CYBERPOCKET · WHITEPAPER

# Security & Privacy Whitepaper

How CyberPocket handles your alerts, your data, and your trust.

Defensive-only alert analysis for the 71% of organizations without a SOC.
Deterministic redaction · zero-retention by default · no model training on customer data.
Companion to the investor pitch deck and the public Trust & Transparency page.

**Version 1.0 · June 2026 · trust@cyberpocket.org**

# Contents

# 1 · Executive summary

**Kicker**

CyberPocket is a defensive-only security alert analyzer for the 71% of organizations that do not operate a Security Operations Center. Users paste an alert; the product returns a redacted, plain-English risk assessment, evidence breakdown, and recommended next steps.

This whitepaper documents the security and privacy controls we operate today, the assurance work we are progressing (SOC 2 Type II readiness, ISO 27001 planning), and the boundaries of our shared-responsibility model. Where a control is in progress, we say so plainly — we do not claim certifications we have not earned.

| | |
|---|---|
| Default posture | Zero-retention for anonymous analyses; 30-day retention for authenticated workspaces; deterministic PII/secret redaction before any model call. |
| Encryption | TLS 1.3 + HSTS in transit; AES-256 at rest on managed Postgres and object storage; per-tenant keys on the Business tier. |
| No model training | Customer alert content is contractually excluded from training, fine-tuning, and evals across every model provider we use. |
| Assurance roadmap | SOC 2 Type II readiness in progress; ISO 27001 in Year-2 roadmap; HIPAA / FedRAMP demand-gated. |

# 2 · Product context & threat model

CyberPocket sits between an end-user (often a non-specialist business owner, MSP analyst, or junior IT operator) and one or more hosted large language models. The primary trust boundaries are: (a) the user's browser, (b) the CyberPocket backend on a serverless edge runtime, (c) the managed database and object storage, and (d) the model provider.

**Layer 4 · User-controlled**
Browser, paste action, workspace settings, redaction overrides.

**Layer 3 · CyberPocket backend**
Auth, RLS, deterministic redaction, safety harness, audit logging.

**Layer 2 · Managed data plane**
Postgres (RLS-enforced), object storage, KMS-managed keys.

**Layer 1 · Hosted model provider**
Inference only; no-training contractual posture; redacted input.

### In-scope threats

• Accidental disclosure of PII, secrets, or internal infrastructure detail inside pasted alerts.

- Cross-tenant data access via the application or database paths.
- Misuse attempts — prompts that solicit offensive payloads or attacker tradecraft.
- Compromise of a model-provider API key or webhook secret.
- Insider access to customer alert content by CyberPocket personnel.

### Explicitly out of scope (handled by the customer)

- Endpoint security on the operator's device.
- What content the operator chooses to paste into the analyzer.
- Identity provider configuration and password hygiene on the customer side.

# 3 · Data flow & deterministic redaction

Every analysis follows the same six-stage pipeline. Redaction runs on the CyberPocket backend before any byte leaves our trust boundary toward the model provider. The user sees a side-by-side diff of what was redacted and why before the prompt is dispatched.

**ANALYSIS PIPELINE**


### What the redaction layer removes

| | |
|---|---|
| Identity | Email addresses, full names tagged in alert templates, usernames matching known directory patterns. |
| Network | IPv4 / IPv6 addresses, MAC addresses, internal hostnames, FQDNs matching configured internal domains. |
| Secrets | JWT / bearer tokens, AWS / GCP / Azure access key formats, generic high-entropy strings, credit-card / SSN-shaped numerics. |
| Custom | Per-workspace regex rules (Business tier) — for example, a customer ID format unique to the organization. |

# 4 · Encryption — in transit, at rest, in use

### In transit

All client–server traffic uses TLS 1.3 with HSTS, modern cipher suites only (no RC4, no 3DES, no TLS 1.0 / 1.1), and HTTP/2. The published domain serves an HSTS max-age of 1 year with includeSubDomains. Internal service-to-service calls inside the managed runtime are mutually authenticated.

### At rest

Managed Postgres and object storage encrypt all data at rest with AES-256 using provider-managed KMS. Backups are encrypted with the same posture and follow a 7-day rolling window. The Business tier supports per-tenant encryption keys; BYOK (customer-managed KMS) is on the roadmap.

### In use

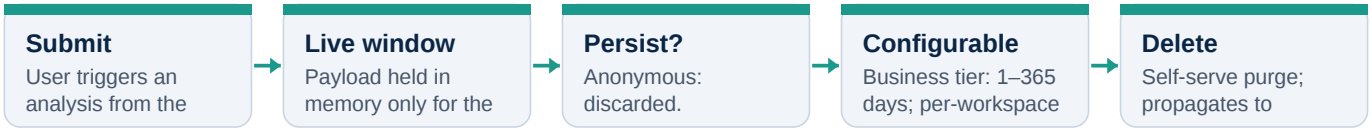
The redaction layer executes inside the CyberPocket backend before the prompt is constructed, so raw PII and secrets never reach the model provider. Process memory holding alert content is short-lived and is not written to disk outside the optional history record.

# 5 · Access control & privileged operations

| | |
|---|---|
| RBAC | Three roles per workspace: Owner, Analyst, Viewer. Permissions are enforced server-side on every mutation, not in the client. |
| Tenant isolation | Row-level security policies on every tenant table; deny-by-default; security-definer helper functions for role checks to prevent recursive policy evaluation. |
| Enterprise SSO | SAML / OIDC SSO and SCIM provisioning on the Business tier; IP allow-listing and session-lifetime controls. |
| Privileged ops | Internal admin actions require SSO + hardware MFA, are scoped by least privilege, and emit immutable audit events. No engineer has standing access to customer alert content. |
| Break-glass | Customer-content access requires two-person approval; the workspace Owner receives an automatic notification within 1 hour with actor, scope, and justification. |

# 6 · Retention, data subject rights, deletion

RETENTION LIFECYCLE




Anonymous one-off analyses are zero-retention — discarded immediately once the response is returned to the browser. Authenticated workspace analyses default to a 30-day rolling window and are configurable down to 1 day or up to 12 months on the Business tier.

Data subject access requests (DSARs) are handled in ≤ 30 days. Workspaces can self-serve export their full history as JSON, purge a single analysis, or purge the entire workspace from in-app settings. Deletions propagate to encrypted backups within the 7-day rolling backup window.

# 7 · Model-provider handling & no-training guarantee

CyberPocket routes between several hosted LLM providers based on latency, cost, and capability. For every provider we use, the integration is configured with the no-training / zero-data-retention flag set, and the contractual posture excludes customer content from training, fine-tuning, and evaluation datasets.

Provider API keys live in the platform secret store. They are never written to repository code, environment files distributed with builds, or log aggregators. Rotation is one-click from the admin console. Webhook callbacks are verified with HMAC-SHA-256 and constant-time comparison.

# 8 · Compliance & assurance — SOC 2 / ISO readiness

We do not currently hold SOC 2, ISO 27001, or HIPAA attestations. The controls below are operational today; the attestations are sequenced on the roadmap with honest status markers.

| | |
|---|---|
| SOC 2 Type II | Readiness phase. Control mapping complete across Security, Availability, and Confidentiality TSCs. Policies in final review. Auditor selection underway. Observation window planned for the next quarter. Pre-audit gap assessment available under NDA. [Placeholder for Type II report.] |
| ISO 27001 | Planned for Year 2. ISMS scope drafted around the analyzer service and supporting infrastructure. Statement of Applicability and risk register maintained internally and reviewed quarterly. [Placeholder for certificate.] |

| | |
|---|---|
| HIPAA | Demand-gated. Will be pursued when an anchor design partner with a covered-entity use case commits. Today, customers should not submit PHI into CyberPocket. |
| GDPR / UK GDPR | DPA available on request for paid workspaces. EU data residency option on the Business tier. DSAR turnaround ≤ 30 days. No sale of personal data. |
| Penetration testing | Annual third-party application pen test; remediation tracked to closure with severity-based SLAs. Latest summary available under NDA. |
| Vulnerability mgmt | Dependency scanning on every PR; critical CVEs patched within 7 days; high within 30 days; quarterly review of internal controls effectiveness. |

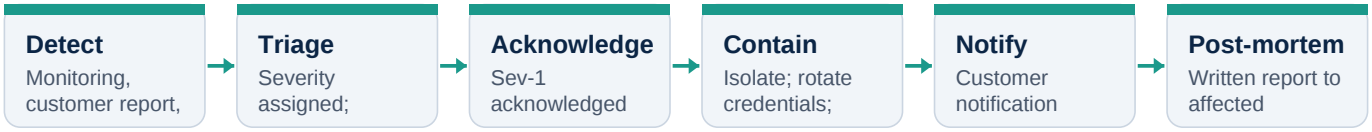
# 9 · Audit logging

Privileged and tenant-administrative actions are written server-side to public.audit_logs. The table is append-only from the application path; writes carry the actor identity, target object, timestamp, source IP, request ID, and a structured metadata payload. Logs are retained for 12 months and streamed to the customer's SIEM on the Business tier.

| | |
|---|---|
| Authentication | Sign-in success / failure, password reset, SSO assertion processed, session revocation. |
| Authorization | Role grants / revocations, workspace invitations, SCIM events. |
| Data | Export, deletion, redaction-rule edit, retention-policy change. |
| Billing & config | Plan change, payment method update, secret rotation, webhook re-configuration. |
| Break-glass | Internal access to customer content (with justification); auto-notification to the Owner. |

# 10 · Incident response & business continuity

**INCIDENT LIFECYCLE**




| | |
|---|---|
| On-call | 24×7 rotation; primary + secondary; runbook reviewed after every Sev-1 or Sev-2. |
| Sev-1 SLA | Acknowledgment within 1 hour; status page updated within 30 minutes of acknowledgment. |
| Notification | Affected customers notified within 72 hours of confirmation; regulator notification per applicable law. |
| Tabletop | Quarterly exercises covering data exposure, credential compromise, and provider outage. |
| Continuity | Multi-region failover for the inference and database tiers; RPO ≤ 24 h, RTO ≤ 4 h for the analyzer path. |

# 11 · Shared responsibility

| | |
|---|---|
| CyberPocket | Platform security, redaction layer, encryption, RBAC enforcement, audit logging, incident response, vendor due diligence on model providers, assurance roadmap. |
| Customer | What content is pasted into the analyzer, SSO / IdP configuration, role assignment within a workspace, endpoint security on the operator's device, internal escalation on findings. |
| Model providers | Bound by no-training contractual posture, regional inference where elected, provider-side encryption in transit and at rest; CyberPocket validates posture at onboarding and on renewal. |

# 12 · Contact & disclosure

Security questions, DPA requests, pen-test summaries, and SOC 2 readiness evidence packages: **trust@cyberpocket.org**.

Coordinated vulnerability disclosure: **security@cyberpocket.org** and .well-known/security.txt on the production domain. We commit to acknowledgment within 72 hours, status update within 7 days, and to never pursue good-faith researchers.

This document is a public companion to the in-app Trust & Transparency page and the investor pitch deck. It is reviewed at least quarterly and on every material change to the platform's security posture.