Kav AI — Security & Deployment Brief

For IT and OT security review: where the data goes, who controls the model, and how read-only is enforced.

Kav AI maintains a **strictly read-only OT posture**: it reads SCADA data via OPC UA and never writes to control systems, actuates valves, or commands field equipment. This boundary is enforced technically, not by policy alone: **(1)** OPC UA server user access rights restricted to Read and Subscribe services only; **(2)** the DMZ proxy/aggregator configured to reject all Write/Call requests; **(3)** firewall TCP/IP port restriction (4840/4843). Kav AI operates from the IT side of the OT/IT boundary (Purdue Level 3.5 and above) and supports outbound connection initiation from the OT side to satisfy strict no-inbound firewall policies.

Deployment tiers

Tier	Data residency	Availability
Cloud SaaS	Kav AI-managed GCP tenant (configurable region); multi-tenant with row-level security	Delivered (alpha)
Customer cloud tenant	Your own Azure or AWS environment — container package supplied by Kav AI, infrastructure controlled by you; no data leaves your environment	Q4 2026 target
On-premise / air-gapped	Your hardware, 100% air-gapped, self-hosted LLM inference; **no heartbeat, telemetry, or license callbacks of any kind**	H1 2027 roadmap

Hardware for the air-gapped tier (full stack including self-hosted LLM inference): min 128 GB RAM, 16 vCPU, 2× NVIDIA A100 80 GB or equivalent. The platform-only footprint excluding LLM inference is min 8 cores, 32 GB RAM, GPU optional.

Connectivity and firewall requirements

Connection	Protocol / port	Direction
Platform ⟶ OPC UA server	OPC UA (TCP) 4840	IT DMZ ⟶ OT DMZ; operator provisions the rule
Platform ⟶ PI Web API	HTTPS 443	IT ⟶ IT/OT DMZ, read-only
Operator browser ⟶ platform	HTTPS/WSS 443	Internet (SaaS) or internal (on-prem)
Platform ⟶ LLM API	HTTPS 443	Cloud tiers only — **absent air-gapped**
Platform ⟶ update service	HTTPS 443	Cloud tiers only — replaced by manual container delivery air-gapped

OPC UA sessions require SignAndEncrypt (minimum policy Basic256Sha256); authentication by username/password or X.509 certificate — anonymous connections are not permitted in production. Up to 10,000 monitored tags per facility in the initial release; subscription mode preferred over polling.

Data protection and identity

- **Encryption:** AES-256 at rest, TLS 1.2/1.3 in transit.
- **Secrets:** no plain-text credentials in container configuration; managed secrets store (HashiCorp Vault or cloud-native equivalent) with rotation through the store, and mTLS via X.509 client certificates for OPC UA where deployed.
- **Identity:** JWT with short-lived tokens; multi-tenant row-level security verified via independent pentest; SSO via SAML 2.0 / OIDC (delivered).
- **Tenant governance:** analysis data — suggestions, decisions, derived datasets — stays tenant-scoped and deletable, and is excluded from any shared training without written authorization.

Supply chain

All production container images are signed (Sigstore/cosign) and verified at runtime; CI/CD runs deep-layer vulnerability scanning (Trivy) before image promotion; an SBOM (SPDX format) ships with every production release for operator-side audit.

Compliance posture

Framework	Position	Target
SOC 2 Type II	Controls mapped; observation period underway	Q4 2026 target, pending audit window
IEC 62443	Read-only OPC UA, no OT writes, Purdue-compatible segmentation	H1 2027 (on-premise tier)
Data sovereignty	Customer-tenant and on-prem tiers keep all telemetry and imagery inside your perimeter	Q4 2026 (customer cloud)
Breach notification	Within 72 hours, per the MSA	Standard

Data deletion on contract termination within 30 days with certificate; backups purged within 90 days. Annual audit rights over Kav AI's data-handling practices, or following any security incident.

Kav AI Platform · PRD v3.12 · 2026-09-03 · Full detail: PRD Deployment Architecture + Appendix F (incl. F.7 Security Technical Reference) · vendor-qualification contact: security@kavai.net