World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Token-as-a-Service (TaaS) Procurement: An enterprise governance framework for consumption-based AI

Prajkta Waditwar *

Senior Strategic Sourcing Manager at Box, Inc. (Independent Research), United States.

World Journal of Advanced Research and Reviews, 2026, 31(01), 1873–1895

Publication history: Received on 22 June 2026; revised on 29 July 2026; accepted on 31 July 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.31.1.2029

Abstract

The rapid adoption of Large Language Models (LLMs) is transforming enterprise technology procurement from traditional software licensing toward consumption-based artificial intelligence services. Unlike perpetual licenses or Software-as-a-Service (SaaS) subscriptions, modern AI platforms increasingly employ token-based pricing models in which organizations procure computational intelligence according to usage. Existing procurement frameworks were developed for fixed commercial models and provide limited guidance for forecasting demand, negotiating consumption-based contracts, governing multi-vendor AI portfolios, and optimizing enterprise-wide token utilization throughout the AI lifecycle.

This paper introduces **Token-as-a-Service (TaaS) Procurement**, a procurement paradigm that treats AI tokens as strategic enterprise resources requiring continuous commercial governance rather than one-time acquisition. To operationalize this paradigm, the paper proposes the **Enterprise Intelligence FinOps (EIFO)** framework, which integrates business demand management, token forecasting, commercial sourcing, supplier governance, runtime consumption monitoring, financial accountability, and continuous optimization into a unified enterprise operating model.

The paper further develops quantitative models for enterprise AI cost estimation, procurement optimization, governance compliance, and business value measurement, demonstrating how procurement organizations can align AI investments with organizational objectives while maintaining commercial flexibility and governance oversight. By positioning procurement as a continuous enterprise capability rather than a transactional purchasing function, this research establishes a governance foundation for managing consumption-based AI services within evolving Large Language Model ecosystems.

Keywords: Enterprise Procurement; Token-as-a-Service (TaaS); Large Language Models; Consumption-Based AI; AI Governance; Strategic Sourcing; Enterprise Intelligence FinOps; Technology Procurement.

1. Introduction

The emergence of Large Language Models (LLMs) has fundamentally transformed how enterprises acquire and consume artificial intelligence capabilities. [1]–[5] Organizations are increasingly embedding generative AI into procurement, software engineering, finance, legal operations, customer service, and knowledge management. [10]–[15] Unlike traditional enterprise software, which is typically acquired through perpetual licenses or subscription-based Software-as-a-Service (SaaS) models, commercial LLM providers increasingly monetize AI through consumption-based pricing [25]–[31], where organizations pay according to token usage, inference requests, or other usage metrics. This shift represents a fundamental change in enterprise technology procurement.

* Corresponding author: Prajkta Waditwar.

Traditional enterprise procurement has historically focused on evaluating software functionality, negotiating licensing agreements, managing supplier relationships, and optimizing technology portfolios. [16]–[22] These practices assume relatively predictable commercial models with stable costs over the duration of a contract. Consumption-based AI challenges these assumptions. Enterprise spending is no longer determined primarily by user counts or subscription tiers but by dynamic variables such as prompt complexity, context length, reasoning depth, model selection, retrieval operations, and user interaction patterns. As AI adoption scales across business functions, procurement organizations must govern highly variable commercial consumption that cannot be effectively managed using conventional sourcing methodologies.

This evolution extends procurement responsibilities well beyond contract negotiation. [16], [17], [20] Procurement leaders must forecast enterprise AI demand, evaluate multiple pricing models, negotiate flexible commercial agreements, govern multi-vendor AI ecosystems, and continuously assess whether AI consumption delivers measurable business value. Unlike traditional software procurement, where governance largely concludes after contract execution, consumption-based AI requires procurement to remain engaged throughout the operational lifecycle.

Existing enterprise governance frameworks only partially address this challenge. Cloud Financial Operations (FinOps) introduced valuable practices for managing variable cloud infrastructure costs through cross-functional financial accountability and continuous optimization. [14], [15] Similarly, governance standards such as the NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0) and ISO/IEC 42001 provide comprehensive guidance for trustworthy AI, organizational accountability, and risk management. **[5]** [7] However, these frameworks do not specifically address procurement governance for AI services delivered through consumption-based commercial models. They provide limited guidance for supplier evaluation, token demand forecasting, commercial optimization, contract governance, vendor portfolio management, or procurement performance measurement within enterprise AI ecosystems.

At the same time, the commercial AI market is rapidly evolving. Organizations increasingly deploy heterogeneous AI portfolios spanning proprietary foundation models, open-source models, specialized domain models, and internally developed AI capabilities. [10]–[15] Procurement decisions therefore involve balancing commercial pricing, technical performance, security, regulatory compliance, data residency, supplier risk, interoperability, and long-term business value. Managing these trade-offs requires procurement capabilities that extend beyond traditional sourcing or cloud cost optimization.

This paper addresses this emerging challenge by introducing **Token-as-a-Service (TaaS) Procurement**, a procurement paradigm for governing consumption-based AI services throughout their commercial and operational lifecycle. Rather than treating AI tokens merely as billing units generated by service providers, TaaS Procurement recognizes token consumption as a strategic enterprise resource requiring demand forecasting, supplier governance, commercial optimization, runtime visibility, and continuous value realization.

To operationalize this procurement paradigm, this paper proposes the **Enterprise Intelligence FinOps (EIFO)** framework. EIFO serves as the enterprise governance model that integrates procurement, finance, enterprise architecture, AI operations, and organizational governance into a continuous operating model for managing AI consumption. Within this framework, procurement becomes an ongoing enterprise capability that aligns commercial decisions with operational intelligence and measurable business outcomes.

The contributions of this paper are fivefold.

- It identifies the limitations of existing enterprise procurement approaches for governing consumption-based AI services.
- It introduces **Token-as-a-Service (TaaS) Procurement** as a new enterprise procurement paradigm for Large Language Model ecosystems.
- It proposes the **Enterprise Intelligence FinOps (EIFO)** framework as the governance model supporting TaaS Procurement.
- It develops quantitative models for AI procurement planning, commercial optimization, governance compliance, and business value measurement.
- It demonstrates the applicability of the proposed framework through a representative enterprise evaluation, illustrating how procurement organizations can improve governance, commercial transparency, and long-term value realization within multi-vendor AI environments.

The remainder of this paper is organized as follows. Section II reviews the literature on enterprise procurement, AI governance, consumption-based computing, and commercial AI adoption while identifying the research gap addressed by this work. Section III presents the proposed Token-as-a-Service (TaaS) Procurement framework and the Enterprise Intelligence FinOps governance model. Section IV introduces the quantitative procurement models and governance metrics. Section V presents the enterprise architecture supporting the framework, followed by an evaluation, discussion, and concluding remarks.

2. Related Work and Industry Context

The emergence of generative artificial intelligence has fundamentally changed the commercial model of enterprise technology. Unlike previous waves of digital transformation that emphasized software ownership or cloud infrastructure consumption, Large Language Models (LLMs) are increasingly delivered as consumption-based services, where organizations purchase intelligence according to usage. This transition is reshaping enterprise procurement by introducing dynamic pricing models that require continuous commercial governance rather than one-time technology acquisition. [8], [9]

Industry research indicates that enterprise AI adoption is moving rapidly from experimentation to large-scale operational deployment. Gartner projects that by 2027 more than half of enterprise generative AI models will be domain- or industry-specific, reflecting increasing organizational maturity and specialization. Similarly, IDC forecasts worldwide AI spending to exceed US$600 billion by 2028, with enterprise software and AI-enabled services representing the fastest-growing investment categories.[10] These trends indicate that AI procurement is becoming a strategic enterprise capability rather than an isolated technology initiative.

Unlike traditional software procurement, AI services increasingly employ consumption-based pricing. Commercial providers charge according to token usage, inference requests, embeddings, reasoning operations, or API consumption, causing enterprise costs to fluctuate with business activity instead of contractual entitlements. Consequently, procurement organizations must forecast variable demand, negotiate flexible commercial agreements, evaluate supplier pricing models, and continuously monitor consumption to ensure sustainable commercial performance. These characteristics differ fundamentally from procurement practices designed for perpetual licenses or subscription-based software.

Industry surveys further highlight governance as a primary barrier to enterprise AI adoption. Deloitte's *State of Generative AI in the Enterprise* identifies governance, cost management, regulatory compliance, and measurable return on investment as persistent organizational challenges. [11] Microsoft's *Work Trend Index* similarly reports that enterprises are rapidly embedding AI into knowledge-intensive work while simultaneously seeking mechanisms to measure business value and optimize organizational adoption. [13] Collectively, these studies suggest that successful enterprise AI adoption depends as much on procurement and governance capabilities as on technical innovation. [12]

2.1. Procurement Evolution in the AI Economy

Enterprise procurement has continuously evolved alongside changing technology delivery models. Traditional procurement focused on acquiring software licenses and hardware assets through predictable contractual arrangements. [16]–[22] The rise of Software-as-a-Service (SaaS) shifted procurement toward subscription management, supplier performance, and lifecycle governance. [20] Cloud computing introduced infrastructure consumption as a commercial model, leading to Cloud FinOps practices that improved financial visibility and operational cost optimization. [14] [15]

Generative AI represents the next stage of this evolution. Organizations are no longer purchasing software functionality or infrastructure capacity alone—they are procuring computational intelligence delivered through dynamic, usage-based commercial models. Procurement decisions therefore extend beyond vendor selection and contract negotiation to include token demand forecasting, supplier portfolio optimization, commercial risk management, runtime consumption visibility, and continuous value realization.

This transition fundamentally changes the role of procurement. AI procurement becomes a continuous governance capability that spans the entire commercial lifecycle rather than a transactional sourcing function completed at contract award.

2.2. Existing Procurement and Governance Frameworks

Traditional procurement literature provides mature approaches for strategic sourcing, supplier relationship management, category management, contract governance, and commercial negotiations. Likewise, enterprise architecture research emphasizes technology portfolio management, governance, and organizational alignment.

Cloud Financial Operations (FinOps) introduced practices for managing consumption-based cloud infrastructure by integrating finance, engineering, and business stakeholders. These practices—including cost allocation, forecasting, showback, chargeback, and continuous optimization—offer valuable concepts for governing variable technology expenditure. However, FinOps primarily addresses infrastructure resources such as compute, storage, and networking rather than the procurement of metered intelligence services.

Similarly, governance standards including the NIST Artificial Intelligence Risk Management Framework[5] (AI RMF 1.0) and ISO/IEC 42001 provide comprehensive guidance for trustworthy AI, organizational accountability, and risk management. While these frameworks establish critical governance principles, they do not explicitly address enterprise procurement for consumption-based AI, supplier portfolio governance, token forecasting, commercial optimization, or procurement performance measurement.

2.3. Enterprise AI Procurement Gap

Academic research has extensively explored transformer architectures, prompt engineering, retrieval-augmented generation, responsible AI, explainability, model optimization, cloud computing, and enterprise architecture. Parallel research has advanced procurement theory in areas such as strategic sourcing, supplier governance, and technology acquisition. [1]–[4]

However, these bodies of knowledge remain largely disconnected.

Existing procurement models assume relatively stable pricing mechanisms based on licenses, subscriptions, or infrastructure capacity, whereas AI services increasingly exhibit dynamic pricing driven by operational usage. Conversely, existing AI governance research emphasizes technical performance, risk management, and responsible AI without addressing the commercial governance required to manage continuously consumed AI services. [16]–[22]

As enterprises adopt heterogeneous AI ecosystems consisting of proprietary foundation models, open-source deployments, specialized domain models, and AI agents, procurement organizations face new responsibilities including supplier diversification, commercial optimization, vendor lock-in mitigation, pricing model evaluation, token demand forecasting, and ongoing contract governance. These procurement challenges remain underrepresented in both academic literature and industry governance frameworks. [23]

2.4. Research Gap

The literature reveals a consistent gap: organizations possess mature guidance for procurement, AI governance, and cloud financial management independently, yet no unified framework exists for governing the commercial lifecycle of consumption-based AI services.

Specifically, current procurement methodologies provide limited support for:

- forecasting enterprise token demand,
- evaluating token-based commercial models,
- governing multi-vendor AI portfolios,
- measuring procurement performance for consumption-based AI,
- linking commercial decisions with runtime AI consumption,
- optimizing supplier strategies throughout the operational lifecycle.

To address this gap, this paper introduces **Token-as-a-Service (TaaS) Procurement**, a procurement paradigm that treats AI tokens as strategic enterprise resources requiring continuous commercial governance. To operationalize this paradigm, the paper proposes the **Enterprise Intelligence FinOps (EIFO)** framework, which integrates procurement governance, financial accountability, AI operations, enterprise architecture, and organizational oversight into a unified operating model for managing consumption-based AI services.

3. Token-as-a-Service (TaaS) Procurement: An Enterprise Governance Framework for consumption-Based AI

3.1. Conceptual Foundation

The commercialization of Large Language Models (LLMs) represents one of the most significant shifts in enterprise technology procurement since the emergence of cloud computing. For decades, procurement organizations have governed technology acquisitions through commercial models based on perpetual software licenses, subscription services, or infrastructure capacity. These models assume that enterprise costs can be estimated before contract execution and remain relatively stable throughout the procurement lifecycle. Procurement activities therefore concentrate on supplier evaluation, commercial negotiation, contract management, and periodic vendor performance reviews.

Consumption-based artificial intelligence fundamentally challenges these assumptions. Rather than purchasing software functionality, organizations increasingly procure computational intelligence delivered through dynamic pricing models in which commercial expenditure is determined by operational usage. The cost of enterprise AI is influenced by variables such as prompt complexity, reasoning depth, retrieval operations, model selection, context length, and user interaction patterns, all of which fluctuate continuously after deployment. Consequently, commercial commitments cannot be fully governed at the time of procurement because enterprise expenditure evolves alongside business demand.

This transition requires procurement to move beyond its traditional role as a transactional sourcing function. Instead, procurement must become an ongoing governance capability responsible for forecasting enterprise demand, selecting appropriate suppliers, monitoring commercial performance, optimizing consumption, and continuously aligning AI investments with organizational objectives. Existing procurement methodologies provide mature guidance for supplier selection and contract governance but offer limited mechanisms for governing continuously consumed AI services whose commercial characteristics change throughout their operational lifecycle.

To address this challenge, this paper introduces **Token-as-a-Service (TaaS) Procurement**, a procurement paradigm that treats AI tokens as strategic enterprise resources requiring continuous commercial governance. Within this paradigm, tokens are no longer viewed merely as technical billing units generated by AI providers. Instead, they represent measurable units of enterprise intelligence whose acquisition, allocation, utilization, and optimization require the same level of strategic oversight traditionally applied to software assets, cloud infrastructure, and critical supplier relationships.

To operationalize this procurement paradigm, the paper proposes the **Enterprise Intelligence FinOps (EIFO)** framework. Rather than positioning FinOps as the primary contribution, EIFO serves as the governance operating model through which procurement organizations collaborate with finance, enterprise architecture, AI engineering, and business leadership to manage the commercial lifecycle of consumption-based AI services. In this context, FinOps principles provide financial visibility and operational accountability, while procurement remains the primary organizational function responsible for commercial governance and supplier strategy.

3.2. Enterprise Governance Framework

The proposed Token-as-a-Service Procurement framework consists of seven interdependent governance capabilities that collectively support the commercial lifecycle of enterprise AI consumption. Unlike traditional procurement processes that emphasize sequential purchasing activities, these capabilities operate as an integrated governance system in which operational intelligence continuously informs future procurement decisions.

The framework begins with **Business Demand Management**, where business requirements are evaluated before any procurement activity is initiated. Rather than assuming that artificial intelligence represents the appropriate solution for every business challenge, organizations first establish measurable objectives, expected business outcomes, regulatory constraints, and strategic alignment. This initial assessment ensures that procurement activities are driven by validated organizational demand instead of technology availability or vendor promotion.

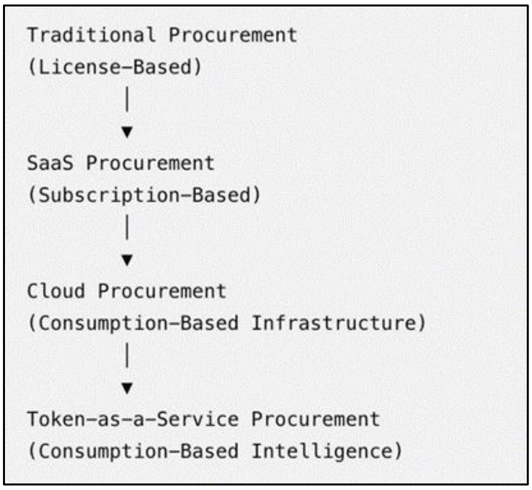

Figure 1 Evolution of Enterprise Technology Procurement (Keep but Redesign)

Once business demand has been established, governance proceeds to **AI Portfolio Governance.** [16]–[22] Modern enterprises rarely rely on a single AI provider. Instead, organizations increasingly operate heterogeneous AI ecosystems consisting of proprietary foundation models, open-source deployments, internally developed models, and specialized domain-specific solutions. Procurement therefore assumes responsibility for governing the enterprise AI portfolio rather than individual supplier relationships. Portfolio governance evaluates suppliers according to commercial competitiveness, technical capability, security, regulatory compliance, interoperability, sustainability, and long-term strategic alignment. This approach minimizes unnecessary supplier proliferation while maintaining flexibility to adopt emerging AI technologies.

The third governance capability introduces **Token Demand Forecasting**, representing one of the principal innovations of the proposed procurement model. Traditional technology procurement estimates future expenditure using relatively stable variables such as employee counts, software licenses, or infrastructure capacity. Consumption-based AI requires a fundamentally different forecasting methodology because enterprise costs depend upon operational behavior after deployment. Procurement organizations must therefore estimate expected interaction volumes, prompt complexity, reasoning requirements, context utilization, autonomous agent activity, and anticipated business demand. Forecasting becomes a continuous analytical capability that supports budgeting, supplier negotiations, reserved-capacity purchasing, and long-term sourcing strategy.

Following demand estimation, procurement advances to **Commercial Sourcing and Supplier Governance**. [17]–[20] Although supplier evaluation remains a core procurement responsibility, consumption-based AI significantly expands its scope. Procurement decisions extend beyond comparing unit prices to evaluating token pricing structures, committed-use agreements, enterprise discount models, service-level commitments, intellectual property protections, data residency requirements, contractual flexibility, supplier portability, and long-term exit strategies. Because organizations increasingly consume AI services from multiple providers simultaneously, procurement must continuously balance commercial competitiveness with operational resilience while minimizing supplier concentration risk.

Once AI services are operational, governance transitions to **Runtime Consumption Management**. Unlike conventional software procurement, where governance activities often diminish after implementation, Token-as-a-Service Procurement requires continuous visibility into operational AI utilization. Runtime governance captures enterprise-wide consumption patterns, monitors supplier performance, measures workload distribution, attributes token expenditure to business functions, and evaluates commercial performance against contractual expectations. This operational intelligence establishes the foundation for evidence-based procurement decisions throughout the commercial lifecycle.

Operational insights subsequently support **Commercial Optimization and Financial Governance**. At this stage, procurement collaborates closely with finance and technology organizations to translate operational consumption into commercial intelligence. Consumption data informs supplier renegotiations, pricing optimization, model substitution decisions, workload routing, enterprise chargeback mechanisms, budget allocation, and portfolio rationalization. Unlike traditional procurement reviews conducted annually or at contract renewal, optimization becomes a continuous activity that adapts commercial strategy according to changing organizational demand.

The framework concludes with **Executive Procurement Governance**, where senior leadership evaluates enterprise AI performance through integrated commercial, financial, operational, and governance metrics. Executive oversight ensures that procurement decisions remain aligned with organizational strategy while maintaining transparency across supplier performance, commercial risk, governance compliance, forecasting accuracy, and business value realization. Insights generated through executive governance are subsequently incorporated into future sourcing strategies, establishing a continuous learning cycle that strengthens enterprise procurement maturity over time.

Collectively, these governance capabilities transform procurement from a transactional purchasing function into a continuous enterprise capability responsible for governing the commercial lifecycle of artificial intelligence.

3.3. Governance Lifecycle

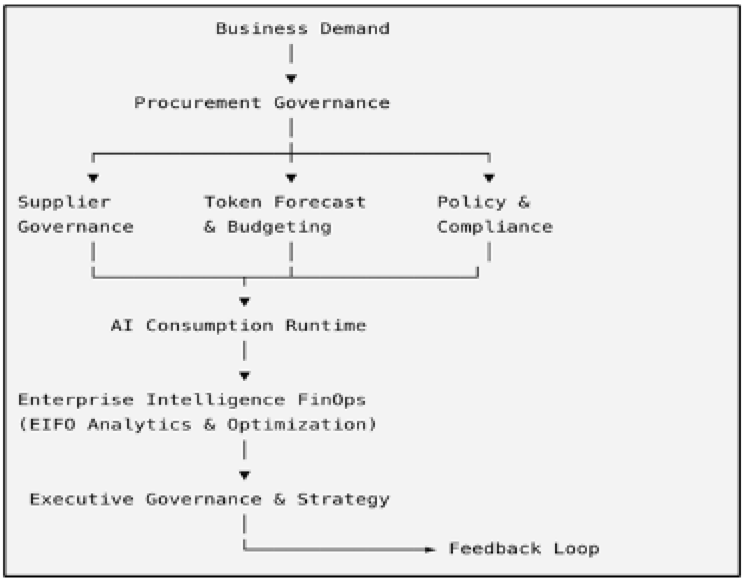

Figure 2 Token-as-a-Service Procurement Governance Framework (Replace Existing EIFO Diagram)

Unlike conventional procurement methodologies, which typically conclude once contracts are executed and technology is deployed, Token-as-a-Service Procurement operates as a continuous governance lifecycle. Procurement decisions are repeatedly refined using operational intelligence generated throughout AI deployment, enabling organizations to respond dynamically to changing business demand, supplier performance, and commercial conditions.

The lifecycle begins by identifying validated business demand and translating organizational objectives into procurement requirements. These requirements inform enterprise AI portfolio decisions, supplier selection, and token demand forecasts before commercial agreements are negotiated. Following deployment, runtime monitoring continuously captures operational consumption patterns and supplier performance, generating commercial intelligence that supports financial governance and optimization. Rather than representing the final stage of procurement, operational deployment becomes the beginning of an ongoing feedback loop through which procurement strategies, forecasting models, supplier relationships, and governance policies are continuously refined.

This closed-loop lifecycle distinguishes Token-as-a-Service Procurement from traditional sourcing methodologies by recognizing that enterprise AI procurement does not end with contract award. Instead, procurement becomes a continuous governance discipline responsible for ensuring that commercial decisions remain aligned with evolving organizational demand throughout the operational lifecycle.

3.4. Governance Principles

The proposed framework is founded upon five governance principles that collectively define the philosophy of Token-as-a-Service Procurement.

The first principle is **business value orientation**, which recognizes that procurement should optimize organizational outcomes rather than simply minimize token consumption. Higher AI expenditure may be commercially justified when it generates proportionally greater business value through productivity improvements, innovation, customer experience, or strategic differentiation.

The second principle is **continuous commercial governance**. Consumption-based AI requires procurement organizations to remain actively engaged throughout contract execution, continuously evaluating supplier performance, commercial effectiveness, and operational demand instead of limiting governance activities to sourcing events.

The third principle is **enterprise-wide transparency**. Every unit of AI consumption should be attributable to a specific supplier, application, business function, and organizational objective. Transparent consumption enables accurate financial accountability while supporting evidence-based procurement decisions and executive oversight.

The fourth principle is **adaptive supplier governance**. Procurement strategies should evolve according to operational intelligence, allowing organizations to adjust sourcing decisions, negotiate improved commercial terms, optimize supplier portfolios, and respond to changing market conditions without disrupting business operations.

Finally, the framework embraces **continuous optimization.** [17], [19] Procurement is no longer viewed as a sequence of isolated purchasing events but as an iterative governance capability in which commercial decisions, operational performance, financial accountability, and business value are continuously balanced throughout the AI lifecycle.

3.5. Formal Definition of Token-as-a-Service (TaaS) Procurement

Within the proposed framework, Token-as-a-Service (TaaS) Procurement is formally defined as:

"An enterprise procurement capability responsible for the strategic sourcing, commercial governance, demand forecasting, supplier management, financial accountability, operational oversight, and continuous optimization of consumption-based artificial intelligence services delivered through tokenized or usage-based commercial models."

This definition distinguishes TaaS Procurement from traditional software procurement by recognizing that enterprise intelligence is no longer acquired as a static technology asset but consumed dynamically according to organizational demand. Procurement responsibilities therefore extend beyond supplier

selection and contract negotiation to encompass the complete commercial lifecycle of enterprise AI consumption, including forecasting, runtime governance, supplier optimization, financial accountability, and continuous value realization.

3.6. Research Propositions

To facilitate future empirical validation, the proposed framework is expressed through five research propositions.

- **Proposition 1** states that organizations implementing Token-as-a-Service Procurement will demonstrate greater visibility into enterprise AI consumption than organizations relying on conventional software procurement methodologies.
- **Proposition 2** proposes that integrating continuous token demand forecasting into procurement governance will improve the accuracy of enterprise AI budgeting and commercial planning.
- **Proposition 3** suggests that centralized supplier governance will reduce vendor fragmentation while improving commercial performance across heterogeneous AI portfolios.
- **Proposition 4** argues that integrating procurement governance with operational intelligence through the Enterprise Intelligence FinOps framework will improve Business Value per Token without compromising governance or organizational compliance.
- **Proposition 5** proposes that enterprises adopting Token-as-a-Service Procurement will exhibit higher procurement maturity for consumption-based AI than organizations applying traditional procurement practices to AI services.

4. Quantitative Procurement Model for Token-as-a-Service (TaaS)

4.1. Motivation

The transition from traditional software acquisition to consumption-based artificial intelligence fundamentally changes how enterprise procurement evaluates commercial performance. Conventional procurement models estimate expenditure using relatively stable variables such as software licenses, subscribed users, or infrastructure capacity. In contrast, AI services are consumed dynamically, causing enterprise expenditure to fluctuate according to operational behavior rather than contractual commitments.

Consequently, procurement organizations require quantitative models that extend beyond cost estimation. They must forecast future demand, compare supplier pricing models, evaluate procurement efficiency, monitor commercial risk, and continuously assess whether AI consumption delivers measurable business value. The proposed quantitative procurement model provides a mathematical foundation for these governance activities by integrating operational consumption data with procurement decision-making throughout the AI lifecycle.

Unlike traditional financial models that primarily measure expenditure, the proposed model treats procurement as a continuous optimization problem in which commercial decisions are repeatedly refined using operational intelligence.

4.2. Enterprise Token Demand Forecasting

Demand forecasting represents the starting point of Token-as-a-Service Procurement because supplier negotiations, budget allocation, and contract design depend upon reliable estimates of future AI consumption.

Enterprise token demand can be expressed as

$$T = \sum_{i=1}^{n} U_i \times I_i \times P_i$$

where:

- U_i represents the number of users within business unit i,
- I_i denotes the expected number of AI interactions,
- P_i represents the average number of tokens consumed per interaction.

Unlike conventional software procurement, where licensing requirements remain relatively constant, this model allows procurement organizations to incorporate evolving operational behavior into commercial planning. Forecasts can be updated periodically as business demand, application usage, or organizational adoption changes, enabling procurement teams to negotiate commercial agreements using expected consumption rather than static assumptions.

4.3. Procurement Cost Estimation

Following demand forecasting, procurement organizations estimate the expected commercial cost of AI consumption across one or more suppliers.

The enterprise procurement cost is expressed as

$$C = \sum_{j=1}^{m} T_j \times R_j$$

where:

- T_j denotes token consumption allocated to supplier j,
- R_j represents the negotiated commercial rate per token.

This formulation enables procurement teams to compare supplier proposals under different pricing structures while evaluating the financial implications of multi-vendor sourcing strategies. Because pricing models frequently include volume discounts, reserved-capacity commitments, or tiered pricing arrangements, the model provides a baseline that can be extended to incorporate more sophisticated commercial contracts.

4.4. Supplier Portfolio Optimization

One of the distinguishing characteristics of enterprise AI procurement is that organizations increasingly rely on multiple providers rather than a single technology vendor. Procurement therefore seeks to distribute workloads across suppliers in a manner that balances commercial cost, operational performance, governance requirements, and supplier risk.

The procurement optimization objective may be expressed as

$$\min \sum m\, C_j$$

subject to

- service-level requirements,
- regulatory compliance,
- security policies,
- supplier capacity,

business continuity constraints.

Unlike traditional supplier selection, portfolio optimization is performed continuously as operational consumption patterns evolve. Procurement organizations may therefore adjust workload allocation among suppliers whenever commercial conditions or business priorities change.

4.5. Procurement Efficiency Index

Cost reduction alone does not indicate procurement effectiveness. An organization may reduce expenditure while simultaneously limiting innovation or degrading operational performance. Accordingly, procurement performance should be evaluated using measures that balance commercial efficiency with business outcomes.

The **Enterprise AI Procurement Efficiency (EAPE)** index is defined as

$$EAPE = \frac{BV}{PC}$$

where

- BV represents measurable business value generated through AI,
- PC denotes procurement cost.

Higher values indicate greater commercial efficiency because procurement expenditure produces proportionally larger organizational outcomes. This metric encourages procurement organizations to optimize value realization rather than minimizing expenditure alone.

4.6. Supplier Diversification Metric

Supplier concentration represents an important commercial risk in enterprise AI ecosystems. Excessive dependence on a single provider may increase pricing risk, reduce negotiation leverage, and limit organizational flexibility.

Supplier diversification is measured using the normalized Herfindahl–Hirschman Index (HHI)

$$HHI = \sum_{j=1}^{m} s_j^2$$

where s_j represents the proportion of enterprise AI expenditure allocated to supplier j.

Lower concentration values indicate more diversified procurement portfolios, while higher values signal increased supplier dependence that may require mitigation through sourcing strategies or contract renegotiation.

4.7. Procurement Forecast Accuracy

Forecasting accuracy directly influences procurement performance because inaccurate demand estimates result in over-provisioning, under-utilization, or unexpected commercial expenditure.

Forecast accuracy is calculated as

$$FA = 1 - \frac{| T_f - T_a |}{T_a}$$

where

- T_f is forecast token demand,
- T_a is actual consumption.

Continuous comparison between forecasted and observed demand enables procurement organizations to improve budgeting models while supporting increasingly accurate supplier negotiations over time.

4.8. Business Value per Token

Because tokens represent the fundamental commercial unit of AI consumption, procurement should evaluate the value generated by each unit of intelligence acquired.

Business Value per Token is defined as

$$BVT = \frac{BV}{T}$$

where

- BV represents quantified organizational value,
- T denotes total token consumption.

This metric enables procurement leaders to compare applications, departments, and suppliers according to the value generated from AI expenditure rather than focusing solely on consumption volume.

4.9. Governance Performance Model

The quantitative framework concludes with an integrated procurement governance score that combines financial, operational, and commercial performance into a single maturity indicator.

$$PGS = w_1 FA + w_2 EAPE + w_3 BVT + w_4 (1 - HHI)$$

where

- PGS is the Procurement Governance Score,
- w_i are organizational weighting coefficients satisfying

$$\sum_{i=1}^{4} w_i = 1$$

Organizations may adjust these weights according to strategic priorities. For example, highly regulated industries may assign greater importance to supplier diversification, whereas organizations pursuing aggressive AI adoption may prioritize business value generation.

4.10. Discussion

Collectively, these quantitative models transform procurement from a reactive purchasing function into a continuous decision-support capability. Demand forecasting informs supplier negotiations, commercial pricing influences portfolio optimization, operational consumption validates procurement forecasts, and governance metrics provide evidence for executive decision-making. Rather than viewing procurement as a sequence of isolated transactions, the proposed framework models procurement as a dynamic optimization process in which commercial decisions continuously evolve alongside enterprise AI consumption.

Unlike existing cloud cost models or traditional software procurement metrics, the proposed quantitative framework explicitly integrates demand forecasting, supplier governance, procurement efficiency, portfolio diversification, and business value measurement into a unified mathematical model for consumption-based artificial intelligence.

5. Evolution of Enterprise Technology Procurement: From Licenses to Token-as-a-Service (TaaS)

Enterprise technology procurement has evolved through successive commercial models, each requiring new governance capabilities. During the era of on-premises enterprise software, procurement primarily managed perpetual software licenses and maintenance contracts. Financial planning emphasized capital expenditure, software asset management, and vendor lifecycle governance because costs were largely fixed throughout the contractual period.

The widespread adoption of cloud computing fundamentally altered this procurement model. Organizations no longer purchased computing infrastructure outright but instead consumed infrastructure services based on operational demand. This transition gave rise to Cloud Financial Operations (FinOps), which introduced continuous monitoring, cost allocation, forecasting, and optimization of cloud infrastructure consumption. The governance focus shifted from software ownership to resource utilization.

Generative Artificial Intelligence introduces a third transformation. Rather than purchasing software functionality or infrastructure capacity, organizations increasingly consume computational intelligence through token-based commercial models. Unlike cloud infrastructure, where costs are primarily determined by computational resources such as processors, memory, and storage, AI expenditure depends upon linguistic interactions including prompt complexity, context length, reasoning depth, retrieval operations, model selection, and autonomous agent execution.

Table 1 Evolution of Enterprise Technology Procurement

Era	Primary Commercial Model	Procurement Unit	Governance Focus	Primary Optimization Objective
Traditional Enterprise Software	Perpetual License	Software License	Software Asset Management	License utilization
SaaS and Cloud Computing	Subscription and Infrastructure Consumption	User, VM, Storage, Compute	Cloud FinOps	Infrastructure cost optimization
Generative AI (Proposed)	Token-as-a-Service (TaaS)	AI Tokens	Enterprise Intelligence FinOps (EIFO)	Business value per token

Recent developments across the AI industry illustrate this transition. Commercial AI providers increasingly expose consumption-based pricing models in which organizations purchase intelligence according to token utilization rather than solely through user licenses. In parallel, AI-native development platforms have begun complementing or replacing traditional seat-based subscriptions with usage-based pricing for advanced AI capabilities. These market developments indicate that enterprise procurement is gradually transitioning from managing software licenses toward governing intelligence consumption.

This evolution fundamentally changes the responsibilities of enterprise procurement. Traditional procurement negotiated licensing agreements and software subscriptions. Cloud procurement optimized infrastructure consumption through FinOps practices. Token-as-a-Service (TaaS) Procurement extends these capabilities by introducing continuous governance over enterprise intelligence consumption. Procurement decisions therefore encompass token demand forecasting, commercial sourcing, runtime financial monitoring, AI portfolio optimization, governance compliance, and business value realization.

Consequently, Enterprise Intelligence FinOps should not be interpreted as a replacement for FinOps. Instead, it represents an extension of FinOps principles designed specifically for the economic characteristics of consumption-based artificial intelligence.

6. Enterprise Architecture for Token-as-a-Service (TaaS) Procurement

6.1. Architectural Overview

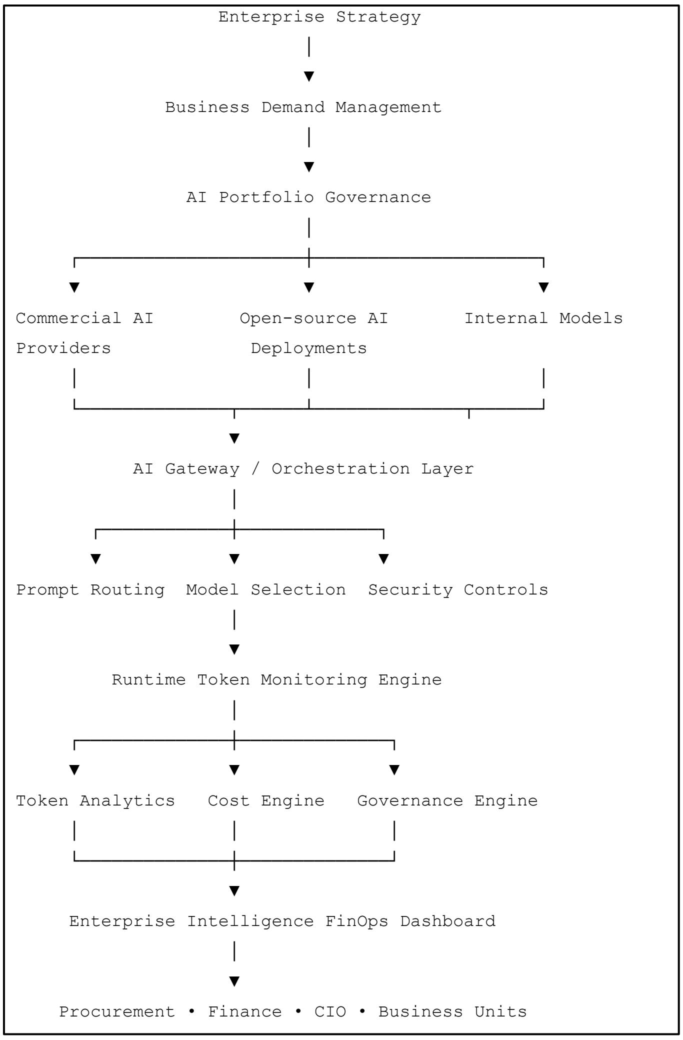

Figure 3 Enterprise Intelligence FinOps Enterprise Architecture

The Token-as-a-Service (TaaS) Procurement framework requires an enterprise architecture capable of integrating commercial governance with operational AI consumption. Traditional procurement systems are designed primarily to support sourcing activities, contract administration, supplier management, and financial reconciliation. While these capabilities remain essential, they provide only limited visibility into the dynamic consumption patterns that characterize Large Language Model (LLM) services. Consequently, procurement organizations often lack the operational intelligence necessary to govern AI expenditure throughout its lifecycle.

The proposed enterprise architecture extends conventional procurement capabilities by establishing a continuous flow of information between business operations, AI platforms, financial systems, procurement functions, and executive governance. Rather than separating procurement from runtime operations, the architecture enables commercial decisions to be continuously informed by operational AI consumption. This integration transforms procurement into an adaptive governance capability that evolves alongside enterprise AI adoption.

The architecture follows a layered design in which each layer performs a distinct governance function while maintaining continuous communication with adjacent layers. Collectively, these layers establish a closed-loop procurement ecosystem that supports demand forecasting, supplier governance, financial accountability, operational monitoring, and strategic decision-making.

6.2. Business and Demand Layer

The architecture begins with the Business and Demand Layer, where enterprise AI requirements originate. Business units identify operational challenges, define expected outcomes, and specify performance objectives before procurement activities commence. This layer ensures that AI investments are initiated by measurable business demand rather than technology availability or supplier offerings.

Demand generated at this level includes anticipated user populations, application characteristics, expected transaction volumes, regulatory requirements, and strategic priorities. These inputs establish the foundation for procurement planning by providing the context required for supplier evaluation, commercial negotiations, and demand forecasting. Because organizational priorities evolve over time, business demand remains an active input throughout the procurement lifecycle rather than a one-time requirement captured during project initiation.

6.3. Procurement and Commercial Governance Layer

The Procurement and Commercial Governance Layer represents the core decision-making component of the architecture. Its primary responsibility is translating business demand into commercially sustainable sourcing strategies. Procurement professionals evaluate alternative AI providers, negotiate commercial agreements, assess contractual obligations, and establish governance policies that align supplier capabilities with organizational objectives.

Unlike traditional software procurement, governance within this layer extends beyond supplier selection. Procurement continuously evaluates token pricing models, committed-use agreements, reserved-capacity contracts, service-level commitments, portability provisions, intellectual property considerations, and supplier concentration risks. These commercial activities remain active throughout contract execution, allowing procurement strategies to evolve as enterprise AI consumption changes.

The governance policies established within this layer also define the financial and operational controls that regulate enterprise AI usage. Budget thresholds, supplier approval policies, allocation mechanisms, and compliance requirements are propagated to downstream operational systems, ensuring that runtime AI consumption remains consistent with enterprise procurement strategy.

6.4. AI Service and Consumption Layer

The AI Service and Consumption Layer represents the operational environment in which enterprise applications interact with LLM providers. Business applications, intelligent assistants, AI agents, workflow automation platforms, and analytical systems consume AI services through standardized interfaces, generating continuous streams of token-based transactions.

This operational layer provides the data required to govern consumption-based procurement. Every interaction produces measurable information regarding token utilization, model selection, application demand, response characteristics, and supplier usage. Rather than serving only technical monitoring purposes, these operational metrics become procurement intelligence that supports financial governance and commercial optimization.

The architecture intentionally remains supplier-neutral, enabling organizations to integrate multiple proprietary and open-source AI platforms within a unified governance environment. Such flexibility reduces dependence on individual vendors while allowing procurement organizations to optimize commercial decisions according to changing market conditions.

6.5. Financial Intelligence and Enterprise Intelligence FinOps Layer

Operational consumption data is consolidated within the Enterprise Intelligence FinOps (EIFO) layer, which functions as the analytical engine of the proposed architecture. Rather than acting as a replacement for procurement, EIFO transforms operational telemetry into financial and commercial intelligence that supports procurement governance.

Within this layer, token consumption is attributed to business units, applications, projects, suppliers, and organizational cost centers. Analytical models evaluate expenditure trends, forecast future demand, identify commercial optimization opportunities, and generate procurement performance indicators. Runtime information is continuously reconciled with contractual commitments, enabling procurement organizations to determine whether negotiated commercial agreements remain aligned with actual enterprise consumption.

EIFO therefore provides the visibility required to transform operational AI activity into evidence-based procurement decisions. Instead of relying on periodic financial reporting, procurement organizations gain continuous insight into supplier performance, commercial efficiency, forecast accuracy, and business value generation.

6.6. Executive Governance Layer

The Executive Governance Layer provides strategic oversight of enterprise AI procurement. Senior leadership requires integrated visibility across commercial performance, financial accountability, operational effectiveness, governance compliance, and organizational value creation. This layer consolidates analytical outputs generated throughout the architecture into executive dashboards that support enterprise decision-making.

Governance activities include evaluating procurement performance, monitoring supplier concentration, assessing commercial risk, reviewing forecast accuracy, and measuring the business outcomes achieved through AI investment. Executive oversight also informs long-term sourcing strategies by identifying opportunities for supplier diversification, contract renegotiation, technology modernization, and portfolio optimization.

Unlike traditional procurement reporting, which often emphasizes expenditure alone, executive governance evaluates AI procurement according to its contribution to organizational strategy, operational efficiency, innovation, and enterprise competitiveness.

6.7. Continuous Governance Feedback Loop

A defining characteristic of the proposed architecture is the continuous feedback loop connecting operational AI consumption with procurement governance. Information generated during runtime does not remain isolated within technical monitoring systems. Instead, it continuously informs demand forecasting, supplier evaluation, commercial negotiations, financial planning, and governance policy refinement.

For example, unexpected increases in token consumption may trigger revised forecasting models, additional supplier negotiations, or changes in workload allocation across AI providers. Similarly, improvements in supplier performance or pricing may influence future sourcing strategies and contract renewals. This continuous exchange of operational and commercial intelligence enables procurement organizations to adapt proactively to evolving business requirements rather than responding only during periodic procurement cycles.

The feedback mechanism distinguishes the proposed architecture from traditional enterprise procurement systems by recognizing that procurement governance continues throughout the operational lifecycle of AI services. As organizations accumulate operational experience, the architecture continuously improves procurement decision-making, strengthens supplier governance, and enhances enterprise AI maturity.

6.8. Architectural Implications

The proposed architecture demonstrates that effective procurement for consumption-based AI cannot be achieved through isolated procurement systems or standalone financial monitoring tools. Instead, governance requires a tightly integrated enterprise architecture in which procurement, finance, AI operations, enterprise architecture, and executive leadership collaborate through shared operational intelligence.

By positioning procurement as the central coordinating function and Enterprise Intelligence FinOps as the analytical governance layer, the architecture operationalizes the Token-as-a-Service (TaaS) Procurement framework proposed in this paper. This integration enables organizations to move beyond transactional technology acquisition toward continuous commercial governance of enterprise intelligence services.

7. Framework Evaluation: Enterprise Case Study

7.1. Evaluation Approach

The proposed Token-as-a-Service (TaaS) Procurement framework is evaluated through an illustrative enterprise scenario that demonstrates how the governance model supports procurement decision-making throughout the lifecycle of consumption-based artificial intelligence services. Because the framework introduces a conceptual governance model rather than a software artifact or predictive algorithm, the objective of the evaluation is to assess its logical consistency, practical applicability, and ability to address the procurement challenges identified in the literature.

The evaluation follows a scenario-based methodology frequently employed in enterprise architecture and design science research. A representative enterprise environment is used to demonstrate how procurement activities evolve when AI services are acquired through token-based commercial models. The scenario highlights the interaction between procurement governance, operational AI consumption, supplier management, and Enterprise Intelligence FinOps (EIFO), illustrating how the proposed framework enables continuous commercial oversight. [14]

Rather than validating numerical performance improvements, the evaluation examines whether the framework provides governance capabilities that are absent from conventional software procurement methodologies.

7.2. Illustrative Enterprise Scenario

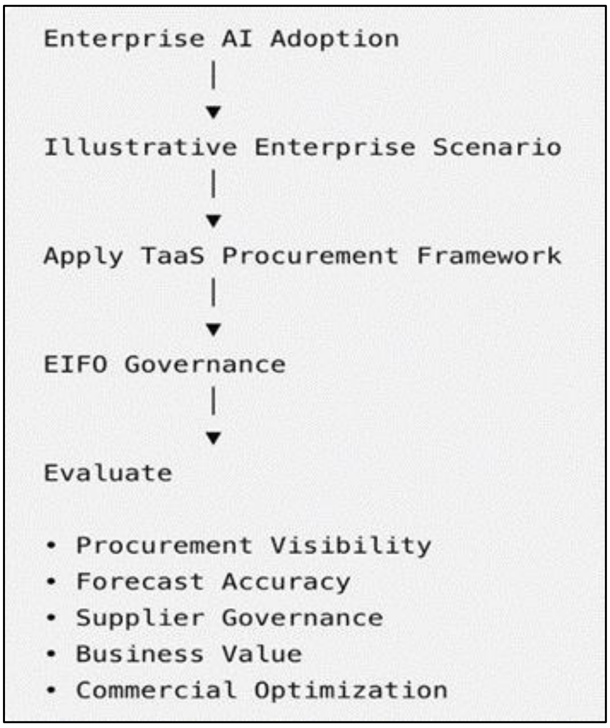

Figure 4 Evaluation Methodology

Consider a multinational financial services organization deploying generative AI across customer support, software engineering, regulatory compliance, legal document analysis, and internal knowledge management. The organization consumes AI services from multiple providers, including proprietary foundation models, cloud-based AI platforms, and internally hosted open-source models. Each business unit exhibits distinct usage characteristics, resulting in highly variable token consumption and commercial expenditure.

Under a conventional procurement approach, suppliers are selected primarily through contract negotiations based on projected licensing costs and service-level agreements. Once contracts are executed, procurement involvement decreases significantly, while operational teams independently manage AI consumption. As enterprise adoption expands, token utilization becomes increasingly unpredictable, making it difficult to forecast expenditure, compare supplier performance, or determine whether AI investments generate proportional business value.

Applying the proposed Token-as-a-Service Procurement framework fundamentally changes this process. Procurement remains actively involved throughout the commercial lifecycle, collaborating with finance, enterprise architecture, AI operations, and business leadership. Demand forecasting is continuously refined using operational consumption data, supplier performance is evaluated against commercial objectives, and financial intelligence generated through EIFO informs contract optimization and sourcing strategies. Rather than responding to unexpected expenditure after deployment, procurement organizations proactively adapt supplier portfolios and commercial agreements as enterprise demand evolves.

This scenario demonstrates how procurement transitions from a transactional sourcing function to a continuous governance capability capable of managing enterprise AI consumption across heterogeneous supplier ecosystems.

7.3. Framework Evaluation

The effectiveness of the proposed framework can be assessed by comparing its governance capabilities with those provided by conventional enterprise procurement approaches.

Table 2 Comparison of Traditional Enterprise Procurement and the Proposed Token-as-a-Service (TaaS) Procurement Framework

Evaluation Dimension	Traditional Procurement	TaaS Procurement + EIFO
Procurement lifecycle	Contract-centric	Continuous lifecycle governance
Cost visibility	Periodic financial reporting	Real-time consumption visibility
Demand planning	Static budgeting	Continuous token forecasting
Supplier management	Individual contracts	Portfolio governance
Financial accountability	Budget tracking	Consumption-based optimization
Governance	Procurement policies	Integrated commercial and operational governance
Decision support	Historical reporting	Continuous procurement intelligence

The comparison illustrates that the proposed framework extends procurement governance beyond sourcing activities by integrating operational intelligence directly into commercial decision-making. Procurement becomes an adaptive capability capable of responding to changes in enterprise AI consumption rather than relying solely on predefined contractual assumptions.

7.4. Governance Capability Assessment

Table 3 Mapping of Enterprise AI Procurement Challenges to the Proposed TaaS Governance Capabilities

Enterprise Challenge	Governance Capability	Framework Component
Variable AI demand	Continuous forecasting	Token Demand Forecasting
Dynamic pricing	Commercial optimization	Procurement Governance
Multi-vendor environments	Supplier portfolio management	AI Portfolio Governance
Limited financial visibility	Runtime analytics	EIFO
Governance compliance	Policy enforcement	Executive Governance
Business value measurement	Procurement KPIs	Quantitative Model

The framework was evaluated against the governance challenges identified in the literature review, including demand uncertainty, supplier diversification, financial accountability, commercial optimization, and executive oversight. Table VI summarizes the alignment between these enterprise challenges and the governance capabilities introduced by the proposed framework.

The assessment indicates that each identified challenge is explicitly addressed through one or more governance capabilities within the proposed framework. Unlike traditional procurement models, the framework establishes clear relationships between commercial decision-making, operational consumption, and enterprise governance.

7.5. Discussion of Evaluation Results

The illustrative scenario demonstrates that Token-as-a-Service Procurement provides governance mechanisms that extend beyond conventional technology procurement practices. By integrating procurement activities with runtime AI consumption, organizations gain continuous visibility into commercial performance and supplier effectiveness. This visibility enables procurement teams to refine forecasting models, optimize commercial agreements, and align AI expenditure with measurable business outcomes.

The evaluation also highlights the role of Enterprise Intelligence FinOps (EIFO) as the analytical governance layer rather than the primary procurement function. Operational telemetry collected during AI execution is transformed into procurement intelligence, allowing commercial decisions to be based on observed enterprise behavior rather than static contractual assumptions. This continuous feedback loop represents one of the distinguishing characteristics of the proposed framework.

Although the evaluation is conceptual, it demonstrates the internal consistency of the framework and its applicability to complex enterprise environments where multiple AI providers, dynamic pricing models, and evolving business requirements must be governed simultaneously.

Limitations of the Evaluation

The evaluation presented in this paper is intentionally illustrative and does not constitute empirical validation through longitudinal industrial deployment. Consequently, the quantitative improvements associated with procurement efficiency, forecasting accuracy, or commercial optimization should be interpreted as expected outcomes derived from the proposed governance model rather than statistically verified results.

Future research should validate the framework through enterprise case studies, controlled organizational pilots, and cross-industry comparative analyses. Such studies would enable empirical assessment of procurement performance, supplier governance, forecasting accuracy, and business value realization across different organizational contexts and AI deployment strategies.

8. Discussion and Research Implications

8.1. Theoretical Contributions

This research contributes to the growing body of knowledge on enterprise AI governance by positioning procurement as a continuous governance capability for consumption-based artificial intelligence. Existing research has extensively examined cloud financial management, enterprise architecture, AI governance, and responsible AI, yet comparatively little attention has been given to how organizations should govern the commercial lifecycle of AI services delivered through token-based pricing models.

The primary theoretical contribution of this paper is the introduction of **Token-as-a-Service (TaaS) Procurement** as a procurement paradigm specifically designed for consumption-based AI. Unlike traditional procurement models, which assume that commercial decisions are largely completed before technology deployment, TaaS Procurement recognizes that AI expenditure evolves dynamically according to operational behavior. Procurement therefore becomes a continuous enterprise function that extends beyond supplier selection and contract negotiation to include demand forecasting, runtime governance, supplier portfolio management, commercial optimization, and business value realization.

A second contribution is the development of the **Enterprise Intelligence FinOps (EIFO)** framework as the governance model that operationalizes this procurement paradigm. Rather than replacing existing procurement practices, EIFO integrates procurement, finance, AI operations, enterprise architecture, and executive governance into a coordinated operating model. This integration extends FinOps principles beyond infrastructure cost optimization toward enterprise intelligence governance, while preserving procurement as the central organizational discipline responsible for commercial decision-making.

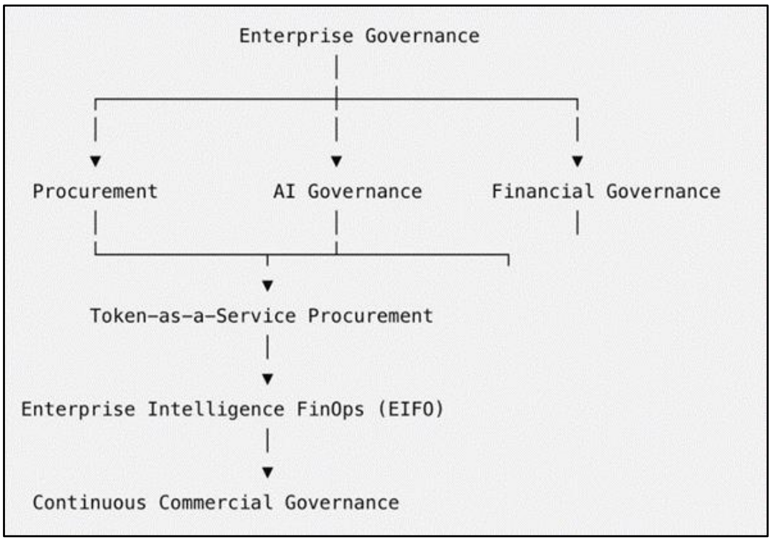

Figure 5 Positioning of Token-as-a-Service Procurement within Enterprise Governance

The quantitative procurement model presented in this paper constitutes a third contribution by introducing metrics that link operational AI consumption with procurement governance. Existing cost models generally focus on estimating expenditure, whereas the proposed metrics incorporate demand forecasting, supplier diversification, procurement efficiency, governance performance, and business value generation into a unified analytical framework. These measures provide a foundation for future empirical studies evaluating procurement maturity within enterprise AI environments.

8.2. Implications for Enterprise Procurement

The proposed framework has important implications for procurement organizations adopting Large Language Models at enterprise scale. Traditional procurement processes were developed for software products with relatively predictable pricing structures and contractual obligations. Consumption-based AI fundamentally alters these assumptions by introducing continuously variable expenditure driven by organizational behavior rather than fixed licensing agreements.

Consequently, procurement organizations must expand their responsibilities beyond sourcing activities. Procurement professionals require capabilities in demand forecasting, AI supplier portfolio management, commercial analytics, token consumption governance, and continuous contract optimization. Procurement decisions increasingly depend on operational intelligence generated after deployment, making collaboration between procurement, finance, enterprise architecture, and AI operations essential for effective governance.

The framework also encourages organizations to view AI suppliers as strategic partners within a continuously evolving ecosystem rather than isolated technology vendors. Commercial decisions therefore extend beyond selecting the lowest-cost provider to balancing performance, resilience, regulatory compliance, innovation, contractual flexibility, and long-term strategic alignment across multiple AI providers.

8.3. Implications for Enterprise Governance

Beyond procurement, the proposed framework contributes to broader enterprise governance by demonstrating that commercial, operational, and financial decision-making cannot be managed independently within consumption-based AI environments. Governance requires continuous information exchange between business stakeholders, procurement teams, AI engineering, financial management, and executive leadership.

The Enterprise Intelligence FinOps framework provides this integration by transforming operational AI telemetry into commercial intelligence that supports evidence-based procurement decisions. Instead of relying on periodic expenditure reports, organizations gain continuous visibility into supplier performance, demand trends, financial efficiency, and business value realization. Such transparency strengthens executive oversight while enabling procurement strategies to evolve as enterprise AI adoption matures.

The framework is also compatible with emerging governance standards, including ISO/IEC 42001 and the NIST AI Risk Management Framework, by complementing technical and ethical governance with commercial governance. In this respect, Token-as-a-Service Procurement addresses an organizational dimension that is not explicitly covered by existing AI governance frameworks.

8.4. Relationship to Existing Research

The proposed framework extends several existing research streams without replacing them. Cloud FinOps demonstrated the value of continuous financial accountability for cloud infrastructure, while strategic sourcing research established mature approaches for supplier management and procurement governance. Similarly, enterprise architecture research has emphasized organizational alignment, and AI governance frameworks have focused on trustworthy, secure, and responsible AI deployment.

Token-as-a-Service Procurement integrates these perspectives into a unified governance model for consumption-based AI. Rather than treating procurement, financial management, and operational governance as independent disciplines, the framework demonstrates that effective enterprise AI governance requires continuous interaction among these capabilities. This integration represents a natural evolution of enterprise procurement as AI transitions from a software product to an operational service consumed on demand.

```
Existing Research Streams
────────────────────────────────────────
• Procurement
• Cloud FinOps
• AI Governance
• Enterprise Architecture
           |
           ▼
Integration Through This Research
           |
           ▼
Token–as–a–Service Procurement
           |
           ▼
Enterprise Intelligence FinOps
           |
           ▼
Continuous Enterprise AI Governance
```

Figure 6 Research Contribution Model

8.5. Limitations

Several limitations should be considered when interpreting the findings of this research. First, the proposed framework is conceptual and has been evaluated through an illustrative enterprise scenario rather than longitudinal industrial implementation. Although the framework is grounded in established procurement principles and AI governance practices, empirical validation across multiple organizations remains necessary to assess its effectiveness under different commercial and regulatory conditions.

Second, the quantitative models introduced in this paper are intended to provide a generalized analytical foundation for procurement governance. Organizations operating in highly regulated industries or specialized technical environments may require additional variables reflecting sector-specific procurement requirements, regulatory obligations, or contractual constraints.

Third, the commercial AI market continues to evolve rapidly. New pricing mechanisms, autonomous AI agents, multimodal foundation models, and open-weight models may introduce procurement challenges that extend beyond the scope of the current framework. Accordingly, the proposed governance model should be viewed as an evolving foundation capable of incorporating future developments in enterprise AI procurement.

8.6. Future Research Directions

The framework presented in this paper establishes several opportunities for future investigation. The most immediate priority is empirical validation through longitudinal case studies involving organizations that have adopted Large Language Models at enterprise scale. Such studies would enable quantitative assessment of procurement efficiency, forecasting accuracy, supplier governance, and business value realization under real operational conditions.

Future research may also investigate procurement maturity models specific to AI, comparative analyses of multi-vendor procurement strategies, optimization algorithms for token allocation across heterogeneous AI providers, and decision-support systems that automate procurement recommendations using operational consumption data. Additional research could examine the interaction between Token-as-a-Service Procurement and emerging regulatory requirements governing AI transparency, accountability, sustainability, and cross-border data management.

Finally, the rapid emergence of autonomous AI agents introduces new procurement challenges. As organizations increasingly procure AI systems capable of independently initiating workflows and consuming computational resources, procurement governance will require new approaches for budgeting, authorization, accountability, and commercial risk management. Extending the proposed framework to support agentic AI therefore represents a promising avenue for future research.

9. Conclusion

The rapid adoption of Large Language Models (LLMs) [1]–[5] and other generative AI technologies is fundamentally transforming enterprise technology procurement. Unlike traditional software licensing or subscription-based Software-as-a-Service (SaaS) models, consumption-based artificial intelligence [10]–[15] introduces dynamic commercial structures in which enterprise expenditure is determined by operational usage rather than fixed contractual commitments. This transition requires organizations to rethink procurement as a continuous governance capability capable of managing evolving commercial, operational, and financial relationships throughout the AI lifecycle.

This paper addressed this challenge by introducing **Token-as-a-Service (TaaS) Procurement**, a procurement paradigm specifically designed for consumption-based artificial intelligence. The proposed paradigm extends traditional procurement beyond supplier selection and contract negotiation to encompass demand forecasting, supplier portfolio management, runtime commercial governance, financial accountability, and continuous optimization. By recognizing AI tokens as strategic enterprise resources rather than simple billing units, the framework establishes procurement as an ongoing enterprise function responsible for governing the commercial lifecycle of artificial intelligence services.

To operationalize this procurement paradigm, the paper proposed the **Enterprise Intelligence FinOps (EIFO)** framework. EIFO integrates procurement, finance, enterprise architecture, AI operations, and executive governance into a unified operating model that transforms operational AI consumption into commercial intelligence. This integration enables organizations to continuously evaluate supplier performance, refine procurement strategies, optimize commercial agreements, and align AI expenditure with measurable business outcomes.

The paper further contributes a quantitative procurement model that links enterprise demand forecasting, commercial cost estimation, supplier portfolio optimization, procurement efficiency, governance performance, and business value measurement within a unified analytical framework. Collectively, the conceptual framework, governance architecture, and quantitative models provide organizations with a structured approach for managing the emerging commercial realities of consumption-based AI services.

From a theoretical perspective, this research contributes to enterprise procurement, AI governance, and technology management by establishing Token-as-a-Service Procurement as a distinct procurement discipline for enterprise artificial intelligence. While previous research has addressed cloud financial management, strategic sourcing, and responsible AI independently, this paper demonstrates the need for an integrated procurement model that governs the commercial lifecycle of tokenized AI services. The Enterprise Intelligence FinOps framework extends these existing research streams by providing the governance mechanisms required to operationalize continuous procurement within enterprise AI ecosystems.

From a practical perspective, the proposed framework provides procurement professionals, enterprise architects, technology leaders, and financial managers with a governance model for managing increasingly complex AI supplier ecosystems. As organizations adopt multi-vendor AI strategies and autonomous AI capabilities become more prevalent, procurement will play an increasingly strategic role in balancing commercial performance, operational efficiency, governance compliance, and long-term business value.

Although the framework has been evaluated through an illustrative enterprise scenario, additional empirical validation remains an important direction for future research. Longitudinal case studies, cross-industry comparisons, procurement maturity assessments, and decision-support systems for automated AI procurement governance represent promising opportunities to further refine and validate the proposed model. Future investigations may also explore procurement strategies for autonomous AI agents, multimodal foundation models, federated AI ecosystems, and emerging token-based commercial mechanisms as enterprise AI continues to evolve.

In conclusion, this paper argues that the next evolution of enterprise technology procurement is not simply the adoption of new pricing models but the emergence of a new governance discipline. **Token-as-a-Service (TaaS) Procurement** provides the conceptual foundation for governing consumption-based artificial intelligence, while **Enterprise Intelligence FinOps (EIFO)** offers the operational governance model through which that vision can be realized. Together, they establish a comprehensive framework for managing the commercial lifecycle of enterprise intelligence in the next generation of AI-enabled organizations.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems (NeurIPS), vol. 30, pp. 5998–6008, 2017.

[2] T. B. Brown et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 1877–1901, 2020.

[3] OpenAI, GPT-4 Technical Report, 2023.

[4] M. Bommasani et al., "On the Opportunities and Risks of Foundation Models," Stanford Center for Research on Foundation Models, 2021.

[5] NIST, Artificial Intelligence Risk Management Framework (AI RMF 1.0), National Institute of Standards and Technology, Gaithersburg, MD, USA, 2023.

[6] OECD (2019) The OECD Artificial Intelligence (AI) Principles, oecd.ai. OECD. Available at: https://oecd.ai/en/ai-principles (Accessed: July 31, 2026)

[7] European Union (2025) The EU Artificial Intelligence Act, The Artificial Intelligence Act. Available at: https://artificialintelligenceact.eu/ (Accessed: July 31, 2026)

[8] Gartner (2023) Gartner Identifies the Top 10 Strategic Technology Trends for 2024, Gartner. Available at: https://www.gartner.com/en/newsroom/press-releases/2023-10-16-gartner-identifies-the-top-10-strategic-technology-trends-for-2024 (Accessed: July 31, 2026)

[9] Gartner, Inc (2026). Buying AI: A CIO's Guide to Vendor Decisions and Cost Control. [online] Govevents.com. Available at: https://www.govevents.com/event/24829/buying-ai-a-cio-s-guide-to-vendor-decisions-and-cost-control [Accessed 31 July 2026]

[10] Noble, J. (2024, December 13). Generative AI. IDC. https://www.idc.com/resource-center/generative-ai/?utm_source=google-ads&utm_medium=search-paid&utm_campaign=22870383810&utm_content=&utm_term=&utm_adgroupid=&utm_matchtype=&utm_network=x&utm_device=c&gad_source=1&gad_campaignid=22876160600&gbraid=0AAAAAoKe52tFPAfeAKZBWtr97F2Hean96&gclid=Cj0KCQjwjvfSBhDpARIsAEiOpSva9LGv4h3evI-MEA_HdyWeabvyzDmfCkXbj67lrNEoqt7w4gP82DIaApcZEALw_wcB

[11] Dutt, D., Beena Ammanath, Costi Perricos, & Sniderman, B. (2024). State of Generative AI in the Enterprise 2024 | Deloitte Global. In *Deloitte United Kingdom*. Deloitte. https://www.deloitte.com/ce/en/services/consulting/research/state-of-generative-ai-in-enterprise.html

[12] McKinsey & Company. (2025). The State of AI: How Organizations Are Rewiring to Capture Value. In *McKinsey & Company*. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

[13] R. Monczka, R. Handfield, L. Giunipero, and J. Patterson, Purchasing and Supply Chain Management, 7th ed. Boston, MA, USA: Cengage Learning, 2020.

[14] P. Kraljic, "Purchasing Must Become Supply Management," Harvard Business Review, vol. 61, no. 5, pp. 109–117, 1983.

[15] Ellram, L.M. (1990), The Supplier Selection Decision in Strategic Partnerships. Journal of Purchasing and Materials Management, 26: 8-14. https://doi.org/10.1111/j.1745-493X.1990.tb00515.x.

[16] Cox, A. (2004). The Art of the Possible: Relationship Management in Power Regimes and Supply Chains. Supply Chain Management, 9, 346-356. https://doi.org/10.1108/13598540410560739

[17] Competitive Advantage: Creating and Sustaining Superior Performance: Michael E. Porter: 9780029250907: Amazon.com: Books. (2026). Amazon.Com. https://www.amazon.com/dp/0029250900?lv=shuf&channelId=500&plpRedirect=mhFallback

[18] Porter, M. E. (1996). What is strategy? In *Hbs.edu*. https://www.hbs.edu/faculty/Pages/item.aspx?num=10698

[19] *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution ^ 8398*. (2026b). HBR Store. https://store.hbr.org/product/enterprise-architecture-as-strategy-creating-a-foundation-for-business-execution/8398?srsltid=AfmBOoo7M6rsn0Y4tD0Q2KNHmuSK8GXN9qctXZlEkbbmypLpZ_hi2bdq

[20] Microsoft. (2018). Pricing Overview—How Azure Pricing Works | Microsoft Azure. In *Microsoft.com*. Microsoft Azure. https://azure.microsoft.com/en-us/pricing

[21] Google Cloud, Vertex AI Pricing, Google Cloud Documentation, 2024 https://cloud.google.com/products/document-ai/pricing?hl=en

[22] Pricing. (2025). In *Claude Docs*. https://platform.claude.com/docs/en/about-claude/pricing

[23] Pricing | OpenAI API. (2026c). In *Openai.com*. https://developers.openai.com/api/docs/pricing

[24] Introduction | Mistral AI Large Language Models. (2024). In *docs.mistral.ai*. https://docs.mistral.ai/

[25] T. H. Davenport and N. Mittal, *All-In on AI: How Smart Companies Win Big with Artificial Intelligence*, Harvard Business Review Press, 2023.