The Impact of AI on Collaboration Application Pricing in Telco B2B

The inevitable and extensive infusion of AI into collaboration application workloads such as Unified Communications-as-a-Service (UCaaS), Contact Center-as-a-Service (CCaaS) and Communications Platform-as-a-Service (CPaaS) will have a significant impact on the pricing models for each of these services. As an increasingly strategic element of telco B2B portfolios, telco B2B teams will need to work closely with both software vendors and enterprise customers to adapt to a new, dynamic, real-time pricing reality with a shift from traditional per-seat pricing to on-demand consumption-driven pricing. The cost and complexity of deploying AI infrastructure in terms of Graphics Processing Units-as-a-Service (GPUaaS) and the associated software, combined with the ephemeral and unpredictable nature of some AI application services, are driving a new era of consumption-driven offerings.

That shift is twinned with both an opportunity and a drive towards value or outcome-based pricing as enterprise organizations seek greater proof of value from their supply chain.

Telco B2B commercial teams should be building short-term commercial strategies that focus on:

- Working with the "as-a-service" (XaaS) software vendors to both accelerate this transition and to assess how their pricing developments and changes should be integrated into telco B2B integrated collaboration offerings.
- Working closely with their internal IT teams to demonstrate the importance of this shift in pricing to their competitiveness and the need for a charging/billing capability and strategy that can deliver this new development at scale.
- Taking a big picture review of the role of dynamic, real-time, on-demand pricing and end-user observability across the complete B2B portfolio. AI will be a major change agent in this space, but it also comes at a time when enterprises are much more demanding of suppliers for greater commercial transparency, insights and observability around their offerings. The adoption of dynamic pricing and value-based outcomes are key to meeting this need.

The Shifting Economics of AI in Enterprise Platforms

AI workloads are not just computationally expensive; they are unpredictable. An AI-enabled query could generate a one-line summary of a knowledge-based article, the full transcript of a customer service call or a detailed multi-paragraph report of a customer's behavior and activity across multiple channels. These different activities consume vastly different resources.

These developments will directly impact the deployment of collaboration applications and telco B2B teams must now factor in:

- High variability of compute: GPU utilization is critical and varies widely across the "AI lifecycle," from training to inference to massive parallel processing of data and tasks.
- Third-party infrastructure costs: Applications providers might incur real-time costs from AI providers, large language models (LLMs) and clouds like OpenAI, AWS, Azure or Google Cloud Platform (GCP).
- Spiky, job-based usage patterns: AI services aren't used evenly; they spike based on manual user or batch activity, time of day, day of week or seasonal demand.

The fundamental model for charging/billing for AI-based services via LLMs or agentic AI is via tokens, where both input and output activities are charged based on the amount of information processed, which could be textual, graphical or audio-based. As many telcos will be reselling or white-labelling collaboration apps, they will need to consider the whole value chain flow. This stretches from the infusion of variable costs from those suppliers, how service margin is layered on by B2B teams and, finally, how that is passed on in near real-time to enterprise customers to maintain that goal of transparency, observability and pricing for business value outcomes, not just technology transactions.

The Coming Wave of AI Bill Shock

While token-based AI pricing will help accelerate adoption by making it easy for customers to experiment, it also introduces new risks, foremost among them, AI bill shock. As close observers of the telco industry can relate from experience, early adopter enterprises may find that these initial token-based offers obscure rather than clarify cost. Customers with pre-purchased bulk credits are not likely to understand how quickly they are consumed by complex queries, data processing or autonomous agent actions.

Anecdotal evidence from early deployments suggests a common theme: usage explodes, bills follow and customers are caught off guard. The lack of transparency into the actual cost per AI action, especially in agent-based models,

makes it difficult to forecast spend or assess ROI. Worse, many enterprises may discover that what appeared to be an efficiency play (e.g., replacing call center agents with AI) may not yet offer a cost advantage, particularly in regions with lower labor costs.

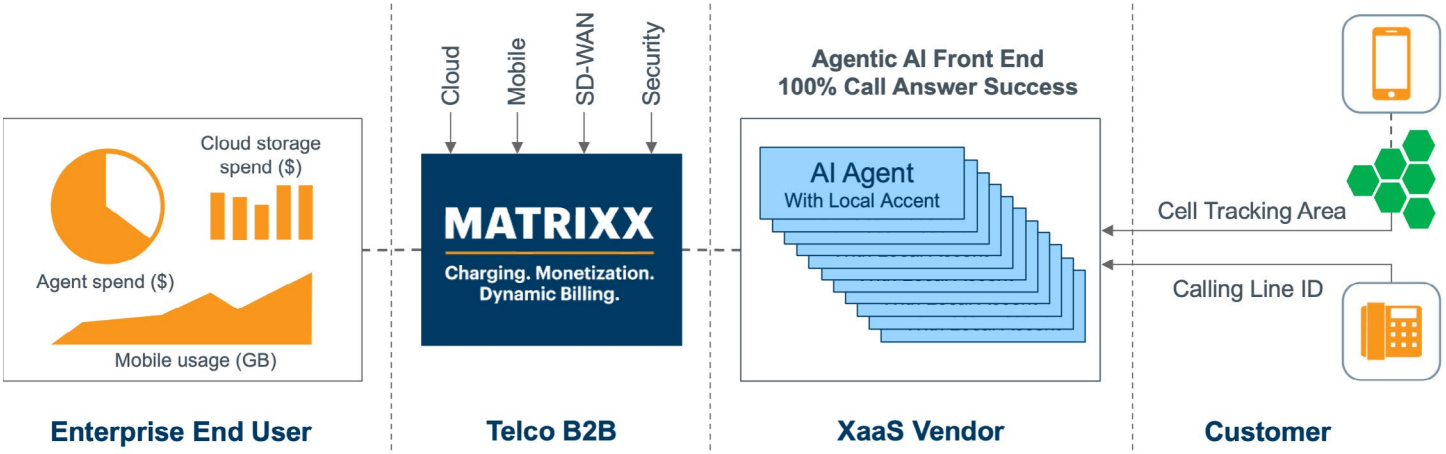
This pricing opacity will likely affect the software providers themselves, who will have to account for real-time usage fees from OpenAI, AWS, GCP or other infrastructure and LLM partners and pass this on to telcos. In this context, real-time monetization systems aren't just about billing; they're a strategic necessity. The ability to meter usage, calculate cost and convey pricing in real-time helps prevent customer backlash and ensures that providers can maintain margins while scaling AI services.

Example Use Case

Consider a small to medium-sized business who relies on inbound calls for a significant percentage of their revenue. They want to maximize the efficiency and effectiveness of their call reception to minimize lost calls, reduce time in queues or reroutes to voicemail, and decrease customer frustration.

With the deployment of agentic AI-based CCaaS, telcos could build out a "100% answer success" guaranteed call reception offering, where no call gets routed to voicemail or sits in a queue, but always gets answered by an agent. Furthermore, based on calling line ID or cell site area ID for mobiles, those calls can be routed to "local accent" AI agents to further improve customer experience.

Example Agentic AI Collaboration Application Use Case



Each call will be rated and a charge raised by the software vendor based on call duration and interaction with the agent, plus any additional features (such as local language treatment and no call lost guarantee), resulting in several tokens being charged to the telco by the XaaS vendor.

Telcos will need to ingest this real-time data per customer and transaction and collate it into bigger picture costs and charges for that customer and, crucially, show that data in real-time to the enterprise via a dashboard or similar. The changing psychology around AI-centric apps in terms of enterprise buyer expectations set by the broader market will dictate that telcos must take this step if they are to deliver competitive, differentiated offerings to the market.

This use case outlines three key elements in the way B2B services are delivered:

- It outlines an example of value-based outcomes and pricing for value. Offering a "no call lost" guarantee is of real business value to several enterprises, particularly small to medium enterprises. Telcos can then market price that service based on conditions, opportunity and unique value.
- The token-based model of AI pricing is here to stay and will be dynamic and consumption-oriented, representing a significant change from current per-seat type software models. How telcos adapt and represent that to their enterprise end-users through transparent, real-time reporting will be crucial to their success. Enterprises will become more expectant of this consumption model as the market shifts. Telcos, and particularly B2B teams, need to adapt to this model quickly.
- This is a significant change for telco B2B teams who have typically been used to per-seat licensing models for software. Early and rapid consultation with both the software supply chain and, crucially, internal IT teams responsible for charging and billing infrastructure will be at the heart of early breakthrough offerings in this space. This market won't wait, and enterprises will be watching and influenced by broader market developments.

The MATRIXX Role

Telco B2B teams must adapt quickly to new, dynamic, consumption-oriented pricing and billing models based on value outcomes and deliver that with transparency, accuracy and real-time observability to their enterprise customers. MATRIXX Dynamic Billing is a patented real-time rating, charging and monetization engine that has supported flexible dynamic and real-time commercial value chains across complex B2B portfolios in telcos across the globe, covering consumption, subscription, recurring and contract-based business models. By adopting dynamic, value-based pricing models in support of AI-centric workloads and applications, telco B2B teams will have the opportunity to drive differentiation, opportunity and growth across the enterprise market.

About MATRIXX Software

MATRIXX Software delivers a dynamic billing, monetization and charging solution proven at scale. Global service providers like Telefónica, IoT providers like Tata Communications and network-as-a-service providers like DISH rely on MATRIXX to overcome the limitations of existing billing applications. MATRIXX provides a unified platform that transforms and simplifies billing operations across consumer, enterprise and wholesale businesses. With MATRIXX, operators can rapidly configure, deploy and monetize personalized offerings, enabling commercial innovation and real-time customer experiences that drive revenue and growth.

matrixx.com