VERITY COMMAND

PUBLIC TECHNICAL NOTE

Public Threat Model and Security Boundaries

What the architecture protects, detects and cannot guarantee.

DOCUMENT	VERSION	PUBLISHED	EDITION	PDF
VC-PUB-002	1.1	16 September 2026	Public	Download (346 KB)

PUBLIC EDITION

CONTENTS

1 Governing principle

> Verity Command does not make intelligence inherently true. It makes the basis, authority and consequences of acting on intelligence inspectable and enforceable.

Every statement in this note follows from that sentence. A platform cannot verify reality; it can record what was asserted, by whom, on what basis, under whose authority and with what effect. Where this note describes a control, it describes a constraint on action and a record of decision — not a guarantee about the world.

2 Method and scope

The analysis follows the thirteen trust boundaries defined in the Trust, Authority and Evidence Architecture (VC-PUB-001): a spoofing, tampering, repudiation, disclosure, denial and elevation review at each boundary; attack trees for the highest-value objectives; and abuse cases for misuse by authorized actors. It describes the architecture's controls. It does not report an independent assessment, audit or penetration test, and none is claimed.

> **Out of scope by declaration.** The security of customer infrastructure, the physical security of customer facilities, and the conduct of people after an instruction leaves the enforcement boundary. A threat model that quietly omits its boundaries is misleading, so they are named here.

3 Assets and adversaries

Four assets shape most of the architecture — authority state, cryptographic roots, the evidence record and communications payloads — because an adversary holding any one of them can act as, or authorize, anything else.

ASSET	WHY AN ADVERSARY WANTS IT
Authority state — grants, leases, policy	Ability to act as, or authorize, anything
Cryptographic roots and keys	Total compromise of an authority domain
Evidence record	Erase or fabricate the record of what happened
Communications payloads and metadata	Orders and intelligence in transit; tempo and relationships
Mission content and source details	Intelligence value; exposure of methods and people
Coalition-released material	Diplomatic damage; loss of partner trust
Model routing and AI outputs	Manipulation of analysis and decisions
Software supply chain	Persistent, wide compromise
Availability of the decision path	Deny decision-making at a critical moment

Adversaries range from state-sponsored actors with supply-chain reach to the careless authorized user. Three hold legitimate credentials and receive particular attention: the malicious insider, the malicious administrator and the hostile federation partner. A design that only resists outsiders is not a design for this domain.

4 Residual risks

These are the risks an evaluator should weigh most carefully. They are stated first and in full, not left to footnotes.

RESIDUAL RISK	STATEMENT
Sensor spoofing	A spoofed signal received by a genuine sensor is a genuine observation of a false signal. Contradiction detection and source reliability tracking raise the chance of noticing; they do not guarantee it.
Authenticated but materially false data	Source authentication establishes who supplied data, not whether it is true. A false report from a genuine source remains a correctly attributed false report.
Insider with valid keys	Authority is scoped by mission, direction, duration and label ceiling, and use is recorded. An insider acting within a valid grant is acting legitimately; the architecture bounds reach, it does not read intent.
Compromised customer roots of trust	The customer holds the roots, so a compromise of them compromises that authority domain. The vendor cannot recover it, because the vendor holds no key that can authorize, decrypt or sign inside it.
Compromised endpoints	A captured, unlocked device with content displayed is not protected by cryptography. Lease duration, minimal local data and truthful duress signaling reduce scope. The Distribution Endpoint, if compromised, exposes the packages it is distributing at that time.
Data and model poisoning	Validation, reliability tracking, contradiction surfacing and statement typing raise the cost of poisoning. Patient, sophisticated poisoning is not claimed to be detectable. Model responses are untrusted input: recorded, attributable, and possibly wrong.
Model and dependency supply chain	Signed builds, a bill of materials, provenance attestation, recorded model versions and restricted model routes reduce exposure. A permitted route to a compromised model, or a compromised upstream dependency, remains a risk.
Quorum collusion	Quorum across people and organizations raises the cost of a fraudulent authorization. Coercion or collusion of an entire quorum is possible and is not claimed to be solved.
Traffic analysis	Separating key material per mission does not defeat traffic analysis. No claim of anonymity or unlinkable mission activity is made.
Metadata observation	Connection metadata, timing and volume remain observable to network infrastructure, including the relay. Content is protected; the existence and pattern of communication is not hidden.
Loss of trusted time	Authority decisions fail closed when time uncertainty exceeds the bound. That is a denial-of-service surface by construction: degrading time degrades the ability to authorize. The trade is deliberate.
Stale revocation state	Revocation reaches an enforcement point when it receives the signed state. A disconnected point acts on what it last received, bounded by lease expiry; the residual window is shown to the commander at revocation.
Prolonged isolation	An isolated Verity Edge node enforces its lease and then fails closed. Isolation ends in loss of capability, which is itself an operational risk to plan for.
Human error under valid authority	Quorum, digest-bound approval and a reconstructable view of what was known reduce and expose error. A well-formed, correctly evidenced authorization can still be the wrong decision.

RESIDUAL RISK	STATEMENT
Physical actions outside the enforcement boundary	The platform governs decisions, authorizations and distribution. It does not govern what a person or system does in the physical world after receiving an instruction.

5 Prevented or constrained

Threats the architecture stops, or bounds so tightly that success yields little.

THREAT	CONSTRAINT	WHAT REMAINS
Elevation through delegation	Delegated permissions never exceed the delegator's at decision time; an AI agent is bounded by grant, delegator and mission	A delegator with excessive authority passes on a share of too much
Approval laundering	Approvals and authorizations bind to an exact content digest; any edit invalidates them	Low while digest binding holds
Replay of a consumed authorization	Single-use consumption, nonces and expiry	Low
Mission substitution	The mission comes from the authorizing grant, never from the caller	Low
Central interception	Only endpoints and the Distribution Endpoint hold payload keys; the relay carries ciphertext	Distribution Endpoint exposure (section 4)
Data reaching an unauthorized model	Classification-aware routing; no permitted route means refusal, not rerouting	A permitted route to a compromised model
Prompt injection	Content is data, never instruction; tools run from a registry, in a sandbox without default egress	Degraded answer quality
Caveat stripping on release	Release decision and acceptance required; caveats bound; withdrawal honored	Partner misuse after acceptance
Unrecorded action	Evidence admitted before effect; no record, no action	Actions taken outside the platform
Authority surviving revocation indefinitely	Leases and validity windows cannot be extended locally	A bounded residual window offline

6 Detected or evidenced

Threats that cannot be prevented but can be made visible, with a record an investigator can rely on.

THREAT	HOW IT IS SURFACED
Evidence tampering	Hash chain and signed checkpoints, optionally exchanged with a partner; continuous verification raises a critical incident and freezes writes on any break

THREAT	HOW IT IS SURFACED
Over-release to a partner	Release decisions and projections are recorded: what, to whom, under whose decision
Vendor access	No standing access; support access is customer-approved, time-bound and evidenced
Misuse by an authorized user	Purpose recording, mission-scoped grants and a reconstructable decision context
Contradictory or repeated sources	Contradictions preserved and shown; lineage exposes repetition presented as corroboration

7 What Verity Command does not protect against

- **A compromised customer root of trust.** Sovereignty means the customer holds the roots and carries that risk.
- **A captured, unlocked endpoint with content on screen.** Cryptography does not help at that point.
- **A coerced or colluding quorum.** Quorum raises cost; it does not make fraudulent authorization impossible.
- **Traffic analysis and metadata observation.** The pattern of communication remains observable.
- **A lying source.** Authentication establishes origin, not truth.
- **Patient data poisoning.** Raised cost is not detection.
- **Incorrect AI output.** Governance constrains what AI may do, not whether it is right.
- **A lawful decision that is simply wrong.** The record makes it reconstructable, not preventable.
- **Physical consequences beyond the enforcement boundary.**
- **Denial of availability.** Failing closed means that degrading identity, policy, evidence or time degrades the ability to act. That is the chosen trade.

8 Properties not claimed

NOT CLAIMED	WHY
Revocation reaches every endpoint at once	False for disconnected endpoints; the lease is the bound
The relay learns nothing	Metadata, timing and volume are observable
Mission activity is unlinkable	Key separation does not defeat traffic analysis
The system cannot be misconfigured into insecurity	Defaults, tests and quorum reduce the risk, not eliminate it
AI cannot produce a wrong answer	Evaluation bounds quality; it does not guarantee correctness
A validated cryptographic module	None is claimed
Any certification, accreditation or approval for classified information	None is claimed
Hardware-backed, non-exportable keys on every platform	Platform support varies; assurance is stated per platform and algorithm

NOT CLAIMED	WHY
Guaranteed secure deletion of media	Media remanence is outside software control
Independent assessment or cryptographic review	None is claimed

9 Related documents and revisions

VC-PUB-001 · Trust, Authority and Evidence Architecture

VC-PUB-003 · Hormuz Reference Scenario — When the Correct Decision Is Refusal

VERSION	DATE	CHANGE
1.0	2026-09-16	First public edition.
1.1	2026-09-16	Residual risks consolidated into one section; aligned to the canonical platform architecture; related documents linked.

Document VC-PUB-002 · Version 1.1 · Published 16 September 2026 · PUBLIC EDITION