NodeLoom

Security Overview

A comprehensive overview of NodeLoom's security architecture, encryption, access controls, AI safety guardrails, compliance automation, and operational security practices.

Version 1.0 — March 2026

NodeLoom, Inc. — San Francisco, CA

nodeloom.io | contact@nodeloom.io

1. Executive Summary

NodeLoom is an AI agent operations platform that enables organizations to deploy, monitor, and govern AI agents in production. Security is built into every layer of the platform — from AES-256-GCM encryption and container-based execution sandboxing to tamper-evident audit logs and automated incident response playbooks.

This document provides a detailed overview of NodeLoom's security architecture for enterprise security teams, compliance officers, and engineering leadership evaluating the platform for regulated environments.

Security at a Glance

ENCRYPTION

AES-256-GCM

All credentials and secrets encrypted at rest with AES-256-GCM authenticated encryption. TLS enforced for all data in transit. Encryption keys validated at startup with minimum length enforcement.

ACCESS CONTROL

5 Roles + SSO + SCIM 2.0

Fine-grained RBAC with Admin, Builder, Operator, Viewer, and Compliance Officer roles. Enterprise SSO via SAML 2.0 and OIDC. Automated user provisioning via SCIM 2.0.

EXECUTION

Container Sandboxing

All code execution runs in ephemeral containers with network disabled, read-only filesystem, all capabilities dropped, strict memory limits, and enforced timeouts.

AI SAFETY

6 Guardrail Types

Prompt injection detection, PII redaction, content filtering, schema validation, custom rules engine, and semantic similarity drift detection with configurable enforcement actions.

AUDIT

Tamper-Evident Logging

Every action logged with user identity, IP address, timestamps, and contextual details. Cryptographic hash chain for tamper detection. SIEM export to Splunk, Datadog, Elasticsearch, or webhook.

COMPLIANCE

6 Frameworks

Report generation tooling aligned with SOC 2, GDPR, HIPAA, SOX, ISO 42001, and NIST AI RMF frameworks.* Configurable data retention policies. Automated adversarial testing.

2. Defense-in-Depth Architecture

Every request passes through six independent security layers before reaching your data.

1

Network Layer

TLS enforced on all connections. HSTS enabled with long-duration max-age and includeSubDomains. Redis-backed sliding window rate limiting across all API endpoints, authentication, webhooks, and widget interactions. Explicit CORS origin allowlisting with no wildcards when credentials are used. Comprehensive security response headers on every request.

2

Authentication Layer

Signed token-based authentication with short-lived access tokens and longer-lived refresh tokens, stored in httpOnly Secure cookies to prevent client-side access. CSRF protection on all state-changing requests. Enterprise SSO via SAML 2.0 and OIDC. Automated user provisioning and deprovisioning via SCIM 2.0 with IP-based access control. Email verification with rate-limited resend.

3

Data Layer

All credentials and secrets encrypted at rest using AES-256-GCM authenticated encryption with unique initialization vectors per operation. Encryption keys are environment-configured, validated at startup, and subject to minimum length enforcement. Secrets are never written to logs and are masked in the UI after initial entry.

4

AI Safety Layer

Six configurable guardrail types protect every AI agent interaction: prompt injection detection with severity scoring, PII redaction, content filtering with external moderation API and local fallback, JSON schema validation, a custom rules engine (regex, keyword, sandboxed JavaScript, and LLM-based evaluation), and semantic similarity drift detection. Token budget enforcement per workflow.

5

Execution Layer

Container-based sandboxing for all user code execution. Strict resource limits on CPU and memory. Network access disabled. Filesystem mounted read-only. All Linux capabilities dropped. Non-root execution enforced. Containers are fully ephemeral — created, executed, and destroyed per invocation with enforced timeouts and output size limits.

6

Audit Layer

Comprehensive audit logging of every action with user identity, IP address, session context, request identifiers, operation duration, and result status. Tamper-evident hash chain links each entry to the previous, enabling integrity verification. Audit data exportable to enterprise SIEM systems (Splunk, Datadog, Elasticsearch, or custom webhook).

3. Access Controls & Identity Management

Five roles with granular permissions, enterprise SSO, and automated provisioning.

Role-Based Access Control (RBAC)

Permission	Admin	Builder	Operator	Viewer	Compliance
Manage team & users	Y	—	—	—	—
Manage billing	Y	—	—	—	—
Create / edit workflows	Y	Y	—	—	—
Delete workflows	Y	Y	—	—	—
Execute workflows	Y	Y	Y	—	—
View workflows & executions	Y	Y	Y	Y	Y
View audit logs	Y	—	—	—	Y
Compliance dashboard	Y	—	—	—	Y
Start adversarial scans	Y	Y	—	—	—
View incident playbooks	Y	Y	—	—	Y

Single Sign-On (SSO)

Enterprise SSO via **SAML 2.0** and **OpenID Connect (OIDC)**. Domain verification required for SSO configuration. Session management with cooldown tracking. Users authenticate through their corporate identity provider, enforcing existing identity policies.

SCIM 2.0 Provisioning

Automated user lifecycle management via the SCIM 2.0 protocol. When employees join or leave, access is provisioned or revoked automatically through your identity provider. Tokens are hashed before storage. IP-based access control via CIDR notation. Rate limited per team.

OAuth 2.0 Credentials (12 Providers)

- Google, GitHub, Slack, Microsoft
- Salesforce, HubSpot, Shopify, Zoom
- Asana, Linear, Jira, Notion

All tokens encrypted at rest. Automatic token refresh. Status tracking for expired, revoked, and reauthorization states. Credentials scoped to minimum required permissions.

Password & Session Security

- Passwords hashed with BCrypt (adaptive cost)
- httpOnly, Secure cookies with SameSite attribute
- CSRF protection on state-changing requests
- Automatic session expiry and re-authentication
- Email verification with rate-limited resend

4. Data Protection & Encryption

AES-256-GCM encryption, credential vault, environment isolation, and data sovereignty options.

Encryption at Rest

All stored credentials and secrets are encrypted using AES-256-GCM authenticated encryption. Each encryption operation uses a unique, cryptographically random initialization vector. The GCM authentication tag provides both confidentiality and integrity verification.
Encryption keys are configured via environment variables, validated at application startup, subject to minimum length requirements, and checked against known default values to prevent misconfiguration.

Encryption in Transit

All network communication encrypted via TLS. HSTS enforced with long-duration expiry and subdomain inclusion. Internal service communication is also encrypted. WebSocket connections secured with encrypted transport and periodic heartbeats.

Credential Vault

- Supports: OAuth tokens, API keys, database credentials, service accounts, custom secrets
- All credentials encrypted before storage
- Secrets never written to application logs
- Secrets masked in UI after initial entry
- Team-scoped and environment-scoped access
- Creator tracking and deletion auditing

Data Residency & Self-Hosted

NodeLoom offers a full self-hosted deployment option for organizations requiring complete data sovereignty. Deploy on your own infrastructure using Docker Compose, Kubernetes manifests, or Helm charts. All data remains within your network perimeter with no external dependencies required.

Data Retention Policies

Configurable retention periods per data type with automatic purge:

- **Audit logs:** 365 days (default)
- **Execution logs:** 90 days
- **Policy violations:** 365 days
- **Anomalies & drift alerts:** 90 days
- **Token usage history:** 365 days

All retention periods configurable. Manual purge available for immediate deletion.

AI Data Policy

NodeLoom never uses customer workflow data, credentials, or execution outputs to train AI models. Data is processed only to deliver the service and is never shared with third-party providers for training purposes.

Data Portability

Full data export capability at any time. Export workflows, credentials (encrypted), and execution history. No vendor lock-in.

5. Application Security

Container sandboxing, multi-layer XSS prevention, prompt injection detection, and workflow policies.

Container Sandbox Isolation

All user-defined code runs in fully isolated, ephemeral containers with hardened security controls:

- **Network:** Disabled by default — no inbound or outbound connectivity
- **Filesystem:** Read-only mount prevents persistent changes
- **Capabilities:** All Linux capabilities dropped
- **Resources:** Strict CPU and memory limits enforced
- **User:** Non-root execution enforced
- **Timeout:** Configurable execution deadline with automatic termination
- **Output:** Maximum size enforced with truncation

Containers are created, executed, and destroyed per invocation. No state persists between executions.

XSS & Injection Prevention

Multi-layer XSS prevention with independent server-side and client-side sanitization. Each rendering context (API, dashboard, embeddable widget) applies its own sanitization pass with context-appropriate rules. Dangerous CSS patterns, event handlers, and script vectors are stripped at multiple stages before content reaches the DOM.

Prompt Injection Detection

Pattern-based detection engine that identifies instruction override attempts, jailbreak patterns, system prompt extraction, role reassignment, context manipulation via fake markup, and delimiter-based attacks. Severity-scored (Critical, High, Medium) with configurable thresholds and enforcement actions.

Additional Protections

Parameterized database queries (SQL injection prevention), CSRF token protection on all state-changing requests, Content Security Policy with restrictive directives, comprehensive security response headers (HSTS, X-Frame-Options, X-Content-Type-Options, Referrer-Policy, Permissions-Policy), and input validation with type checking and length limits.

6. AI Safety Guardrails & Adversarial Testing

Six guardrail types, automated red team scanning, and configurable enforcement actions.

Guardrail Types

Prompt Injection Detection

Detects instruction overrides, jailbreaks, system prompt extraction, role reassignment, and context manipulation. Severity-scored with configurable thresholds.

PII Redaction

Automatically detects and redacts personally identifiable information from AI agent inputs and outputs before storage or downstream processing.

Content Filtering

Moderation API integration with local fallback. Covers hate, harassment, self-harm, sexual, and violence categories including graphic variants.

Schema Validation

Validates AI output against user-defined JSON schemas. Non-conforming outputs are blocked before reaching downstream systems.

Custom Rules Engine

Four rule types: regex patterns, keyword lists, sandboxed JavaScript execution, and LLM-based evaluation with configurable confidence thresholds.

Semantic Similarity Detection

Embedding-based drift detection measuring semantic distance between expected and actual outputs. Detects behavioral changes over time.

Enforcement actions per guardrail: Block (prevent output), Redact (strip sensitive content), Flag (log and alert), or Warn (notify user). Violation severity: Low, Medium, High, Critical.

Workflow Execution Policies

- **Tool Whitelist:** Restrict which workflow nodes can execute
- **SQL Scope:** Enforce read-only, read-write, or full SQL access
- **HTTP Domain Control:** Allowlist or blocklist external domains
- **Credential Scope:** Restrict which credentials a workflow can access
- **Execution Rate Limits:** Max executions per configurable time window

Policy violations are tracked with severity, blocking status, and full detail logging.

Red Team Scanning & Behavioral Monitoring

Automated adversarial testing across six attack categories: Prompt Injection, Jailbreak Attempts, Data Exfiltration, Tool Abuse, PII Leakage, and Harmful Output Generation. Configurable attacker model with severity-scored findings and remediation recommendations. Critical findings automatically dispatch incident response playbooks. Continuous behavioral monitoring with drift detection, anomaly scoring, sentiment analysis, per-workflow token budget enforcement, and configurable alert thresholds.

7. Compliance, Incident Response & Environment Isolation

Automated compliance reports, incident playbooks, SIEM integration, and multi-environment security.

Compliance Report Generation

SOC 2 Service organization controls	**GDPR** Data privacy & portability	**HIPAA** Protected health information
SOX Financial controls & audit	**ISO 42001** AI management systems	**NIST AI RMF** AI risk management

Reports generated on-demand with configurable date ranges. Stored as structured data with optional file export. Generated-by tracking for accountability.

Note: NodeLoom provides tooling to generate reports aligned with these frameworks. This does not constitute certification or attestation under any framework. Organizations should work with qualified auditors to achieve formal compliance.

Incident Response Playbooks

Automated playbooks triggered by anomalies, guardrail violations, execution failures, or manual trigger. Configurable severity threshold, workflow filter, auto-response actions, and cooldown periods. Full lifecycle tracking (Triggered → Running → Completed/Failed) with event data captured for post-incident analysis.

SIEM Integration

Export audit logs and security events to **Splunk**, **Datadog**, **Elasticsearch**, or **Custom Webhook**. Configurable per team with sequence tracking and delivery error monitoring.

Multi-Environment Security

Three isolated environments with independent security controls:

- **Development** — Default sandbox for building and testing
- **Staging** — Pre-production validation environment
- **Production** — Live environment with full monitoring

Each environment maintains separate workflow definitions, environment-scoped credentials, independent anomaly baselines, separate audit trails, and promotion tracking.

Widget Embed Security

Token-based authentication per widget, per-visitor session isolation, three-layer XSS sanitization, configurable rate limiting, and separate Content Security Policy for embed contexts.

Questions, Security Inquiries, or Vulnerability Reports?

contact@nodeloom.io | (415) 340-1981 | nodeloom.io/security | nodeloom.io/contact