AMAIMA Enterprise Pilot Program: A Solution Overview

1.0 Introduction: Evaluating AI Governance Without Operational Risk

For enterprises in regulated industries like finance and healthcare, the potential of generative AI remains largely inaccessible, locked behind non-negotiable requirements for compliance, security, and operational risk. Deploying unproven AI systems directly into sensitive environments is not a viable option.

The AMAIMA Enterprise Pilot is a purpose-built solution to this dilemma. AMAIMA is not another AI model; it is an **AI Control Plane** and **Decision Governance Layer** that sits between your enterprise applications and various AI model providers. It provides an intelligent, auditable, and secure framework for routing queries and managing AI interactions.

The core purpose of this pilot is to enable your organization to evaluate the full governance capabilities of AMAIMA within your own environment, using your own workflows, but without any operational risk, production exposure, or vendor lock-in. The program is designed to earn your trust not with promises, but with tangible evidence. This document provides a comprehensive overview of the pilot's objectives, architecture, security posture, and engagement model.

2.0 Core Pilot Objectives: Validating Trust and Transparency

We measure the pilot's success not by conventional performance metrics like throughput or latency, but by the quality and integrity of its decisions, the transparency of its operations, and the compliance evidence it generates. The pilot is designed to validate four key capabilities that are critical for any regulated enterprise.

- **Intelligent Routing** Leveraging a 5-level taxonomy (TRIVIAL to EXPERT), AMAIMA ensures that resource allocation is both efficient and defensible, preventing the use of powerful, costly models for trivial tasks while ensuring expert models are reserved for appropriate, high-value scenarios.
- **Decision Explainability** For every query, AMAIMA provides an inspectable, audit-ready rationale explaining *why* a specific routing decision was made. This transparency is crucial for satisfying regulatory reviews and internal audits, transforming the "black box" of AI into a clear, explainable process.
- **Governance Enforcement** The pilot demonstrates how AMAIMA produces reproducible, versioned decisions that create a verifiable and immutable audit trail. This capability directly supports compliance with stringent frameworks like the Sarbanes-Oxley Act (SOX), where traceable decision logic is a non-negotiable requirement.

- **Safe Integration** AMAIMA acts as a secure intermediary between your internal applications and external model providers. This architectural separation ensures that sensitive enterprise data is never directly exposed, providing a critical layer of insulation and control.

These objectives are achieved through a carefully designed architecture that prioritizes safety and control above all else.

3.0 The AMAIMA Architecture: A Dual-Plane Approach to Safety

AMAIMA's dual-plane architecture is the core of its risk mitigation strategy. By strictly separating decision-making from execution, it provides the verifiable control required in regulated environments. This design is fundamental to de-risking AI evaluation and meeting the continuous control expectations of modern enterprises.

The Decision Plane (Always On)

The Decision Plane is the core governance engine of AMAIMA and is active throughout the pilot. Its key attributes are designed for maximum safety and auditability:

- **Stateless and Deterministic:** Processes each request independently without retaining memory of previous interactions, ensuring predictable and repeatable outcomes.
- **Side-Effect Free:** The plane's operations are confined to analysis and decision-making; it cannot alter external systems, write data, or trigger actions.
- **Inspectable Decisions:** Its sole function is to evaluate a query's intent, complexity, and risk profile to produce a clear, versioned, and auditable routing decision.

The Execution Plane (Disabled by Default)

The Execution Plane, which passes the query to the selected AI model, is intentionally disabled at the start of any pilot. This ensures that no external calls are made until your organization provides explicit, written approval.

- **Explicitly Gated:** Can only be activated through a specific configuration change, providing a hard-stop against unintended model interactions.
- **Fully Observable:** When enabled, every action taken by the Execution Plane is logged and monitored to provide complete visibility.
- **Optional:** The pilot can deliver its full value by operating exclusively in decision-only mode, making execution an optional extension rather than a requirement.

This strict separation is a deliberate philosophical choice. The simulation-first default directly aligns with an enterprise's risk-averse posture, demonstrating a shared understanding of operational reality and ensuring no pilot activity can inadvertently impact production systems.

4.0 Security and Data Handling by Design

The AMAIMA pilot is built on a foundation of privacy-by-design. Our data handling protocols are specifically tailored for organizations managing highly sensitive information, such as Personally Identifiable Information (PII) or electronic Protected Health Information (ePHI). The following table outlines the key controls that protect your data throughout the pilot.

Control Measure	Description and Business Impact
No Raw Query Persistence	Minimizes the data attack surface and dramatically simplifies audit scope, satisfying data minimization principles and accelerating compliance reviews.
No PII or ePHI Retention	The system is designed to not index or retain any PII or ePHI, ensuring alignment with the technical safeguards required by frameworks like HIPAA and GDPR.
Telemetry with Hashed Identifiers	Enables operational insight and analytics without compromising user privacy or exposing sensitive data in logs, simplifying security reviews.
Ephemeral Outputs	Model outputs are treated as temporary and are not stored, preventing accidental data persistence and reducing the scope of data governance policies.
Scoped and Revocable Credentials	All API keys are narrowly scoped and can be instantly revoked, providing your security team with complete and granular control over access.
Isolated Deployment Support	The backend can be deployed entirely within your network, ensuring no data leaves your

	control boundary and satisfying the strictest data residency requirements.

These specific controls are designed not only for operational security but also to provide tangible evidence for satisfying broad regulatory frameworks.

5.0 Alignment with Regulatory Frameworks

While AMAIMA does not claim pre-certification against any specific regulation, its architecture is engineered to provide evidence-ready controls that directly support your organization's existing compliance programs. The platform’s features map clearly to the requirements of major regulatory frameworks across key industries.

- **Finance**
    - **SOX:** The generation of immutable, versioned, and traceable decision logic provides a robust audit trail, directly supporting internal controls over financial reporting.
    - **FINRA:** The decision explainability feature provides the clear rationale necessary to support supervisory review obligations, ensuring that AI-driven actions are defensible.
- **Healthcare**
    - **HIPAA/HITECH:** The platform's privacy-by-design principles—including no PII/ePHI retention and minimal data exposure—provide technical safeguards to protect patient information.
- **Cross-Industry**
    - **EU AI Act:** For organizations classifying AI as a high-risk system, AMAIMA provides the governance readiness, auditability, and control framework required for compliance.
    - **General Frameworks:** The pilot also provides control mappings and evidence to support broader standards, including the **NIST AI Risk Management Framework (RMF)**, **SOC 2** controls, and principles from **GDPR** and **CCPA**.

This strong alignment with established frameworks smooths the path from a technical pilot to a fully compliant production deployment.

6.0 Pilot Scope and Engagement Model

The pilot is a structured, time-bound engagement, typically lasting 14–21 days. It is designed to deliver clear, evidence-based outcomes with no future obligation.

6.1 Scope of the Pilot

The pilot scope is clearly defined to ensure focus and safety.

Included in the Standard Pilot:

- Decision-only routing
- Full telemetry and metrics access (Grafana)
- Use of `/v1/simulate` and `/v1/query` endpoints
- Frontend inspection UI for decision review
- Versioned decision schema and audit logs
- Access to the user feedback loop for confidence tuning

Explicitly Excluded (Requires Prior Written Approval):

- Model execution (i.e., sending queries to an LLM)
- Persistent storage of raw queries or PHI/ePHI
- Autonomous self-modification of routing rules
- Any form of external data exfiltration

6.2 Phased Timeline (14–21 Days)

1. **Week 1: Setup and Alignment**
   - Kickoff meeting and architecture walkthrough
   - Pilot environment setup and API key provisioning
   - Mutual confirmation of success criteria
2. **Week 2: Simulation and Review**
   - Live usage of the system in simulation mode
   - Joint decision review and confidence tuning sessions
   - Deep dive into telemetry and audit logs
3. **Optional Week 3: Controlled Execution and Readout**
   - Enablement of controlled execution (if approved)
   - Performance of an incident or audit replay exercise
   - Final executive readout session and pilot summary

6.3 Defining Success

The pilot is considered successful when the following criteria are met:

- The system's routing confidence score meets or exceeds a pre-defined threshold of 0.85.
- All routing decisions are confirmed by stakeholders to be explainable and reproducible.
- No security or governance blockers are identified by your internal IT, security, or compliance teams.
- Stakeholders confirm a tangible increase in trust in governed AI compared to direct model usage by demonstrating that governance provides predictable, explainable, and safe outcomes.

6.4 Flexible Deployment Options

We offer two deployment models to accommodate your organization's specific security and operational requirements.

- **Option A: Hosted Pilot** This is the fastest path to getting started. It includes a read-only frontend for inspection and an AMAIMA-hosted backend accessed via scoped, secure API keys.
- **Option B: Customer-Hosted Backend** This option provides full network isolation by deploying the AMAIMA backend directly into your cloud infrastructure (e.g., EC2/Kubernetes). This ensures no data ever leaves your environment.

This structured engagement model is designed to provide maximum value and clarity, leading to a clear and informed path forward.

7.0 Pilot Conclusion and Path Forward

The AMAIMA pilot is offered on a no-obligation basis. Its primary goal is to provide your team with unequivocal evidence, empowering you to make an informed, confident decision about the future of AI governance in your enterprise.

At the conclusion of the pilot, there are four potential outcomes:

- Proceed to a full production deployment of AMAIMA.
- Extend the pilot scope for further evaluation of specific use cases.
- Retain AMAIMA in a decision-only governance mode for ongoing audit and oversight.
- Discontinue use with no residual dependency, vendor lock-in, or further obligation.

The initial 14-21 day pilot is offered completely free of charge.

This pilot exists to earn trust through evidence, not promises.