **appends, never edits** — the register's rule since v0.1.26. The answer to *what APIs do we need* is mostly: none that we build. Interoperability lives in the documents; the API is whatever each party puts in front of them.

Drawn. One loose end is preserved exactly as the corpus preserved it, because it is a model of the discipline: the memo's final sentence cuts off mid-thought — *"a 24-hour period where within that window"* — and the doctrine declines to finish it, on the grounds that guessing the end of a sentence and rendering it as the project lead's position is the same error as guessing a survey's findings. The transcript ends where it ends. The schema for warranties, meanwhile, has never been written down; doctrine 10's does-not-prove keeps that on the record, and chapter 17 counts it.

14 · Not in line

Part four — The machinery

Every architecture this corpus builds eventually states where its author refuses to stand, and for the insurance pivot the statement arrives in memo 5 with a self-correction in the middle of it — the most consequential eight words in the series:

> our job is to be the broker, right? Not the broker.

Stated — memo 5, verbatim. The sentence corrects itself because one word had been carrying three jobs, and the audit later confirmed the disambiguation as the reading that kept the doctrine coherent. The **execution broker** holds the credential and performs the action — in line, and foreclosed. The **insurance broker** places risk with carriers and holds the client relationship — in the money path, regulated, and foreclosed. The **connective tissue** — schemas, flows, mappings, connectors, evidence — is beside the line, and is the job:

> we basically are basically in in that loop, right? And and that's that's our bit

Stated. Everything open source, everything Creative Commons, integrate with whatever maturity a company already has. That is the position. The rest of the chapter is what it costs and what it buys, because the corpus refuses to let a business model pose as a neutrality.

What the position forecloses, by the estate's own test

A party outside the path of the transaction cannot enforce anything. By the estate's own tier test — a control bounds a grant only when it is enforced by something the grant does not include — a schema, a connector and a published derivation are none of those things:

> **This project can never itself be a boundary.** It ships a check that *becomes* one when installed by somebody who is in line.

Stated — doctrine 05. Compute the rating and publish the derivation: instrumentation, expectation tier, this project. Install the gate in a pipeline the deploying team controls: setting, the customer. Install it where the deploying team cannot reach: boundary — the customer, or a broker. Every tier above the first is somebody else's action. The doctrine turns the concession into the honest pitch — *we can tell you what you are carrying; whether anything stops it is your install* — and chapter 12 recorded the one mechanism that squares the circle: the relying party, who is in line by construction and enforces for reasons of its own.

Drawn. Read as strategy, the foreclosure is the moat's price. The party defining a market's disclosure format decides what *good* is measurable as, while carrying none of the market's capital or liability — that is what "we define what a broker must be able to show" buys. The cost is that the definer can never be the enforcement it defines, and every deployment of its work depends on somebody else caring enough to install it. The corpus chose the schema seat with both eyes open, and wrote the eyes-open part down.

Openness is load-bearing

Three separate arguments in the corpus converge on the same requirement, and their convergence is the strongest structural claim in part four. **Openness is the substitute for capital**: chapter 3 established that insurance's virtue is creating a party with money at stake in the data being true; stage 1 has no such party, so the demand for trustworthy data must come from the method being attackable — which requires it to be public. A closed rating engine in a stage with no money has no honesty mechanism at all. **Openness is the monoculture mitigation**: chapter 5's correlation problem applies to the rating standard itself — a single standard everyone adopts is a concentration risk one altitude up, and a standard anyone can fork, audit and dispute is a monoculture that can be broken on purpose. **And openness is the only way the position pays**: a schema's value is proportional to adoption; a proprietary interchange format nobody else implements is not connective tissue but a product with an integration problem.

Drawn. Note what kind of claim this is. Most open-source commitments are values statements; this one is a dependency graph. Remove the openness and stage 1 loses its honesty mechanism, its systemic-risk mitigation and its revenue logic simultaneously. The corpus does not say "we believe in open"; it says the position is insolvent without it, which is a stronger and more falsifiable thing to say.

Acceptance, returned to its seat

The pivot began, in chapter 1, by replacing acceptance with the policy at the foundation of the pyramid. Memo 5 quietly completes the thought by putting acceptance back — not as the foundation, but as the thing a level makes possible:

> A rating does not replace acceptance. It makes acceptance specific.

Stated — doctrine 05. *We accept the risk of this agent* is a sentence nobody can check. *We accept a level 4 placement, for this quarter, on this service* has a subject, a threshold and a date — which is what the register was built to hold. *Drawn.* The pivot's arc thus closes without discarding the thing it pivoted from: the empty seat of chapter 1 gets two possible occupants, an insurer accepting for money in some eventual stage 2, and an owner accepting a *named level for a dated interval* today. Either way the seat stops being null — and the difference between the old acceptance and the new is precisely that the new one has a number in it that somebody else computed and anybody can recompute.

15 · The world model

Part five — The proof and its absence

By the eighth memo the site agent processing this series had proposed the same MVP four releases running: a placement rated 1–5 with its derivation, plus a which-control-buys-the-most view. Memo 8 turns it down, and gives the reason twice — *users is going to be the first important thing*, and:

> we need to now create this sort of gamification, this sort of environment, this sort of game we have that fundamentally allow us to simulate this

Stated — memo 8, verbatim, and the purpose arrives a minute later in the same transcript — *"ultimately that's what we can use to explain this"* — in the register the memo itself proposes: somewhere between Monkey Island, Minecraft and SimCity — places a person moves between, assets visibly moving, missions, some sequences replayed and others connected live. The doctrine compresses the correction into the distinction this chapter turns on:

> An instrument answers a question somebody already knows how to ask. An explainer creates the person who can ask it.

Stated — doctrine 08, GM-D67. Nobody currently holds a mental model of grant-minus-mandate priced as insurance, so a rating engine has no audience yet; building it first would be building the second thing first. The rating still exists underneath — as the thing the world demonstrates rather than the thing that ships. And the reframe indicts the estate's own inventory on the way past: the simulator, the experiments and the workbench are instruments, excellent ones, for somebody who already has the vocabulary. **This estate has instruments and a card game. It has no world.**

What stage 1 needs from assets, and what it refuses

The memo binds cost to assets — time, money, recoverability, liability — and the doctrine keeps the boundary chapter 3 drew: valuing assets is outside this estate's competence. The reconciliation is clean enough to state as arithmetic: **a level is relative; ranking two placements needs no absolute value; pricing one does.** So stage 1 needs asset **class** — production or not, personal data or not, customer-facing or not: a declared fact, marked as such — and does not need asset **value**, which is stage 2's problem and somebody else's data. The four dimensions are recorded now anyway, as the columns a future valuation would fill, because naming them stops a resource count being mistaken for a cost.

Three actors join the twins, under rules already in force: the **underwriter**, an identity issuing policy statements; the **insurer**, absent in stage 1 by construction *and shown as absent rather than greyed-in as though pending*; and the **claim** — `loss-event/v0`, still undrafted, the one primitive the whole pivot lacks, exactly where chapter 1 left it.

Each declares fixture-or-real like every identity here: an underwriter in a demonstration whose private half is published is not an underwriter, it is a fixture, and the world says so as prominently as the record pages do.

The rule that keeps it honest

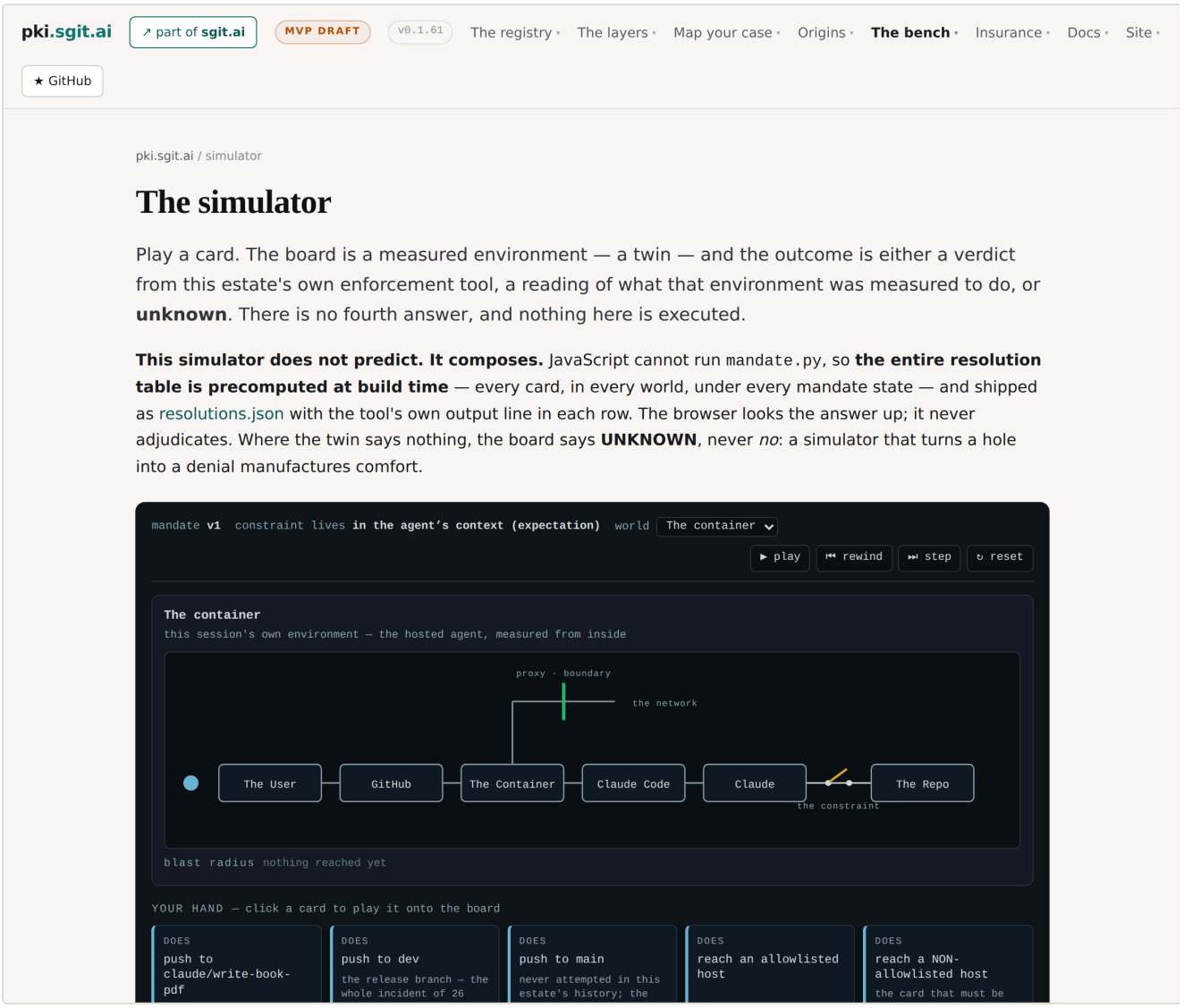
Then the load-bearing requirement, the one this book regards as part five's contribution to the whole corpus and not just to the MVP:

> A polished simulation is the most effective mechanism yet devised for making a demonstration look like a product.

Stated — doctrine 08. A world with departments and data visibly flying between them implies a working system — smooth, populated, confident. Of everything this estate would depict: ten of eleven register identities are fixtures with published private keys; both twins are servers and nobody has measured a desktop agent; the claim shape does not exist; the world-state feed exists nowhere for anybody; no insurer, underwriter or relying party has ever been implemented. A rendered world is a continuous, wordless claim about how much of this exists — apparent authority arriving in the user interface. So:

> unmeasured places look unmeasured. Fixture identities look like fixtures. Mechanisms that do not exist are visibly absent, not smoothly rendered.

Stated — GM-D69. A player sees which parts of the city are built and which are drawn, without reading a caveat. The doctrine's better sentence is the aesthetic one: **a city with construction sites and empty lots explains where this work actually stands better than a finished skyline does** — and it makes the roadmap part of the fiction rather than a footnote under it.



2D first, and the walkthrough

One disagreement with the memo is preserved, labelled, and defended as engineering sequencing rather than smuggled in as interpretation: the memo asks for 3D, and the doctrine argues most of the explanatory value is spatial rather than dimensional — places, actors, things moving along edges all work in 2D, which this estate already draws by hand in SVG with no third-party requests. A 2D world tests whether the explanation lands, which is the whole risk, at low cost; 3D tests whether immersion adds to an explanation that already works, at high and irreversible cost. **Build the 2D world first.** A 3D world that explains badly is expensive to discover and expensive to abandon. The audit singled this handling out — the site agent's advice, carried into does-not-prove rather than blended into the memo's position — as the disagreement done correctly.

The MVP, then, as specified: a 2D world walking one worked example end to end. Places are the actors — the environment where the twin is measured, the operator, the underwriter, the relying party, and an empty lot marked *insurer — stage 2*. Moving things are documents the estate already publishes: a grant, a mandate, the delta, a policy statement, a verification. The walkthrough runs: measure the environment → see the grant → read the mandate →

watch the delta appear → get it rated → install a control and watch the level move → a zero day lands and the level moves again → the policy is revoked → the relying party refuses. Every step is a document or a computation the estate already has, except two — the claim, and the world-state feed — and those appear as construction sites.

Drawn. This book cannot help noticing that it is itself the object the memo warned about: a coherent narrative is a polished rendering of a mostly unbuilt thing. The defence is the same as the world's — the emptiness is shown. Chapter 17 is this book's construction-site district, and it is the longest chapter in part five by design. Nothing in this chapter is built either; the world model is a specification with a rule attached, the rule is the best part, and the first thing that will be argued away when somebody wants to impress a room is the empty lots. The doctrine says that last sentence about its own rule. It is probably right, and writing it down is the only defence anyone has invented.

16 · Make them insurable

Part five — The proof and its absence

Memo 9 hands the work its name:

> almost like the tagline is "make your agents insurable, which is quite an interesting angle, because it also allows us to do a nice market research

Stated — memo 9, verbatim. As positioning it is honest twice over: it places the work as *preparation for a market* rather than a product in one, which is true, since the market does not exist; and it does not require the market to exist before the work has value, which is the not-in-line position's commercial survival condition. But the memo's real cargo is the research programme underneath the tagline — because to say what makes an agent insurable you first have to establish what agent insurance *exists*:

> what insurance exists for agents? Like, can you insure Claude? Can you insure you know my Claude code session? Can you insure this session? And how much does it cost? And what does it cover?

Stated. Can you insure this session — asked, in a voice memo, about the class of session that transcribed the memo, processed it into doctrine, and wrote this book. The doctrine names what that question is:

> The survey is the first thing in this series that could be wrong in a way the world would correct.

Stated — doctrine 09. As of memo 9 the pivot had produced seventy-one decisions and no external evidence, and the count has only grown; every document in the folder is reasoning, checked against the corpus, entirely internal. The survey's answer exists outside the estate and nobody here has looked. For a body of work whose own rule is that a claim must be checkable, that matters more than any position in the preceding chapters — and this book's front matter carries it as position three because a reader who missed it has missed the epistemic status of everything else.

The guard-rail, and the sentence nobody wants on a landing page

The positioning's hazard mirrors the folder's rule about levels:

> "Make your agents insurable" must not become "make your agents look insurable."

Stated — doctrine 09. Insurability is achieved by narrowing the delta or covering it — never by documenting it. And the honest counterpart, said out loud where a marketing document would bury it: **some placements should not be insurable.** A desktop agent under a user identity with network egress and no containment may be uninsurable at any

level, and the useful output is that sentence. Finding one is a success of the method, not a failure of the customer. *Drawn.* This is chapter 8's badge hazard given its operational form, and the only structural defence the corpus has is the one it keeps reaching for: a programme that can output *no* is the only kind whose *yes* means anything.

The mapping that upgrades a rule

The memo's inverse question — *what can you do that breaks the insurance?* — defines cover more sharply than cover does, and insurance has three classical answers. Two the estate already carries: **exclusions** are the mandate's prohibitions; **warranties** are the facts, failing as chapter 13 described. The third is the finding of part five:

> **A control declared on the card that the twin cannot find is not merely a weak input. It is the shape of material non-disclosure** — the thing that voids a policy rather than merely worsening a level

Stated — doctrine 09, GM-D73. Chapter 6 planted this quantity as a rating input; here it is promoted. Material non-disclosure — you did not disclose something that would have changed the terms — is the classical ground on which claims are refused, and the card-versus-twin gap has exactly its shape. Which changes what the gap is worth computing for: it has been an accuracy device; it is also **the single most consequential quantity an underwriter would want**, because it decides whether a claim gets paid. And the estate can compute it today, from two documents it already holds. *Drawn.* Whether any carrier or court would treat a computed card-versus-twin diff as legal non-disclosure is unknown and outside this estate's competence — the doctrine says so — but the direction of the finding survives the caveat: the pivot's most litigable quantity is one the measurement infrastructure already produces as a by-product.

The survey is measure.py pointed at a market

The memo wants the market's pulse taken repeatedly — *as the market evolves, you can keep track of it* — and the doctrine's design move is to refuse to invent a methodology when the estate already has one:

> The survey is measure.py pointed at a market instead of a container

Stated — doctrine 09, GM-D74 — inheriting the tool's discipline unchanged: record what a product states publicly, never characterise what it privately offers, the same self-restraint that has the tool report presence and never contents; every finding carries its evidence class — a published price is observed, a policy wording is read, a vendor claim is documented; **unknown is never absent** — *we could not establish whether X covers agent placements* is the finding, not *X does not*; the survey is a floor, not a census; and it is dated, because an insurance observation goes stale fast, and a survey with no date is a claim about now that was true once.

And the one prediction is published in the only honest tense available:

> the survey will find no product insuring an agent placement as a unit

Stated — doctrine 09 §5, stated once so it can be falsified. If it holds, it is the strongest argument for the positioning; if it does not, the positioning changes — which is what evidence is for. The doctrine is explicit that predicting the survey's answers and publishing them as research would be the estate's cardinal sin, committed in the one document whose entire value is that somebody went and looked.

Where the survey pays

Drawn. The commercial punchline is quieter than the epistemics and worth the chapter's last word. Every process the memo lists — web form, spreadsheet, Word document, API — collects evidence by questionnaire: the declared channel, exclusively. This estate produces measured evidence with provenance, which none of them asks for. Half an advantage — because an underwriter does not want a JSON document they cannot interpret. The gap is not producing evidence; it is producing evidence in a form an existing process can consume. So the survey's most valuable output is not prices or terms but **the format the market accepts** — because that format is the connector specification, and chapter 6's rule already governs it: the rendering must carry the evidence classes across, or a measured fact and a declared one arrive at the underwriter looking identical, and our own tooling will have manufactured the exact non-disclosure this chapter exists to detect.

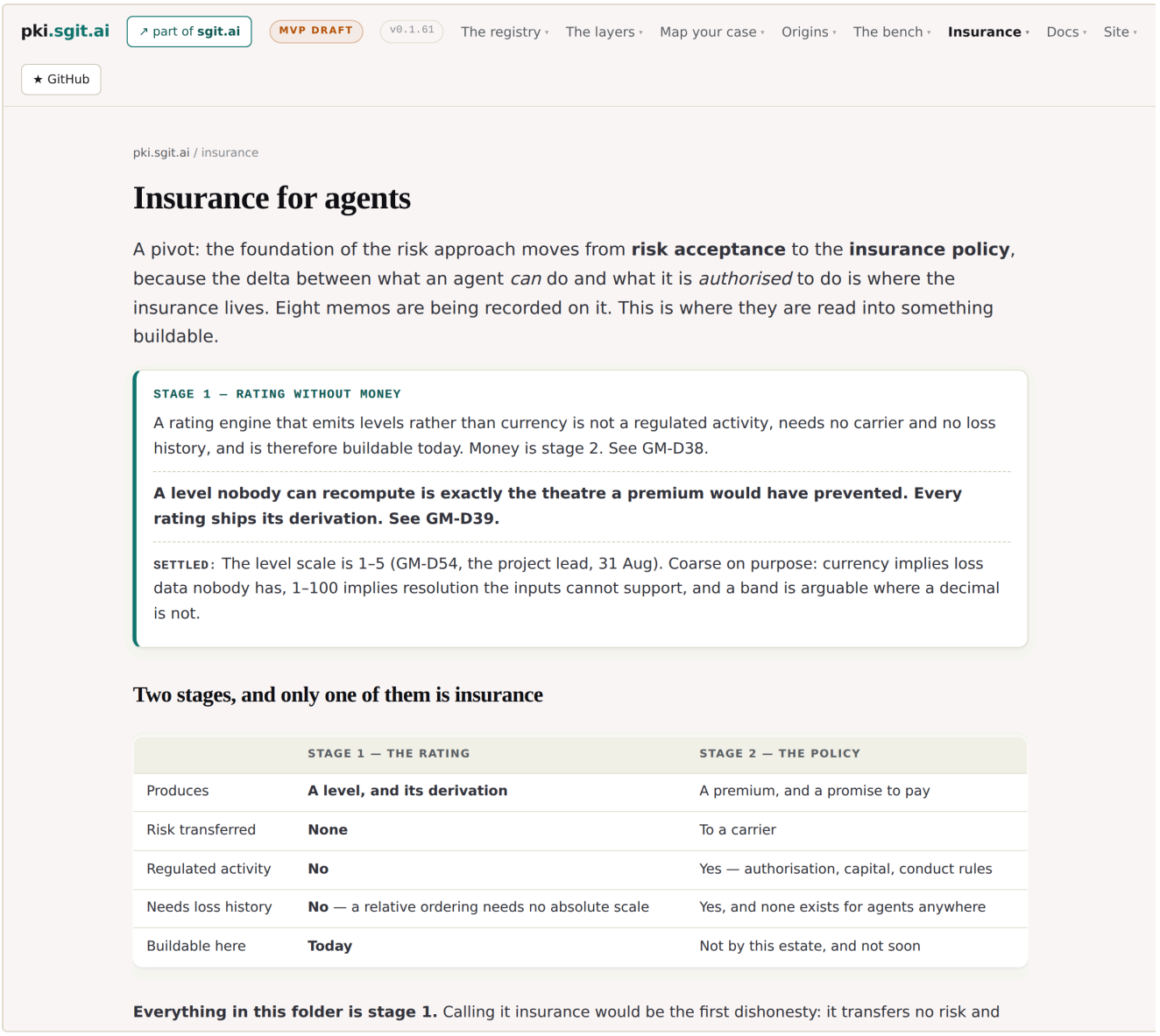
17 · What none of this proves

Part five — The proof and its absence

This chapter is the book's construction-site district: the accounting of what the corpus does not prove, what it got wrong about itself, and what the numbers actually are. Its counts are computed at build time from the repository, because the defect this chapter spends most of its length on is what happens when they are typed.

The state of the work, in computed numbers

The series stands at 10 memos plus the pivot briefing, read into 11 doctrine documents. The pivot has proposed 51 decisions onto a change-control log that now holds 85 — and of the pivot's decisions, exactly one is settled: the 1–5 scale. The insurance folder's own machine surface carries 20 does-not-prove entries, which this book has distributed across sixteen chapters rather than quarantined in one. The register the rating would read holds 11 records, 10 of them fixtures with published private keys. The excess-authority view still shows a grant of 41 resources against a mandate of 1 — an excess of 40, acceptor still null. The insurance arc ran from v0.1.50 to v0.1.60 — 11 releases in 5.1 hours — inside an estate that has now cut 60 releases in total. And the number that outranks all of these: **zero MVPs built, zero external facts gathered.**



pki.sgit.ai ↗ part of sgit.ai MVP DRAFT v0.1.61 The registry The layers Map your case Origins The bench Insurance Docs Site ★ GitHub

pki.sgit.ai / insurance

Insurance for agents

A pivot: the foundation of the risk approach moves from **risk acceptance** to the **insurance policy**, because the delta between what an agent *can* do and what it is *authorised* to do is where the insurance lives. Eight memos are being recorded on it. This is where they are read into something buildable.

STAGE 1 — RATING WITHOUT MONEY

A rating engine that emits levels rather than currency is not a regulated activity, needs no carrier and no loss history, and is therefore buildable today. Money is stage 2. See GM-D38.

A level nobody can recompute is exactly the theatre a premium would have prevented. Every rating ships its derivation. See GM-D39.

SETTLED: The level scale is 1–5 (GM-D54, the project lead, 31 Aug). Coarse on purpose: currency implies loss data nobody has, 1–100 implies resolution the inputs cannot support, and a band is arguable where a decimal is not.

Two stages, and only one of them is insurance

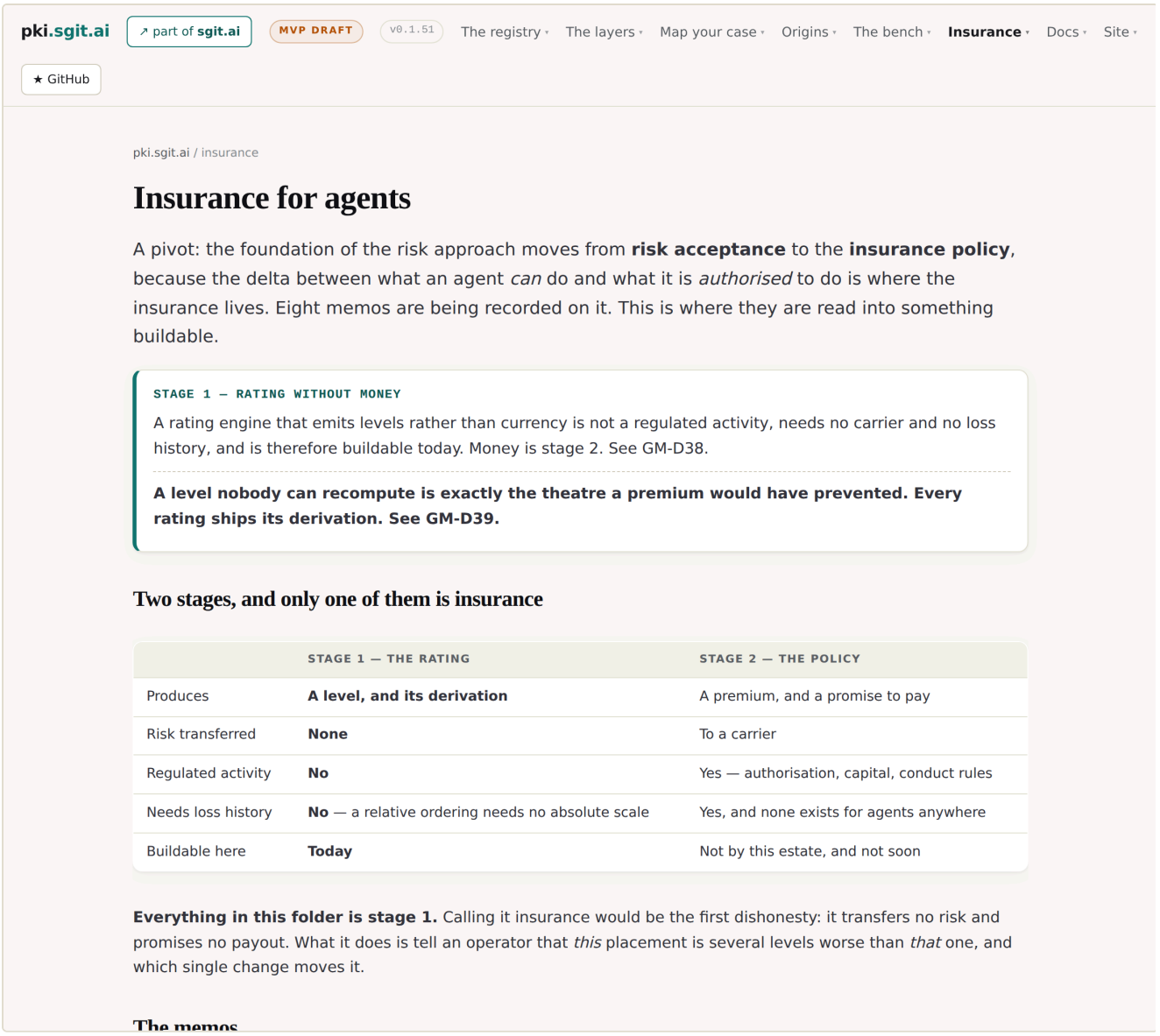
	STAGE 1 — THE RATING	STAGE 2 — THE POLICY
Produces	**A level, and its derivation**	A premium, and a promise to pay
Risk transferred	**None**	To a carrier
Regulated activity	**No**	Yes — authorisation, capital, conduct rules
Needs loss history	**No** — a relative ordering needs no absolute scale	Yes, and none exists for agents anywhere
Buildable here	**Today**	Not by this estate, and not soon

Everything in this folder is stage 1. Calling it insurance would be the first dishonesty: it transfers no risk and

The audit: six defects, and what they teach

Every one of the eleven readings was re-read against the transcript it came from, at v0.33.82, before this book was written. The verdict first, because it is the part a summary drops: the interpretation held. The two-stage split is real; the three-brokers disambiguation was caught; the Fibonacci repair saved a settled decision from a good instinct; all four load-bearing corpus quotations verify verbatim at their claimed dates. What did not hold was the arithmetic and the housekeeping — 6 defects:

Three stale counts. Doctrine documents said *eight memos* after there were ten, said *memo N of 8* in their footers, and twice called memo 8 *the last of the series* while memos 9 and 10 sat in the folder beside it. The pattern behind all three is the one this estate keeps re-learning: a number typed into prose by a writer who knew it, in a folder whose own rule is that a claim must be checkable. The fix was not the three edits — it was a build gate, which then caught real text on its first run and was kept rather than loosened.



The pair of figures above is this defect made visible. The hub at v0.1.51, preserved at its tag, believed the series was eight; the hub today computes its count from the manifest because believing turned out to be the wrong verb.

One claim identified with the wrong quantity. The doctrine had equated the enforcement tier with impact reduction — the quantity memo 4 asked for. The audit narrowed it: the tier ranks reachability, which is blast-radius reduction, an adjacent quantity and not the same one. Chapter 3 carries the corrected version; the uncorrected one would have credited a boundary control for post-incident recovery it has nothing to do with.

One over-claim. Stage 1 was called immune to moral hazard because it has no payout. Chapter 8 carries the correction: removing the payout removes one channel and opens another — the badge. Immunity claims are exactly the over-statements this corpus exists to refuse, and it briefly made one about itself.

One dropped question. Memo 3's model-vendor question — would Claude carry a policy on its own mistakes — was flagged when the memo was read and then never landed in doctrine, until the audit put it in chapter 9's place in the argument. A forward reference that does not land is the same defect as a count that does not add up.

Drawn. What the six defects have in common is that none of them is an interpretation failure. The method — transcripts filed before reading, doctrine naming its memo, corrections recorded in both documents — held under audit. What failed was everything the method did not yet mechanise: counts, footers, promised cross-references. The estate's response was to mechanise them. This book's response is this chapter's markers, every one generated.

What nothing here proves, aggregated

The chapters have carried these individually; assembled, they are the honest silhouette of the whole pivot. Nothing proves that **any of this is insurance** — stage 1's whole design is that it is not. Nothing proves **the placement orderings** — the estate cannot measure its own leading example. Nothing proves **a level means the same thing twice** — no calibration without loss data, and no loss data exists. Nothing proves **aggregation works** — correlation is named as a graph problem and unimplemented. Nothing proves **the delta classes can be told apart automatically** — the platform-granularity library does not exist, so today the classification is a judgement. Nothing proves **dynamic re-rating is close** — the event-to-grant-node mapping exists nowhere, for anybody. Nothing proves **the handshake is a boundary anywhere today** — it reaches that tier only under an independence condition nothing has tested. Nothing proves **a world explains better than a document** — the memo's hypothesis and the site agent's agreement are two opinions, not evidence. Nothing proves **the survey's answers** — and the corpus published its one prediction in falsifiable form rather than as a finding. And nothing proves **the estate's own measurement is complete** — memo 3 found a grant node the tool misses, by conversation rather than by tooling, and the honest reading is that there are others.

Drawn. The front matter promised that several of these positions *are* the argument, and this is where the promise cashes out. A rating whose first product is a checkable derivation only matters in a world where most assessments cannot be checked; a corpus that publishes twenty ways it might be wrong is only interesting if the twenty are specific, dated, and wired to gates that fire. The alternative posture — confidence, in the standard idiom of product announcements — was available at every step and would have been cheaper. The bet this book documents is that in a market whose last iteration died of unfalsifiable numbers, the auditable posture is not the ethical alternative to the commercial one. It is the commercial one.

END MATTER

Reference card

End matter — one page, written to be pasted into an agent session

---

What this is. The insurance pivot of pki.sgit.ai, in one card. Sources: briefs v0.33.71–v0.33.81 (the memos, verbatim), `/insurance/src/` (the doctrine), the audit at v0.33.82. The transcript outranks any summary — including this one.

The vocabulary

- **The delta** — grant minus mandate: what an agent *can* do minus what it was *authorised* to do. The insurable exposure of this whole programme, computed and never stored. The register publishes it with `"acceptor": null`.
- **A placement** — an agent, in an environment, under an identity. The unit that gets rated. A library entry *is* a placement; "rating Claude" is meaningless.
- **A rating** — a level 1–5 (settled, GM-D54) plus its derivation: inputs with evidence channels, what moved the level and which way, the twin's age, and what it does not prove. Stage 1's entire product.
- **The rule** — *a level nobody can recompute is exactly the theatre a premium would have prevented.* Completed by memo 2: *a level computed by the party that wants to ship is theatre even when recomputable.* Method and separation, both.
- **Two channels** — measured (full weight) and declared (discounted, rendered as declared) never merge; unknown is never scored as absent. The questionnaire is a declared-fact collector; the card-versus-twin gap is a rating input — and the shape of material non-disclosure.
- **Delta classes** — elective (operator can close; carries a cost-to-close, whose high region is *latent*), structural (platform's finest grain is coarser than the mandate; the pool's, or nobody's), defect (a vulnerability; temporary). A rating that does not separate them is unfair and useless in one move.
- **A policy** — a mandate-shaped statement issued by a rater. It never signs; its subject signs, and the policy says what that signature is worth. Amending, activating, ending: appends, never edits.
- **A warranty** — a fact plus a maximum age. Fails three ways: false, stale, unknown — and unknown counts as failure, the deliberate opposite of the rating rule.
- **Three clocks** — the policy interval; twin and world freshness; each warranty's check interval. A policy in force against a fresh twin with a warranty last checked eleven days ago is not covered.
- **The handshake** — the relying party refuses the request unless the subject proves possession and a policy is in force. The first mechanism in the pivot that can reach tier *boundary*, because the relying party is outside the requester's grant — where genuinely independent.

The postures

- **Stage 1, not insurance**: a rating transfers no risk and promises no payout — which is why it needs no carrier and can be built today. Money is stage 2, and stage 2 is regulated and not this estate's.

- **Not in line**: schemas, flows, connectors, evidence — never the carrier, never the execution broker. This project can never itself be a boundary; every tier above instrumentation is somebody else's install.
- **Openness is load-bearing**: with no money at stake, an attackable public method is the only honesty mechanism left.
- **Any party claiming to reduce somebody else's exposure should carry cover against being wrong about it** — brokers, and one day model vendors.
- **The world model MVP explains rather than calculates, and shows its own emptiness**: fixtures look like fixtures, unmeasured places look unmeasured, the missing claim primitive is a construction site.

Carry these limits with any summary

Nothing here is insurance. Nothing here is built — two MVPs specified, zero built. No external evidence exists — the survey has not been run, and the one prediction (no product insures an agent placement as a unit) is published to be falsified. The register is fixtures. The placement orderings are unmeasured judgements. The readings were audited: six defects found, fixed, and half-mechanised into gates. Dropping these limits misrepresents the work.

END MATTER

Colophon

End matter — how this book was made, and the gates it must pass

---

This book was commissioned in a single sentence on 31 August 2026 — *sync this repo, take a look at /insurance/ and its llms.txt, and in a new base folder write a book that incorporates all the ideas in those briefs/voice-memos into a coherent book/pdf about insurance* — and the sentence is preserved verbatim in `BRIEF.md` beside the decisions the writing session added on its own authority: the title, the five-part structure, and the folder. It is the second volume from this estate, by the method of the first, and it was written by the same class of agent the corpus proposes to rate — a fact the reader is entitled to keep in view throughout.

The sources, and their order of precedence

The memos outrank everything. Ten voice memos and a pivot briefing were filed verbatim as briefs v0.33.71 through v0.33.81 *before they were read* — transcription stumbles, mid-sentence corrections and one truncated final thought included — because a transcript outranks any summary of it. The doctrine under `/insurance/src/` is derived from the memos and names which memo each document came from; where doctrine and memo disagree, the memo wins. The audit at v0.33.82 re-read all eleven readings against their transcripts and found six defects, which chapter 17 walks. This book quotes all three layers, and quotes the estate's pre-insurance corpus where the doctrine leans on it.

The gates

Every discipline is inherited from the first volume, and each fails the build rather than warning.

Quotes. Every *stated* quotation is extracted from the chapters, its source discovered rather than asserted — searched for across the briefs, the doctrine, the machine surfaces and the estate's artefacts — and re-read out of that source on every build. A quotation not found where it claims to be fails the build.

Stats. Every count in the prose is a `gen:stat` marker computed from the repository at build time: memo counts from the manifest, decision counts from the change-control sources, the excess from the register's own view, release spans from the tags. In check mode, a drifted count fails the build. This is the mechanised form of the audit's largest lesson, and this book watched the mechanism work: its own draft carried wrong values in several markers, and the first build corrected them.

Figures. Every figure is taken from the version its caption names — a `git worktree` at the tag, a one-shot server on a port used once, a headless browser killed in a `finally`. A figure of the past must re-derive at its tag; a figure of the present must match the live page, and the build fails when it stops matching, which it will on the next release. The terminal figures are real transcripts, executed by `shots/transcripts.sh` at build time, never pasted.

Hashes and captions. `book.json` records the SHA-256 of every chapter's markdown, and the build fails on a mismatch. Every figure carries a caption that says what to *notice*, and a caption under forty characters fails.

The writing session's own findings

Recorded here because a colophon is where a book confesses. First: the corpus is unusually quotable because it was built to be — filing transcripts before reading them is what makes a verbatim-quote gate possible at all, and this book's 76 verified quotations are downstream of that one filing habit. Second: the hardest structural choice was where to put the audit; it earned the closing chapter because the corpus's most distinctive property is not any position but the fact that it checked itself and published the result. Third: the book's own first build caught its own stale numbers, which is simultaneously embarrassing and the entire point — the writer who typed those numbers had read, hours earlier, an audit brief about writers typing numbers.

Set in the estate's own chrome, rendered to HTML by the site's reader and to PDF by the harness in `gen_pdf.mjs`; the PDF embeds every figure as a data URI and reads start to finish offline with no link followed. The markdown under `content/` is the source of truth. CC BY 4.0, like everything else here.