Algorithmic Accountability Guideline

PROCUREMENT READY | PUBLIC TRUST | SD4

Public declaration of AI logic, API dependencies, data sovereignty, and dataset governance

This guideline sets out CyberPlug's obligations and operational standards with respect to algorithmic transparency, data governance, and national data sovereignty. It is intended to satisfy the accountability requirements of applicable AI governance frameworks and to inform regulators, auditors, and enterprise customers.

1. External API Dependencies

API / Service	Provider	Purpose	Data Sent	Governance Status
NVD CVE Feed	NIST (US Gov)	Vulnerability intelligence enrichment	None — pull only	Approved
OSV.dev Feed	Google Open Source	Open-source package vulnerability data	Package names only	Approved
GitHub Advisory DB	GitHub (Microsoft)	Advisory enrichment for code dependencies	None — pull only	Approved
Shodan API	Shodan LLC	External attack surface mapping	IP ranges only	Approved — DPA in place

2. Internal Platform APIs

API	Function	Authentication	Rate Limits	Data Classification
Scan Orchestration API	Initiates and manages scan jobs	mTLS + service token	50 req/min per tenant	Internal — security metadata
Findings API	CRUD operations on vulnerability findings	OAuth 2.0 + RBAC	200 req/min per tenant	Internal — vulnerability data
Remediation AI API	Generates AI-powered fix recommendations	Internal service mesh (mTLS)	100 req/min per tenant	Internal — AI model output
Report Export API	Generates and streams compliance reports	OAuth 2.0 + RBAC	10 req/hr per tenant	Internal — compliance data
Webhook Delivery API	Pushes alerts to tenant-defined endpoints	HMAC-SHA256 signing	500 events/hr per tenant	External — alert metadata

3. National Data Sovereignty

Regional Isolation

CyberPlug operates independent deployment stacks per designated region: West Africa (Ghana), Europe, and North America. All scan data, findings, and audit logs for a tenant are processed and stored exclusively within the tenant's selected region. No cross-regional data transfer occurs without explicit written consent.

Storage & Encryption Controls

Tenant vulnerability data is stored in dedicated, logically isolated PostgreSQL instances within the relevant regional cloud zone. AES-256 encryption at rest. TLS 1.3 minimum in transit.

Critical Security Infrastructure

Threat intelligence feeds (CVE, OSV) are cached and processed within the regional stack. The AI remediation model is deployed per-region — model weights are not shared across regional boundaries. All audit logs are tamper-resistant and regionally independent.

4. Training Datasets

Dataset	Source	Purpose	PII / Sensitive Data
NVD CVE Historical (2002–present)	NIST	Vulnerability classification & severity prediction	None
OWASP Benchmark Suite	OWASP Foundation	False positive/negative calibration	None — synthetic data
CWE Enumeration Corpus	MITRE Corporation	Weakness classification & remediation mapping	None
Anonymised Remediation Outcomes	CyberPlug (internal)	Model fine-tuning on remediation effectiveness	Stripped — tenant identifiers removed before training

5. Model Transparency Controls

EXPLAINABILITY: Every AI recommendation includes a plain-language rationale explaining why the finding was flagged, which rule or pattern it matched, and what remediation action is recommended

CONFIDENCE SCORING: Every recommendation includes a confidence score (Low / Medium / High). Low-confidence outputs are automatically flagged for mandatory human review

BIAS MONITORING: Systematic over- or under-flagging monitored quarterly across asset types, technology stacks, and industry sectors

RIGHT TO EXPLANATION: Tenants may request a detailed explanation of any finding via the support portal. Explanations provided within 5 business days

OPT-OUT: Tenants may configure CyberPlug to operate in rules-only mode, disabling AI-generated recommendations while retaining full scan functionality