# Scaling AI in government: Cost strategies with the FinOps framework

The better the question. The better the answer.
The better the world works.

Shape the future
with confidence

# Table of contents

# Executive summary

Governments are increasingly adopting artificial intelligence (AI) to enhance public services, improve decision-making and drive efficiency.

However, scaling AI in the public sector brings significant financial and operational complexities. Agency leaders must integrate AI into their strategic vision and technical architecture, manage new risks and secure a tangible return on investment (ROI) for taxpayers. Key cost drivers — from expensive cloud GPU resources to data management and skilled talent — demand careful planning. Decisions about how to deploy AI (buy vs. build, cloud vs. on-premises) have far-reaching cost implications, and pricing trends are rapidly evolving. A comprehensive cost strategy is needed, encompassing detailed application cost breakdowns, robust forecasting of future spending and smart capacity planning to meet demand without waste.

This whitepaper explores the financial side of scaling AI in government. It examines major cost drivers of AI initiatives and considerations for selecting deployment solutions that balance capability with cost-effectiveness. We review current pricing trends and the layered cost structure of AI applications to help forecast expenses. Effective cost forecasting methods and capacity planning approaches are discussed, acknowledging the unpredictability of AI workloads. We also address the challenges unique to AI — such as surging usage and difficulty tying cost to mission value — that complicate traditional IT budgeting. To keep spending under control, we outline Financial Operations or FinOps best practices as "guardrails," including cross-team cost governance and real-time spend visibility. Finally, we suggest key indicators of success that agencies can use to gauge the effectiveness of their AI cost management strategies. Throughout, FinOps is a model to help government organizations develop effective cost strategies and confidently scale their AI enterprises for maximum public value.

# Introduction: AI adoption and the cost management imperative

AI has the potential to transform federal agencies by streamlining operations and enabling data-driven policies. Yet alongside the promise of AI comes a pressing need to manage its costs. Government IT budgets are limited and every dollar spent must deliver clear value. Notably, public sector cloud spending has surged – the US government cloud market grew from about $15 billion in 2019 to a projected $41.8 billion by 2025. As cloud and AI expenditures grow, agencies face greater scrutiny to manage, budget, forecast and optimize these costs. FinOps enables engineering, finance and program teams to collaborate on data-driven spending decisions that maximize value.

A strong cost management framework is essential when scaling AI projects in government. Agencies must plan for significant up-front investments in data and infrastructure as well as ongoing operational expenses for model training, cloud services and software licensing. At the same time, they need to remain agile – experimenting with AI capabilities in pilot projects – without letting "experimentation" turn into uncontrolled costs. The risk of AI projects overrunning budgets is real: AI cloud workloads often require premium hardware (GPUs and TPUs) that cost much more per hour than standard computing and may run for days during model training. In fact, one industry survey found average monthly AI cloud spend per company was $63,000 in 2024 and projected to jump to $85,000 in 2025 – a 36% year-over-year increase in AI spending. Such growth, if left unmanaged, could strain federal IT budgets. However, when AI is applied with a structured approach focused on cost efficient, workforce productivity and operational streamlining, AI investments can yield dramatic returns even in complex enterprises like federal agencies.

The imperative for government leaders is to adopt AI to fulfill agency missions more effectively with rigorous cost oversight. The FinOps framework provides a pathway to achieve this balance by integrating financial stewardship into the AI lifecycle. In the following sections, we delve into specific cost drivers for AI, strategies for deployment and pricing, and how FinOps practices – from forecasting to chargeback – can support a sustainable, value-driven expansion of AI in government.

# Key cost drivers of AI initiatives in government

Implementing AI solutions entails a variety of cost factors that agencies must account for. Unlike traditional software, AI systems often involve intensive computation and large-scale data usage, which drive up costs. Understanding these key cost drivers is the first step in effective AI financial management:

1. **Infrastructure costs:** Training modern AI models, especially large language models or deep neural networks, requires powerful GPUs or TPUs running for extended periods with hourly rates far above standard servers. In production, inference or serving AI predictions also incurs ongoing compute costs. In both cases, computational capacity is a primary cost driver. If an agency adopts a third-party AI API, compute cost is embedded in the "per API call" pricing (often per token of input and output), but those fees essentially reflect the provider's GPU costs plus margin. For self-hosted or open-source models, the agency directly incurs the cloud compute costs for both training and inference.

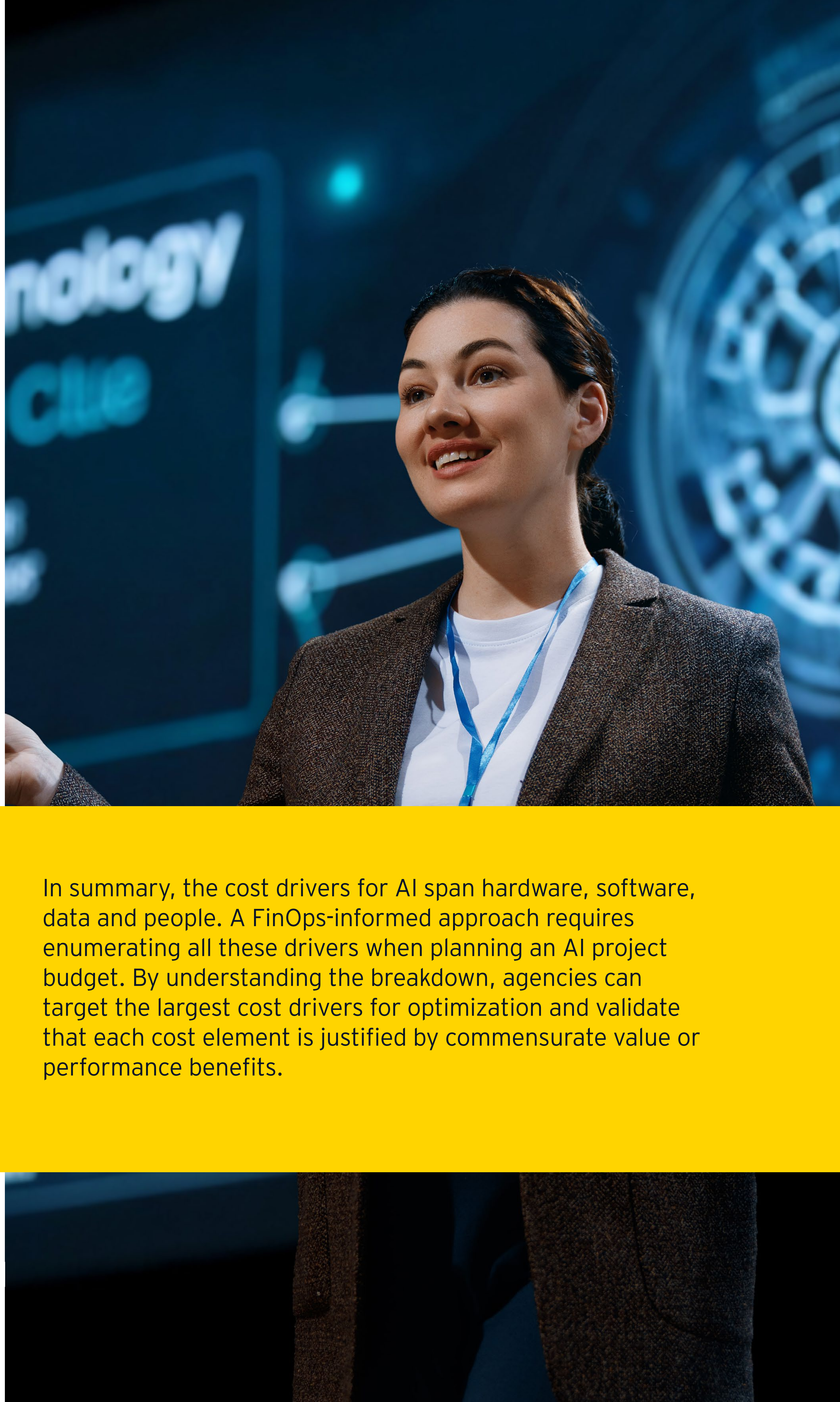
2. **Data acquisition and management:** AI is data intensive. Agencies must store large datasets securely, often in cloud storage or data lakes, and incur costs for data retrieval and processing. Data pipelines for AI (ETL jobs and feature generation) consume storage and compute as well. In AI systems that learn continuously or retrain, new data ingestion and data versioning add to storage needs. All these data-related activities contribute to the overall cost. Even when using a third-party AI service, agencies might pay for data egress or for prepping data to send to the service. Managing data quality also has costs, whether through contracting data preparation services or using internal staff.

3. **Software licensing:** Many agencies experiment with commercial AI services and usage-based pricing. The primary cost driver is usage volume, generally measured in tokens or API calls. If usage surges, costs scale accordingly. Some vendors offer tiered pricing or committed-use discounts, which can mitigate per-unit costs but often require upfront contracts. Additionally, agencies might pay licensing fees for AI software, pre-trained models or extra security tools and services.

4. **Training and development:** AI solutions are rarely "set and forget." For open-source or in-house models, there are ongoing costs for customization, fine-tuning and maintenance. Fine-tuning a model on agency-specific data may require additional compute and machine learning (ML) engineering effort. Over time, models need updates, incurring further training or data collection costs. Even for third-party models, integration and operations costs can arise – setting up pipelines, monitoring model performance, content filtering or bias mitigation. Furthermore, AI systems require monitoring and observability, which may involve additional tooling or cloud services with associated costs.

5. **Talent acquisition:** The human factor is often overlooked in cost discussions but is critical. AI talent – data scientists, AI engineers, ML operations specialists – are in short supply and command high salaries or contractor fees. Government agencies may face particularly high labor costs if they need to hire scarce expertise or rely on vendor professional services to implement AI. The cost of training staff or upskilling existing IT personnel in AI is another consideration.

6. **Auxiliary infrastructure and integration:** An AI application rarely exists in isolation – it must integrate with existing agency systems and user interfaces. There may be costs for additional infrastructure such as databases (for logs or metadata), content delivery networks (if serving AI features to many users) or microservices that wrap around the AI model. Integration with legacy systems can incur development and middleware costs. Also, requirements for redundancy and high availability (especially for mission-critical uses) mean duplicating resources across regions or failover systems, raising costs.



In summary, the cost drivers for AI span hardware, software, data and people. A FinOps-informed approach requires enumerating all these drivers when planning an AI project budget. By understanding the breakdown, agencies can target the largest cost drivers for optimization and validate that each cost element is justified by commensurate value or performance benefits.

# Cost forecasting strategies for AI services

Accurate cost forecasting for AI projects is notoriously challenging – yet it is essential in the public sector to avoid budget surprises and to plan for scaling. AI workloads have unpredictable usage patterns and evolving costs, making traditional fixed IT budgeting methods insufficient. The following are methods to improve cost forecasting:

1 **Use scenario planning for scale:** AI cost does not scale linearly in all cases; therefore, it is advised to model multiple scenarios: a conservative or base case, an expected case and a high-growth case. Especially consider peak usage scenarios because infrastructure often must be provisioned for peak capacity, which costs more. For example, if you forecast an average of 100,000 requests per day but peak could be 200,000, you might need to secure capacity for 200,000. FinOps advises basing forecasts on expected peak usage for each period when dealing with provisioned capacity, since buying capacity typically requires sizing for peak to safeguard performance. This might inflate cost forecasts, but it is realistic for budgeting – any unused capacity can be seen as contingency. Additionally, scenario modeling can help decision-makers understand cost-risk trade-offs: for example, "If adoption doubles (high scenario), do we have budget? Should we invest in a more efficient model to mitigate that cost?"

2 **Leverage historical data and unit cost metrics:** If the AI service is already running (even in pilot form), use actual cloud billing data to inform forecasts. FinOps practitioners encourage tracking unit economics – cost per unit of work, which is analogous to the more familiar "cost per transaction" models in activity-based costing. For instance, measure cost per 1,000 inferences, or cost per user session or cost per gigabyte (GB) of data processed. With these metrics, forecasting becomes a matter of multiplying by planned activity volumes. If currently it costs $0.05 per request and you expect 1 million requests next quarter, forecast approximately 50,000 (then adjust for any known changes like price or efficiency improvements). Over time, strive to improve these unit costs.

3 **Use tools and benchmarks:** Given the complexity, specialized tools can help. Cloud providers offer cost calculators for AI services. Many FinOps platforms provide features to forecast based on historical trends and simulate changes. Moreover, industry benchmarks can be useful to sanity check – for example, knowing that similar organizations saw AI costs double when going from pilot to production can inform your multiplier.

In practice, one forecasting challenge unique to AI is distinguishing research and development (R&D) from production usage. Many AI costs start in research then transition to steady production use. FinOps encourages clearly tagging or separating environments – forecast R&D or experimentation costs separately from production costs. R&D might be more unpredictable and could be time-bound, while production should have more predictable patterns after initial ramp. If not separated, experimental costs can blur the picture and make it seem like the cost per user is wildly erratic.

Finally, once forecasts are made, tracking forecast vs. actual is vital. FinOps process treats this as a feedback loop – on a set cadence, compare actual spending to the forecast and understand variance. Is the model being used more than expected? Did a cost optimization not materialize? Continuous improvement of the forecasting model will result. A key success metric is achieving high forecast accuracy. Hitting such accuracy consistently for AI spend is challenging, but even moving from, say, 50% error to 10% error is a huge win for financial predictability.

Forecasting AI costs requires agility, communication and the use of detailed metrics. It is as much an ongoing process as a one-time task. Governments must get comfortable with revisiting forecasts frequently and using them actively in decision-making. With FinOps practices, forecasting becomes a way to bridge the gap between technical plans and budget realities, leading to fewer surprises and allows leadership to have increased trust that AI initiatives will stay within manageable cost bounds.

# FinOps guardrails and best practices for cost control

To keep AI spending in check while still enabling innovation, governments can implement a series of FinOps-based guardrails and best practices. Think of these as guiding principles and mechanisms that help teams stay within budgetary boundaries and make cost-conscious decisions without stifling the potential of AI. Here are several key guardrails and practices:

1. **Tagging and cost attribution:** Tag every cloud resource or AI service usage with the proper label, including project, environment, owner and purpose. This allows for granular cost allocation and accountability. For shared services – like a central data lake used by multiple AI projects – implement a cost allocation scheme so each project "sees" its portion. Tools can then break down the bill by these tags so that every dollar is traced to a team or function. This not only curbs the common "no one owns this cost" problem but also encourages teams to clean up resources they no longer need.

2. **Budgets, quotas and alerts:** Set budgets at multiple levels – for example, per project or per month – and configure cloud platforms to alert when spending approaches those limits, rather than after overruns have occurred. Some agencies implement hard quotas for nonproduction environments; for instance, an experiment environment might have a quota that stops further usage if a certain dollar amount is reached in a month. This prevents runaway experiments from infinitely burning money. However, be careful with hard stops on production – better to alert and investigate than to cut off a service. Alerts should be actionable in real time: if an anomaly is detected (such as a sudden spike), notify the responsible engineer and FinOps analyst to check it. A culture of "no surprises" is the goal.

3. **Cost-aware engineering:** Invest in making cost part of the engineering ethos. This involves training developers and data scientists on how cloud billing works, what things cost and how to design cost-efficient solutions. Encourage a mindset of "cost-aware experimentation" – engineers should experiment with AI but also be mindful of the resource impact. One approach is to provide teams with cost dashboards that they can easily access (not hidden in finance systems). When an engineer runs a big job, they could see its cost the next day on their dashboard. Some organizations even implement integrated development environment (IDE) plugins or command-line interface (CLI) tools that estimate cost of a job before it runs. The idea is to bring cost feedback as close as possible to the moment of decision.

4. **Regular FinOps review meetings:** Conduct regular cost review meetings that include stakeholders from engineering, product and finance. In these meetings, present the latest spend, compare to forecasts, highlight any anomalies and discuss upcoming changes. These meetings are meant to be collaborative forums to course correct as needed. For example, if one project's cost spiked, the team can explain why and discuss whether budget adjustments or optimizations are needed. This also raises awareness across teams – one team's expensive pitfall can be a lesson shared, or a clever cost optimization by another can be adopted widely. Keeping finance in the loop helps to make sure they understand the context of spend.

5. **Automated waste elimination:** Implement automation to clean up known waste patterns, like automatically shutting down idle GPU instances or setting lifecycle policies on storage. Government may not be able to automate fully if certain approvals are needed, but even automated suggestions followed by manual action are beneficial.

6 **FinOps dashboard and KPIs:** Develop a dashboard of key cost and usage metrics that is shared openly with all stakeholders of AI projects. Key metrics could include: current month spend vs. budget, cost per unit (token, user, and so on), GPU utilization rates, forecast vs. actual trend and more. By monitoring KPIs, the team can set targets and track progress. Achieving these targets then becomes part of project success criteria.

7 **Chargeback models:** Implementing chargeback or showback in government agencies can enforce discipline. If each program or bureau gets "billed" from a central IT fund for their AI usage, they will pay more attention to that usage. It essentially treats internal teams like cloud customers who must pay for what they consume. Intragovernmental billing can be complicated, but even a notional chargeback can have psychological impact. Chargeback works best when combined with the transparency and optimization support FinOps provides.

8 **Predeployment cost assessments:** Incorporate a cost review as a gate in the AI project lifecycle. Before a new AI service goes live or a new model is deployed to production, conduct a robust cost assessment and address cost-related failure modes. This is analogous to security reviews or performance testing before go-live. It forces consideration of cost early.

9 **Optimize before scale:** A guardrail philosophy to adopt is "optimize before you amplify." If a pilot is working and slated to scale to a larger audience, take the time to optimize the cost while still in the pilot phase to avoid scaling inefficiencies. Sometimes timelines are tight, but even small optimizations early can pay huge dividends once multiplied by many users.

10 **Continuous education and FinOps engagement:** Technology and pricing change frequently in AI, so one best practice is continuous education. FinOps teams should stay up to date on new features and work with engineering to evaluate cost-saving architectures and patterns.

Implementing these guardrails transforms cost management from a reactive task to a proactive part of the AI development process. With such controls, government agencies can pursue ambitious AI projects with less fear of budgetary fiascos, maintaining public trust in their stewardship of funds.

# Considerations for selecting AI deployment solutions and optimizing AI costs (build vs. buy)

A critical strategic decision is how to deploy AI solutions while considering costs: purchase managed service APIs, leverage commercial off-the-shelf (COTS) solutions, or develop in-house solutions. Each of these has different cost implications, advantages, and trade-offs which must be weighed against control, security and budget constraints. Below we outline key considerations for each:

1 **Third-party commercial AI services (COTS):** This refers to using hosted AI solutions from vendors.

- **Pros:** The vendor provides a fully managed service with high-quality pre-trained models and support. This is attractive if an agency needs quick results and lacks AI expertise. Updates and improvements are handled by the provider.
- **Cons:** Limited customization – the agency cannot tweak the model beyond perhaps minor fine-tuning, and it depends on the vendor's feature roadmap. Data security and privacy could be concerns, especially with sensitive data. Cost-wise, these services often have premium pricing per API call or monthly subscription, and costs can accumulate unpredictably with usage. There is also vendor lock-in risk – switching providers could be difficult after integration.
- **Cost drivers:** Primarily the usage fees, generally measured by token. These costs are easy to start incurring and can scale quickly if usage grows. Additionally, vendors may charge for premium features or higher tiers of performance. Because costs are tied to usage, cost forecasting is tricky – one must estimate how often the service will be called and with how many tokens, which can vary with user demand. Another consideration is that newer model versions often cost significantly more – for instance, the latest, most powerful generative model might cost five to twenty times more per token than a previous version. Agencies need to choose models wisely by workload need to avoid overpaying for unnecessary sophistication.

2 **Third-party hosted open-source models (managed platforms):** In this model, open-source AI models (such as various open large language models (LLMs) or vision models) are hosted by a third-party platform or service. It is a middle ground – you get more model flexibility than closed APIs, but someone else manages the heavy lifting of infrastructure.

- **Pros:** Greater control and flexibility than closed services. You can often select specific models, fine-tune them or use niche models that better fit agency needs. Many open-source models mean no licensing cost for the model itself, and competition among hosting providers can drive prices down for the compute serving costs. Some platforms also allow deploying within a virtual private cloud or with stricter data handling, which can be important for government privacy requirements.
- **Cons:** Still requires considerable technical expertise to choose and fine-tune the models. The support and maturity of these platforms may be lower than those of large vendor services, potentially leading to more engineering effort on the agency's part. The agency is often responsible for tasks such as maintaining prompt safety and monitoring model outputs, which are additional burdens.

Performance and reliability might not match the big providers unless carefully managed. Time to value can be slower than a turnkey API because you might need to experiment with different open models and configurations.

- **Cost drivers:** Key cost components include infrastructure usage (GPU, CPU and memory) for both training and inference, which the platform will pass through in pricing. Some providers might charge a platform fee or markup on top of raw compute charges. If fine-tuning or custom training is performed on open models, training charges may be incurred. Also, the volume of data that is both processed and stored will incur costs – if you are embedding a lot of data or handling large datasets, these costs are important to understand. Compared to closed APIs, you have more line items – this means more control to optimize but also poses additional risk for cost overruns if not closely managed. This approach carries a higher risk of unexpected costs, as you – not the provider – are responsible for resource management. FinOps practices like tagging and monitoring are crucial here to avoid waste. Lastly, the complexity of the solution can make it harder to predict the total cost without robust cost estimation tools.

## 3 **Self-hosted AI on cloud or on-premises:** This is the option where the agency's IT team builds and deploys AI systems using cloud providers' services or even on-premises hardware in some cases.

- **Pros:** Full control over models, data and infrastructure. This is important if agencies require strict data governance or need to implement custom algorithms. It allows integration with existing cloud infrastructure, security controls and DevOps processes the agency already uses. The agency retains intellectual property of the trained models and has the opportunity to optimize the system for specific workloads.
- **Cons:** This approach demands significant technical acumen and increases the need for retaining a skilled workforce. ML engineers and cloud architects are needed to design, build and maintain the solution, which can be challenging for government agencies competing with the private sector for talent. As development and deployment responsibilities increase, longer lead times are often associated, which can slow down the agency's ability to deliver timely results. There is also a greater burden in maintaining performance, reliability and security. Any mistakes in architecture could lead to inefficiencies or security vulnerabilities.
- **Cost drivers:** In-house solutions require agency management of all cost components directly. This includes cloud compute instances for training and inference, storage, networking costs and even costs for pipelines and development or test environments. There may be additional software or license costs if using enterprise tools. One advantage is the potential to optimize – for example, using reserved instances or savings plans to lower compute unit costs or adopting spot instances for noncritical training jobs. In fact, many cloud providers offer AI-tailored services (managed distributed training, serverless ML inference) that can lower the effort and sometimes cost by scaling resources up or down automatically. However, leveraging these requires careful planning. Another driver of cost arises when the scale of use grows, demanding more capacity within the instances and potentially requiring additional purchases of higher-tier hardware. If an AI solution becomes very popular – for example, powering a citizen service – the agency might need to invest in significant cloud resources or even commit to capacity contracts with vendors to secure GPUs at scale. This can increase costs sharply if not anticipated. The do-it-yourself (DIY) model has the widest cost variance: a well-optimized deployment can be very cost-efficient, whereas a poorly managed deployment will introduce waste. Indeed, studies indicate that between a fully optimized self-hosted solution and a naive usage of a leading vendor API, there could be a thirty to two hundred times cost difference for essentially similar outcomes. This underscores how important FinOps discipline is – matching the right technology and purchase model to the workload can produce orders-of-magnitude savings.

When deciding among these options, agencies should consider factors such as the nature of the application, required level of model sophistication, data sensitivity and available skills or budget. For example, for a simple public-facing chatbot with no sensitive data, a third-party API might suffice. But for a complex predictive analytics platform dealing with confidential data, a self-hosted solution or customized open-source model might be warranted despite higher initial costs. The goal is to align the AI deployment choice with business value – sometimes a cheaper model that is slightly less advanced is perfectly fine and saves money, whereas other times paying more for a high-quality model prevents wasting effort on something that does not meet user needs.

In practice, many government AI initiatives start with a quick win using a commercial service and then evaluate longer-term options such as open-source or in-house models for scale. A FinOps-informed approach during this evaluation can help project teams quantify the cost trade-offs.

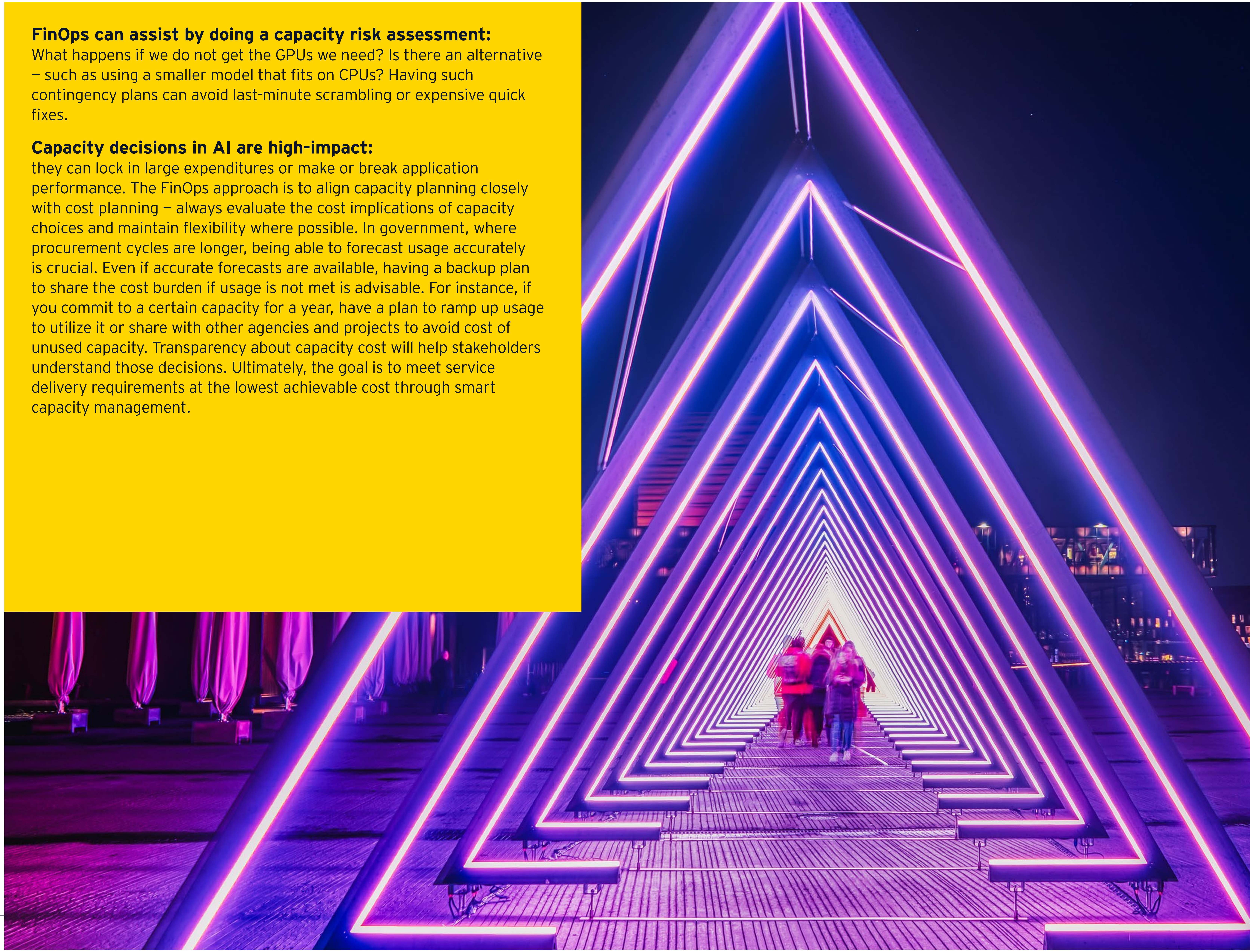
Capacity planning for AI is intertwined with cost forecasting, but it brings its own set of decisions: How much infrastructure should we provision? Should we rely on on-demand resources or secure dedicated capacity? Capacity choices can significantly impact cost, performance and risk. In government settings – where reliability and continuity of service are often paramount – these decisions are critical. Below are key considerations and FinOps-aligned strategies regarding capacity management for AI.

- **On-demand (shared) capacity:** This is the "pay as you go" model – you use what you need when you need it and you pay per unit. On-demand is flexible and requires no long-term commitment. The downside is performance variability and potential availability limits during peak times, since you are sharing resources with others. Cost-wise, on-demand usually has a higher unit price compared to committed capacity. It is suitable for development, low-volume usage, or variable and spiky workloads where you do not want to pay for idle capacity.
- **Dedicated capacity:** Here, you reserve a certain capacity in advance – for instance, a dedicated cluster of GPU instances or a fixed throughput from an AI API. This provides consistent performance and availability, which is important for production applications needing low latency or guaranteed service levels. The cost is typically an upfront or fixed fee covering that capacity, often at a lower marginal rate than on-demand. However, you pay for the capacity whether you use it fully or not. If your utilization is less than the provisioned amount, your effective cost per unit goes up.

For government agencies, the choice often leans toward provisioned capacity for mission-critical AI systems – you cannot afford an outage or slow response due to external demand spikes. However, this must be balanced with cost efficiency. A good practice is to start on-demand in early stages, gather usage data and only move to provisioned capacity when you are confident about steady utilization levels. When opting for provisioned capacity, negotiate the terms so that the minimum commitment is not far above your peak needs.

**FinOps can assist by doing a capacity risk assessment:** What happens if we do not get the GPUs we need? Is there an alternative – such as using a smaller model that fits on CPUs? Having such contingency plans can avoid last-minute scrambling or expensive quick fixes.

**Capacity decisions in AI are high-impact:** they can lock in large expenditures or make or break application performance. The FinOps approach is to align capacity planning closely with cost planning – always evaluate the cost implications of capacity choices and maintain flexibility where possible. In government, where procurement cycles are longer, being able to forecast usage accurately is crucial. Even if accurate forecasts are available, having a backup plan to share the cost burden if usage is not met is advisable. For instance, if you commit to a certain capacity for a year, have a plan to ramp up usage to utilize it or share with other agencies and projects to avoid cost of unused capacity. Transparency about capacity cost will help stakeholders understand those decisions. Ultimately, the goal is to meet service delivery requirements at the lowest achievable cost through smart capacity management.



# Conclusion

Scaling AI in government is a balancing act between ambition and prudence. On one hand, AI offers transformative opportunities for public services – automating tedious processes, providing insights for better decisions and enhancing citizen interactions. On the other hand, AI's underlying costs can be significant and unpredictable.

Understanding key cost drivers – from compute and data storage to third-party service fees and personnel – allows agencies to plan AI projects with all expenses in mind. Agencies can evaluate different deployment models (buy vs. build) that come with trade-offs in cost control, highlighting the importance of aligning the choice with agency capabilities and needs.

Forecasting for AI should be a continuous, collaborative process – not an annual checkbox – due to the fast-changing usage patterns of AI services. Capacity decisions, such as on-demand vs. provisioned and reserved instances, are major levers that need to be pulled with care and data, since they lock in cost structure and affect performance commitments.

Scaling AI will undoubtedly present challenges, from managing bursty workloads and multi-team efforts to marking certain every dollar spent translates to value for citizens. FinOps offers practical solutions to these challenges: cultural change toward cost-awareness, real-time monitoring to catch anomalies and governance mechanisms to allocate costs fairly and spot inefficiencies. In essence, FinOps bridges the gap between the technical engineers pushing AI forward and the financial stewards upholding accountability. By implementing guardrails like tagging, budgets and regular reviews, agencies create a safety net that keeps AI initiatives on track financially. For government, the stakes are not just financial but also public trust. Demonstrating that AI projects are run efficiently and deliver tangible results helps build credibility with oversight bodies and the public.

Ultimately, scaling AI in government does not have to mean scaling costs uncontrollably. With a FinOps framework guiding cost strategy, agencies can innovate rapidly yet responsibly, turning AI from a speculative expense into a well-managed investment with clear returns in efficiency and service quality. The journey involves learning and adaptation – from educating teams on cost drivers to iteratively refining forecasts – but it leads to an AI-enabled government that is cost-effective, outcome-focused and accountable. By viewing cost strategy as integral to AI strategy, public sector leaders can make certain that AI's promise is realized in a sustainable way, delivering maximum value to the citizens they serve.

**EY | Building a better working world**

EY is building a better working world by creating new value for clients, people, society and the planet, while building trust in capital markets.

Enabled by data, AI and advanced technology, EY teams help clients shape the future with confidence and develop answers for the most pressing issues of today and tomorrow.

EY teams work across a full spectrum of services in assurance, consulting, tax, strategy and transactions. Fueled by sector insights, a globally connected, multi-disciplinary network and diverse ecosystem partners, EY teams can provide services in more than 150 countries and territories.

All in to shape the future with confidence.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. Information about how EY collects and uses personal data and a description of the rights individuals have under data protection legislation are available via ey.com/privacy. EY member firms do not practice law where prohibited by local laws. For more information about our organization, please visit ey.com.

Ernst & Young LLP is a client-serving member firm of Ernst & Young Global Limited operating in the US.

2507-10947-CS
ED None
US Score no. 28280-251US

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, legal or other professional advice. Please refer to your advisors for specific advice.

ey.com