

Winning the AI cost game

FinOps strategies for the technology value era

JEREMY CHAPLIN

SNR. DIRECTOR, SOLUTIONS ARCHITECTURE & ADVISORY, FINOPS

flexera™

Welcome!



Jeremy Chaplin

Snr. Director, Solutions Architecture
& Advisory, FinOps



Agenda

1. The Road to Gold: FinOps in the AI Era
2. A New Event: FinOps for AI
3. Training Smarter: AI for FinOps
4. The FinOps Decathlon: Competing across every Scope
5. On the Podium: Q&A / Close



Agenda

- 1. The Road to Gold: FinOps in the AI Era**
2. A New Event: FinOps for AI
3. Training Smarter: AI for FinOps
4. The FinOps Decathlon: Competing across every Scope
5. On the Podium: Q&A / Close

\$2,020,000,000,000

FORECAST FOR AI SPEND IN 2026

Gartner, September 2025

The challenges organizations face

They struggle to manage spending, multi-cloud sprawl, AI workload demands, CO2e reporting



Cloud spending is exploding, driven by AI

Global cloud infrastructure services spending **up 25% year over year.**

Worldwide spending on AI is forecast to total \$2.52 trillion in 2026, a **44% increase YOY.**

Source: [Omdia](#) and [Gartner](#)



AI is adding complexity

Bursty workloads, token-based usage billing, utilization & commitment planning challenges **add time pressure and hurdles** to traditional cloud cost management.

The future of AI cost optimization is **autonomous.**

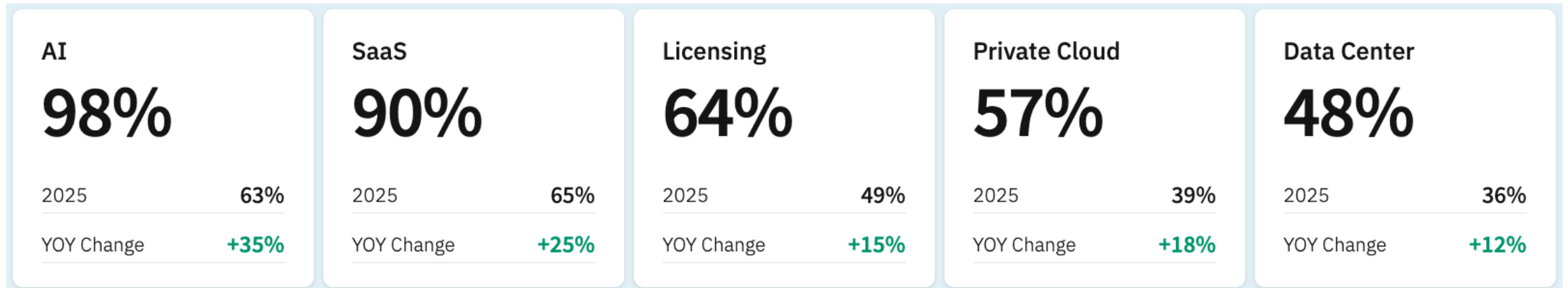


Visibility is a struggle

It's hard to distinguish AI service spend from overall cloud spend, hard to see CO2e impacts and meet compliance reporting needs.

Optimizing costs, cutting waste and tracking CO2e requires **real visibility.**

“Advancing the people who manage the value of *technology*.”



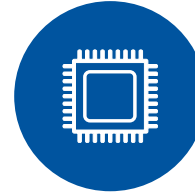
Source: [State of FinOps](#) by [FinOps Foundation](#)

Pricing / Procurement

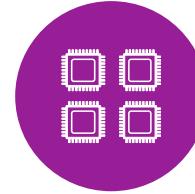
Per Unit Pricing



Per Server



Per CPU



Per Core



Per User

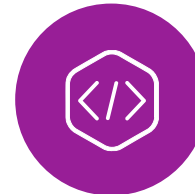
Consumption Based



Public Cloud



Indexed Data



API Calls



Messages

Outcome Based



Per Conversation



Per Assist



Per Resolution



Agenda

1. The Road to Gold: FinOps in the AI Era
- 2. A New Event: FinOps for AI**
3. Training Smarter: AI for FinOps
4. The FinOps Decathlon: Competing across every Scope
5. On the Podium: Q&A / Close

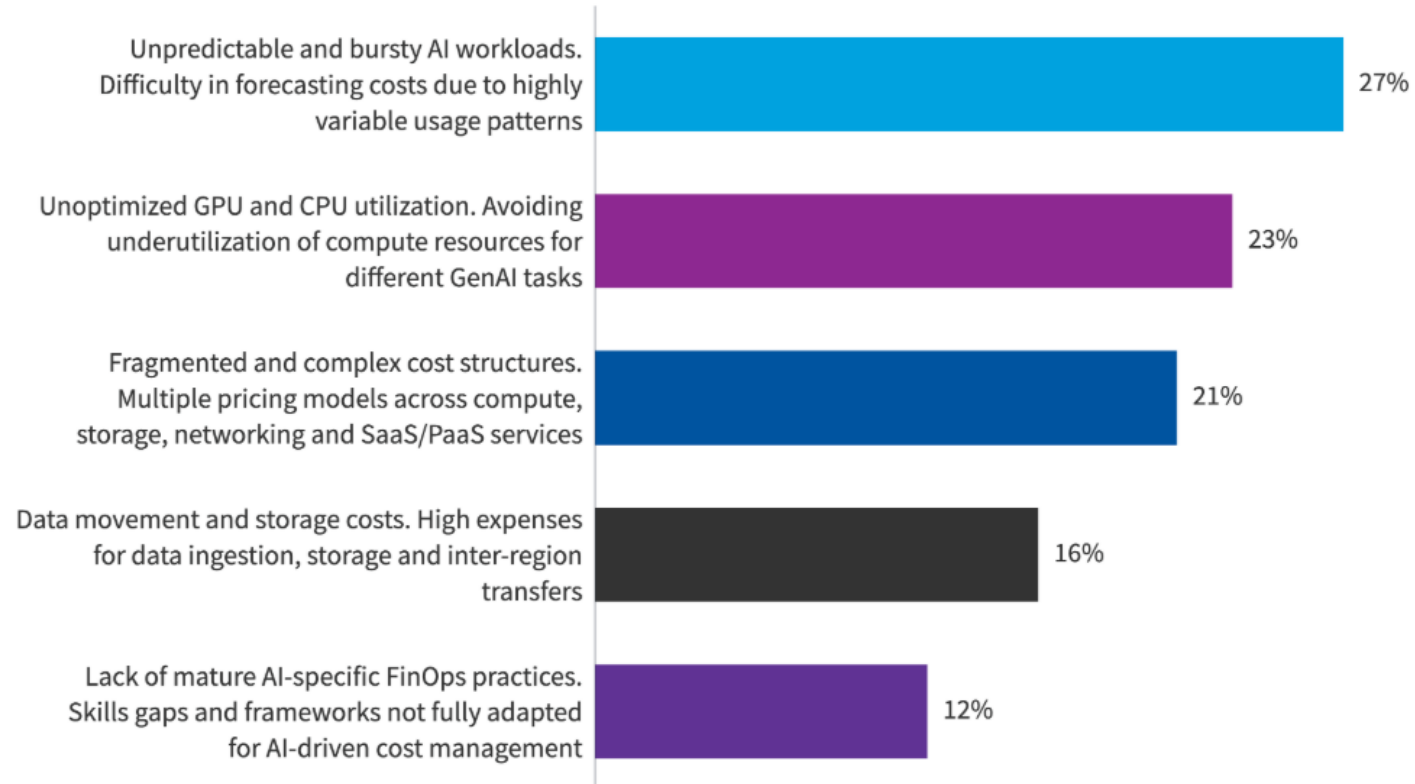
98%

OF FINOPS PRACTITIONERS NOW MANAGE AI SPEND

State of FinOps, 2026

Managing AI Costs is a growing challenge

Top challenges managing AI workload costs



N=753

Source: Flexera 2026 State of the Cloud Report (Figure 44)

flexera

As AI use is exploding, the FinOps visibility gap is growing

Budgeting & forecasting

Has AI increased unplanned spend?



Showback & Chargeback

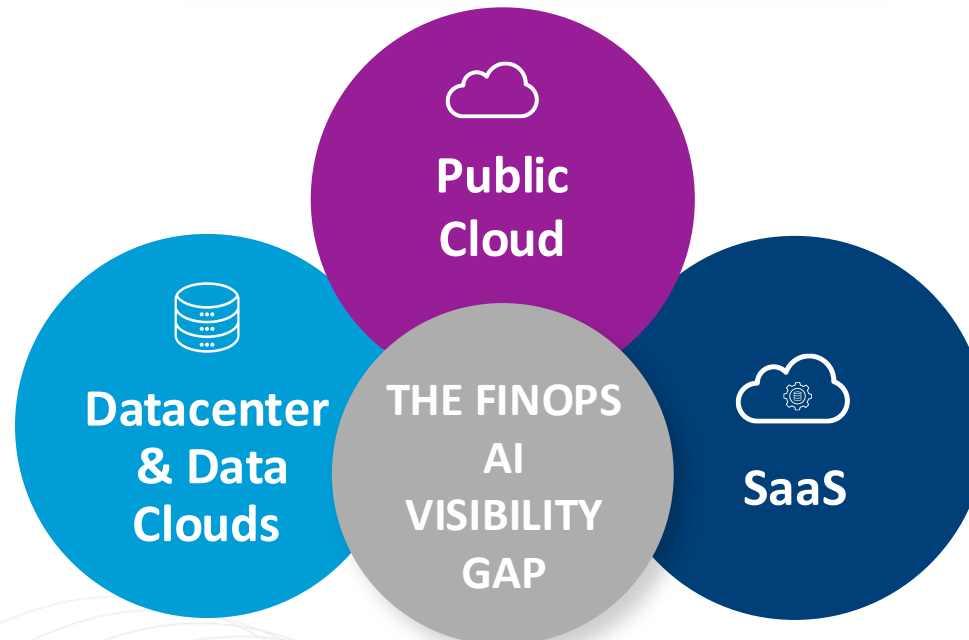
How can we ensure AI spend is accountable?

Cloud governance

Who is using what data for AI, where?

Unmanaged software in the cloud

What AI apps are running where?



Procurement

How can we negotiate better for our AI spend?

License & Usage Optimization

How can we optimize for AI spend and value?

Manage your AI adoption and spend

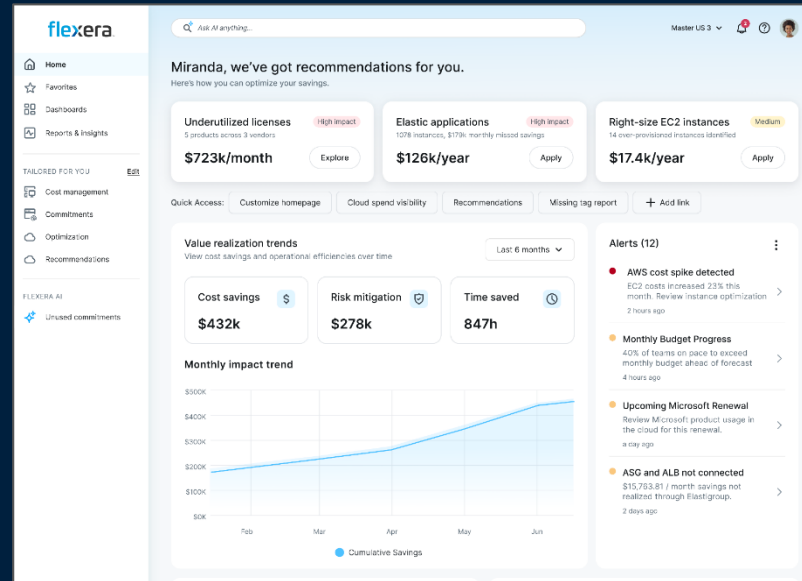
Understand the AI you buy and the AI you build

Bought AI

Generative AI / LLMs



SaaS AI Capabilities



Built AI

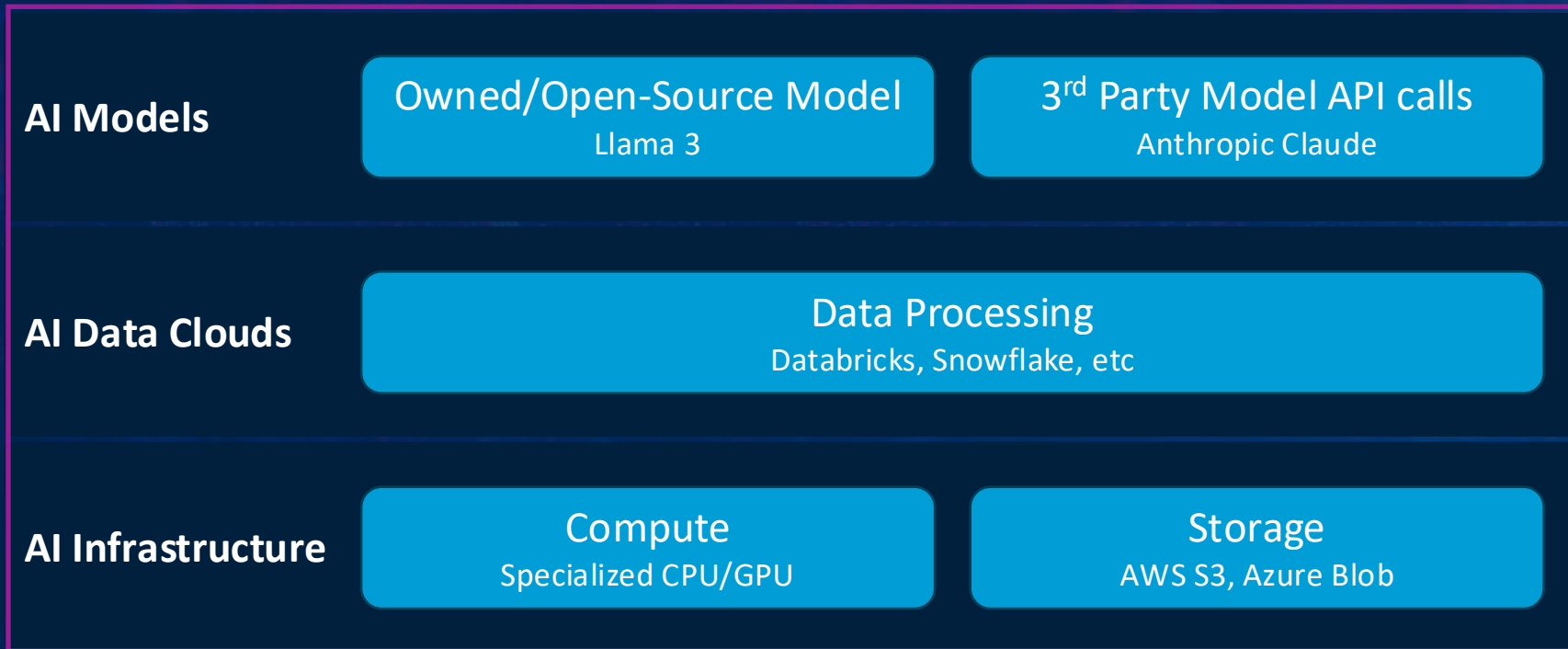
API/Token Usage & Spend



Data clouds (AI PaaS)



Understand what's driving AI spend to optimize investments



Know what's driving AI costs

- See prompts & requests driving token usage
- Identify expensive usage patterns

Run data workloads more efficiently

- Right size Databricks queries & compute
- Improve efficiency at the workload level

Pay only for the compute you need

- Match CPU/GPU resources to usage
- Avoid over-provisioning as demand changes



Agenda

1. The Road to Gold: FinOps in the AI Era
2. A New Event: FinOps for AI
- 3. Training Smarter: AI for FinOps**
4. The FinOps Decathlon: Competing across every Scope
5. On the Podium: Q&A / Close



We've heard: Customers have 3 needs for Flexera AI

1

Power critical AI solutions with clean, curated, contextualized information

2

Provide AI capabilities that reduce spend, risk and effort across my technology portfolio

3

Help me manage my growing AI spend across the IT estate

AI-powered value across Flexera One

AI Insights

Technology Intelligence

Flexera is aware of your business context so you can understand exactly how technology is being used across the organization

AI Agents & Automation

Autonomous Optimization

Flexera agents maximize your results by performing intelligent technology optimization at scale.



Technology Inventory

A complete and accurate inventory of the technology you build and buy

Hardware



Software



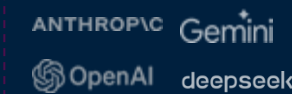
SaaS



Cloud



AI



Enriched with expertly curated technology data

- > Standard Names & Categories
- > Lifecycle & Vulnerability Data
- > ...and much more

Built on the Flexera One platform

Secure

Compliant

Transparent

Reliable

FinOps teams need to leverage AI and provide oversight

AI For FinOps

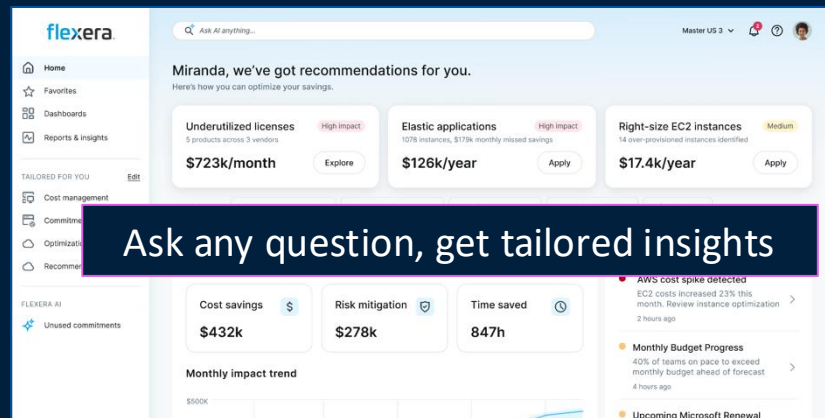
AI powered insights & automation that are aware of your business context to maximize efficiency and impact



FinOps for AI

Comprehensive visibility of your AI estate to accelerate innovation with a standardized, cost optimize & compliant approach

AI insights



AI-powered automation

Rate Optimization

Discounts, Reservations, SPs

Workload Optimization

AI Models

Open-Source
e.g. Llama 3

Commercial
Anthropic Claude

AI Data Clouds

Data Processing
Databricks, Snowflake, etc

AI Infrastructure

Compute
VMs | Containers | GPUs

Storage
AWS S3 | Azure Blob

Flexera turns technology questions into business decisions

Your business objectives

Save Money | Reduce Risk | Increase Efficiency

Your technology assets

Hardware | Software | SaaS | Cloud | AI

Your business context

Business Unit | Region | Department | Team



How can Flexera help today?

We need to free up 10% of spend for innovation.
Can you assess our AI token consumption and
benchmark by application to provide suggestions?



Agenda

1. The Road to Gold: FinOps in the AI Era
2. A New Event: FinOps for AI
3. Training Smarter: AI for FinOps
- 4. The FinOps Decathlon: Competing across every Scope**
5. On the Podium: Q&A / Close

FinOps ownership is expanding *fast*

78% of FinOps teams report to CTO/CIO



Ownership (from 2026 State of FinOps)

100% Manage cloud spend

98% Manage AI spend (prev 31%)

90% Manage SaaS spend (prev 25%)

64% Manage licensing (prev 15%)

57% Manage private cloud (prev 18%)

48% Manage datacenter (prev 12%)

Priorities for technology spend

What capabilities are your top priorities to apply to different areas of technology spend?

SaaS

Allocation

Forecasting & Budgeting

Data Ingestion, and Reporting & Analytics

Planning & Estimating

FinOps Tools and Services

Rate Optimization

Licensing

Allocation

Forecasting & Budgeting

Planning & Estimating

Data Ingestion, and Reporting & Analytics

Rate Optimization

FinOps Tools & Services

Data Center

Allocation

Planning & Estimating

Forecasting & Budgeting

Data Ingestion, and Reporting & Analytics

Workload Optimization

FinOps Tools & Services

Data Cloud Platforms

Allocation

Data Ingestion, and Reporting & Analytics

Forecasting & Budgeting

Planning & Estimating

Rate Optimization

FinOps Tools & Services

FinOps Domains:

Optimize Usage & Cost

Manage the FinOps Practice

Understand Usage & Cost

Quantify Business Value

Source: [State of FinOps](#) by [FinOps Foundation](#)

Flexera One covers every FinOps Scope, domain and capability

FinOps Portfolio



Flexera maximizes savings across your IT investments

Visibility, Governance & Compliance

Budgets & Allocation

Monitor Budgets & Forecast Accurately

AI Spend Visibility

Optimize & Invest More Into AI Strategy

Sustainability & Carbon Footprint Reporting

Compliance Mitigation & Faster Reporting

Software/License Visibility

Up to 40% License Savings

Rate Optimization

Discount Management

Spot Instance Management

Up to 65% Effective Savings Rate

Workload Optimization

Rightsizing & Bin Packing

Auto-Scaling

Data Cloud Optimization

Up to 15% Workload Savings

Up to 30% Data Cloud Savings

Thank you