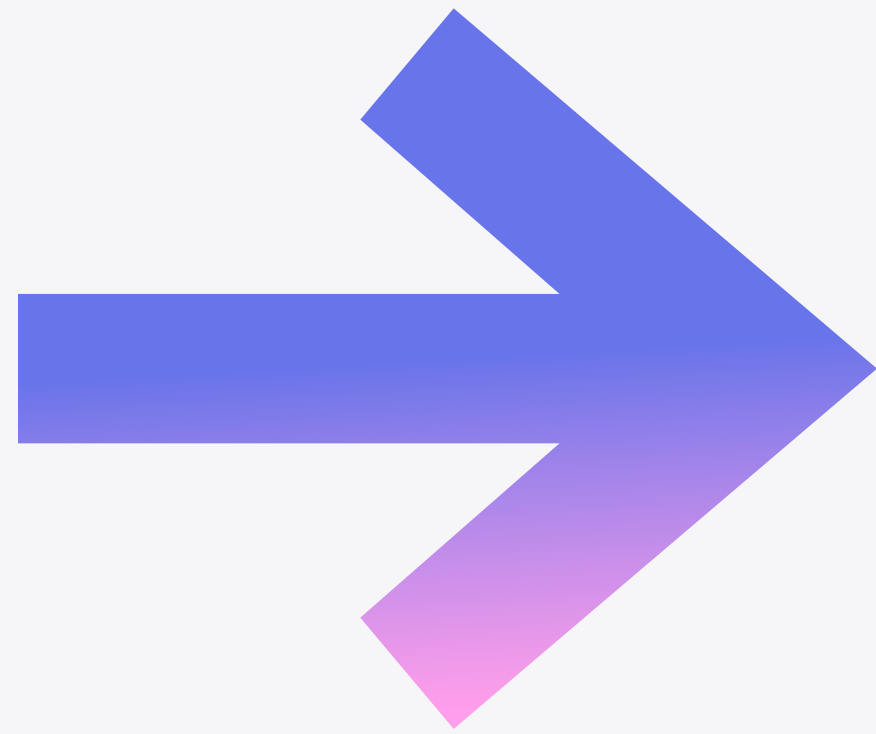




Azure Essentials: Cost Efficiency Series

A Comprehensive Guide to Achieving Success on Azure AI



03 /

Introduction

- 03 Overview
- 05 What is FinOps?

08 /

Establish AI foundation

- 09 Create business case
- 11 Secure funding
- 13 Organizational readiness

16 /

Build GenAI solutions efficiently

- 17 Establish governance policies
- 19 Efficiently design and deploy AI solutions

22 /

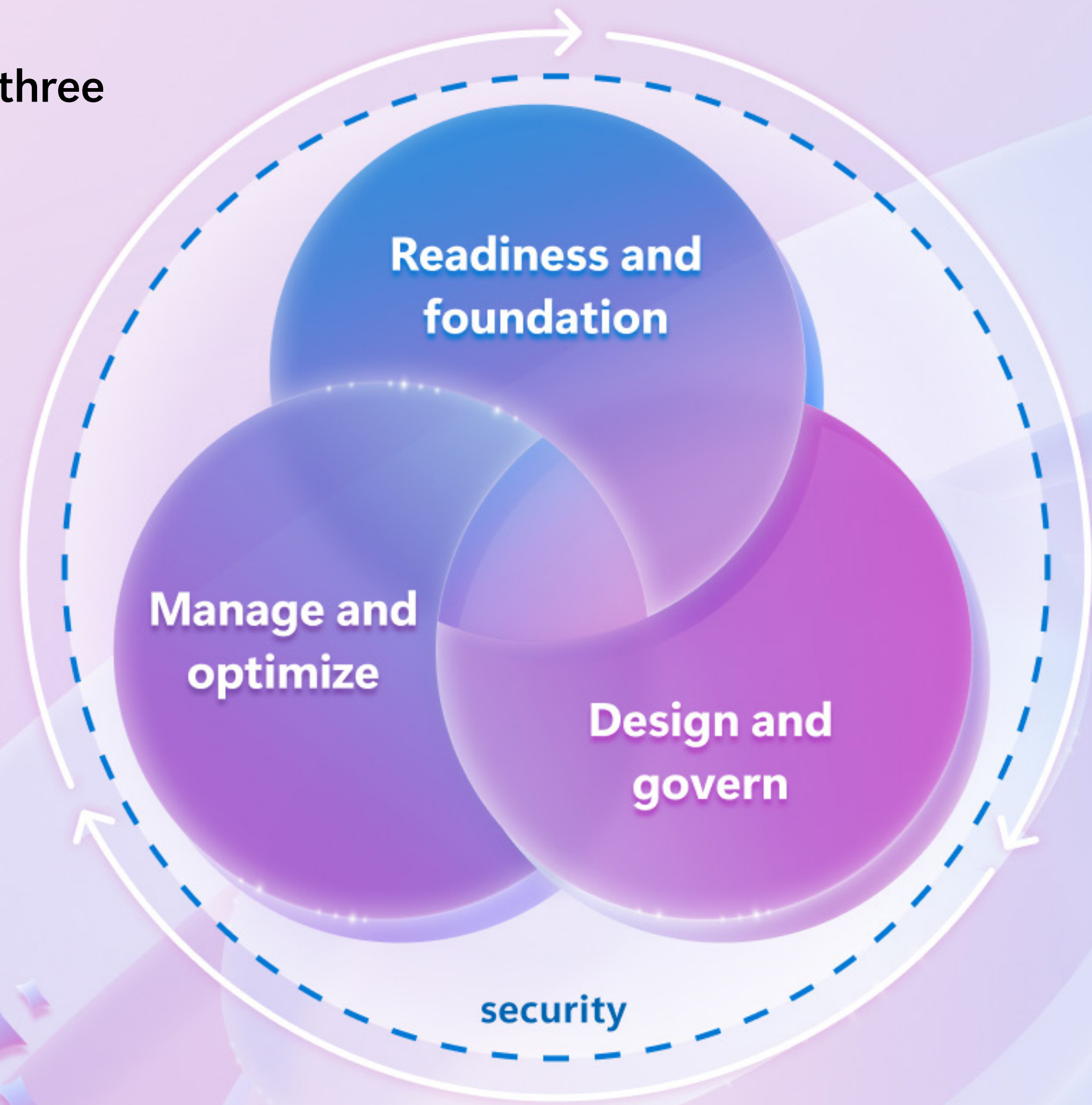
Manage AI investments

- 23 Quantify AI business value
- 25 Monitor usage and patterns
- 27 Capitalize on efficiency opportunities

Introduction

Adopting artificial intelligence (AI) can be a transformative step for any organization, offering the potential for innovation and competitive advantage. However, the journey to AI integration comes with its own set of financial complexities and organizational challenges. This e-book is designed to help you navigate the intricacies of planning and managing AI projects, providing a detailed exploration of financial considerations, best practices, and actionable solutions to ensure a cost-effective and successful AI implementation.

Your journey is structured into three critical phases:

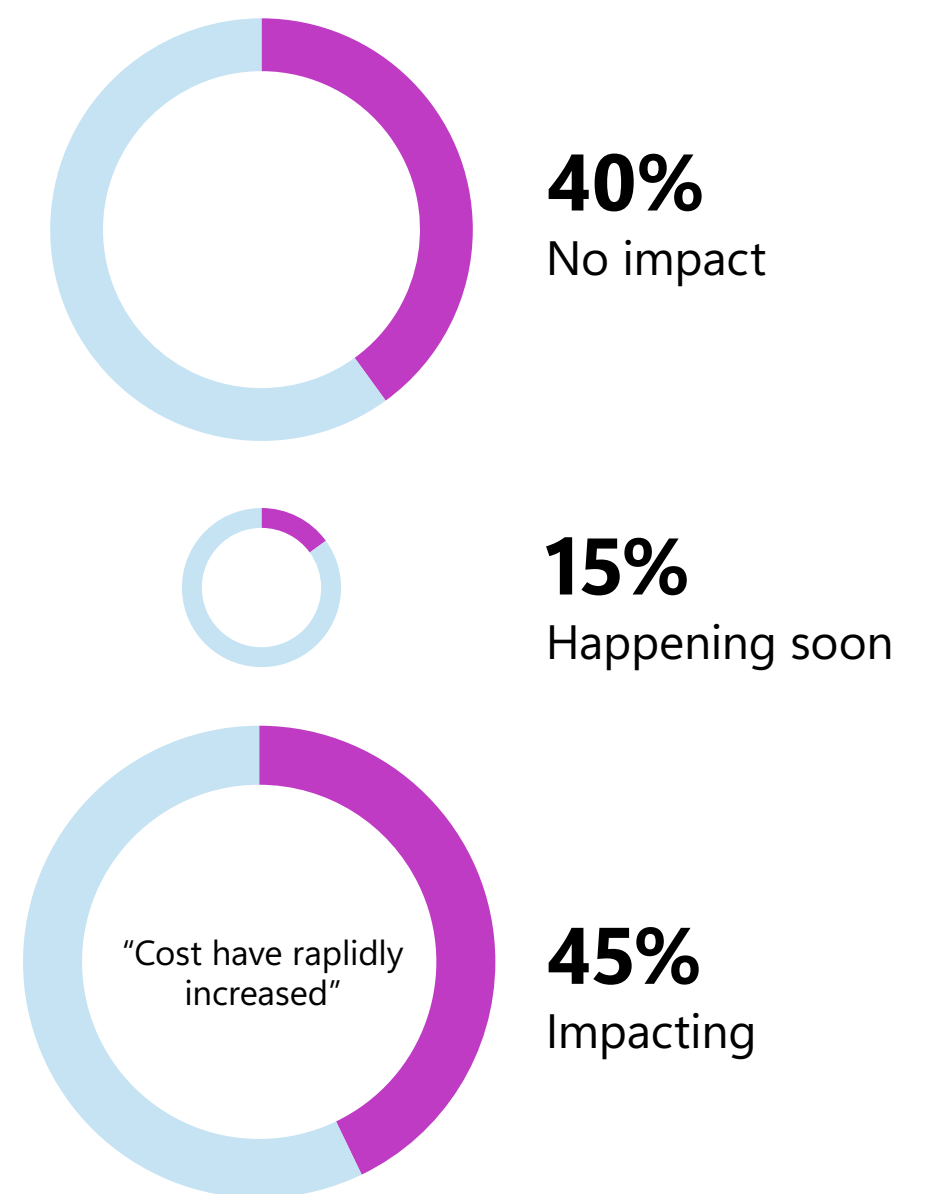




Since effective financial management extends beyond merely having access to tools, this e-book aligns with financial best practices of the FinOps Framework from the **FinOps Foundation**. These practices support long-term objectives by fostering organizational and cultural shifts toward financial responsibility and efficiency. This alignment helps ensure that your AI initiatives deliver a positive return on investment and are aligned into your broader financial strategy.

Leverage this e-book to equip your team with the knowledge, best practices, and tools necessary to strike a balance between innovation and cost-efficiency. Whether you're at the initial stages of AI exploration or looking to refine your existing AI strategies, this comprehensive guide can provide valuable insights, practical steps, and resources to help you navigate the financial complexities of AI adoption.

How are AI/ML costs currently impacting your FinOps practice? (\$100M+ annual cloud spend)



<https://data.finops.org/#11757>
State of FinOps by FinOps Foundation

What is FinOps?

“FinOps is an operational framework and cultural practice which maximizes the business value of cloud, enables timely data-driven decision making, and creates financial accountability through collaboration between engineering, finance, and business teams.”

FinOps definition by FinOps Foundation
Technical Advisory Council



Some ways in which FinOps supports the AI adoption journey

- **Organizational and cultural alignment:** FinOps promotes a culture of financial accountability, ensuring AI adoption aligns with long-term business value and efficiency.
- **Cross-functional alignment:** Collaboration between finance, engineering, and operations is fostered by FinOps, aligning AI strategies with business objectives to balance innovation and financial responsibility.
- **Understanding costs and usage:** FinOps offers clarity on AI service lifecycle costs, enabling informed decisions on scaling and resource allocation.
- **Governance, accountability, visibility:** FinOps helps ensure that AI investments are tracked, optimized, and aligned with cost efficiency through governance and visibility into usage.
- **Rate optimization:** FinOps can help secure optimal rates for AI services, minimizing costs while maintaining performance.
- **Quantify AI business value:** FinOps provides insights into AI service costs, helping businesses prioritize high-return AI projects.
- **Continuous monitoring and optimization:** FinOps calls for the optimization of AI services, ensuring efficient spending and alignment with evolving business needs.

FinOps Framework overview

FinOps is cloud agnostic and based on best practices identified and codified throughout the industry. The FinOps Framework encompasses a comprehensive set of ideas and best practices designed to maximize cloud value. The core components, of the FinOps Framework are:

- **Principles:** FinOps principles act as a north star, guiding the activities of your FinOps practice.
- **Personas:** Personas represent broad stakeholder groups that must collaborate, not individual people. Each has an important part to play in supporting FinOps.
- **Phases:** FinOps is adopted by working iteratively on the Framework capabilities through three phases: Inform, Optimize, and Operate.

- **Maturity model:** A “Crawl, Walk, Run” approach to performing FinOps enables organizations to start small, and grow in scale, scope, and complexity as business value warrants maturing a functional activity.
- **Domains:** The domains of the FinOps Framework describe the fundamental business outcomes organizations should achieve from the FinOps practice.
- **Capabilities:** The capabilities of the FinOps Framework represent functional areas of activity in support of their corresponding FinOps domains.

Together, these components form a robust framework that helps organizations achieve financial accountability and efficiency in their cloud operations, ultimately driving business value and sustainable growth. [Learn more about the FinOps Framework.](#)

[FinOps Framework](#) by [FinOps Foundation](#)

Get started or advance you FinOps practice on Azure

FinOps solutions

Explore the range of Microsoft products and services designed to bolster your efforts in implementing FinOps best practices.

FinOps review assessment

Use FinOps guidance to assess your organization’s capability gaps. Get recommendations for maximizing cloud business value using FinOps best practice

FinOps interactive guides

Gain hands-on experience with Microsoft solutions that enable various FinOps capabilities and best practices.

FinOps training

Gain a fundamental understanding and develop practical skills for adopting FinOps on Azure with Microsoft Learn content.

FinOps documentation

Find technical documentation and guidance to help make informed decisions and follow FinOps best practices.

FinOps e-book

This e-book guides you through the FinOps principles to highlight Microsoft solutions that can support your cloud journey.

Establish AI foundation

In this phase, organizations should concentrate on selecting optimal use cases, securing funding for AI projects, establishing a solid and sustainable foundation, and preparing their teams for successful AI adoption. This comprehensive approach ensures that AI projects are well-funded, efficiently managed, and aligned with organizational goals and capabilities in the long run.





Create business case

Viewing AI as an end rather than a means to an end can be counterproductive. This approach can cause strategic misalignment and result in ineffective implementation, where AI is pursued for its own sake rather than to achieve specific business goals.

To avoid this pitfall, it's essential to carefully select your initial AI use cases to secure quick wins that can serve as a foundation for further development. Start by treating these use cases as you would any technical project, prioritizing and ranking them based on their potential impact and feasibility. Securing executive buy-in is crucial at this stage, as leadership support can significantly influence the success of AI initiatives.

Identifying business goals is another key step. Focus on use cases that offer substantial value and align with your organization's strategic objectives, such

as cost reduction, revenue generation, or improving customer experience. It's important to avoid one-off projects or those with limited value, as these can divert resources from more impactful initiatives.

Next, assess the feasibility and technical capability of each use case. Determine whether the use case is technically viable and whether your organization has the necessary data, talent, and infrastructure to support it.

Finally, define clear metrics and criteria for success, and establish ways to implement measurement mechanisms. If you're struggling to define what success looks like for a particular use case, it's a sign that it may not be a good fit for your initial AI efforts. By following these guidelines, you can ensure that your AI adoption is both strategic and effective, setting the stage for long-term success.

Explore how other organizations are putting Microsoft AI to work for them



Johnson & Johnson transforms business through intelligent automation

[Learn more](#)



Lionbridge disrupts localization industry using Azure OpenAI Service and reduces turnaround times by up to 30%

[Learn more](#)



Volvo Group streamlines invoice and claims processing with Azure AI and AI Document Intelligence

[Learn more](#)



INSPIRE Environmental uses Azure AI to improve efficiency and streamline data analysis to meet the pace of offshore wind development

[Learn more](#)



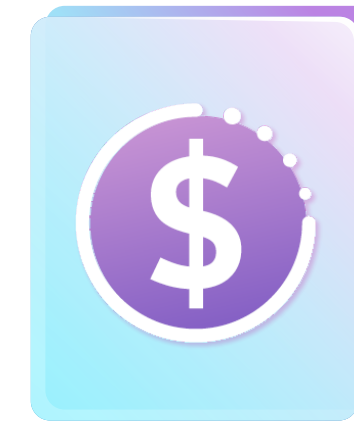
John Hancock supports call center teams with Microsoft conversational AI tools

[Learn more](#)



Mercy lays the foundation for a future of innovation in patient care using Microsoft Azure to unlock insights from data in the cloud

[Learn more](#)



Secure funding

Securing funding for AI projects is pivotal for organizations aiming to leverage advanced technologies effectively. The choice between optimizing current spending and seeking new budget approvals reflects strategic decisions that can significantly impact the success and scalability of AI initiatives. In most cases, a combination of both approaches ensures that organizations can fund AI projects effectively. By balancing cost optimization with strategic investment, businesses can unlock the transformative potential of AI, and drive innovation, efficiency, and sustainable growth in an increasingly competitive landscape.



Optimizing current spending



Funding AI projects



New budget approvals



Optimizing current spending

In FinOps, decisions are driven by the business value of the cloud. Implementing FinOps best practices may represent a proactive approach to funding AI projects without requiring additional budget allocations. FinOps focuses on enhancing accountability and visibility into costs across the organization. By adopting this framework, businesses undergo organizational and cultural shifts that prioritize cost efficiency and financial discipline. This often leads to identifying redundant expenses, optimizing cloud usage, and negotiating better pricing. As a result, significant savings might be realized, giving organizations the option to reallocate freed resources to AI projects. This approach not only minimizes the need for new funds but also ensures that existing financial resources are utilized optimally, enhancing overall operational efficiency and competitiveness.

New budget approvals

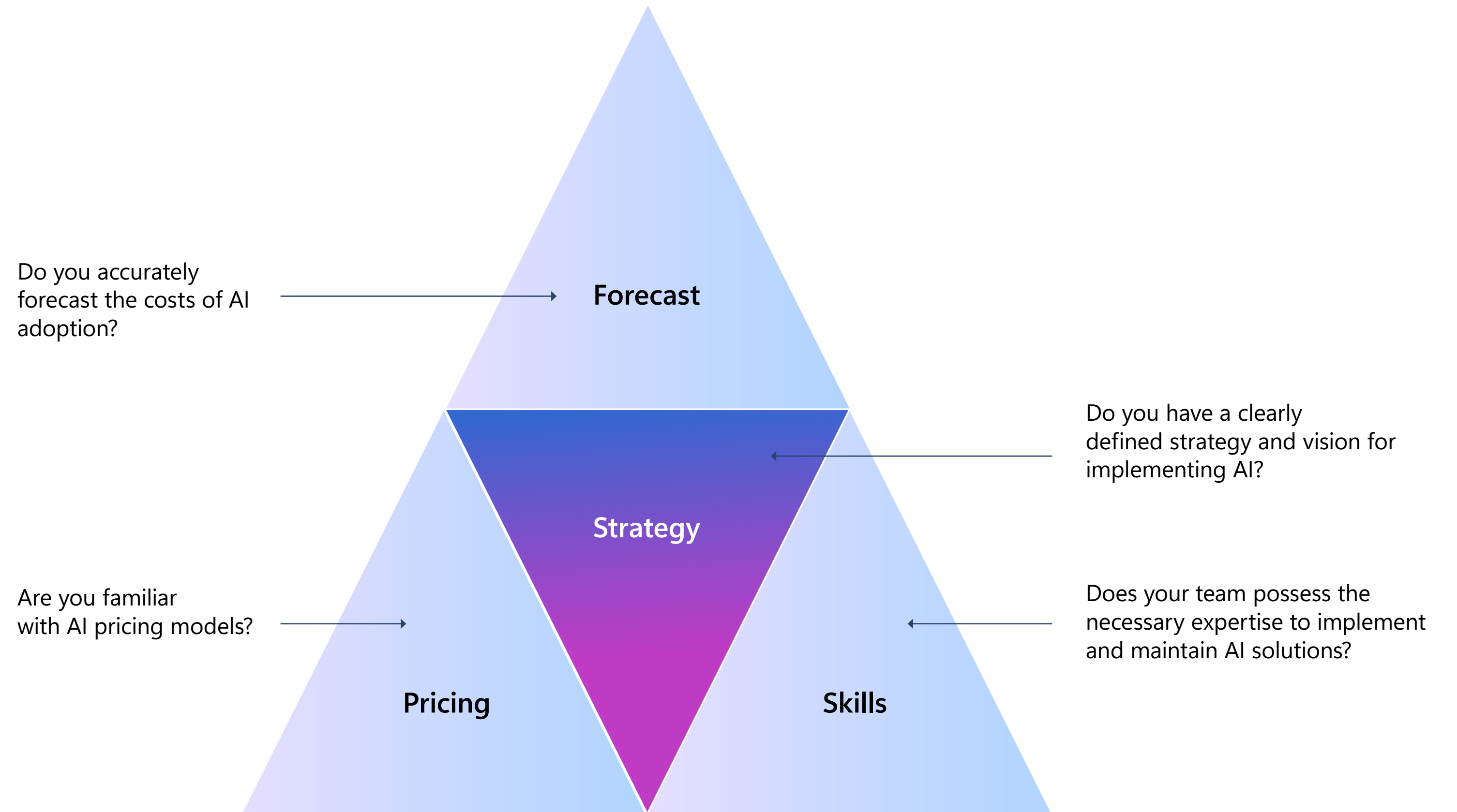
While optimizing current spending is crucial, certain AI initiatives may demand new budget allocations to achieve their full potential. In such cases, presenting a robust business case becomes imperative. A compelling business case outlines the strategic importance of AI projects, emphasizing their potential return on investment and alignment with organizational goals. It provides stakeholders with a clear understanding of how AI technologies can drive innovation, improve operational efficiencies, enhance customer experiences, or capture new market opportunities. Moreover, demonstrating the scalability and long-term benefits of AI implementations can bolster confidence in new budget approvals. This strategic approach not only secures the necessary funding but also aligns financial resources with the organization's broader strategic objectives, ensuring sustained growth and competitive advantage in the market.





Organizational readiness

Establishing a solid foundation for AI initiatives is crucial not only for preventing unnecessary waste but also for ensuring that your investment yields a healthy and efficient return. As organizations increasingly turn to AI to drive innovation and maintain a competitive edge, the importance of laying the right groundwork cannot be overstated. Without a strong foundation, AI projects are at risk of falling short of their potential, leading to suboptimal performance, unforeseen costs, and missed opportunities. By carefully planning and aligning AI efforts with broader business objectives, organizations can fully harness the transformative power of AI, achieving sustainable growth and maximizing the value of their investments. To build this foundation, it's important to carefully consider the following key aspects:



Strategy

Begin by defining a clear and measurable objective for your AI initiative. This objective should align closely with your overall business goals to ensure that the AI project contributes directly to your organization's success. Additionally, consider how these objectives will guide the selection of AI use cases and the allocation of resources to ensure the initiative's success.

Pricing models

It's critical to understand how to purchase AI services, and to then estimate, build budgets, and optimize those once deployed. Understanding the differences between standard (pay-per-token) or provisioned (provisioned throughput units) models is foundational to make pricing decisions with confidence and for a successful AI adoption on Azure: Get started by: 1. **Evaluate what Azure OpenAI Service models are best for you** 2. **Determine your requirements and deployment type** 3. **Learn about Provisioned Throughput Units (PTUs).**

Cost forecasting

Accurately forecasting the total cost of AI adoption is critical. This includes not only the initial investment but also ongoing costs. A comprehensive cost analysis will help you avoid unexpected expenses and ensure a smoother implementation. Get started with the **Azure pricing calculator** to turn anticipated usage into estimated costs for planning and budgeting purposes and the **capacity calculator** to estimate value of PTU units required for the provided workload inputs.

Skill development

To effectively manage and leverage AI investments, your team must be equipped with the necessary skills. This may involve training current employees, hiring new talent, or partnering with external experts. Ensuring your organization has the right expertise in place will maximize the potential of your AI initiatives and drive successful outcomes. Get started with **Azure AI fundamentals** and **Overview FinOps best practices**.



More resources

Find resources

Azure Essentials brings together all Microsoft's best practices, guidance, training, products, and assessments in one place.

[Learn more](#)

Create business case

Strengthen your business case by examining how organizations across various industries and sizes are successfully leveraging Microsoft AI.

[Explore](#)

Secure funding

Optimize current spending by accelerating efficiency through Microsoft solutions that support FinOps best practices.

[Learn more](#)

Establish a solid foundation

Transform your business with Microsoft AI. In this learning path, you will find the knowledge and resources to adopt AI in your organization.

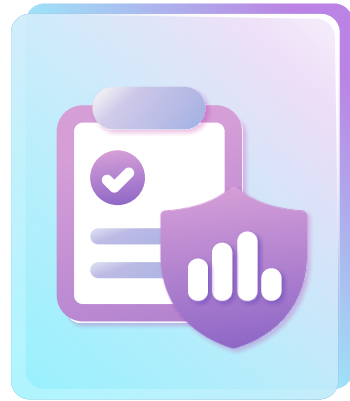
[Get started](#)



Build GenAI solutions efficiently

In this phase, it is essential to follow best practices and establish comprehensive governance policies to guide the design, development, and deployment of AI solutions. Doing so will support long-term business goals, ensuring investments are well-managed, cost-effective, and aligned with organizational objectives.





Establish governance policies

Governance policies provide a structured framework for managing AI investments in alignment with strategic objectives. By clearly defining roles, responsibilities, and decision-making processes, these policies help ensure that AI initiatives are prioritized based on their potential to deliver measurable business value. This strategic alignment helps in optimizing resource allocation, directing financial investments towards projects that promise the highest return on investment, and avoiding wasteful spending on initiatives that do not align with organizational goals.



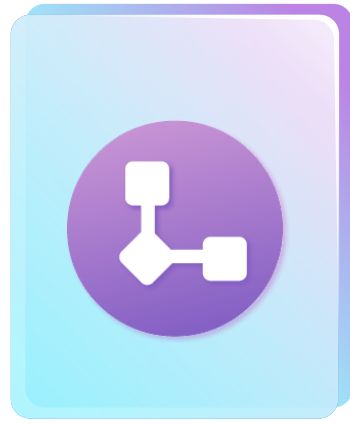
By establishing governance frameworks that include risk assessment, mitigation strategies, and compliance protocols, organizations can proactively address these challenges. This helps safeguard financial and enhances trust among stakeholders, including customers, regulators, and investors.

Setting policies

In Azure, governance policies provide a structured framework for managing investments in line with your strategic objectives. Using Azure Policy, you can define and enforce specific rules for your AI resources, ensuring that they align with your organization's goals.

[Learn more](#)

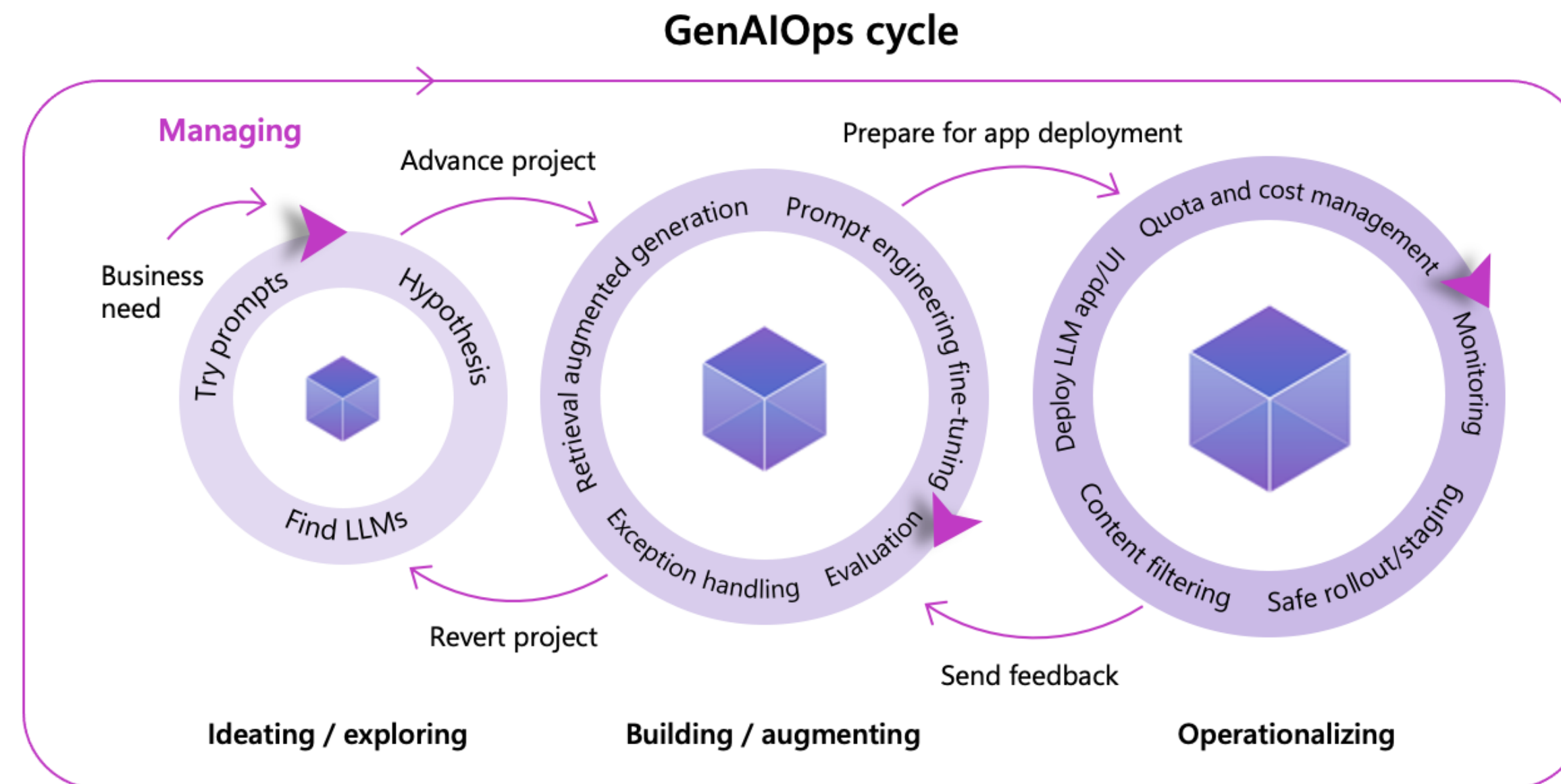
Governance policies mitigate financial risks associated with AI implementations. AI technologies often involve complex data handling, algorithmic decision-making, and integration with existing systems, all of which pose potential risks such as data breaches, regulatory non-compliance, and operational disruptions. To effectively manage these risks, it's crucial to identify, prioritize, and document them, ensuring a consistent methodology across the organization. [Learn more](#)



Efficiently design and deploy AI solutions

Efficiently designing, building, and deploying AI solutions is paramount from a financial perspective due to several critical factors that directly impact cost-effectiveness and return on investment.

A streamlined development process ensures that AI projects are completed on time and within budget. Time-to-market is crucial in gaining competitive advantages and delays can result in increased costs and missed opportunities. By leveraging prebuilt models and adopting best practices, organizations can accelerate the delivery of AI solutions while improving cost predictability. This not only minimizes operational expenses but also allows businesses to start realizing the benefits of AI sooner, enhancing overall ROI. Many organizations are adopting the GenAIOps framework to achieve this objective.



What is GenAIOps?

GenAIOps encompasses a comprehensive set of practices, tools, foundation models, and frameworks designed to integrate people, processes, and platforms to manage the lifecycle of generative AI models effectively.

[Learn more](#)

Moreover, efficient design and development practices help in optimizing resource utilization. This includes effectively managing human resources, infrastructure, and tools throughout the AI project lifecycle. By aligning resource allocation with project requirements and milestones, organizations can prevent underutilization or overprovisioning, thereby controlling unnecessary costs. Additionally, efficient resource management ensures that teams remain focused on high-priority tasks, reducing idle time and maximizing productivity.

From a financial perspective, the benefits of efficient AI solution development extend beyond cost containment to encompass enhanced operational efficiency, improved customer satisfaction, and sustainable growth. By prioritizing efficiency throughout the AI project lifecycle, organizations can achieve strategic objectives, maximize ROI from technology investments, and establish a foundation for continued innovation and success in an increasingly digital world.

Leveraging prebuilt models

Azure AI includes a robust and growing catalog of frontier and open-source models from OpenAI, Microsoft, Meta, Mistral, Hugging Face, and more partners that can be applied over your data. You can even compare models by task using open-source datasets and evaluate the model with your own test data to see how the pretrained model would perform to fit your own use case.

[Learn more](#)



More resources

Responsible AI

Microsoft empowers responsible AI practices by its commitment to the advancement of AI driven by ethical principles.

[Learn more](#)

Establish governance policies

The Azure landing zone accelerator provides organizations with mature governance controls.

[Learn more](#)

Efficiently design, build, and deploy

Understanding Azure AI Pricing

Azure AI offers flexible pricing options tailored to meet your project needs. Whether you're using prebuilt models or training custom solutions, you can choose a pricing plan that aligns with your usage and budget. Explore different pricing tiers, based on factors like data processing volume, compute power, and deployment scale.

Ensure you're getting the best value by selecting the right plan for your AI solution. [Explore Azure AI Pricing Plans.](#)

Microsoft Well-Architected Framework

Architectural recommendations to help you make informed decisions when you use Azure OpenAI as part of your workload's architecture.

[Learn more](#)



Manage AI investments

Managing and optimizing AI investments post-deployment is critical from a financial perspective as it helps ensure ongoing efficiency, cost-effectiveness, and maximization of return on investment over the long term.





Quantify AI business value

The first step is to accurately quantify the **unit economics** to understand the business value of AI and its impact across various dimensions of the organization. This may involve examining how AI influences: business processes, customer experiences and revenue generation.



Aspects

Business processes

AI can streamline and automate operational workflows, enhancing efficiency and reducing manual labor. By analyzing how AI optimizes processes, such as supply chain management or customer service operations, organizations can identify cost savings and productivity gains.

Customer experiences

AI tools, such as chatbots and recommendation engines, significantly enhance customer interactions by providing personalized and timely responses. Understanding how AI improves customer satisfaction and engagement can help businesses refine their strategies, foster loyalty, and potentially increase revenue.

Revenue generation

AI can drive revenue growth through predictive analytics, targeted marketing, and new product innovations. By evaluating how AI contributes to increased sales and market expansion, organizations can better justify investments and allocate resources to high-impact areas.

By gaining insights into these aspects, leaders can align AI investments with strategic priorities, ensuring that resources are focused on initiatives with the greatest potential for return. This alignment not only helps in anticipating future needs but also encourages proactive innovation and adaptation.

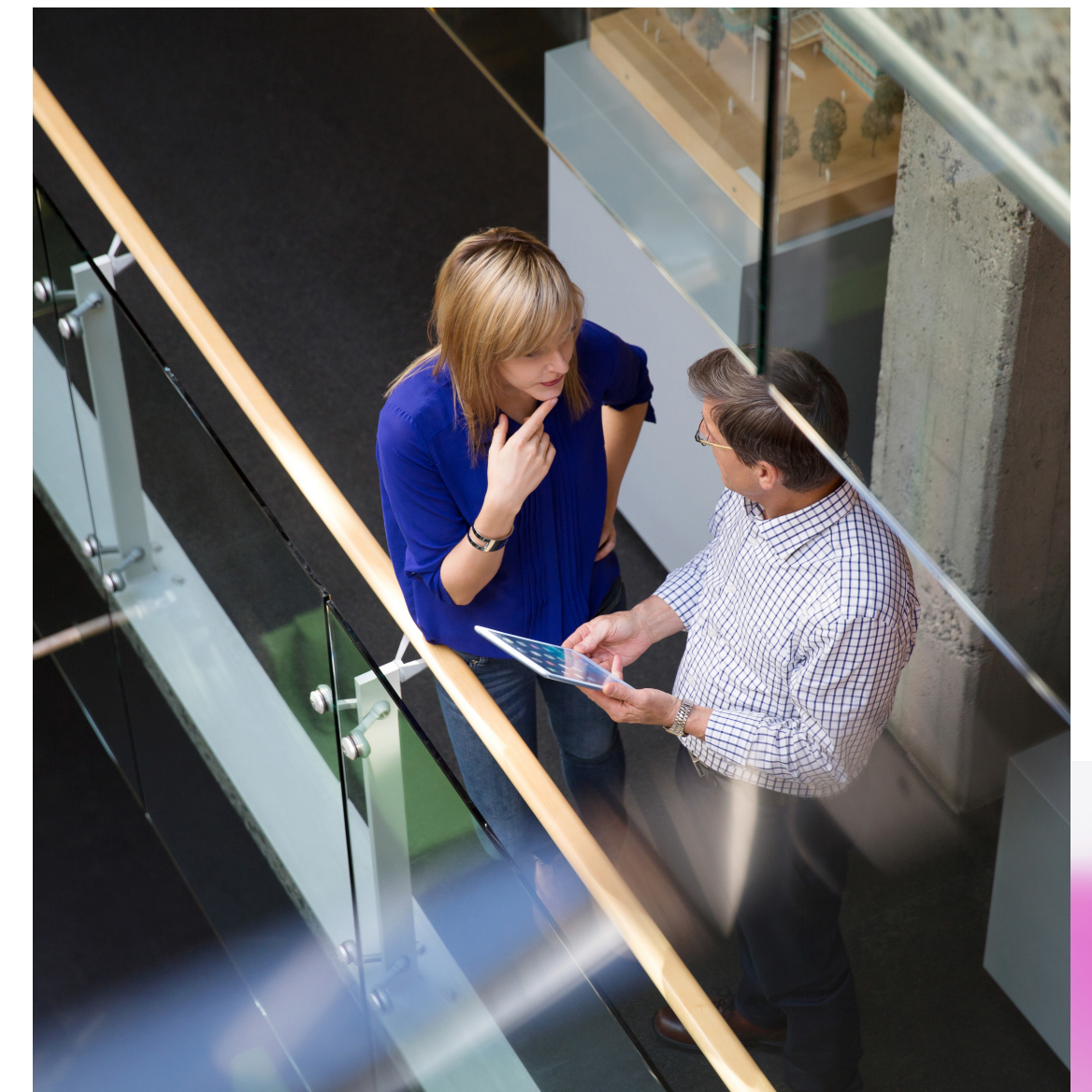
Such foresight empowers organizations to stay ahead of competitors by leveraging AI to fuel growth initiatives and optimize operations. Measuring and demonstrating the tangible benefits of AI enhances strategic decision-making, improves overall efficiency, and reinforces the organization's position as a leader in its industry.

Research

What can Copilot's earliest users teach us about generative AI at work?

A first look at the impact on productivity, creativity, and time.

[Learn more](#)





Monitor usage and patterns

After deploying AI solutions, it's essential to continuously monitor usage and analyze patterns to gain deep insights into how effectively the technology is being utilized across the organization. This process involves tracking and evaluating various metrics which are useful for assessing resource utilization and overall system performance.

Key performance indicators that might be monitored include:

<p>Pay-as-you-go vs. commitment discounts</p> <p>Analyzing the cost-effectiveness of different pricing models, such as standard versus provisioned deployments and reservations, can provide insights into financial efficiency.</p>	<p>Resource utilization</p> <p>Tracking how resources are used can identify underutilized assets or potential bottlenecks. By optimizing resource allocation, organizations can reduce waste and ensure that AI workloads are running efficiently.</p>	<p>Identifying inefficiencies</p> <p>Monitoring for inefficiencies, such as redundant processes, idle resources, or suboptimal data flows, can highlight areas where improvements can be made.</p>	<p>Usage trends</p> <p>Understanding how AI tools and applications are being used over time allows organizations to identify trends, such as increased demand during certain periods or shifts in user behavior.</p>
---	---	---	---

By closely analyzing these metrics and usage patterns, organizations can uncover opportunities to optimize resource allocation, streamline workflows, and enhance operational efficiency. This proactive approach is key to maximizing the value derived from AI investments. When resources are aligned with actual needs and business goals, AI initiatives are more likely to deliver significant returns and drive meaningful business outcomes.

Moreover, integrating these insights with broader business strategies helps ensure that AI deployments are not only cost-effective but also strategically aligned. By continuously refining AI operations based on real-time data and KPIs, organizations can stay agile and respond to changing demands to maintain a competitive edge in their industry. Ultimately, this approach helps to ensure that AI investments contribute to long-term growth, innovation, and success.

Understanding workloads can help you to identify potential cost saving opportunities. For development, testing, and lower-volume workloads, the standard pay-as-you-go pricing model may be a good option. For production and high-volume workloads that need guaranteed latency, the provisioned (PTU) pricing model provides a consistent cost structure and scalable throughput. Users can also **save up to 85%*** with Azure OpenAI service provisioned reservations.

*The 85% savings are based on the Provisioned Throughput Hourly rate of approximately \$2/hour, compared to the reduced rate of a 1-year reservation at approximately \$0.3028/hour. Azure pricing as of July 29, 2024 (prices subject to change). Actual savings may vary depending on the specific Azure OpenAI model and region availability.



Capitalize on efficiency opportunities

Identifying inefficiencies or bottlenecks in AI operations is just the beginning. The real value comes from capitalizing on these opportunities to enhance efficiency and drive growth. Once inefficiencies are spotted, the next step is to implement strategies that not only streamline operations but also maximize the return on your AI investment.

One effective approach is to shift from a standard offer to provisioned offer. By committing to a certain level of usage, your organization can secure discounts that significantly reduce costs. These savings can then be reallocated to other high-impact projects, enabling further innovation and expansion. However, it's essential to approach this strategy with a growth-oriented mindset rather than focusing solely on cost reduction.

While cost reduction is an important outcome of efficiency improvements, the primary objective should be to drive growth by eliminating waste and optimizing resource utilization. This means that any efficiency gains should be closely aligned with broader business goals.



In addition to financial strategies, organizations should also consider process optimizations, such as automating repetitive tasks, improving data management practices, or refining AI model training processes. These optimizations may help reduce operational costs and improve the speed and quality of AI-driven outcomes, contributing directly to business growth.

By aligning efficiency improvements with strategic business objectives, organizations can create a virtuous cycle where cost savings are reinvested into growth-driving initiatives. This approach

ensures that AI investments continue to deliver increasing value over time, supporting both immediate operational goals and long-term business success.

Ultimately, the goal is to create a lean, agile operation that maximizes the potential of AI technologies while continually pushing the organization towards greater innovation and market leadership. When efficiency improvements are strategically aligned with business goals, they become a powerful lever for growth, enabling the organization to scale and compete more effectively in an ever-evolving market landscape.

More resources

Find resources

Azure Essentials brings together all Microsoft's best practices, guidance, training, products, and assessments in one place.

[Learn more](#)

Quantify AI business value

Understand the Quantify Business Value domain within the FinOps Framework and how to implement its capabilities in the Microsoft Cloud.

[Learn more](#)

Monitoring usage and patterns

Leverage Microsoft Cost Management to monitor costs for Azure OpenAI Services.

[Learn more](#)

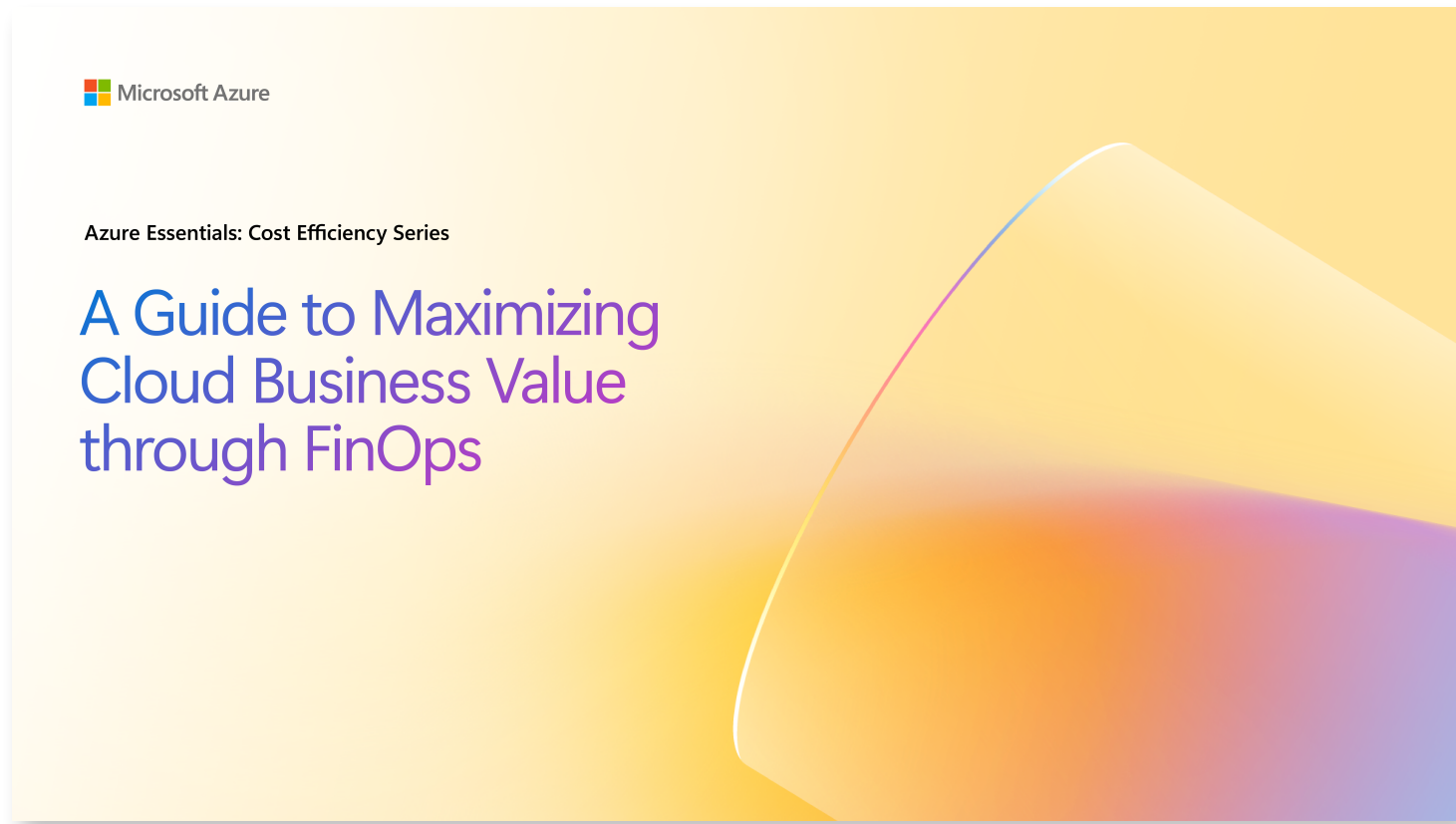
Capitalizing on efficiency opportunities

Save costs with Microsoft Azure OpenAI Service Provisioned Reservations.

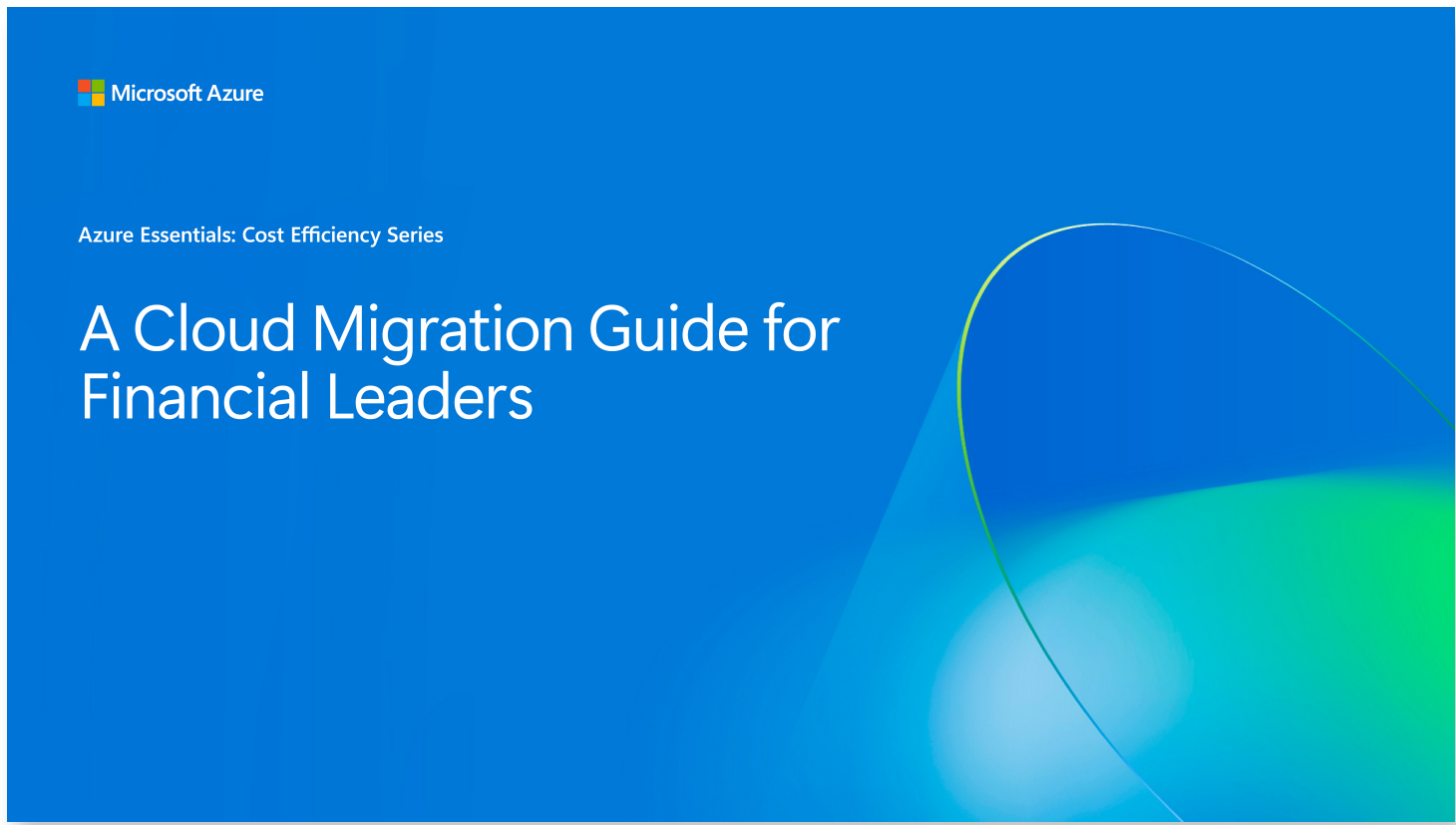
[Learn more](#)

Azure Essentials: Cost Efficiency Series

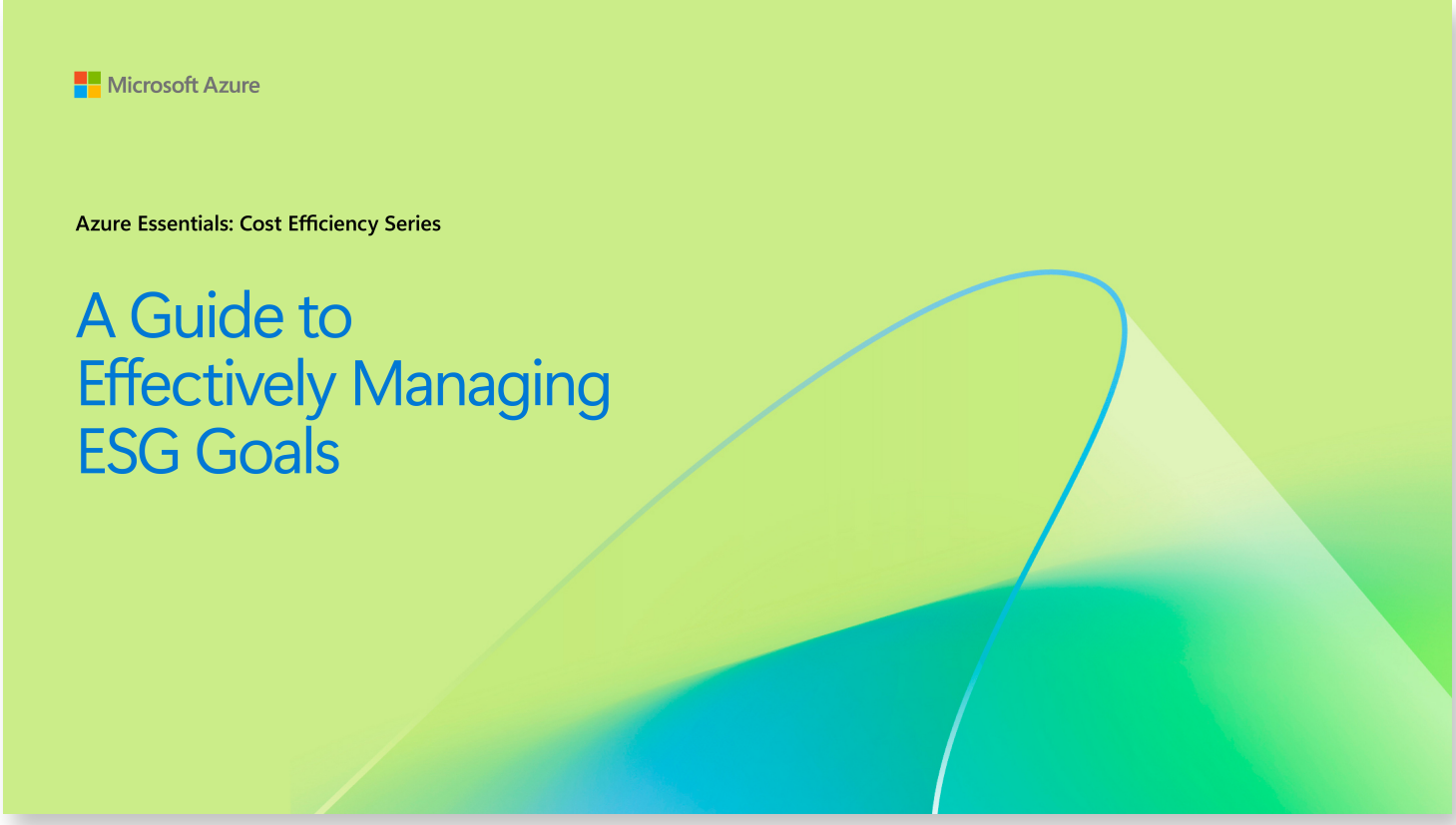
This series of e-books explores key financial topics to help organizations prepare for and enhance efficiency in their cloud journey. Each e-book provides actionable insights and guidance from a financial perspective to empower decision-making to optimize resources and maximize investments. Here are the e-books included in this series:



[Read more](#)

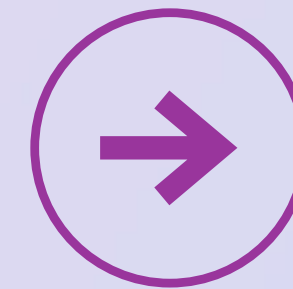


[Read more](#)



[Read more](#)

Azure Essentials



Learn more and get started at
aka.ms/AzureEssentials

Azure Essentials helps elevate the reliability, security, and ongoing performance of your cloud and AI investments to enable you to maximize the value of your cloud and empower your digital transformation.

Azure Essentials brings together all Microsoft's best practices, guidance, training, products, and assessments in one place. It connects you to our suite of pricing offers to help you manage your investment effectively. It can also help you connect to experts to help you implement your business priorities across a range of scenarios for new or existing deployments.

Learn more about **Azure Innovate and Azure Migrate and Modernize and Azure Essentials** to understand how they can help you achieve secure migrations.

©2025 Microsoft Corporation. All rights reserved.

Disclaimer

This document is provided "as-is." Information and views expressed in this document, including URL and other Internet Web site references, may change without notice. You bear the risk of using it. Examples herein may be for illustration only and if so are fictitious. No real association is intended or inferred. This document does not provide you with any legal rights to any intellectual property in any Microsoft product. You may copy and use this document for your internal, reference purposes. This document is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY. Some examples are for illustration only and are fictitious. No real association is intended or inferred.

Microsoft assumes no responsibility for any errors, omissions, or inaccuracies in this document or for any actions taken by the reader based on the information provided. We also do not endorse or guarantee the products or services of the FinOps Foundation. The views and opinions expressed in this document are those of Microsoft employees or the FinOps Foundation and do not necessarily reflect the views or opinions of our company. The information provided in this document should not be considered as legal, financial, or professional advice, and should not be relied upon as such.