# Adopt AI with Confidence:
# 5 Strategies for Managing Generative AI Cost

By Andrew Bush (research@trustbolt.ai)

TrustBolt

Adopt AI with Confidence

www.trustbolt.ai

Trust Bolt

Adopt AI with Confidence

www.trustbolt.ai

# Adopt AI with Confidence: 5 Strategies for Managing AI Costs

*By Andrew Bush, TrustBolt.ai Email: research@trustbolt.ai*
*Research for this report was conducted with the assistance of various AI services, including Perplexity, ChatGPT, and Copilot, and was edited using Grammarly.*


## 1 Executive Summary

**Generative AI is more than just "another technology tool" and has the potential to change many aspects of business strategy and how employees bring value to companies. Enterprises are increasingly learning to integrate this latest technology into their automation, decision-making, and content generation processes.**

**A well-structured and thoughtful cost-management strategy is critical to enabling companies to continue increasing their adoption with confidence, which in turn leads to further innovation.**

**This paper identifies five common AI cost drivers and approaches to managing these costs effectively.**

## 2 Understanding Cost Drivers

To effectively control costs, enterprises must first understand the five key drivers contributing to the total ownership cost of third-party generative AI platforms. This knowledge will empower you to make informed decisions and implement cost-effective strategies, ensuring you are prepared for the challenges ahead.

1. **Subscription and Licensing Fees** – The amount that 3rd party AI vendors invoice you each period. Costs often vary based on several factors, including
    1. Subscription Model - i.e., Usage-based or User Based
    2. Volumes - i.e., companies with larger contracts and commitments can often negotiate tiered pricing structures based on expected volumes
    3. Prompt/Response Complexity - The more complicated the prompt and the more in-depth the resulting answer, the higher the costs.
    4. Model Complexity - Providers continually deliver new and improved models at a rapid pace of development. With these new models come increased capabilities and often higher prices.
2. **Data Storage and Processing** – Additional expenses arise from storing, securing, and managing AI-generated content.
3. **Compliance and Security Investments** – Necessary to meet regulatory standards, protect intellectual property, and safeguard user privacy.

4. **Infrastructure and Integration Costs** – Expenses associated with integrating AI solutions into existing enterprise systems and workflows.
5. **Scalability and Performance Optimization** – Scaling AI usage across business units may introduce additional computational costs and performance considerations.

# 3 Cost Management Strategies

After identifying some of the key drivers behind the cost of adapting AI, enterprises can develop strategies to utilize these capabilities effectively and efficiently, ultimately controlling costs.

## 3.1 Apply AI Thoughtfully

- Define clear AI use cases aligned with strategic goals to ensure targeted, cost-effective deployment.
- Prioritize AI investments in high-value, high-volume areas. Examples include customer service automation, fraud detection, market research, and real-time business intelligence.
- Establish governance and monitoring programs to assess the long-term cost-effectiveness and sustainability of AI use is essential to promoting responsible deployment.
- Promote transparency throughout the AI adoption decision-making processes to build trust among employees, stakeholders, and customers. Employees are more likely to adopt AI when they are part of the change.
- Implement role-based access controls to limit AI usage to essential personnel and business units based on approved use cases.
- Develop guidelines for AI utilization to avoid over-reliance on expensive cloud-based generative AI models for non-essential tasks.
- Avoid using AI where it is not needed. Apply AI to problems that it can uniquely solve. For instance, use AI for complex data analysis, but for routine administrative tasks, traditional solutions may suffice. Eliminate first, then automate!

## 3.2 Negotiate Enterprise Agreements

- Engage in volume-based pricing discussions with AI vendors to secure more favorable contracts.
- Leverage enterprise-tier solutions that offer predictable and scalable pricing models.
- Explore hybrid approaches, such as integrating open-source AI models for non-critical tasks to reduce reliance on high-cost proprietary platforms.
- Conduct periodic contract reviews to ensure cost-effective licensing and avoid underutilized service subscriptions.

### 3.3 Monitor and Control AI Expenditure

- Establish a governance framework for AI usage, budget allocation, and cost tracking.
- Deploy AI usage dashboards to monitor real-time expenses, detect inefficiencies, and identify cost-saving opportunities.
- Set spending thresholds, enforce approval workflows for high-cost AI functions, and implement automated alerts for API calls that exceed budgeted limits.
- Provide transparency of costs back to individuals, teams, and functions leveraging AI services.
- Conduct periodic cost audits to assess AI efficiency and align spending with business goals.
- Leverage lower-cost models for more straightforward, lower-value activities.
- Deploy internally hosted solutions for general en-mass rollouts.

### 3.4 Optimize AI Through Fine-Tuning and Hybrid Approaches

- Train AI models on proprietary company data to improve relevancy, accuracy, and efficiency while reducing excessive API usage.
- For cost-sensitive applications, utilize a combination of third-party AI and internally hosted models, striking a balance between performance and affordability.
- Deploy smaller, fine-tuned AI models for domain-specific tasks, rather than relying solely on general-purpose AI systems.
- Evaluate the cost-benefit ratio of on-demand AI services versus pre-trained models for recurring enterprise use cases.
- Leverage prompt-catching mechanisms for the frequency used or similar prompts and responses to reduce redundant queries.
- Introduce prompt batching to group multiple queries into a single request, especially where multiple responses relate to the same large data sets.
- Invest in training and prompt engineering to ensure efficient query structuring to minimize redundant calls.

### 3.5 Address Compliance and Security Efficiently

- Adopt AI solutions with built-in compliance and security features to minimize regulatory overhead.
- Store and process sensitive data within your company boundaries, while leveraging external AI selectively for less sensitive operations.
- Implement privacy-preserving AI techniques such as differential privacy, encryption, and secure multi-party computation to mitigate security risks.
- Conduct regular AI risk assessments to ensure that cost savings do not override data security and compliance obligations.

## 3.6 Conclusion

By implementing structured cost-management strategies, enterprises can mitigate excessive AI expenses while harnessing the full potential of generative AI platforms. Through usage optimization, governance frameworks, and strategic investments in AI efficiency, businesses can maximize ROI while maintaining financial control over AI adoption.

To learn more about how TrustBolt can help you deploy AI with confidence, visit us at https://www.trustbolt.ai/learn.