



MODERN AI
smart. secure. transformative

Reduce AI Token Costs

Modern AI Token Cost Optimization & AI FinOps Roadmap

Reduce LLM and AI application costs while improving performance, control, and scalability.



EXECUTIVE CHALLENGE

! Oversized prompts and context windows

! Expensive model overuse and poor right-sizing

! Poor RAG design and inefficient retrieval

! Inefficient agents and uncontrolled loops

! Lack of caching and reuse strategies

! Limited visibility into AI spend and usage

OUR SERVICES



1. Assess & Baseline

Inventory AI apps, models, prompts, RAG flows, and usage to establish a cost and performance baseline.



2. Prompt & Model Optimization

Refactor prompts, reduce context, tune models, and optimize output to cut token spend.



3. RAG & Agent Efficiency

Tune retrieval, route smartly, and set agent step & tool limits to eliminate waste.



4. AI FinOps Monitoring

Implement dashboards, alerts, and anomaly detection for real-time spend visibility.



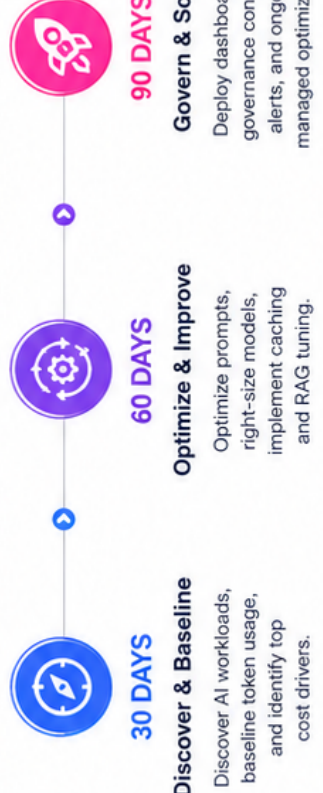
5. Governance & Managed Service

Establish policies, budgets, approvals, and exception workflows with ongoing managed optimization.

COST REDUCTION LEVRS



30 / 60 / 90 DAY ROADMAP



30 DAYS

Discover & Baseline

Discover AI workloads, baseline token usage, and identify top cost drivers.

60 DAYS

Optimize & Improve

Optimize prompts, right-size models, implement caching and RAG tuning.

90 DAYS

Govern & Scale

Deploy dashboards, governance controls, alerts, and ongoing managed optimization.

BUSINESS OUTCOMES



Lower AI Operating Costs

Reduce token spend and improve unit economics.



Faster Responses & Better UX

Optimize models and contexts for speed and quality.



Stronger Governance & Budget Control

Policies, budgets, and alerts drive accountability and prevent overruns.



Scalable Enterprise AI Operations

Repeatable FinOps processes enable confident scale across the enterprise.

TYPICAL OPPORTUNITY:

**20-35%
TOKEN COST REDUCTION***

*Typical range based on client engagements. Results vary by use case and environment.



Modern AI helps organizations **measure**, **optimize**, and **govern** AI usage at scale.



Modernaims.com