IRONGAP TECHNOLOGIES

Deploy Frontier AI Inside Your Own Perimeter.

Vault-OS is air-gapped, zero-trust AI infrastructure built for organizations that cannot legally send data to the cloud.

PREPARED FOR	FOUNDER	CONTACT	STATUS
Enterprise Partners	Taha Halakooei	taha@iron-gap.com	Piloting Now

BRIEFING AGENDA

What We'll Cover

01 The compliance reality you operate under

02 Why cloud AI and DIY on-prem both fail

03 Introducing Vault-OS

04 Architecture & security model

05 Identity, access & data lifecycle

06 Deployment models & compliance mapping

07 Cost & risk comparison

08 Engagement path & next steps

SECTION 01 / 04

The Challenge

The regulatory and operational reality that makes mainstream enterprise AI a liability, not a tool.

Your Compliance Obligations Didn't Slow Down For AI

PIPL / CSL / DSL

China's cross-border data-transfer regime treats most cloud AI calls as a regulated export event.

GDPR / CCPA

Subprocessor exposure from model-vendor telemetry and training pipelines is increasingly rejected outright.

CMMC 2.0 Level 3

Defense Industrial Base contracts require demonstrable CUI protection cloud AI cannot provide.

Sector Mandates

Financial core infrastructure and SCADA/critical-infrastructure environments carry their own isolation rules.

Every mainstream “enterprise AI” product asks you to trust a subprocessor agreement. Your regulator does not accept a subprocessor agreement as a control.

It's Not A Feature Gap. It's An Architecture Problem.

The Path Is The Risk — Every prompt and document is an API call across a border you don't control — the network path is the risk, not the vendor's intentions.

Subprocessor Exposure — Vendor telemetry and model-improvement pipelines create subprocessor exposure regulators increasingly reject outright, no matter the contract language.

Contracts Don't Reroute Packets — No enterprise-tier terms-of-service clause changes the physical route the data actually takes to reach the model.

Air-Gapped Isn't The Same As Secure

Missing Lifecycle Controls — Hand-built local LLM deployments typically skip RBAC, audit ledgers, hardware licensing, and patch discipline — exactly the controls your auditors test for.

No Tamper-Evident Record — No tamper-evident record of who asked what, when, and at what clearance level.

No Anti-Clone Guarantee — No path to prove the install hasn't been cloned onto unauthorized hardware after the fact.

SECTION 02 / 04

The Solution

Vault-OS: a sealed enterprise AI appliance, engineered around four non-negotiable invariants.

Vault-OS: A Sealed Enterprise AI Appliance, Not A Science Project

On-premise LLM inference and retrieval-augmented generation, built from day one around four invariants.

Air-Gapped

No outbound network path exists in the application logic.

Zero-Trust

Every request re-validates identity, status, and token version.

Fail-Deadly

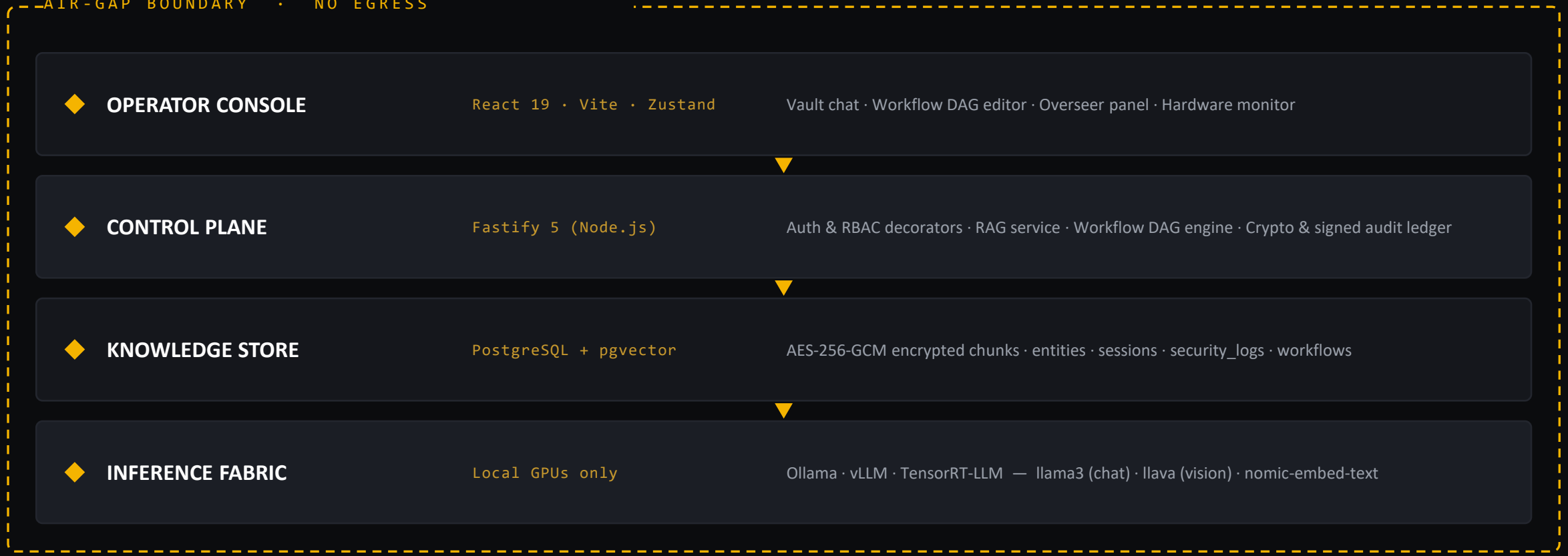
Missing secrets or unrecognized hardware halt the boot — never a silent degrade.

Hardware-Tethered

The install is cryptographically bound to its own TPM 2.0 silicon.

Every document, embedding, and conversation stays inside your walls — encrypted, audited, and destroyed on schedule.

Three Tiers, One Air-Gap Boundary



React 19 · Fastify 5 · PostgreSQL + pgvector · Ollama / vLLM / TensorRT-LLM — all co-resident on a single air-gapped host or an isolated internal segment.

Four Invariants, Enforced On Every Request

A request crossing the vault traverses every stage above, in order — the failure of any single layer does not collapse the model.

Air-Gapped

Local-only inference; internal-only workflow nodes; no telemetry phone-home; CORS anchored to internal hosts only.

Zero-Trust

Per-request JWT + DB status + token-version check; RBAC-filtered retrieval; ephemeral dropzone storage.

Fail-Deadly

Absent secrets halt the process. Failed migrations halt boot. Decorators fail closed on missing users.

Hardware-Tethered

SHA-256 fingerprint of machine UUID + MAC + TPM key, checked against a signed license before any service starts.

Nine Clearance Tiers. Every Action Signed.

9-Tier RBAC Lattice

From Chairman down to Standard — composable decorators enforce clearance on every route, not just at login.

TOTP With Replay-Proof Tickets

Two-factor login issues a single-purpose, 5-minute pre-auth ticket; no protected route will ever accept it as a session.

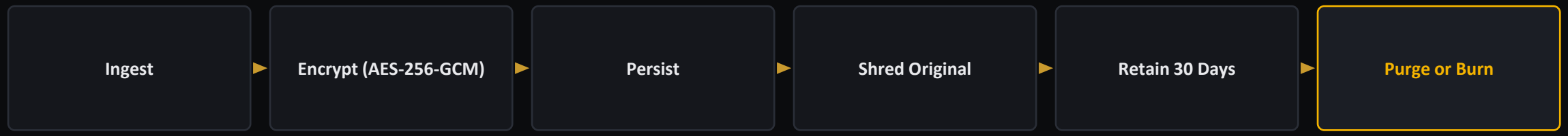
Instant Global Revocation

Each session carries a token version checked live against the database — terminate a user, and every outstanding token dies instantly.

Hash-Chained, Signed Ledger

Every privileged action is written to a tamper-evident audit chain; any retroactive edit breaks every downstream link.

Data That Proves Its Own Destruction



Zero-Trust Storage

Ingested originals are shredded immediately after encryption — no lingering plaintext copy anywhere on the appliance.

Encrypted Backups

Backups are streamed through AES-256-GCM under a separate recovery key, held only in the sealed environment.

The Burn Protocol

An irreversible, NIST SP 800-88–aligned multi-pass wipe of the dropzone, archive, models and database for physical-compromise scenarios.

SECTION 03 / 04

Why IronGap

How you deploy, what it costs, how it maps to your compliance obligations — and where we're still hardening.

Bring Your Hardware, Or Let Us Bring The Box

PATH A

BYOH — Bring Your Own Hardware

A software license deployed on your existing accelerators. Fastest path to a pilot; lowest CAPEX; keeps procurement inside your existing hardware policy.

PATH B

Turnkey Appliance

Hardware + Vault-OS, pre-sealed and shipped, with on-site installation. Highest margin path for us, fastest time-to-boot for you if you have no GPU fleet of your own.

Either path, same guarantees: the same 9-tier RBAC, hash-chained audit ledger, and hardware-tethered licensing ship on day one — deployment model changes only where the box sits.

Built To Answer The Questions Your Auditors Actually Ask

Regulation / Framework	How Vault-OS Addresses It
PIPL / CSL / DSL (China)	Zero cross-border transfer by architecture — all inference and storage stay on local silicon.
GDPR / CCPA	No subprocessor exposure; no third-party data-processing agreement is required.
CMMC 2.0 Level 3	Hardware-tethered, air-gapped deployment aligns with CUI-protection expectations for the Defense Industrial Base.
NIST SP 800-88	The Burn Protocol implements multi-pass sanitization aligned to media-destruction guidance.

The Real Comparison Isn't Software Price. It's Risk Exposure.

	Vault-OS	Cloud AI	DIY On-Prem
Upfront Investment	Moderate — license or appliance	Low — pay-as-you-go	High — build from scratch
Ongoing Operating Cost	Predictable — flat license tiers	Variable — scales with usage	High — in-house maintenance team
Compliance Fine Exposure	Structurally minimized	Material — cross-border transfer risk	Unverified — no audit trail
Data Breach Liability	Encrypted at rest, signed audit ledger	Shared responsibility, vendor-controlled	Depends entirely on internal discipline
Time-to-Value	Weeks — turnkey or BYOH	Days	12-18+ months to production-grade

Directional comparison for planning purposes — final commercial terms are quoted per deployment.

We Publish Our Own Residual Risk

Most vendors hide open findings. IronGap tracks them in the published technical whitepaper.

Constraint	Status & Plan
js_sandbox uses Node vm	Not a true security boundary. Admin-only today; slated for isolated-vm or removal.
Burn Protocol authority	Reachable at VIP clearance with no re-auth; under review for Chairman-only + confirmation token.
Self-update surface	Spawns a child server and copies staged files over root; slated for signed, verified update bundles.

The presence of a tracked-constraints register — not its absence — is the security signal.

SECTION 04 / 04

Getting Started

How to move from first call to a production deployment your security team has personally validated.

From First Call To Production In Four Steps

1

Assess

A compliance and security fit review against your existing regulatory obligations.

2

30-Day Pilot

A paid, on-site pilot running live on your data, on your hardware or ours.

3

Audit & Sign-Off

Your security team validates the ledger, RBAC lattice, and boot chain directly.

4

Enterprise Contract

An SLA-backed annual contract, expandable by seat and department over time.

The Cost Of Waiting Is A Compliance Finding

The Law Is Tightening — Cross-border data law is tightening, not loosening — PIPL, CSL/DSL, and the EU AI Act are all moving in one direction.

Competitors Are Already Moving — Competitors in your sector are already piloting sovereign AI infrastructure; the compliance-driven adoption curve rewards early movers.

Productivity Compounds Early — The earlier you deploy, the earlier your teams get the productivity gain your competitors are being denied by their own cloud-AI bans.

Request A Secure Audit

Book a 30-minute architecture review. Bring your compliance team — we'll bring the whitepaper.

Taha Halakooei · Founder/CEO & Chief Architect

taha@iron-gap.com · www.iron-gap.com