Nexoraa Security & Governance Architecture

Technical White Paper for Secure AI-Governed SOP Execution

Prepared for Arizona State University | July 2026 | Version 1.0

Core position: Nexoraa separates AI reasoning from enterprise authority. Models can classify, draft, summarize, and recommend. The Haluvance enforcement layer determines what is allowed, who must approve it, what can be sent to connectors, and what is preserved for audit.

Document control

Document type	Security architecture white paper
Audience	CISO, security architecture, enterprise architecture, privacy, compliance, and platform engineering stakeholders
Status	Discussion draft for technical and security review
Scope	Nexoraa Studio, Haluvance enforcement layer, workflow runtime, agent/model execution, connector broker, audit/evidence controls
Classification	Proprietary and Confidential - restricted distribution

Contents

1. Executive summary

Nexoraa converts standard operating procedures into governed AI workflows. The security question is not whether the model can reason over a process. The security question is whether the platform can control what the model sees, what it is allowed to decide, which actions require approval, and how the result is recorded for review.

The Nexoraa platform uses Haluvance as its security and governance enforcement layer. Haluvance sits between workflow intent and enterprise authority. It applies tenant isolation, role-based access, policy binding, model boundary controls, connector authorization, human approval rules, and audit evidence capture before a workflow can execute an output or external action.

Operating principle: the model is not the control plane. The governance layer is the control plane.

For a CISO, this distinction matters. AI can assist with classification, drafting, routing, and decision support. It must not independently create uncontrolled enterprise changes, disclose protected records, bypass approval, or use connectors beyond their approved scope. Nexoraa is designed around that separation.

This paper provides a technical view of the platform security layer, how workflows execute, how AI risk is contained, how connector access is controlled, and how evidence is preserved. It also maps the design to commonly used security and AI governance references, including NIST AI RMF, OWASP Top 10 for LLM Applications, CSA AI Controls Matrix, AICPA SOC reporting concepts, CISA agentic AI guidance, and FERPA-sensitive handling requirements.

2. Platform overview: from SOP to governed execution

Nexoraa starts with existing SOPs, policy documents, decision tables, ticketing practices, and operating rules. The platform converts these artifacts into dynamic workflows that can intake work, classify cases, route to the right owner, generate safe responses or ticket payloads, and preserve an evidence trail.

A workflow is not just a prompt. It is a controlled execution object with explicit inputs, versioned logic, role assignments, approval rules, model configuration, connector permissions, output schemas, and audit metadata. This makes the platform suitable for operational processes that need both automation and control.

- SOP interpretation creates a structured process map from the source procedure.
- Workflow generation converts the process map into executable steps, decision branches, output schemas, and exception paths.
- Haluvance binds the workflow to identity, policy, data, model, connector, and approval controls.
- Runtime execution applies deterministic checks before and after model reasoning.
- Audit capture records the decision path, model use, approval state, and connector activity.

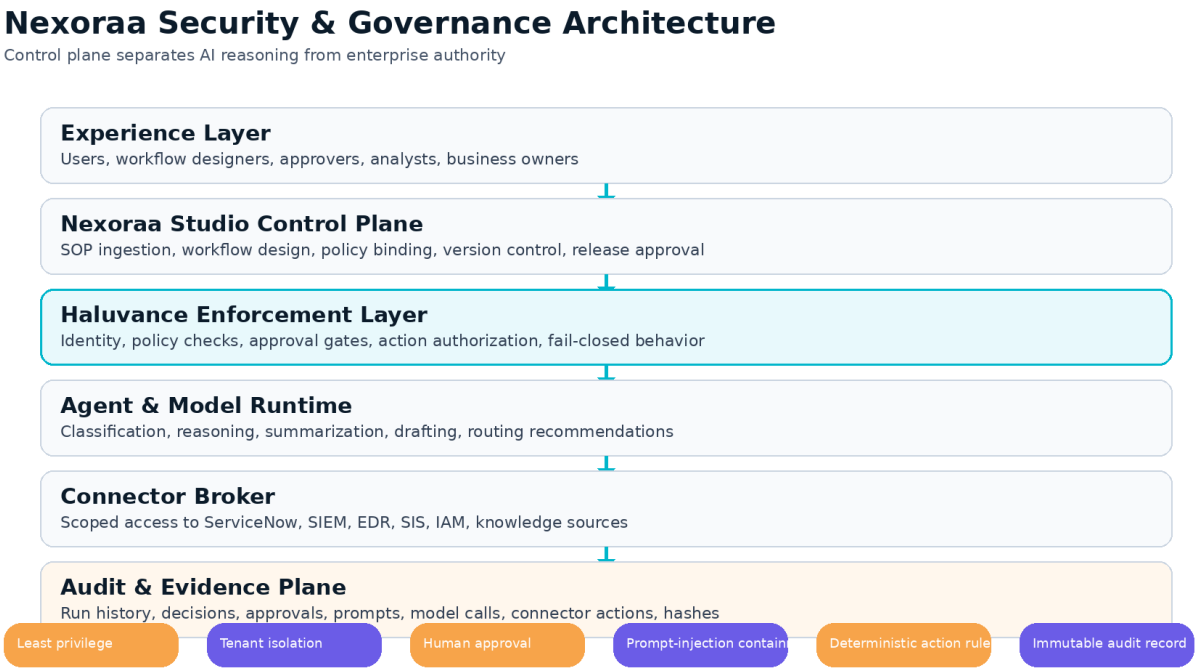

Figure 1. Nexoraa architecture view - security and governance controls wrap the workflow runtime and model layer.

3. Security architecture principles

Principle	Operational meaning
Least privilege	Users, agents, workflows, and connectors receive only the access required for the approved process. Read, draft, create, update, approve, and execute permissions are separated.
Tenant isolation	Customer workflows, runtime state, configurations, secrets, audit logs, and outputs are tenant-scoped. Cross-tenant visibility is not a normal execution path.
Policy-bound execution	Workflows run against explicit policy and SOP rules. A model recommendation does not override policy.
Human authority over high-risk actions	AI may recommend, classify, summarize, or draft. Sensitive actions require human approval based on configured thresholds.
Fail-closed behavior	Missing inputs, missing policy mappings, low confidence, connector errors, or missing approvals route to review or stop the run.
Evidence by default	The platform records the workflow version, input, model use, decision path, approvals, connector calls, outputs, and audit hash.

4. Security and governance layer

The Haluvance enforcement layer is the primary control point between a workflow and the systems it may influence. It is responsible for policy evaluation, authorization, runtime safety, and audit capture. It is not a passive logging layer. It actively determines whether the workflow can continue, whether a branch is allowed, whether a connector operation is permitted, and whether human approval is required.

4.1 Identity and access control

Nexoraa supports role-based control over design, deployment, execution, review, approval, and administration. The same person or service account should not automatically receive all rights. For example, a workflow designer can be allowed to edit a workflow but not approve a production execution. A security analyst can run a cybersecurity assessment but not approve a containment action unless assigned the approver role.

- SSO and enterprise identity integration for user authentication.
- Tenant-level roles for administrator, workflow designer, operator, reviewer, approver, and auditor.
- Environment separation across development, test, and production workflows.
- Workflow-specific execution permissions and output visibility rules.
- Separation of duties between authoring, approval, and production execution.

4.2 Tenant and data isolation

Tenant isolation is a structural requirement, not a display preference. The tenant boundary contains workflow definitions, runtime state, uploaded SOPs, generated outputs, logs, policies, role bindings, and connector configuration. Tenant-scoped configuration prevents a workflow in one tenant from using data, connector permissions, or policies assigned to another tenant.

4.3 Data security

Nexoraa minimizes the data provided to workflow steps and model calls. The runtime assembles only the context needed for the workflow step. Sensitive fields can be masked, excluded, or routed to a human review path depending on policy. Data in transit and data at rest should be encrypted according to the selected deployment pattern and customer requirements.

- Data minimization before model invocation.
- Sensitive-field labeling for student, financial, HR, security, and regulated data.
- Optional redaction or masking before prompts are assembled.
- Retention controls for uploaded files, workflow state, and audit logs.
- Output controls that prevent unauthorized disclosure of sensitive records.

Customer data does not need broad exposure to the model. Each workflow can be configured to provide only the fields and context required for the approved process step.

5. Technical platform architecture

Nexoraa separates platform responsibilities across four planes: control, runtime, data, and audit. This separation helps security reviewers understand where policies are configured, where workflows execute, where data is handled, and where evidence is preserved.

Plane	Primary responsibility	Typical controls
Control plane	Workflow design, policy configuration, role assignment, connector setup, versioning, and deployment promotion.	RBAC, change control, versioning, release approval, environment separation.
Runtime plane	Execution of workflow nodes, model calls, branch logic, approval gates, connector calls, and failure paths.	Deterministic rules, policy checks, timeout controls, retries, fail-closed routing.
Data plane	Input records, SOP files, workflow variables, model context, enrichment data, generated outputs.	Tenant isolation, encryption, redaction, retention, output visibility, data minimization.
Audit plane	Evidence collection for each workflow run and external action.	Decision ID, actor, timestamp, workflow version, model version, prompts, outputs, approvals, connector logs, audit hash.

Figure 2. Runtime decision flow - AI reasoning is wrapped by validation, enforcement, approval, and audit controls.

6. Agent, model, and prompt security

Nexoraa uses models for bounded reasoning tasks: classification, summarization, extraction, drafting, recommendation, and explanation. The model is not granted blanket authority to perform enterprise actions. Its output is treated as an input to the governance layer, not as an automatic command.

6.1 Model boundary controls

- Model selection can vary by workflow risk and task complexity.
- Prompt templates are versioned and bound to the workflow version.
- Model input context is assembled from approved variables and policies only.
- Model output is validated against expected schemas before downstream use.
- High-risk or low-confidence outputs route to human review rather than execution.

6.2 Prompt injection and untrusted content

Inputs from users, websites, emails, public threat feeds, documents, and ticket descriptions are treated as untrusted content. They cannot override workflow policy. The platform separates system instructions, SOP rules, user content, model output, and connector actions. If external content attempts to change policy, ignore previous instructions, or trigger unauthorized action, the workflow follows the configured review or stop path.

This design follows the practical lesson from modern LLM security guidance: prompt injection should be addressed at the system level, not only by asking the model to behave safely. The platform constrains what the model can observe, what it can decide, and what consequences can follow from its output.

6.3 What the model can and cannot do

Allowed model role	Not allowed by default
Classify a case, signal, document, or SOP step.	Bypass policy, role, or approval rules.
Generate a draft response or ticket payload.	Disclose sensitive records to unauthorized users.
Recommend a next action and explain the reasoning.	Execute a connector action beyond the workflow permission scope.
Identify missing information or research gaps.	Convert untrusted external text into an enterprise command.
Summarize audit-relevant decision rationale.	Change workflow policy, connector permissions, or approval thresholds.

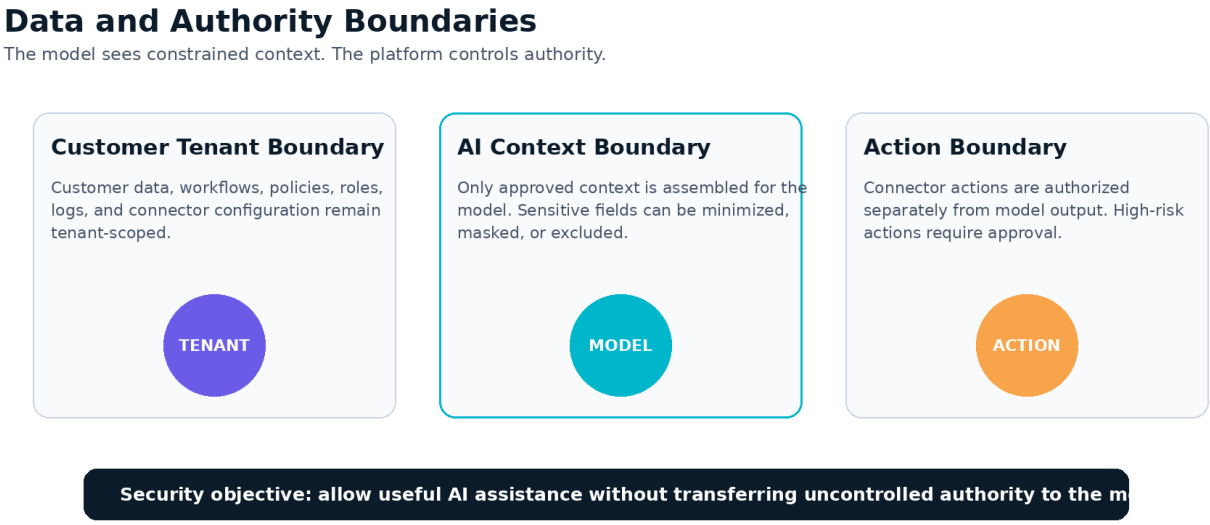

Figure 3. Data and authority boundaries - model context is limited and connector authority is separately controlled.

7. Connector and enterprise system security

Enterprise workflows often need to interact with ticketing systems, SIEM, EDR, identity platforms, student information systems, document repositories, communication tools, and internal knowledge sources. Nexoraa treats connectors as controlled enterprise interfaces, not general agent capabilities.

- Connector credentials are scoped to the tenant and workflow.
- Read and write actions are separated.
- Connector operations are explicitly allow-listed.
- High-risk connector operations can require approval before execution.
- Every connector call is logged with actor, workflow run, action, target, timestamp, and outcome.
- Connector errors preserve workflow state and route to retry or human review.

For example, a cybersecurity workflow may be allowed to create a ServiceNow remediation ticket but not isolate a host. A student support workflow may be allowed to draft a safe response but not disclose student-specific aid or billing details to a parent without confirmed authorization. This is an action boundary, not just a prompt instruction.

8. Human approval, exception handling, and fail-closed behavior

Human approval is configured as part of the workflow policy. It can be triggered by data sensitivity, priority, action type, confidence, role, system, or business impact. Approval does not have to be a generic yes/no step; it can require a specific role, reason, and evidence package.

8.1 Approval triggers

Trigger	Example	Result
Sensitive data	FERPA-related financial aid, student account, or transcript information.	Safe response only; route to authorized review.
Critical security exposure	Internet-facing system with known exploited vulnerability and restricted data impact.	Escalate to CISO or assigned security approver.
Write-back action	Create, update, close, or escalate a system-of-record ticket.	Require connector permission and optional approval.
Low confidence or policy conflict	Input lacks CVE, vendor advisory,	Route to insufficient-evidence or intake

	authorization status, or asset match.	correction branch.
Connector failure	ServiceNow, SIEM, SIS, or IAM connector is unavailable.	Preserve state; retry or route to manual queue.

8.2 Failure modes

Failure scenario	Platform behavior
Missing required input	Stop and route to intake correction. Do not infer missing facts for high-risk decisions.
Low-confidence classification	Route to human review with the reason and missing evidence.
Policy conflict	Fail closed and require an authorized reviewer.
Approval missing	Do not execute the high-risk action.
Model unavailable	Retry according to policy or route to manual review.
Connector unavailable	Preserve state, log failure, retry, or assign manual queue.
Suspicious external instruction	Do not follow external instructions that conflict with system policy; route to review if needed.

9. Audit, evidence, and traceability

Nexoraa is designed to make AI-driven process execution reviewable after the fact. The platform records not only what happened, but why the workflow took a specific path. This is important for security operations, privacy review, process quality, dispute handling, and compliance evidence.

9.1 Run-level evidence

- Tenant ID, workflow ID, workflow version, environment, and run ID.
- User or service actor, role, timestamp, and invocation channel.
- Input payload, document references, normalized variables, and data classification.
- Prompt template version, model identifier, model parameters, and output schema.
- Decision branch, policy checks, approval status, and exception reasons.
- Connector actions, target system, target object, request metadata, response status, and error handling.
- Final output, ticket payload, generated response, audit hash, and retention class.

9.2 Example audit record

```
{
  "tenant": "asu-demo",
  "workflow": "known-threat-signal-to-action",
  "workflow_version": "1.0",
  "run_id": "ASU-SIG-2026-0001",
  "decision_id": "a58b9306-6ebe-4b42-b53d-e11edcc0ea89",
  "branch": "ACTION_REQUIRED",
  "highest_priority": "P1",
  "approver_required": "CISO",
  "model_use": ["signal-classification", "ticket-drafting"],
  "connector_actions": ["create-remediation-ticket"],
  "audit_hash": "66e663a02458..."
}
```

10. Application to ASU demo workflows

The ASU demos shown during the meeting illustrate two different operating patterns: security signal triage and high-volume student support triage. Both use the same security architecture: SOP-bound workflow logic, constrained model reasoning, deterministic rules, privacy handling, action gating, and audit capture.

10.1 Cybersecurity known threat signal-to-action

The cybersecurity workflow processes a batch of public threat signals and determines which ones are valid and actionable for ASU, which are watchlist items, which are not applicable, and which lack enough evidence. It creates ticket-ready remediation payloads only for valid actionable threats and escalates when the configured priority threshold is met.

Risk or operational problem	Nexoraa control
Public threat noise	Credibility, applicability, asset match, exposure, and patch-status rules.
False positives	Not applicable and insufficient-evidence branches.
Over-automation	Tickets only for valid actionable threats; remediation actions require permission and approval.
Critical exposure	P1 threshold triggers CISO escalation.
Audit gap	Decision ID, audit hash, policy references, model and workflow metadata.

10.2 Student support and financial aid case triage

The student support workflow processes a batch of student-service cases and separates urgent items, FERPA-sensitive requests, auto-response candidates, duplicate cases, and human-review cases. It produces queue routing, safe draft responses, and case-system payloads.

Risk or operational problem	Nexoraa control
FERPA disclosure risk	Parent and third-party requests route to safe handling unless authorization exists.
Missed deadlines	Priority rules detect same-day exams, registration deadlines, document deadlines, and financial-aid timing.
Wrong queue routing	SOP-bound category and queue mapping.
Duplicate work	Related-case detection and link-to-existing disposition.
Inconsistent responses	Approved response templates and safe-response rules.
Review evidence	Case payload, decision metadata, queue, priority, and audit record.

11. Deployment and integration model

Nexoraa can be introduced in phases. For security review and pilot adoption, the lowest-risk approach is advisory mode: the workflow reads approved inputs, classifies, recommends, drafts, and creates evidence packages. Write-back and direct system actions are added only after connector permissions, approval rules, and production controls are validated.

11.1 Progressive adoption pattern

1. Advisory mode: run workflows on supplied SOPs, sample data, or read-only datasets; generate reports and ticket drafts.
2. Controlled ticket mode: allow creation of tickets or tasks in approved queues; no closure or destructive action.
3. Approved write-back mode: allow selected updates after human approval and connector allow-list validation.
4. Operational mode: expand to additional SOPs, monitoring, dashboards, and continuous improvement loops with production change control.

11.2 Integration targets

- Ticketing: ServiceNow, Jira Service Management, Salesforce Service Cloud, or equivalent case systems.
- Security operations: SIEM, EDR, vulnerability management, threat intelligence, asset inventory, and identity platforms.
- Student operations: student information systems, financial aid systems, document processing systems, knowledge bases, and contact-center platforms.
- Enterprise identity: SSO, role groups, approver groups, and audit-reader groups.

- Data stores: document repositories, SOP libraries, policy libraries, and approved knowledge sources.

12. Industry control alignment

The following table summarizes how the architecture aligns with common security, privacy, and AI governance references. This is an alignment view for security review. It is not a compliance certification or attestation.

Reference	Relevant theme	Nexoraa / Haluvance design response
NIST AI Risk Management Framework (AI RMF 1.0)	Govern, map, measure, and manage AI risk across the lifecycle.	Policy-bound workflows, model-use inventory, risk-based approval gates, audit evidence, and controlled deployment.
OWASP Top 10 for LLM Applications	Prompt injection, sensitive information disclosure, insecure plugin/tool design, excessive agency.	Untrusted-content handling, model context constraints, schema validation, connector allow-lists, least privilege, and human approval for high-risk actions.
Cloud Security Alliance AI Controls Matrix	Vendor-neutral controls for secure and responsible cloud-based AI systems.	Tenant isolation, lifecycle governance, logging, access control, model governance, monitoring, and operational control mapping.
AICPA SOC reporting concepts / Trust Services Criteria	Security, availability, processing integrity, confidentiality, and privacy.	RBAC, change control, encryption, audit logs, workflow versioning, retention controls, and evidence trail.
CISA agentic AI security guidance	Risk-aware adoption of agentic AI services and protection against unexpected or unsafe behavior.	Advisory-first deployment, action boundaries, approval gates, connector scoping, and fail-closed paths.
FERPA student privacy requirements	Protection of student education records and controlled disclosure.	FERPA-sensitive routing, parent/third-party safe handling, authorization checks, and disclosure prevention.

13. What Nexoraa does not do by default

A useful security white paper should define boundaries as clearly as capabilities. The following default constraints are part of the risk-control posture.

- Nexoraa does not allow agents to execute arbitrary actions outside configured workflows.
- Nexoraa does not let model output bypass enterprise identity, approval, or connector permissions.
- Nexoraa does not disclose student-specific, financial, security, or account data to unauthorized users.
- Nexoraa does not treat public web or user-provided content as trusted instructions.
- Nexoraa does not require broad access to every enterprise data store for a workflow to be useful.
- Nexoraa does not automatically remediate high-risk incidents unless the customer explicitly configures and approves those actions.
- Nexoraa does not replace system-of-record permissions; it operates within the permissions granted by the customer.

14. Security review package for ASU

For a formal security review, ASU can request the following artifacts. These materials help validate the architecture against ASU security, privacy, integration, and compliance expectations.

Artifact	Purpose
Architecture diagram	Shows platform components, tenant boundary, model boundary, connector boundary, and audit plane.
Data-flow diagram	Shows what data enters the workflow, what reaches the model, what reaches connectors, and what is logged.
Role matrix	Documents admins, designers, operators, reviewers, approvers, and auditors.
Connector permission matrix	Shows read/write operations, scopes, secrets handling, approval requirements, and audit fields.

Workflow execution trace	Shows a sample run from input through branch decision, approval, output, and audit record.
Prompt and model inventory	Lists prompt templates, model choices, parameters, and version bindings.
Audit log sample	Shows decision ID, input hash, workflow version, branch, model use, connector action, and output hash.
Deployment option summary	Defines SaaS, private tenant, customer cloud, hybrid connector, and advisory-mode options as applicable.
Retention and deletion policy	Defines retention classes for SOPs, inputs, outputs, logs, and audit evidence.
Incident response contact path	Defines support, severity levels, escalation contacts, and customer notification process.

15. References

The following references were used as public industry context for this white paper. Nexoraa control implementation should be evaluated against the final customer deployment model and contract-specific security requirements.

Organization	Reference	URL
National Institute of Standards and Technology	Artificial Intelligence Risk Management Framework (AI RMF 1.0)	https://www.nist.gov/itl/ai-risk-management-framework
OWASP Foundation	OWASP Top 10 for Large Language Model Applications	https://owasp.org/www-project-top-10-for-large-language-model-applications/
Cloud Security Alliance	AI Controls Matrix	https://cloudsecurityalliance.org/artifacts/ai-controls-matrix
AICPA & CIMA	System and Organization Controls: SOC Suite of Services	https://www.aicpa-cima.com/resources/landing/system-and-organization-controls-soc-suite-of-services
Cybersecurity and Infrastructure Security Agency	Careful Adoption of Agentic AI Services	https://www.cisa.gov/resources-tools/resources/careful-adoption-agentic-ai-services
U.S. Department of Education	FERPA - Protecting Student Privacy	https://studentprivacy.ed.gov/ferpa

Appendix A. Glossary

Term	Meaning
Agent	A workflow component that uses a model and approved context to reason, classify, draft, or recommend.
Approval gate	A workflow control requiring an authorized human decision before the workflow can continue or execute an action.
Connector broker	The platform component that mediates scoped access to enterprise systems.
Control plane	The platform area where workflows, policies, roles, connectors, and deployments are configured.
Haluvance	Nexoraa enforcement layer responsible for policy checks, authorization, approvals, action controls, and audit capture.
Prompt injection	An attempt to manipulate model behavior through user-provided or externally sourced content.
Tenant boundary	The logical and operational boundary separating customer data, workflows, configuration, logs, and secrets.