cymulate

Cymulate Vero AI

Security, Privacy & AI Governance

Architecture Overview | Data Handling | Safety Controls | Compliance

This document provides an overview of the security architecture, data privacy controls, AI safety guardrails, and compliance posture of Cymulate Vero AI. It is intended for security, risk and compliance stakeholders evaluating the platform for enterprise deployment.

Cymulate Vero AI is a domain-specific, agentic AI system built for cybersecurity use cases. It is not a general-purpose chatbot. Customer data is never used to train AI models, never shared across tenants and never exposed to public AI services.

Architecture Overview

Vero AI is built on an agentic AI architecture. Each capability, including cybersecurity Q&A, dashboard creation, template generation and assessment creation, is handled by a dedicated AI agent. A supervisor agent orchestrates request routing, ensuring each query is processed by the appropriate domain-specific agent within strict operational boundaries.

AI Approach	Agentic AI with supervisor orchestration, using commercially available OpenAI models (not self-trained)
LLM Inference	Commercially available foundation models (OpenAI) via Azure OpenAI and AWS Bedrock, accessed through private endpoints
Orchestration	LangGraph for workflow management
Observability	Langfuse for tracing and audit
Guardrails Engine	NVIDIA NeMo Guardrails for content filtering
Hosting	Private AWS tenant (no public endpoints, no shared tenancy)

Cymulate does not train or maintain proprietary AI models. The platform uses commercially available OpenAI foundation models, accessed exclusively through private endpoints on Azure OpenAI and AWS Bedrock within the Cymulate secured infrastructure. These are not public-facing AI services. No customer data is sent to or retained by the model providers for training purposes.

Data Privacy & Isolation

Cymulate enforces strict data isolation at every layer. Customer data remains the property of the customer and is governed by Cymulate security and privacy policies throughout its lifecycle.

Data Processing Principles

- **Zero Cross-Tenant Exposure:** Customer data is never shared with third parties and never shared across customer tenants.
- **No Model Training:** Customer data, prompts, and outputs are never used to train or fine-tune any AI model.
- **Private Infrastructure:** All data is processed and stored within the Cymulate private AWS tenant, protected by encryption at rest and in transit (TLS 1.2+), RBAC, SSO and MFA.
- **Retention Policy:** AI interaction data is securely stored for up to 6 months for traceability and auditability, in MongoDB with encryption at rest.

Vero AI User Interactions

User questions and AI-generated responses submitted through Vero AI are securely stored for up to 6 months to support traceability, auditability, and continuous system improvement. Data is stored in MongoDB with encryption at rest and in transit, strict access controls, and audit logging. All records are tied to authenticated user sessions and governed by Cymulate security and privacy policies. No user interaction data is used for model training or fine-tuning.

SIEM Validation Queries: Cryptographic Hashing

For SIEM validation workflows, Cymulate applies an additional layer of data protection. SIEM rules and validation queries are never stored in plain text. Instead, they are processed through a one-way cryptographic hash mechanism before any persistence occurs. This means:

- The original query content cannot be reconstructed from the stored hash. The hashing is irreversible by design.
- The hash is used solely to detect whether changes have occurred in underlying data sources, not to retain, reproduce, or search the query itself.
- All retained metadata (such as customer identifiers) serves audit and traceability purposes only and does not contain query or rule content.
- System logs retain only limited operational metadata for audit purposes. No query content appears in any log.

In practice, this means that even in the event of unauthorized access to the data store, the actual content of SIEM validation queries and rules cannot be recovered. The system retains proof that a validation occurred and whether the underlying data has changed, without retaining the query itself. This is particularly relevant for organizations integrating Cymulate AI-driven SIEM validation, where the confidentiality of detection rules is paramount.

Guardrails & Safety Controls

Cymulate applies defense-in-depth safety controls across four layers: AI/LLM, application, infrastructure, and observability. These controls work in concert to prevent data leakage, unauthorized access, off-topic usage, and unsafe or biased outputs.

Layer	Control	Purpose
AI / LLM	NeMo Guardrails	Content filtering, output sanitization, topic restriction to cybersecurity domain
AI / LLM	Prompt Constraints	System boundaries and tool-restricted agents prevent off-topic or unsafe operations
Application	Input/Output Validation	Detection and blocking of sensitive, confidential, or restricted data patterns
Application	Supervisor Agent Routing	Ensures requests are handled only by authorized, domain-specific agents
Infrastructure	RBAC, SSO, MFA	Users access only data authorized for their tenant and role
Infrastructure	TLS 1.2+ Encryption	All data encrypted in transit; encryption at rest for stored data
Observability	Langfuse Tracing	Full observability and tracing of all AI interactions for audit
Observability	Structured Audit Logs	Automated alerts for flagged outputs; complete interaction traceability

Unsafe Output Handling

When the system detects a potentially unsafe or out-of-scope output, it applies a layered response: real-time filtering via NeMo Guardrails, prompt constraint enforcement by tool-restricted agents, validation layers for PII or policy violations, and safe fallback responses. All flagged interactions are logged with full traceability via Langfuse.

Data Loss Prevention (DLP)

DLP is enforced through NeMo Guardrails (content filtering and output sanitization at the LLM level), input/output validation (detection of confidential or restricted data patterns), access controls (RBAC, SSO, MFA), and full audit logging of all interactions. Currently, AI-generated outputs cannot be exported directly to files. Users may copy selected content manually for local use.

Compliance & Certifications

Cymulate and its hosting providers maintain compliance with industry-standard security frameworks. The platform undergoes regular security audits, penetration testing, and follows a Secure Software Development Lifecycle (SSDLC) that includes code reviews, threat modeling, SAST and DAST.

Standard / Framework	Status
SOC 2 Type II	Compliant
ISO 27001	Certified
GDPR	Compliant
NIST AI RMF	Aligned
EU AI Guidance	Aligned
SSDLC (SAST, DAST, Pen Testing)	Implemented

Further Information

For a comprehensive FAQ covering architecture, data handling, DLP controls, AI training practices, accuracy controls and compliance details, Cymulate customers can refer to the Cymulate Vero AI Knowledge Base:

https://document360.cymulate.com/docs/about-cymulate-ask-ai#service-overview

For additional questions or to schedule a security architecture review, contact your Cymulate account team.