

TECHNICAL BRIEF

# AI Gateway Architecture: How Reign Routes, Guards, and Governs

Every enterprise AI request—from developer coding assistants to production agent workflows—should pass through a governed gateway. This brief explores how Reign's AI Gateway becomes your control point.

Technical Buyers, Enterprise Architects, Security Engineers  
March 2026 | v1.0

**200+**

Enterprise Deployments

**15+**

Model Providers

**< 5 min**

Time to First Route

**100%**

Request Visibility

**OVERVIEW**

# Table of Contents

- ✓ Executive Summary
- ✓ What an AI Gateway Does
- ✓ Core Capabilities
- ✓ Deployment Models
- ✓ Integration & Getting Started
- ✓ Next Steps

**EXECUTIVE SUMMARY**

## Executive Summary

Every enterprise AI request—from a developer's coding assistant to a production agent workflow—should pass through a governed gateway. Reign's AI Gateway is the single control point where routing, guardrails, cost management, and evidence collection converge. Without one, your enterprise faces blind spending, unaudited AI actions, compliance gaps, and data leakage. With Reign, every request is routed intelligently, every response is logged, and every token is accounted for.

*Route Every Request. Govern Every Response. The foundation of AI governance is visibility.*

## ARCHITECTURE

# What an AI Gateway Does

An AI Gateway sits between your users, agents, and AI model providers. Every request passes through it. Every response is logged. The gateway is your control point for routing, safety, cost, and compliance.

**Key Functions:**

- ✓ Intelligent routing to multiple model providers
- ✓ Content filtering and guardrails (prompt injection, topic boundaries, PII redaction)
- ✓ Real-time cost tracking and budget enforcement
- ✓ Complete audit trail and evidence collection
- ✓ Model fallback and automatic load balancing
- ✓ Centralized API key management

**Without vs. With Reign AI Gateway:**

Dimension	Without Gateway	With Reign
Visibility	Blind	Complete request-level visibility
Cost Control	Post-hoc analysis	Real-time budget enforcement
Data Protection	Trust the model	PII redacted before it leaves your perimeter
Compliance	Manual logging	Automated evidence collection
Model Selection	Hard-coded	Dynamic routing by policy

## FEATURES

# Core Capabilities

## 01 Multi-Provider Routing

Route requests to OpenAI, Anthropic, Google, Azure, AWS Bedrock, or self-hosted models. Policy-based routing by use case, team, cost, or performance requirements.

## 02 Content Guardrails

Block prompt injection, enforce topic boundaries, redact PII/PHI before requests reach the model. Custom rules per team and use case.

## 03 Cost Governance

Token-level cost tracking per user, team, project, and model. Budget caps with automatic enforcement. No more surprise AI invoices.

## 04 Evidence Collection

Every request and response logged with metadata: user identity, model, tokens, latency, policy checks. Cryptographically hashed for tamper evidence.

## 05 Model Fallback & Load Balancing

Automatic failover between providers. Request queuing during rate limits. Zero downtime model switches for high-availability deployments.

## 06 API Key Management

Centralized key vault. Teams access models through Reign, never directly. Key rotation, expiration, and revocation on demand.

## ARCHITECTURE

# Deployment Models

Reign AI Gateway is available in three deployment models. All include the same feature set. Choose based on your data sovereignty and operational requirements.

Deployment Model	Description	Best For
Cloud-Hosted	Reign-managed SaaS deployment	Fast start, teams < 500
Private Cloud	Deployed in your VPC (AWS, Azure, GCP)	Data sovereignty requirements
Air-Gapped	Fully isolated, no external connectivity	Regulated industries, defense

*All deployment models include the same feature set. Choose based on your data sovereignty requirements, not capability trade-offs.*

**IMPLEMENTATION**

# Integration & Getting Started

Reign is designed for rapid deployment. Change your base URL. That's it. Reign is API-compatible with OpenAI, Anthropic, and all major model providers.

**Integration Quick-Start:**

- ✓ Drop-in replacement for OpenAI/Anthropic API endpoints
- ✓ SDK support for Python, TypeScript, Go, Java
- ✓ MCP server discovery and governance
- ✓ SSO integration (Okta, Azure AD, Google Workspace)
- ✓ RBAC with team-level and project-level policies
- ✓ Webhook integrations for SIEM and alerting

*Most enterprises route their first request through Reign in under 5 minutes.*

NEXT STEPS

# Route Every Request. Govern Every Response.

Take the next step toward governed AI. Every day without visibility is another day of unaudited AI actions and blind spend.

## 01 Connect

Point your AI traffic to Reign in minutes. No code changes required.

## 02 Configure

Set routing rules, guardrails, budgets, and evidence collection policies.

## 03 Govern

Full visibility from day one. Monitor, audit, and optimize every request.

### Included with Reign

- ✓ Full audit trail of every AI agent action
- ✓ MCP connection governance and approval workflows
- ✓ Real-time cost dashboards with token-level visibility
- ✓ 24/7 enterprise support with dedicated TAM

### Schedule a demo: [itmethods.com](https://itmethods.com)

Join 100s of enterprises building with the Fortress Family | AI-Native backed by 21+ Years Enterprise Trust