California Management Review

INSIGHT

Artificial Intelligence

Financial Planning for Agentic & AI Systems: Managing Volatility in the Age of Autonomy

Ankit Chopra

Image Credit | WrightStudio

A clearer managerial lens for planning in AI and agentic systems

INSIGHT | FRONTIER 17 Feb 2026

Introduction: Planning in the Age of Autonomy

As artificial intelligence becomes more deeply embedded in the enterprise, the craft of planning and forecasting will enter into a new era. Agentic AI systems, capable of learning and acting autonomously, are changing not only how work gets done but also how costs and revenues behave. Every interaction between an AI agent and a user generates a series of financial consequences like token consumption, compute costs, and potential value creation. This differs fundamentally from traditional business operations, where resource utilization and outcomes could be forecasted in relatively stable patterns.

RELATED ARTICLES

Tobias Straube, Olaf J. Groth & Jöran Altenberg, "**AI Initiatives Don't Fail - Organizations Do: Why Companies Need AI Experimentation Sandboxes and Pathways,**" California Management Review Insight, May 2025.

Rebecka C. Ångström et. al, "**Getting AI Implementation Right: Insights From a Global Survey**," California Management Review, vol. 66, no. 1 (Aug. 30, 2023): 5–22.

Oliver Müller Konstantin Hopf et. al, "**Organizational Implementation of AI: Craft and Mechanical Work**," California Management Review, vol. 66, no. 1 (Sept. 19, 2023): 23–47.

RELATED TOPICS

Agentic AI
Finance
Data Analysis
Predictive Modeling

Studies by a16z and Epoch AI show that the unit cost of inference in large-language models has declined by roughly tenfold each year since 2021, while global AI adoption has accelerated at an even faster pace[1, 2]. Cloud infrastructure providers continue to release periodic price reductions on compute and storage capacity, yet the overall spend by enterprises often rises as workloads scale.[3] This environment has made cost forecasting and value measurement increasingly complex, requiring finance and planning teams to evolve from static, control-oriented processes towards agile and intelligent systems capable of learning and responding in real time.

The Business Need: From Control to Orchestration

Traditional budgeting systems rely on a clear separation between planning and execution. Finance teams build annual forecasts, allocate budgets, and monitor variance. In agentic AI environments, this rhythm breaks because operating costs and value creation happen continuously and often unpredictably. Each AI task consumes compute and generates outcomes at a variable pace, introducing volatility into both expense and revenue streams.

The role of finance extends beyond monitoring expenditure; it involves interpreting behavioral signals embedded in usage patterns, customer interactions, and model performance. Planning becomes an orchestration exercise, continuously aligning learning systems, data pipelines, and financial models. The ultimate challenge is to design processes that preserve discipline and accountability while remaining flexible enough to adapt to evolving business contexts. In practice, the quality of any forecast depends not on algorithmic sophistication alone but on how accurately it represents the underlying business dynamics: how the product generates value, how customers engage with it, and how technology evolves.

The Financial Planning Engine: From Usage to ROI

Building a financial model for AI systems requires integrating usage, cost, pricing, and ROI into a single decision engine. Usage signals demand and system behavior, cost translates it into economics, pricing defines monetization, and ROI measures value creation. Together, these form the financial engine of intelligent planning, which links operations to outcomes and enables continuous adaptation as data, efficiency, and business conditions evolve.

Usage as the Backbone of Planning

In agentic AI systems, usage is the most direct representation of both demand and cost. Each token processed or model invocation translates into measurable resource consumption. For finance and planning teams, this means usage becomes the new backbone metric that connects technical operations to financial performance. Forecasts are only as credible as the visibility that exists between operations and outcomes. To understand why a certain increase in token volume leads to higher revenue in one product but not another, one needs to trace the chain of causality which includes data input, model decision, customer interaction, and financial return. This value-chain transparency converts raw usage data into business insight and reduces forecast error. It also emphasizes that forecasting is never just a mathematical problem; it is a strategic interpretation of how value actually flows through the business model. The most effective planning environments maintain an ongoing dialogue between finance, product, and customer teams to ensure that changes in usage reflect genuine drivers of performance rather than temporary fluctuations.

A growing body of research has shown that the success or failure of AI initiatives depends less on the technology itself and more on how organizations interpret and apply the value it creates.[5,6,7] A recent Gartner's study predicts that 40% of Agentic AI projects will fail to move beyond pilots, not because of model accuracy or engineering challenges, but due to poor alignment between technical outcomes and business objectives.[4] For executive and financial planning teams, these findings underscore that understanding value movement

across the AI value chain is the most important determinant of forecast reliability and strategic return. Causality within this chain does not flow linearly from token usage to financial impact. It begins with *adoption and perceived value*, where users decide that AI delivers meaningful assistance or efficiency. This adoption drives *agentic interaction or usage intensity*, which in turn produces measurable outcomes such as time savings, improved accuracy, or customer satisfaction. Those operational gains translate into *financial results* like revenue growth, cost optimization, or margin improvement. The cycle completes as financial outcomes feed back into business learning, informing future investments, pricing models, and feature prioritization.

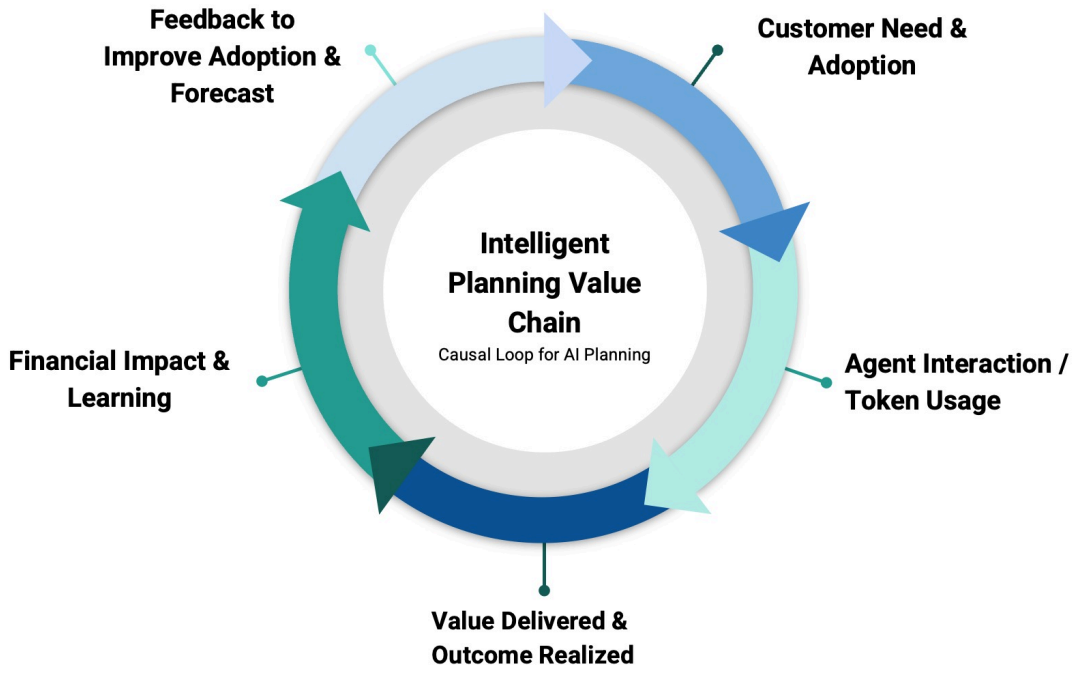

Figure 1. Intelligent Planning Value Chain / Causal Loop for AI Planning

When finance teams map this entire loop, they can see where value accelerates and where it dissipates. For example, a spike in token usage may reflect either healthy engagement or inefficient behavior depending on whether it aligns with positive outcomes in the customer or productivity layer. By establishing visibility across these links, managers can identify the true drivers of return, model adoption-to-value conversion rates, and allocate capital to the segments that amplify financial performance.[9] Ultimately, the lesson is that forecasting in agentic AI systems is not about predicting computational activity but about

tracing how usage moves through the organization. In other words, it is about tracing learning, adoption, and value creation and their interactions to produce measurable business impact.

Cost Modeling: Understanding Trends & Context

Forecasting cost in agentic AI systems requires understanding multiple moving curves rather than assuming a single trajectory. Token prices, compute rates, and model efficiency can all shift independently, influenced by competitive pricing, partnerships, and model optimization. Historical patterns from cloud computing suggest that costs may continue to decline in stepwise fashion, similar to technology learning curves observed in prior decades, but these declines are neither uniform nor guaranteed.

Across both compute infrastructure and model operations, costs are showing sustained deflationary trends. The price of cloud compute has steadily declined with improved hardware efficiency, dynamic instance pricing, and better workload optimization, while the cost per token for large language models has dropped drastically in just a few generations. Together, these trends indicate a shift in AI economics, where unit costs fall even as aggregate spending grows, driven by broader adoption, expanding workloads, and rising model capability

Financial Planning teams should ideally model a range of possible cost scenarios rather than a single forecast line. At the lower bound might be aggressive efficiency gains driven by specialized technical advancement; at the upper bound, a stabilization or even temporary increase if demand spikes faster than supply or if higher-performance models introduce new pricing tiers and unlock new customer cohorts.

The key insight is that forecasting cost is inseparable from business context. For a company integrating AI into an internal process, the relevant cost driver may be compute efficiency and workload scheduling. For a company selling AI-enabled products, the primary factor may be external pricing strategies and customer adoption rates. In every case, the reliability of a forecast depends on how well it captures these contextual nuances.

By treating cost curves as conditional on strategy rather than as universal truths, finance teams can prepare for multiple futures, each with distinct implications for margin, capital allocation, and pricing flexibility.

Pricing Models in Practice

Pricing remains one of the most decisive variables in determining profitability and customer adoption. In practice, most organizations are not inventing entirely new approaches but adapting and combining proven structures.[8] The most common models fall into three families: flat-rate subscriptions, usage-based pricing, and hybrid systems that combine both.

Subscription pricing, familiar to most SaaS businesses, offers predictability and is attractive for customers with stable usage patterns. However, it can limit scalability and may under-monetize heavy users. Pure usage-based models align revenue directly with activity but can create volatility in cash flows and make customer budgeting difficult. Hybrid models balance the two by establishing a base subscription for predictable access while charging variable rates for incremental usage. A smaller but growing group of companies has begun to experiment with outcome-linked pricing, where the fee corresponds to a specific result (such as time saved or accuracy achieved) rather than the quantity of tokens consumed.

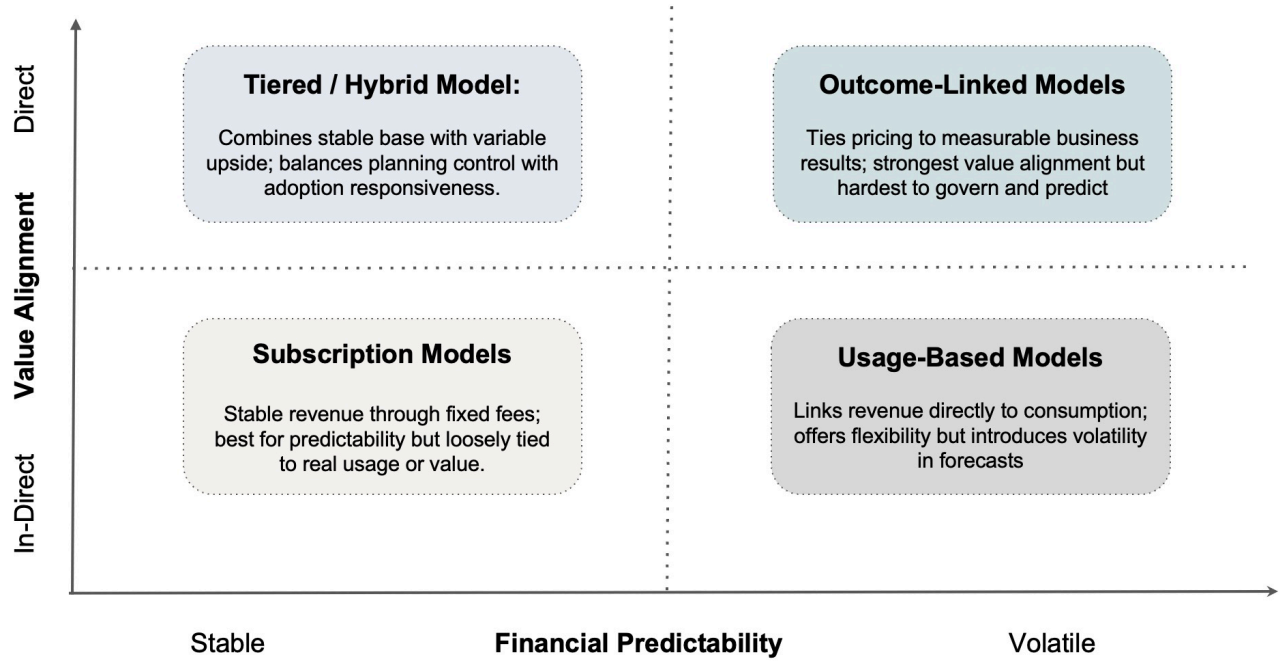

Figure 2. Pricing models vary in financial predictability & value alignment

Each approach carries trade-offs that must be reflected in the financial plan. Subscription models simplify forecasting but may obscure elasticity effects. Usage models provide clear demand signals but can amplify cost-revenue timing mismatches. Outcome-based structures better align price with perceived value but require strong measurement systems and carry greater performance risk. Finance leaders must choose or combine models that align with their business context, maturity, and customer expectations. The goal is not to find a universal formula but to construct a pricing architecture that keeps predictability and value realization in balance.

Risk-Adjusted ROI: Measuring Performance under Uncertainty

Return on investment remains the standard metric for evaluating projects. Traditional formulas, while very valuable, assume that costs and benefits are known and stable. In AI initiatives, both sides of the equation are fluid. Token costs may fall, user adoption may accelerate or stall, and regulatory or ethical considerations may change deployment speed. Backward-looking evaluation, using realized costs and returns, will still follow the

standard ROI formula of net gain divided by investment. However, forward-looking projections should recognize that outcomes lie along a continuum between the traditional calculation and a risk-adjusted interpretation.

The risk-adjusted approach introduces a discount factor that represents uncertainty arising from technical performance, data quality, or adoption behavior. Instead of producing a single ROI figure, finance can present a range bounded by the standard estimate at the upper limit and the risk-adjusted estimate at the lower. This provides decision-makers with a probabilistic view rather than an illusion of precision. In doing so, finance moves from a reporting mindset to a forecasting mindset, using ROI not as a scorecard but as a planning instrument that continuously recalibrates as more information becomes available. Risk-adjusted ROI serves as both a performance evaluation and a feedback loop, translating outcomes into learning signals that guide continual refinement and decision-making within the intelligent financial ecosystem.

Bringing It Together: Designing the Financial Model

The insights from usage, cost, pricing and ROI converge in the design of a financial model that serves both immediate planning needs and longer-term learning. In practice, this can be achieved through two complementary layers.

The Base Model can be built in familiar planning tools or spreadsheets, allowing finance and planning teams to link usage, unit costs, and pricing tiers to revenue and margin forecasts. Using time-series and driver-based methods, it provides a quick yet structured way to test assumptions, run sensitivities, and understand the financial impact of discrete operational changes.[10] This layer grounds decision-making in transparency and discipline.

The Advanced Model extends the same framework through machine-learning techniques (such as ensemble forecasting) that refine usage forecasts and update cost relationships as new data arrives. This layer creates a feedback loop, enabling the model to learn from actual results and improve its predictive accuracy over time.

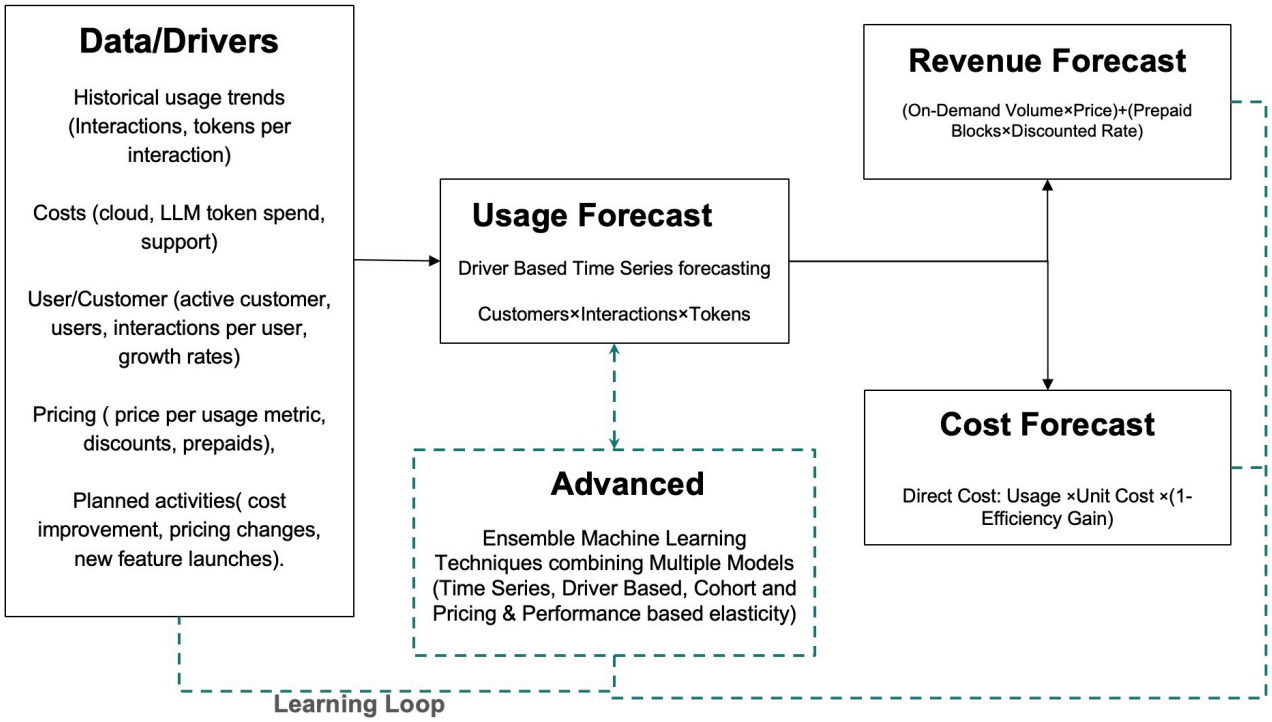

Figure 3. High level design of Financial Planning Model for Usage Based AI Systems

Together, these layers lay the foundations of an intelligent planning system, where models enable finance to bridge control and adaptability. The base model enforces control and clarity; the learning model adds foresight and adaptability. By connecting traditional planning discipline with continuous data-driven learning, organizations can move from static forecasting to intelligent decision orchestration, without losing sight of the most important anchor: the business context that drives value creation.

Managerial Implications and Evolution of Planning Maturity

The transformation of financial planning in AI-driven enterprises is as much organizational as it is analytical. Finance is moving from a backward-looking discipline focused on control to a forward-looking system that learns continuously from the environment it measures. This evolution from static forecasting to strategic orchestration defines the managerial frontier of AI-era finance.

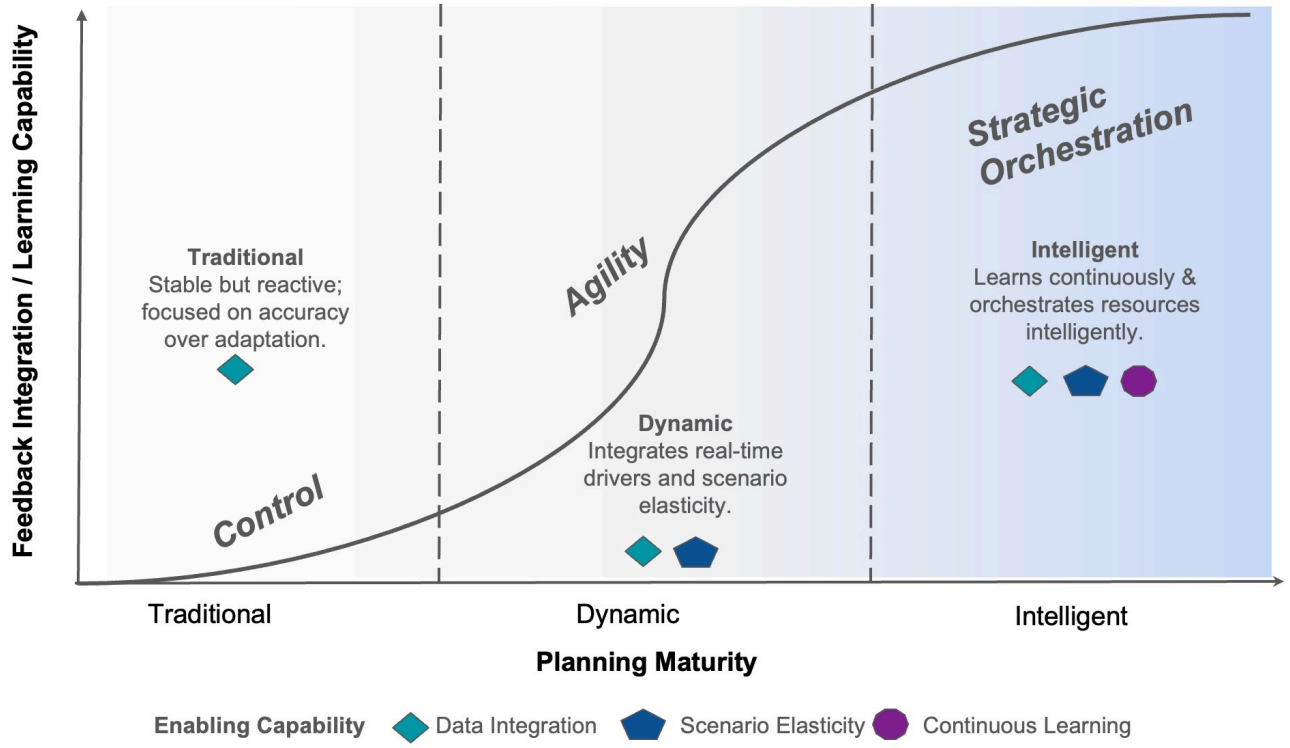

Figure 4. Evolution of Planning Maturity in AI Environments

In the first stage, organizations rely on traditional forecasting, characterized by manual inputs, fixed cycles, and retrospective analysis. Accuracy depends largely on human judgment and historical trends. As AI adoption deepens, finance teams begin integrating operational drivers such as token consumption, compute efficiency, and adoption metrics into driver-based models. Forecasts become dynamic, data pipelines shorten feedback loops, and scenario planning starts to resemble simulation. The most mature organizations would operate at the intelligent stage, where forecasting models themselves learn from outcomes, recalibrating assumptions automatically and orchestrating resources across digital and human dimensions.

Short-Term Considerations: Navigating Immediate Volatility

In the near term, executives and finance leaders must address three converging pressures. First, declining unit costs yet expanding total spend: token and compute prices are dropping, but total consumption is accelerating. This demands models that capture both deflationary and inflationary dynamics simultaneously.

Second, hybrid and usage-based pricing is reshaping revenue recognition. Forecasts must now integrate fixed subscription baselines with variable usage layers, requiring elasticity modeling rather than static budgeting.

Third, planning cadence is evolving. Quarterly cycles may no longer suffice when model parameters, APIs, and usage patterns change monthly. Continuous re-forecasting, supported by live data feeds and learning systems will be the next step in this evolution. There are benefits in taking the dual approach which combines the regular planning cycle and short term rolling forecasts.

At the same time, explainability and auditability are emerging as governance priorities. As AI expenditures scale, boards and investors expect finance to justify spend through transparent linkages between usage, cost, and value. Organizations that establish clear attribution frameworks will command greater credibility and investment agility.

Long-Term Considerations: Building a Learning Model for Planning

Over the longer horizon, financial planning becomes inseparable from organizational learning and strategic renewal. Several enduring shifts define this trajectory.

Finance as a Learning System: The purpose of planning expands from predicting outcomes to *learning from outcomes*. Forecast accuracy becomes a measure of how well the organization understands the drivers of its own performance. This reframes planning from compliance to cognition, the planning and finance function itself becomes a sensing mechanism within the enterprise.

Balancing Efficiency with Optionality: Traditional planning is optimized for efficiency; Future must optimize for adaptability and efficiency. Budgets should include strategic slack which is the capacity that can be redeployed as models evolve or new opportunities emerge. The most resilient organizations maintain an efficient core while preserving an adaptive perimeter that can absorb volatility.

Portfolio-Based Revenue Planning: As AI ecosystems diversify, revenue streams will resemble a portfolio of initiatives each with different risk–return profiles. Financial planning can apply portfolio management principles, balancing near-term cash engines with long-term exploratory bets.

Learning-Curve Cost Advantage: Future cost efficiency will arise less from procurement and more from learning. As systems scale, shared infrastructure and cross-model knowledge reduce unit costs through experience and scale effects. Finance should explicitly model *learning curves* of the rate at which cost per insight or transaction declines with cumulative data and iteration.

Strategy and Capital Allocation: In fast-moving AI markets, long-range plans must evolve through *checkpoints*, the moments where capital allocation decisions are revisited based on new evidence. Rather than treating forecasts as fixed, finance should evaluate both performance and *learning achieved* before reinvesting, ensuring strategy remains aligned with environmental shifts.

Short-term financial planning for AI systems demands agility with faster cycles, hybrid modeling, and transparent cost-to-value linkages. Long-term success, however, depends on institutionalizing learning: embedding feedback, treating financial data as an evolving intelligence, and balancing efficiency with flexibility. The true managerial implication is that financial planning itself must become a *learning system*, one capable of sensing change, learning from experience, and guiding the organization through uncertainty.

Closing Insights: From Control to Intelligence

The financial planning discipline is entering a structural transformation. What began as a practice of projecting stability has become an exercise in navigating volatility. As AI systems increasingly mediate cost, value, and decision-making, the role of financial planning expands beyond stewardship, towards becoming the architecture of organizational learning.

The central insight from this article is that the future of financial planning lies not in precision, but in responsiveness. In the AI economy, assumptions decay quickly; value is created through feedback and learning loops. The finance function must therefore evolve from being a reporting mechanism to serving as an intelligence layer, translating operational signals into strategic guidance in real time. When forecasts are continuously informed by usage data, pricing elasticity, and realized outcomes, planning ceases to be static prediction and becomes active orchestration.

For executives, this evolution demands a shift in mindset and investment in the right resources. Success in the next decade will depend less on perfect models and more on the organization's capacity to learn faster than it spends. Investments in AI, data, and automation must be accompanied by investments in financial agility which includes data architecture, cross-functional governance, and talent capable of interpreting probabilistic outcomes. The companies that thrive will treat every forecast not as a commitment, but as a learning experience - one that improves the enterprise's collective intelligence with each cycle.

Ultimately, planning for agentic AI systems is not about controlling complexity; it is about learning from it. When finance becomes the interpreter of how intelligence turns into value, the organization acquire not just foresight about its costs and revenues, but foresight about organization itself unlocking new capabilities and growth pathways.

Selected References

1. Guido Appenzeller, "**LLMflation: LLM Inference Cost Is Going Down Fast**," *a16z*, November 12, 2024.
2. Epoch AI, "**Data Insights on LLM Inference Price Trends**," 2024.
3. AWS News Blog, "**Announcing Price Reduction for Amazon EC2 NVIDIA GPU-Accelerated Instances**," *Amazon Web Services*, June 1, 2025.
4. Gartner, Inc., "**Gartner Predicts Over 40% of Agentic AI Projects Will Be Cancelled by End of 2027**," press release, June 25, 2025.
5. Tobias Straube, Olaf J. Groth, and Jöran Altenberg, "**AI Initiatives Don't Fail—Organizations Do: Why Companies Need AI Experimentation Sandboxes and**

Pathways," *California Management Review Insights*, May 2025.
6. Rebecka C. Ångström *et al.*, "**Getting AI Implementation Right: Insights From a Global Survey**," *California Management Review* 66, no. 1 (August 30, 2023): 5–22.
7. Konstantin Hopf *et al.*, "**Organizational Implementation of AI: Craft and Mechanical Work**," *California Management Review* 66, no. 1 (September 19, 2023): 23–47.
8. Mohit Khanna *et al.*, "**Upgrading Software Business Models to Thrive in the AI Era**," *McKinsey & Company*, September 22, 2025.
9. Ankit Chopra, "A**doption of AI and Agentic Systems: Value, Challenges, and Pathways**," *California Management Review Insights*, August 15, 2025.
10. Ankit Chopra, "**AI/ML-Based Forecasting in Banking SaaS: From Challenge to Competitive Edge**," *INFORMS PubsOnLine*, September 16, 2025.

Ankit Chopra Follow

Ankit Chopra is a finance and analytics leader specializing in financial planning for cloud, AI, and agentic systems. As Director of FP&A at Neo4j, his work focuses on usage-based forecasting, pricing alignment, and cost optimization to support scalable, data-driven business growth.