The Nihilo Security & Governance Framework: Building Enterprise Cloud Sovereignty for AI

Executive Summary: The Imperative of Enterprise Cloud Sovereignty

The advent of agentic AI, capable of autonomous planning and action, introduces unprecedented opportunities and risks for the enterprise. For companies leveraging cloud platforms like **AWS** and **Azure** to build sophisticated AI workflows and automations, the challenge is not just security, but **Enterprise Cloud Sovereignty**. This is the ability to maintain complete control over data, access, and compliance, even when utilizing third-party AI services.

Nihilo Solutions specializes in bridging this gap. Our core philosophy, **"Compliance by Design,"** ensures that data protection, security, and ethical guardrails are embedded directly into the cloud-native architecture. We transition the discussion from a general AI trust gap to a specific, actionable framework for achieving **data isolation, operational control, and regulatory alignment** within the client's sovereign cloud environment.

I. Data Privacy & Cloud Perimeter Defense

Nihilo's framework is built on the principle that the client's data must remain within their control and never be used to train external models. This is achieved through a combination of strict API protocols and a clear definition of the **Shared Responsibility Model** in the context of AI.

Zero Data Retention (ZDR) API Integration

To guarantee data privacy, Nihilo configures all calls to leading foundational model services to enforce a **Zero Data Retention (ZDR)** policy [1] [2].

- **Azure OpenAI Service:** We ensure that the service is configured to prevent the logging or storage of prompts and completions, making the data invisible to the underlying model provider [3].
- **AWS Bedrock:** We utilize the service's built-in data protection features, which ensure that customer data is not used to improve or train the models [4].

This ZDR integration is a critical component of **Compliance by Design**, ensuring that proprietary data is used only for the immediate inference and is immediately discarded, thereby eliminating the risk of data leakage or model contamination.

The Shared Responsibility Model for AI

In a cloud-based AI deployment, the traditional **Shared Responsibility Model** is adapted to define clear boundaries between Nihilo's service and the client's ownership.

Responsibility Area	Nihilo Solutions (Service Provider)	Client (Data Owner)
Data Sovereignty	Configuration of ZDR, VPC/Private Link, and encryption.	**Full ownership and control of all underlying data.**
Cloud Infrastructure	Configuration of cloud-native security tools (IAM, RBAC, Network ACLs).	Management of the core cloud account, billing, and high-level security policies.
AI Governance	Implementation of technical guardrails, RAG security, and audit logging.	Definition of acceptable use policies and regulatory compliance mandates.

Nihilo configures the cloud-native security tools to protect the AI workflow, while the client retains **full ownership and ultimate responsibility** for the underlying data and its compliance.

II. Technical Guardrails in Cloud Environments

Achieving Cloud Sovereignty requires technical controls that isolate the AI workflow from the public internet and enforce strict identity management.

Network Isolation: VPC & Private Link Implementation

To ensure that sensitive data never touches the public internet during processing, Nihilo implements advanced network isolation techniques:

- **VPC (Virtual Private Cloud) / VNet (Virtual Network):** The entire AI workflow, including the RAG knowledge base and the application layer, is deployed within the client's private cloud network.
- **AWS PrivateLink / Azure Private Endpoint:** We utilize these services to establish a **private, secure connection** between the client's VPC/VNet and the foundational model services (e.g., AWS Bedrock, Azure OpenAI). This ensures that all traffic between the client's environment and the AI service is routed exclusively over the secure, private AWS or Azure backbone, significantly reducing the attack surface [5] [6].

Least Privilege Access: AWS/Azure Native Identity Management

The principle of **Least Privilege Access (LPA)** is enforced using the cloud provider's native Identity and Access Management (IAM) tools for both human users and AI agents.

- **AWS IAM (Identity and Access Management):** We define granular roles and policies that grant AI agents and human users only the permissions necessary to perform their specific tasks. For instance, an AI agent responsible for document summarization will only have read access to the relevant S3 bucket and the ability to call the LLM API, but no write access to production databases.
- **Azure RBAC (Role-Based Access Control):** Similar to AWS, we use RBAC to manage access to Azure resources, ensuring that permissions are tightly scoped and regularly audited. This prevents unauthorized system modifications and limits the "blast radius" of any compromised identity.

RAG Architecture & Knowledge Base Security

Retrieval-Augmented Generation (RAG) systems are central to enterprise AI, but their vector databases are a critical security concern. Nihilo implements security best practices for popular vector databases:

- **Securing Vector Databases:** For managed services like **Pinecone** or **Azure AI Search**, we configure network security groups and access controls to ensure that only the authorized AI application within the private network can query the knowledge base [7] [8]. Data within the vector database is encrypted at rest.
- **Traceability and Integrity:** We ensure the integrity of the RAG knowledge base by protecting it against unauthorized alterations and maintaining a clear audit trail. Every AI response is traceable back to the specific, authorized source documents, which is essential for factual verification and compliance.

III. Compliance & Auditability

Nihilo's Compliance by Design framework provides a clear, auditable path to meet global regulatory requirements by leveraging the cloud environment's native logging capabilities.

Mapping Azure/AWS Logs to Regulatory Requirements

The framework is designed to simplify compliance by mapping the comprehensive logging and monitoring features of AWS and Azure directly to the evidence required by major regulatory frameworks.

Regulatory Framework	Requirement	Cloud-Native Evidence Source
EU AI Act Readiness	Technical Documentation for High-Risk Systems	Configuration-as-Code (e.g., Terraform/CloudFormation), Azure Policy/AWS Config logs, and deployment history.
SOC 2 Framework	Processing Integrity & Confidentiality	AWS CloudTrail / Azure Monitor logs, which record all API calls and data access events.
Audit Trails	Comprehensive logs of sensitive AI decisions.	Structured logging within the AI application, stored securely in Amazon S3 or Azure Blob Storage, with immutable retention policies.

This approach ensures that all sensitive AI-generated decisions and system actions are recorded in an immutable, auditable format, providing the necessary evidence for internal oversight and external audits.

References

[1] OpenAI. *Enterprise privacy at OpenAI*. https://openai.com/enterprise-privacy/ [2] J Kes. *OpenAI's Zero Data Retention Policy*. https://medium.com/@jeffkessie50/openais-zero-data-retention-policy-916ff04a3599 [3] Microsoft. *Securing Azure OpenAI inside a virtual network with private endpoints*. https://learn.microsoft.com/en-us/azure/ai-foundry/openai/how-to/network?view=foundry-classic [4] AWS. *Data protection - Amazon Bedrock*. https://docs.aws.amazon.com/bedrock/latest/userguide/data-protection.html [5] AWS. *Use interface VPC endpoints (AWS PrivateLink) to create a private connection*. https://docs.aws.amazon.com/bedrock/latest/userguide/vpc-interface-endpoints.html [6] Microsoft. *Network and access configuration for Azure OpenAI On your data*. https://learn.microsoft.com/en-us/azure/ai-foundry/openai/how-to/on-your-data-configuration?view=foundry-classic [7] Pinecone. *RAG with Access Control*. https://www.pinecone.io/learn/rag-access-control/ [8] Microsoft. *Security in Azure AI Search*. https://learn.microsoft.com/en-us/azure/search/search-security-overview [9] Secureframe. *2025 Trust Services Criteria for SOC 2*. https://secureframe.com/hub/soc-2/trust-services-criteria [10] AI Act. *Article 11: Technical Documentation*. https://artificialintelligenceact.eu/article/11/