CYE AI
Governance Overview

December 2025

AI Principles

To ensure responsible innovation within CYE's Cyber Exposure Management platform, CYE AI adheres to core AI principles guiding the development and operation of all AI features:

Privacy & Security by Design

AI components operate within CYE AI's secure enclave. Data isolation and strong security controls are foundational.

Transparency

CYE AI clearly marks AI-powered features within the UI with an **"AI" badge**. Users are informed when interacting with AI-driven functionality.

Human-led Supervision

AI augments, but does not replaces, human judgment. Users retain full control and may override AI outputs at any time.

Accountability & Oversight

AI systems are governed by structured, cross-functional human oversight, ensuring safety, compliance, and alignment with CYE AI's policies.

Reliability & Testing

CYE AI validates AI features through security reviews, testing, and monitoring to ensure accuracy and stability.

Overview of Generative AI Features

CYE AI integrates selective Generative AI (GenAI) functionalities that support Cyber Risk & Threat Exposure Management workflows. All features operate under CYE's governance, privacy, and security controls.

Feature	Description	Model/Provider	Purpose	User Interaction
Unstructured Data Ingestion	Secure ingestion and parsing of files (PDF, DOCX, TXT, CSV), contextual embedding, tagging, and enrichment	Private LLMs hosted on AWS Bedrock	Enhance risk and control data with unstructured intelligence	File upload interface
AI Chatbot Agent	Context-aware querying of organizational risk data, explanation of findings, trend summaries	Private LLMs running in CYE AI's secure enclave on AWS Bedrock	Provide accessible insights into complex cybersecurity data	In-app conversational UI
Refine Finding Description	Rephrase or shorten findings description as they are been added manually to CYE AI	Private LLMs hosted on AWS Bedrock	Assist CYE AI users in creating findings that are well explained, standardized and coherent.	Field update

NIST AI RMF Reference:

- **MAP 1.1:** Intended purposes, potentially beneficial uses, context-specific laws, norms and expectations, and prospective settings in which the AI system will be deployed are understood and documented.
- **MAP 2.1:** The specific tasks and methods used to implement the tasks that the AI system will support are defined.
- **GOVERN 1.6:** Mechanisms are in place to inventory AI systems and are resourced according to organizational risk priorities.
- **GOVERN 2.1:** Roles and responsibilities and lines of communication related to mapping, measuring, and managing AI risks are documented and are clear to individuals and teams throughout the organization.

High-Level Architecture

Architecture Summary

- **Frontend:** Role-based authenticated web and API clients
- **Backend:** Containerized AI microservices isolated within CYE AI's secure VPC
- **Inference Layer:** Private LLMs deployed within controlled compute clusters; no external API dependencies
- **Data Stores:** Encrypted relational and vector databases
- **Observability Stack:** MLflow for model traceability and lineage; Datadog for real-time telemetry and anomaly detection

Data Flow

1. User input is received via TLS
2. Input is pre-processed and tokenized
3. Inference executed by a private model endpoint inside CYE AI's secure enclave
4. Prompts and outputs are not retained by the model host after completion
5. All interactions are logged with strict access controls and monitoring

NIST AI RMF Reference:

- **MAP 1.5:** Organizational risk tolerances are determined and documented.
- **MAP 2.2:** Information about the AI system's knowledge limits and how system output may be utilized and overseen by humans is documented.
- **MEASURE 2.1:** Test sets, metrics, and details about the tools used during TEVV are documented.
- **MEASURE 2.4:** The functionality and behavior of the AI system and its components – as identified in the MAP function – are monitored when in production.

User Data Protection

Data Handling & Isolation

- Each tenant's data is isolated within CYE AI infrastructure
- No cross-tenant access
- Temporary inference artifacts are automatically purged

Encryption & Access Control

- TLS enforced for all traffic
- Strong encryption at rest
- Strict RBAC and least-privilege IAM
- Full audit logs of model interactions and administrative action

NIST AI RMF Reference:

- **MANAGE 1.2:** Treatment of documented AI risks is prioritized based on impact, likelihood, and available resources or methods.
- **MANAGE 3.1:** AI risks and benefits from third-party resources are regularly monitored, and risk controls are applied and documented.
- **MANAGE 4.1:** Post-deployment AI system monitoring plans are implemented, including mechanisms for capturing and evaluating input from users and other relevant AI actors, appeal and override, decommissioning, incident response, recovery, and change management.

Data Use & Training Policy

No Customer Data Used for Model Training

- CYE AI operates private, non-training GenAI model instances
- Customer data, chatbot transcripts, and uploaded files are **not** used for training or fine-tuning
- Policy compliance reviewed quarterly

Data Scope

CYE AI processes:

- Explicit user-provided content
- Relevant organizational data stored within the system

No redaction or identifier stripping is performed, as private LLM hosting provides sufficient protection.

User Opt-Out Controls

- Customers may request to disable the chatbot through CYE's support team
- When disabled, the model receives **no data** from that customer's workspace
- All other AI features are optional: users simply choose whether to use them

Transparency & User Notice

AI Labels

AI-powered features include an **"AI" badge** in the UI.

AI Feature Documentation

Each AI feature must include:

- Data processed
- Expected outputs
- Known limitations
- Safety considerations

This documentation is completed during the feature's design phase.

AI Output Disclaimer

AI-generated outputs may occasionally contain inaccuracies or incomplete information. Users should review and, where needed, override results.

Third-Party Provider Governance

Provider Scope

CYE AI relies exclusively on AWS Bedrock configured with **private network access (VPC endpoints), account-level isolation, and non-retentive foundation model invocation**, aligned with AWS security and compliance best practices.

Provider Behavior

- AWS Bedrock instances do **not** retain customer data
- No training or fine-tuning on customer content
- No external safety classifiers or third-party abuse detection mechanisms are permitted

Third-Party Risk Management

- Provider must pass CYE's technical and security evaluation
- Procurement ensures alignment with CYES's privacy and security standards

Security & Privacy Reviews

Secure Development Practices

- Integration with SSDLC
- Mandatory peer review for all code changes
- SAST/DAST scanning
- Inference gateway protections against prompt injection, chaining, and data exfiltration

Testing & Validation

- Security reviews
- Adversarial testing
- Controlled scenario evaluations

Security Assessments

Review Type	Date	Scope	Outcome
Red Team Assessment	Q3 2025	Full CYE AI platform and AI pipeline	No critical vulnerabilities
Threat Modeling (STRIDE / ATLAS)	Q2 2025	Inference and ingestion layers	Controls validated
Privacy Impact Assessment	Q2 2025	Data retention & residency	Compliant
Penetration Testing	Annual	APIs & inference endpoints	All findings remediated

AI Lifecycle & Change Management

Pre-Deployment Requirements

Any new AI feature that processes user data requires:

- Security review
- Privacy/legal review
- Data flow documentation
- Safety and reliability evaluation
- Approval by the AI Governance Committee

Preview Stage

CYE AI may release features in:

- **Private Preview**
- **Public Preview**

Preview features:

- Are labeled as "Preview"
- Undergo review once user data is involved
- Are evaluated for performance, safety, and customer experience before GA

Human Oversight

- Users may disregard or override AI suggestions at any time
- AI does **not** autonomously modify customer data without explicit user confirmation

Compliance Alignment & Governance

AI Governance Committee

The committee is comprised of senior representatives from the following departments:

- IT Security
- Legal
- Tech
- Business Operations

Responsibilities include:

- Review of major new AI features
- Approval authority for releases
- Ownership of the AI Risk Register
- Oversight of compliance with CYE AI's AI Principles

Framework Mapping

CYE AI aligns with the following frameworks and standards:

Framework	Relevant Controls	Evidence / Implementation
NIST AI RMF	Govern, Map, Measure, Manage	AI Risk Register, MLflow audit logs, data flow diagrams
ISO/IEC 42001	AI management system	Policies, assessments, governance controls
ENISA AI Cybersecurity	Integrity & robustness	Datadog dashboards, security monitoring
GDPR / EU AI Act	Transparency & data minimization	Feature documentation, user opt-out
SOC 2 / ISO 27001	Information security controls	Encryption, access controls, monitoring

Monitoring & Incident Response

Continuous Monitoring

CYE AI conducts continuous monitoring of AI systems, including:

- Drift detection
- Anomaly and abuse detection
- Model usage and performance telemetry

Model lineage, parameters, and metadata are recorded via MLflow.

AI-Specific Incident Response

AI incidents receive:

- Special triage classification
- Immediate escalation to Security and the AI Governance Committee
- Customer notification if their data is affected

Incidents follow CYE AI's ISO 27001–aligned incident management protocol, including root-cause analysis and executive reporting.

Continuous Improvement

CYE AI continuously evaluates and enhances AI capabilities. Future improvements **may** include:

- Expanded adversarial testing
- Additional robustness and hallucination evaluation techniques
- Enhanced model monitoring pipelines

Appendix

Service Accuracy & Hallucinations

Generative AI models may occasionally produce inaccurate or incomplete outputs. Users should review all results before applying them in operational decisions.

CYE AI Chatbot Disclaimer

Your interactions with the CYE AI chatbot are designed with **security and privacy as top priorities**. Please note the following key terms:

Data Protection & Model Usage

- **Data Security:** All conversation data, organizational data, and AI responses are processed and maintained **exclusively within CYE AI's secured internal environment**. We utilize Private Large Language Model (LLM) instances, meaning **your data does not leaves our controlled infrastructure.**
- **No Training on Your Data:** CYE AI **does not** use customer data, chatbot conversations, or organizational content to train or improve AI models.
- **User Control:** No data is shared with the AI unless **explicitly provided by the user during** a conversation.

Guidelines & Compliance

- **Confidentiality:** While our platform enforces strong safeguards, users should **avoid sharing highly sensitive personal information** (e.g., passwords, credit card numbers, health identifiers) through the chatbot.
- **Compliance:** CYE AI aligns with industry best practices for cybersecurity, data residency, and regulatory compliance, ensuring your information is protected according to **strict standards**.
- **Opt-Out:** You may opt out of chatbot use at any time by contacting your **Customer Success representative in writing.**