OUTREACH SAFE • 10 slides

NOVACOV

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

Network egress • Privileged I/O • Accelerator dispatch

One-line: Agents stay blocked from real actions (egress / device I/O / dispatch) until they present a verifiable permit + audit-ready receipts.

Goal for this deck: earn a 30-minute technical deep dive with Security present.

WHY

A new security boundary: model output becomes action

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

PROOF

The 60-second proof (what we show live)

6 scenes (kernel-feel enforcement):

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

Meeting outputs: report.md • IAL JSONL receipts • deny_cache table (readable reasons)

NOVACOV in 3 control primitives

1) Verify-to-Activation

Append-only log publishes signed heads.
Runtime validates freshness + inclusion + consistency.

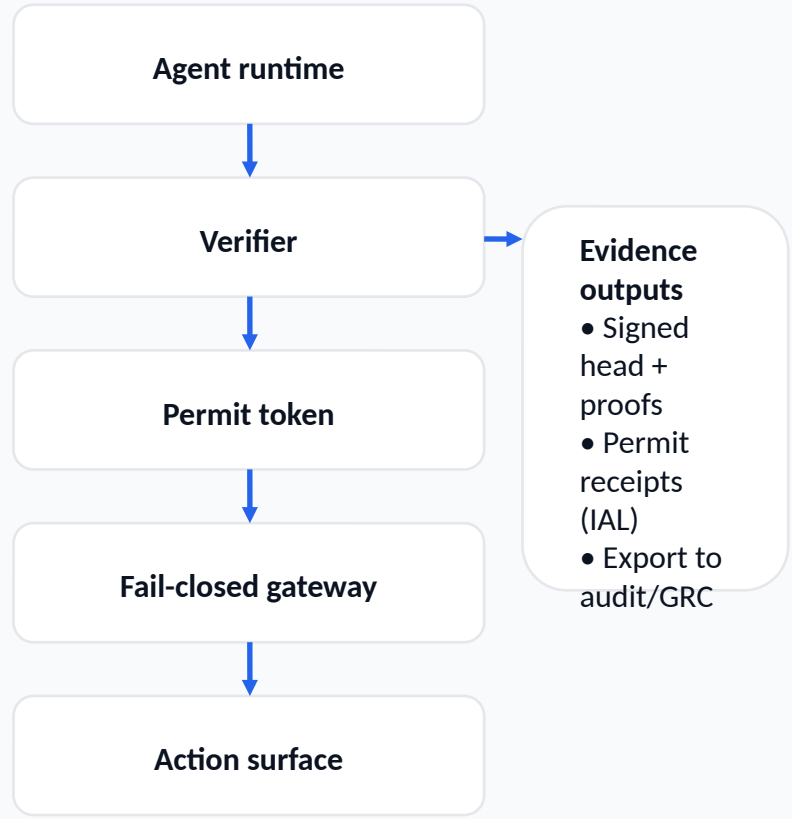
2) Permit-before-Action

On pass: mint signed, expiring permit.
Audience-bound (agent/tenant/mission).
Record receipt in IAL (exportable).

3) Fail-Closed Gateways

On fail: HOLD / QUARANTINE / DENY / ESCALATE.
Gateways require a valid permit ID before actions.

Execution path (non-bypassable)



EVIDENCE

Audit-ready proof artifacts (buyer-facing)

Signed head

Tree size, root hash, timestamp, signatures.
Optional witness co-signatures.

Inclusion proof

Proves ICC commitment is included in the current head.

Consistency proof

Proves append-only evolution when the head advances.

Permit token

Signed; audience-bound; nonce + counter; expiry.

IAL record

Hash-chained receipt per decision; exportable to audit systems.

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

ENFORCEMENT

Where enforcement lives (non-bypassable chokepoints)

Chokepoints (choose one for a pilot)

User space (agent runtime)

Kernel egress / syscall intercept (eBPF/cgroup)

Driver path (queue/doorbell before action)

Firmware / microcode (before dispatch)

Hypervisor intercept (VM-exit / gateway)

Fail-closed default: no valid permit ID → DENY / HOLD / QUARANTINE.

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)

Pick ONE gateway for a pilot. Prove non-bypassable control + receipts. Expand after.

Security invariants (what we promise and test)

Kernel enforcement invariants

✓ Fail-closed default: missing permit map entry → DENY
✓ Revocation overrides allow (even if permit still valid)
✓ TTL enforced in kernel state (short-lived permits)
✓ Loopback policy: verifier-only port 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 (evidence-only)

POLICY PULL

Why now: policy + procurement increasingly require provable controls

Headline mapping (non-exhaustive)

Framework	Timing / buyer ask	NOVACOV mapping
EU AI Act	High-risk rules staged 2026/2027; 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

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

PILOT

Design Partner offer (6–12 weeks)

Week 0–2	Week 3–6	Week 7–12
Align threat model + target gateway	Verifier MVP + permit-before-egress demo	Move enforcement into driver/firmware/hypervisor path

Plan

- Week 0–2: align threat model + target gateway; agree on evidence schema and reason codes
- Week 3–6: integrate verifier MVP in sandbox; demonstrate permit-before-egress + receipts export
- Week 7–12: move enforcement to driver/firmware/hypervisor pathway; production hardening plan + pricing 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
- Integration report + hardening plan + commercial conversion terms

NEXT

The ask: 30-minute technical deep dive (with 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 (agent runtime / infra / kernels/drivers)
- Security (appsec / product security / governance)
- Optional: audit/compliance stakeholder (evidence consumer)

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