

Kurahi Security Overview & Questionnaire Responses v1

Product: Kurahi 2.2.0 (agent runtime) · **Prepared by:** the KODA SōKEN team · **Date:** 25 October 2026

Every answer is taken from the Kurahi documentation or source tree and cites the document it comes from. Paths refer to the public repository github.com/koda-tokyo/kurahi; the same documents are published at <https://kurahi.kodasoken.com/docs/>. Answers marked **Not applicable (self-hosted)** concern services that do not exist because Kurahi runs on the customer's own infrastructure. Answers marked **Not yet** state a gap plainly. Fields marked (to be confirmed by KODA SōKEN) will be completed during procurement.

A. Company and service

#	Question	Answer	Source
A1	Legal name of the supplier	KODA SōKEN, Tokyo, Japan. The contracting entity is named in the written agreement.	Company page, kurahi.kodasoken.com/company
A2	Registered address, corporate number, invoice registration number	(to be confirmed by KODA SōKEN)	—
A3	Security contact	security@kodasoken.com	SECURITY.md
A4	Delivery model	Self-hosted software. One gateway process per identity runs on the customer's machine or server under the OS service manager. No Kurahi-operated cloud component in the evaluation configuration.	docs/security/SECURITY_OVERVIEW.md §1
A5	Does the supplier operate infrastructure that stores customer data?	Not applicable (self-hosted). The gateway does not contact Kurahi or its vendor at runtime.	docs/security/SECURITY_OVERVIEW.md §2 ; docs/security/privacy.md
A6	Sub-processors	None by default. Data goes only to the model provider, channels and business connectors the customer configures.	docs/security/privacy.md ; docs/capabilities.md
A7	Insurance (cyber / professional liability)	(to be confirmed by KODA SōKEN)	—

B. Certifications and assurance

#	Question	Answer	Source
B1	ISO/IEC 27001 certification	Not yet. Not held.	Trust Center, kurahi.kodasoken.com/trust
B2	SOC 2 report	Not yet. Not held.	Trust Center
B3	Independent penetration test	Not yet. No third-party penetration test report is available.	—
B4	Formal verification or conformance testing	A runtime continuity conformance suite (C0 to C9) runs in CI and produces signed reports. It checks retained state, files, permissions and transcripts, not equivalent reasoning across models.	docs/capabilities.md ; docs/continuity.md
B5			

#	Question	Answer	Source
	Japan AI Guidelines for Business	Kurahi is designed to support several common guiding principles. No claim of compliance or certification.	AI governance page, kurahi.kodasoken.com/ai-governance
B6	OWASP Top 10 for Agentic Applications	A self-assessment by the KODA SōKEN team (status 24 October 2026) rates 4 risks addressed, 6 partly addressed and 0 not addressed. It is not a certification and not a third-party audit; “addressed” does not mean the risk is gone.	docs/security/owasp-agentic.md

C. Data protection and privacy

#	Question	Answer	Source
C1	What data leaves the customer’s environment?	Model requests to the configured provider; messages on enabled channels; a version-only update check (npm installs); an mDNS announcement on the LAN; the installer download. Each has an off switch. Business connectors (Microsoft 365, Gmail, Entra ID, kintone, freee, Money Forward, SmarHR, LINE WORKS, Box) are off by default and exchange data with their service only after an administrator turns one on; with enterprise identity, sign-in goes through the organisation’s identity provider.	docs/security/privacy.md ; docs/security/SECURITY_OVERVIEW.md §2 ; docs/capabilities.md ; docs/security/identity.md
C2	Telemetry, analytics or crash reporting	None are sent.	docs/security/privacy.md
C3	Data residency	Chosen by the customer: local model, Japan-hosted model (default profile), external provider, or air-gapped (licensed separately).	docs/deployment.md
C4	Where is data stored?	In one directory under the customer’s control (KOL0_HOME). Nothing is written elsewhere except program files.	docs/security/SECURITY_OVERVIEW.md §1, §8
C5	Retention	Sessions and transcripts until deleted; logs rotated daily; audit chain in monthly files; support bundles only when created. Retention rules per data class (transcripts, work items, evidence, audit, attachments, outputs, memory, connector caches), legal holds and signed eDiscovery exports can be configured; every class is kept indefinitely until a retention rule names it, and deleted evidence leaves a signed tombstone.	docs/security/SECURITY_OVERVIEW.md §8 ; docs/security/data-governance.md
C6	Deletion	<code>kurahi uninstall --all --yes</code> removes configuration, secrets, sessions, audit and logs; <code>--dry-run</code> lists what would be removed.	docs/security/SECURITY_OVERVIEW.md §9
C7	Personal data minimisation towards AI models	The context guard replaces personal data in tool results with placeholders before the model sees it. Limits: images not inspected; detectors are patterns; user-typed chat text is sent as written.	docs/context-guard.md
C8	Does the supplier access customer data for support?	Only if the customer sends a support bundle, created by <code>kurahi support</code> and redacted.	docs/security/SECURITY_OVERVIEW.md §8
C9	Japanese personal identifiers (My Number)	Before a tool that sends or writes outside Kurahi runs, every string in its final arguments is checked for Japanese personal identifiers. By default a valid 個人番号 (My Number,	

#	Question	Answer	Source
		check digit validated) refuses the call and is masked in audit rows and evidence.	docs/security/data-governance.md

D. Access control and identity

#	Question	Answer	Source
D1	Default network exposure	Loopback only. Remote access requires an explicit gateway. bind setting and a token on every connection.	docs/security/SECURITY_OVERVIEW.md §3
D2	Authentication of API calls	Every /api/* route and WebSocket method requires a credential, except health checks. Credential attempts are rate limited (20 per minute per source).	docs/security/SECURITY_OVERVIEW.md §3
D3	Role or scope model	Device pairing with scopes observe < attach < admin < owner, short-lived access tokens, rotating refresh tokens, revocation, and re-consent (default 90 days). With enterprise identity, people hold roles (org-admin, security-admin, auditor, department-manager, approver, operator, member, viewer) within departments and teams, mapped from IdP groups; no role gives the gateway admin scope.	docs/security/SECURITY_OVERVIEW.md §4; docs/security/identity.md
D4	Single sign-on (SAML/OIDC)	Built in: OIDC (for example Entra ID, Google, Okta, Keycloak) or SAML 2.0 sign-in for people, SCIM 2.0 provisioning, departments, teams and roles. Off by default; included in every plan, Free included. With ssoRequired: true, token logins of people become an audited break-glass that needs a stated reason.	docs/security/identity.md; docs/capabilities.md; docs/releases/2.2.0.md
D5	Credential separation between agents	With auth.isolation: "explicit", each agent sees only the credential profiles granted to it; nothing is copied or merged from the main agent.	docs/gateway/security/index.md (Credential isolation between agents)
D6	Multi-factor authentication	Not applicable (self-hosted). Device pairing uses Ed25519 proof of possession; host login is governed by the customer. With enterprise identity, people sign in through the organisation's identity provider.	docs/security/SECURITY_OVERVIEW.md §4; docs/security/identity.md

E. Secrets and cryptography

#	Question	Answer	Source
E1	Storage of secrets	Auth profile store and encrypted secrets store under KOL0_HOME; gateway token file mode 0600.	docs/security/SECURITY_OVERVIEW.md §4, §5
E2	Secrets in logs or command lines	Never printed, never accepted as command-line arguments, redacted from logs.	docs/security/SECURITY_OVERVIEW.md §5
E3	Encryption at rest	Secrets store only. Full-disk encryption of the host is recommended; Kurahi adds no other disk encryption.	

#	Question	Answer	Source
			docs/security/ SECURITY_OVERVIEW.md §8
E4	Encryption in transit	Remote clients should use HTTPS/WSS (for example through a tailnet); TLS fingerprint pinning is available for remote wss://. The default loopback bind keeps traffic on the host.	docs/gateway/security/ index.md
E5	Signatures	Policy documents, evidence bundles and the AI inventory are signed with Ed25519.	docs/security/ SECURITY_OVERVIEW.md §7a; docs/evidence.md; d ocs/inventory.md

F. Agent control and AI-specific risks

#	Question	Answer	Source
F1	How are agent actions restricted?	One signed policy per gateway: tool allow and deny lists, providers, models, channels, budgets, paths, egress and MCP allow lists. Enforced at every tool call, model request, channel message and job start.	docs/ policy.md; d ocs/ security/ SECURITY_OV ERVIEW.md §7a
F2	Human approval of consequential actions	Tool classes can require a human approval, in the terminal or over a chat channel; waits are bounded by the policy timeout.	docs/ policy.md; d ocs/ capabilities .md
F3	Emergency stop	kurahi policy kill --agent <id> --all; running jobs are aborted and the state survives a restart until released.	docs/ security/ SECURITY_OV ERVIEW.md §7a
F4	Third-party agent CLIs	A compatibility contract lists, per adapter, what is enforced before the CLI runs, what is only detected after, and what is not controlled.	docs/agents/ external- adapters.md
F5	Protection of important files from agent edits	snapshots.required refuses an edit to a protected file when its recovery snapshot cannot be captured. Sandboxed edits are not covered; shell commands are not covered by snapshots.	docs/tools/ snapshots.md; docs/ capabilities .md
F6	Inventory of AI systems	kurahi inventory lists agents, models, tools, policy rules, MCP servers, channels and evidence; signed export and verification.	docs/ inventory.md
F7	Ransomware	Detection only: decoy files and mass-encryption detection stop every agent and write an audit entry. Files encrypted before a signal fires are damaged; it stops agents, not the malware; no off-machine copies.	docs/ shield.md; d ocs/ capabilities .md
F8	Prompt injection	Content from outside (web pages and search results, remote agents, external hooks, business connector results) reaches the model inside untrusted-content markers; memory writes after	docs/ security/

#	Question	Answer	Source
		such input are quarantined, and consequential actions need a policy decision or a person's approval. The markers reduce prompt-injection risk; they do not prevent it. Messages from channel users and MCP results are not wrapped in the markers. Rated partly addressed (ASI01).	owasp-agentic.md
F9	Inbound files and malware	Mail attachments are quarantined, checked by a static file policy (executables, scripts, macros, encrypted archives refused), scanned with ClamAV or Microsoft Defender and only then stored in the Project; text is extracted in a separate bounded process. The scan fails closed: a timeout, crash or output without a verdict is never treated as clean, and by default (requireScan) an unscanned file is refused. Residual: unknown (zero-day) malware.	docs/security/attachment-intake.md; docs/capabilities.md
F10	Consumer channels and plugins in a business deployment	The enterprise package omits consumer-channel plugins and consumer logins; under the enterprise profile the runtime refuses an excluded plugin even if someone copies it in, and writes the refusal to the audit log. Some consumer-channel code remains in the core, inert: it runs only when its channel plugin registers it, and those plugins are not shipped.	docs/security/enterprise-profile.md; docs/releases/2.2.0.md
F11	Autonomous actions (Autopilot)	Routines run at an autonomy level set by a person (L0 observe to L3 act externally). The reasoner has no tools; the engine executes only what the level allows, and external actions need a person's standing approval limited to named recipients, a template, a daily maximum and an expiry. A behaviour watchdog pauses a routine acting outside its pattern, and a global stop halts every routine.	docs/autopilot.md; docs/releases/2.2.0.md

G. Logging, monitoring and evidence

#	Question	Answer	Source
G1	Audit logging	Authentication, pairing, attach/detach, admin actions and tool executions are appended to a hash-chained audit log; kurahi audit verify checks the chain.	docs/security/SECURITY_0 VERVIEW.md §6
G2	Tamper resistance	Tampering is detectable, not preventable, unless the files are shipped to write-once storage.	docs/security/SECURITY_0 VERVIEW.md §6
G3	Export of audit to SIEM	Licensed separately (enterprise feature).	docs/capabilities.md
G4	Evidence of what an agent did	Signed, hash-chained evidence bundle per job, verifiable offline with kurahi evidence verify. Every Autopilot run also writes a signed evidence bundle, chained to the routine's previous run. Signed evidence establishes the integrity and provenance of the record, not that a decision was correct.	docs/evidence.md; docs/autopilot.md
G5	Integrity of configuration		docs/security/

#	Question	Answer	Source
		Boot integrity gate re-verifies protected policy files on every start; enforce mode refuses to boot on mismatch.	SECURITY_0 VERVIEW.md §7

H. Secure development and supply chain

#	Question	Answer	Source
H1	Code signing of binaries	Not yet. Builds are not yet code-signed or notarised.	docs/capabilities.md
H2	Release integrity	SHA256SUMS.txt with every release; installers refuse a mismatched download.	docs/security/ SECURITY_OVERVIEW.md §10
H3	SBOM and provenance	CycloneDX SBOM and build provenance attestation (gh attestation verify) with every release.	docs/security/ SECURITY_OVERVIEW.md §10; d ocs/PACKAGING.md
H4	Dependency management	Dependencies pinned in pnpm-lock.yaml; CI runs a secret scanner; Dependabot is configured.	docs/security/ SECURITY_OVERVIEW.md §10; . github/dependabot.yaml
H5	Testing	About 9,700 automated tests (unit, gateway, end-to-end and extension suites) and conformance checks in CI.	Kurahi homepage figures; .gi thub/workflows/ci.yaml

I. Vulnerability and incident management

#	Question	Answer	Source
I1	Vulnerability reporting	Privately to security@kodasoken.com.	SECURITY.md
I2	Response targets	Acknowledgement within 3 working days; triage within 10 working days; fix or mitigation for confirmed high or critical issues within 30 days; coordinated disclosure at most 90 days after the report.	SECURITY.md
I3	Incident notification to customers	(to be confirmed by KODA SōKEN) in the support terms.	—
I4	Business continuity of the service	Not applicable (self-hosted). The software keeps running without KODA SōKEN; the customer controls backups of KOL0_HOME.	docs/ security/ SECURITY_OVE RVIEW.md §1, §8

J. Licensing and evaluation

#	Question	Answer	Source
J1	Licence model	Open core: MIT (protocol, SDK, plugin API, CLI); FSL-1.1-Apache-2.0 (gateway runtime); proprietary (identity engine, enterprise features).	docs/security/ SECURITY_OVERVIEW. md §12
J2	Evaluation terms	90-day evaluation licence; currently a draft pending legal review.	docs/ EVALUATION_LICENSE .md