Meridian Verity Group

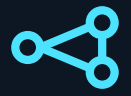
AI Safety & Governance

HALTSEAL

Verifier → Permit → Fail-Closed Gateways for Agentic AI Execution

HALT real actions. SEAL the evidence.

One-line: Block real actions (egress / privileged I/O / accelerator dispatch) until the agent presents a verifiable, expiring permit installed at a non-bypassable gateway.



Network egress

TCP/UDP, IPv4/IPv6

Privileged I/O

Device I/O + privileged syscalls

Accelerator dispatch

Queues / doorbells

Updated 14 Feb 2026

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

WHY

When model output becomes action, the boundary moves

Where agents act

- Network egress (TCP/UDP, IPv4/IPv6)
- Device I/O + privileged syscalls
- Accelerator dispatch (queues / doorbells)
- Actuation / fieldbus / robotics pathways

What breaks in practice

- "Allow-by-default" paths appear under failure, drift, or misconfig
- TOCTOU: checks happen earlier than enforcement
- Local proxies / loopback shortcuts undermine policy
- Revocation must override allow — immediately and provably
- Post-hoc logs are not pre-action control

The fix

Move verification into the execution path, and make the gateway fail-closed by default.

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

ARCHITECTURE

Three primitives that make control non-bypassable

1 Verify-to-Activation

- Append-only verifiable log publishes signed heads
- Runtime checks freshness, inclusion, and consistency
- Verification can run inside a read-only boundary (ELV)

2 Permit-before-Action

- On PASS: mint a signed, expiring permit
- Audience-bound (agent / tenant / mission)
- Record append-only receipt in IAL (exportable)

3 Fail-Closed Gateways

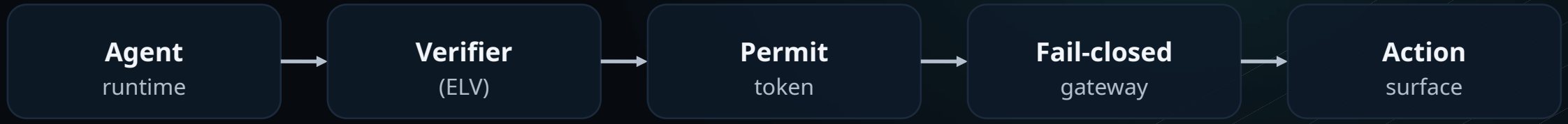
- On FAIL: HOLD / QUARANTINE / DENY / ESCALATE
- Gateways require a valid permit ID before actions
- Revocation overrides allow at the chokepoint

Outputs: signed head + proofs • permit token • IAL receipts (exportable) • stable reason codes

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

EXECUTION PATH

Non-bypassable: verifier → permit → gateway → action



Evidence outputs: signed head + proofs • permit token • IAL receipts (exportable)

Key properties

- Fail-closed by default: no valid permit → DENY
- Permit binds head IDs, audience, nonce+counter, expiry
- Freshness + consistency proofs enforce continuity
- Receipts are append-only and exportable to Audit/GRC

What you can validate in a deep dive

- Gateway is non-bypassable on the chosen surface
- Revocation overrides allow immediately
- TTL expiry + anti-replay work under load
- Receipts verify on buyer side in 60s (keys + proofs)

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

PROOF

The 60-second proof (live)

1	No permit	DENY
2	Loopback / proxy bypass attempt	DENY
3	Agent calls verifier API	mint → install
4	TCP + UDP network egress	ALLOW
5	Revocation overrides allow	DENY
6	TTL expiry	DENY

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

ENFORCEMENT

Where enforcement lives

Choose one chokepoint for a pilot

Chokepoints

User space (agent runtime)
Fast pilot; useful for integration + evidence flow

Kernel egress / syscall intercept (eBPF/cgroup)
Strong control for outbound actions

Driver path (queue/doorbell before action)
Prevents accelerator dispatch without permit

Firmware / microcode gate (before dispatch)
Hard boundary; minimal bypass surface

Hypervisor intercept (VM-exit / network egress)
Central control in multi-tenant environments

Action surfaces you can gate

- Network egress: TCP/UDP, IPv4/IPv6
- Device I/O and privileged operations
- Accelerator dispatch: queues/doorbells
- Actuation pathways (fieldbus/robotics)

Pilot strategy
Pick ONE gateway, prove non-bypassable control + receipts, expand after.

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

Security invariants

What we promise — and test

Gateway invariants

- Fail-closed default: missing permit → DENY
- Revocation overrides allow immediately
- TTL enforced at the gateway state
- Loopback policy: verifier-only allowed; rest denied
- Spam control: deny cache + throttling are mandatory

Verifier + evidence invariants

- Freshness enforced for signed heads
- Inclusion proof ties ICC commitment to current head
- Consistency proof required on head advance
- License tier + m-of-n quorum required to mint permits
- Receipts are append-only (IAL) and exportable
- Optional: canonicalization/context-label checks

The bar: Big Tech security teams can verify the cryptography and reproduce the proof drills.

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

EVIDENCE

Audit-ready artifacts

Buyer-facing

Artifacts

- Signed head (root hash, timestamp, signatures)
- Inclusion proof (ICC commitment → current head)
- Consistency proof (append-only evolution)
- Permit token (signed, audience-bound, expiring)
- IAL receipt (append-only JSONL) + reason codes

Example: permit token (shape)

```
{
  "permit_id": "...",
  "audience": {
    "agent_id": "...",
    "tenant_id": "...",
    "mission_id": "..."
  },
  "expiration_ts": "...",
  "icc_head_id": "...",
  "signed_head_id": "...",
  "license_tier_id": "Alpha",
  "nonce": 4259182012,
  "monotonic_counter": 118,
  "signature":
{"alg":"ed25519","kid":"...","sig":"..."}
}
```

Positioning: receipts bind policy + approvals + execution evidence into artifacts Security/Legal can verify.

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

IP + LICENSING

Patent-backed primitives

Licensing-ready modules

Claim-backed coverage (high level)

- Verify-to-Activation gate: signed heads + proofs
- License authority: tier check + m-of-n quorum (GLG)
- Permit-before-Action: require permit ID before privileged I/O / egress
- Fail-closed dispositions: HOLD/QUARANTINE/DENY/ESCALATE
- Enforcement hooks: driver path, firmware/microcode, hypervisor

License package (under NDA)

- Verifier SDK (signed-head validation + permit mint/verify)
- Gateway hooks (egress / privileged I/O / dispatch)
- Deep-dive kit (JSONL receipts + reason codes + buyer-side verify + bypass drills)

IP snapshot: legacy title “NOVACOV” (U.S. App. Nos. 19/183,800; 19/404,229). NDA unlocks claim charts + code + evidence pack.

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

POLICY PULL

Why now: policy + procurement increasingly require provable controls

Procurement language that wins: “prove pre-action controls + provide audit-ready evidence.”

Framework	Timing / buyer ask	HALTSEAL mapping
EU AI Act	Staged 2025–2027; full roll-out by 02 Aug 2027	Pre-action gating + audit receipts (proofs + IAL)
Korea AI Basic Act	In force 22 Jan 2026	Traceable approvals (tier/quorum) + evidence pack
US Federal (OMB M-24-10)	Minimum risk management practices; procurement wants proof of controls	Controls enforced before action; exportable receipts
NIST AI RMF + ISO/IEC 42001	Common audit language for AI risk management	Receipts bind policy + approvals + execution evidence

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

PILOT

Design Partner offer

6–12 weeks

Week 0–2

Align threat model + pick ONE gateway
Agree on evidence schema + reason codes

Week 3–6

Verifier MVP + permit demo
Buyer-side verify in 60s

Week 7–12

Move enforcement into driver/firmware/hypervisor
Production hardening plan + pilot-to-prod option

Deliverables buyers care about

- Fail-closed intercept on chosen surface (egress or dispatch)
- Stable receipts: IAL lines + policy digest + key IDs
- Revocation + TTL drills with evidence
- Go/no-go: production backlog + ownership + rollout plan

What we need from you

- Platform owner for the chosen chokepoint
- Security partner (AppSec / Product Security)
- Optional: Audit/GRC evidence consumer
- NDA scope for evidence pack + pilot success criteria

Default recommendation: start with network egress.

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA

NEXT

The ask: 30-minute technical deep dive

Security present

In 30 minutes we will

- Map your agent runtime → pick ONE gateway (egress / dispatch / actuation)
- Define evidence acceptance criteria (receipts + reason codes)
- Agree on NDA scope for the evidence pack + pilot success criteria

Who should join

- Platform owner (runtime / infra / kernels / drivers)
- Security (AppSec / Product Security / Governance)
- Optional: Audit/GRC stakeholder (evidence consumer)

Reply with: (1) your top action surface, and (2) the best owner to invite. We will do the rest.

Contact: licensing@meridianverity.com

For business/technical discussion only (not legal advice) • Not a compliance certification • Evidence pack/code under NDA