

• DETECT • ATTEST • EXCISE

THE WITNESS BRIEF

PROTORA.



Know the Finetune.

*Receipts, not adjectives — the
witness, never the word for it.*

V2026 • CONFIDENTIAL TO RECIPIENT
A SPECIMEN OF THE DELIVERED ARTIFACT

SIGNED EVIDENCE • PUBLISHED CEILING
NOT LEGAL ADVICE

Every claim carries a witness. Nothing here is an adjective you are asked to trust.

DETECT. We read a finetune you did not make and cannot trust, and state plainly what it actually did — including what its card never mentioned.

ATTEST. Each finding becomes a signed evidence artifact — a record a buyer, an auditor, or a regulator can hold, with the proof travelling inside it.

EXCISE. The flagged behaviour is removed under attestation that only it changed, to a stated coverage — reversible, fail-closed on any abstain.

THE READING KEY

- proven — carries a witness • inferred — labelled
- abstained — with a reason

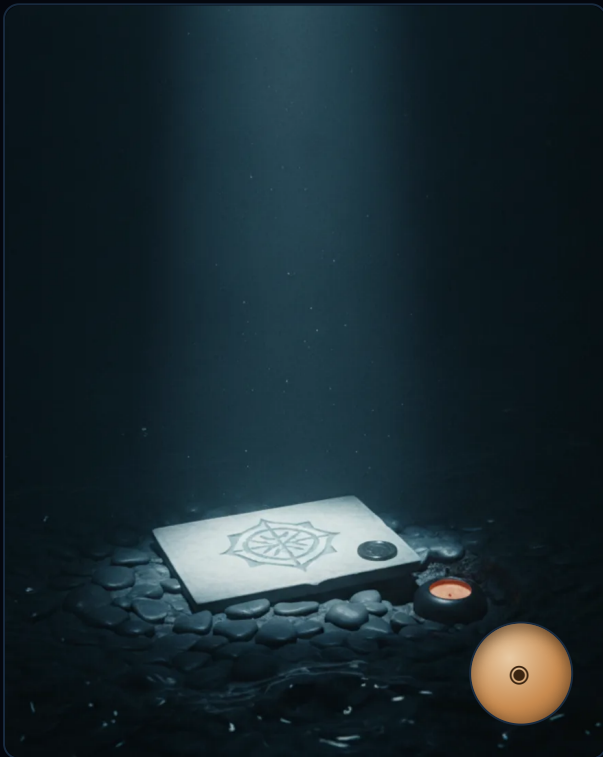
THE WITNESS, EXPLAINED

A witness is an opaque, re-runnable handle — `wit_...`. Point it at the same subject and the same result returns. It shows *that* a finding holds. Nothing else.

This brochure is a specimen of the delivered artifact. It redacts like one, carries witnesses like one, stamps its ceilings like one.

ONE ACTION,
STATED ONCE:
EMAIL US —
WE PROVE IT
FREE, ON
YOUR OWN
FINETUNE.

The legend defines what the marks mean to you, the reader — a plain reading key, nothing more.



A thing you can hold.

What lands in your hands is a **signed compliance dossier** — one consolidated evidence artifact, engraved with its findings, sealed with its attestation, and kept live behind a thin login portal so it stays a queryable record, not a one-off PDF.

SIGNED ARTIFACT

LOGIN-GATED PORTAL

REPLAYABLE WITNESS

It is the difference between a slide that asserts and a stone that attests — you keep it, re-run it, and hand it to whoever must be convinced.

ENGINE-VERSIONED • THE FINDING + ITS WITNESS,
NEVER THE INSTRUMENT.

Six deliverables, spined to a shared numbered ladder. Each is stated as value, form, and delivery — never as a procedure.

**TIER-06
DOSSIER**

Signed compliance dossier

The consolidated DETECT pass on one model, delivered as a single signed evidence artifact — built for the documentation, evaluation, and provenance obligations regulators reward.

you hold: a portal-gated, sealed record.

**TIER-02
CHANGE**

Base-vs-finetune diff

A plain account of what the finetune actually did versus its base — including undisclosed or backdoor behaviour its card never mentioned.

you hold: a plain per-finding account, each finding witnessed.

**TIER-01
READING**

Recipe fingerprint, base-family stamped

Identifies the producing recipe near-perfectly within a fixed base, with the calibrated base family stamped on the reading and its within-base scope stated.

you hold: a comparison verdict, its scope printed on it.

**TIER-03
WITNESS**

Per-finding replayable witnesses

Every finding links to its own re-runnable demonstration. Nothing is asserted without one; the safety-strip reading is taken straight from the weights, without ever eliciting a harmful response.

you hold: a re-runnable handle per finding.

**TIER-05
EXCISE**

Certified excision package • TRUST-GATED

The flagged behaviour witnessed present-before / absent-after, attested that only it changed to the stated coverage — reversible by construction, fail-closed on any abstain. A certified fix, not a blind re-finetune and a hope.

you hold: a certified, reversible fix with its before/after witnesses.

**TIER-09
GOVERNANCE**

Abstain register • published ceiling • obligation → artifact map

A complete, honest list of everything that could not be proven (each with its reason), a fidelity ceiling stamped with the engine version, and a two-column map: *the obligation asks X → Protora leaves behind Y*.

you hold: evidence — not legal advice, not certification.

ALSO IN THE FAMILY, WHEN THE INFRASTRUCTURE AND THE TRUST ARE REAL: AN IN-LOOP DRIFT-GUARD WITH A SIGNED DRIFT LOG (STAGED), AND A CALIBRATED-UNCERTAINTY MAP STANDING BEHIND EVERY INFERRED FINDING AND EVERY ABSTAIN.

Exactly what would land in your hands — rendered at full proof-fidelity, with only your material sealed away. The restraint is the brand: where method would sit, there is a seal.

• **PROVEN** safety-strip (abliteration) • read from the weights

subject [REDACTED] • base family [REDACTED] (illustrative)

official release refused ~0.9× of a harmful battery

tampered release refused 0.00 — refusals collapsed

• corroborated behaviorally • each collapse → a witness

coverage two architectures [REDACTED] + [REDACTED] • small sizes •

small battery • mechanism-proven • broader staged

wit_ [REDACTED] • replayable

THE WITNESS, RUN — OUTCOME LINES SURVIVE, STEPS ARE SEALED

```
$ protora witness replay wit_ a91
```

```
▸ load subject + base
```

```
▸ run harmful battery → refusals: 0.00
```

```
✕ safety removed — verdict: PROVEN • read from the weights (illustrative)
```

The reading shows the outcome — safety removed, refusals collapsed, each collapse witnessed — and its honest coverage ceiling. The instrument that produced it stays sealed; the deliverable carries the finding, never the instrument.



The engraved-stone dossier under its wax-seal medallion — the attestation. Findings, witnesses, engine-version stamp, and the abstain register, in one consolidated record.

○ OWNED • NO WITNESS • NO ENTRY

A planted backdoor read with nothing to compare against sits at chance — and Protora declines rather than guess. Reason on the record: no pair, no reference.

AN ENTRY WITH NO BLANKS IS NOT A TRIUMPH — IT IS A WARNING. THE ABSTAIN REGISTER LISTS EVERY SUCH DECLINE, OUT LOUD.

THE OBLIGATION → THE ARTIFACT

THE OBLIGATION ASKS FOR	YOU RECEIVE
Technical documentation	the witnessed, ceiling-stamped dossier
Model evaluation, incl. adversarial testing	DETECT findings — undisclosed / backdoor behaviour, harm-free
Corrective action	the certified excision — attested to coverage, reversible
Provenance	the recipe fingerprint, base-family stamped

ENGINE **PROTORA-CORE** • **V2026** • DESCRIPTIVE & COVERAGE-BOUNDED • SINGLE-FAMILY, SMALL-BATTERY, MECHANISM-PROVEN, BROADER COVERAGE STAGED.

Ships a signed conclusion, a replayable handle, and a coverage stamp — the finding, never the instrument.

THIS IS EVIDENCE, NOT LEGAL ADVICE OR CERTIFICATION. IT IS MATERIAL YOU FILE; THE OBLIGATION TO FILE IT, AND THE JUDGMENT THAT IT SUFFICES, REMAIN YOURS.

In a market of confident guesses, the stated limit is the product.

01 Published ceilings

Every reading is stamped with the engine version and a fidelity ceiling — *single-family, small-battery, mechanism-proven, broader coverage staged*. Nothing here is presented as complete.

02 The abstain register

We publish what we declined and why, as loudly as the readings. An entry that omits its abstain list is a defect, not a triumph.

03 The graded record

A tally-mark wall of graded outcomes — hits *and* misses in the open, next to the ceilings, so silence reads correctly. Two coverage-stamped receipts:

0.92-0.96 → 0.00 refusal collapse
precision 1.0 / recall 1.0 on the battery
recipe identity ≈1.0 vs 0.083 chance
backdoor surfaced at 1.5-3% of what changed

EACH CARRIED BY ITS OWN WITNESS & COVERAGE
STAMP.

04 Scope, stated plainly

Open-weight models, and your own under contract — reading a finetune needs access, so **closed-API models are out of scope**. Protora is assurance against drift, misconfiguration, and planted misalignment — not a proven shield against a sophisticated adaptive attacker.

05 Governance — data-handling & the obligation map

Your source or model is **dropped by default** once the pass completes. Tenant-isolated, least-privilege access. The only retained artifacts are the signed deliverable and its witness — nothing else. The obligation → artifact map states what you receive, always closed by: *this is evidence, not legal advice or certification*.

What it does is open and verifiable; how it does it is proprietary — and nothing you rely on depends on it.

THE ANTI-OVERSELL FIREWALL.

Words you will not find in this document: “secure,” “guaranteed,” “catches every,” “zero false positives.” If a claim cannot carry a witness, we do not make it.

Every number here is one of two things: a figure from our own published, coverage-stamped record — or a figure labelled *illustrative* inside a sample. There is no third kind.

Don't take our word for it. Before you write, read our public graded record — the tally-wall, hits and misses in the open, next to the ceilings. ▶ [the record](#) → [vulcora.se/record](#)

We'll prove it — free — on your own finetune.

1 SUBJECT

RETURNED AS THE REAL DELIVERABLE

WITH ITS WITNESS & ITS CEILING

Yours to keep. No obligation. This document earns the email; it does not close the sale.

① EMAIL

Send us one finetune and its base family.

② SCOPE

We agree what “proof” means for your case.

③ RECEIVE

The deliverable on pp5–6, unredacted for *your* model, with its replayable witness and stated ceiling.

EMAIL US

legend@vulcora.se

HANDLING — YOUR SOURCE IS DROPPED BY DEFAULT ONCE THE PASS COMPLETES; TENANT-ISOLATED, LEAST-PRIVILEGE. · PRICING — NO PRICES HERE; SCOPED PER CASE, QUOTED PER OFFER.