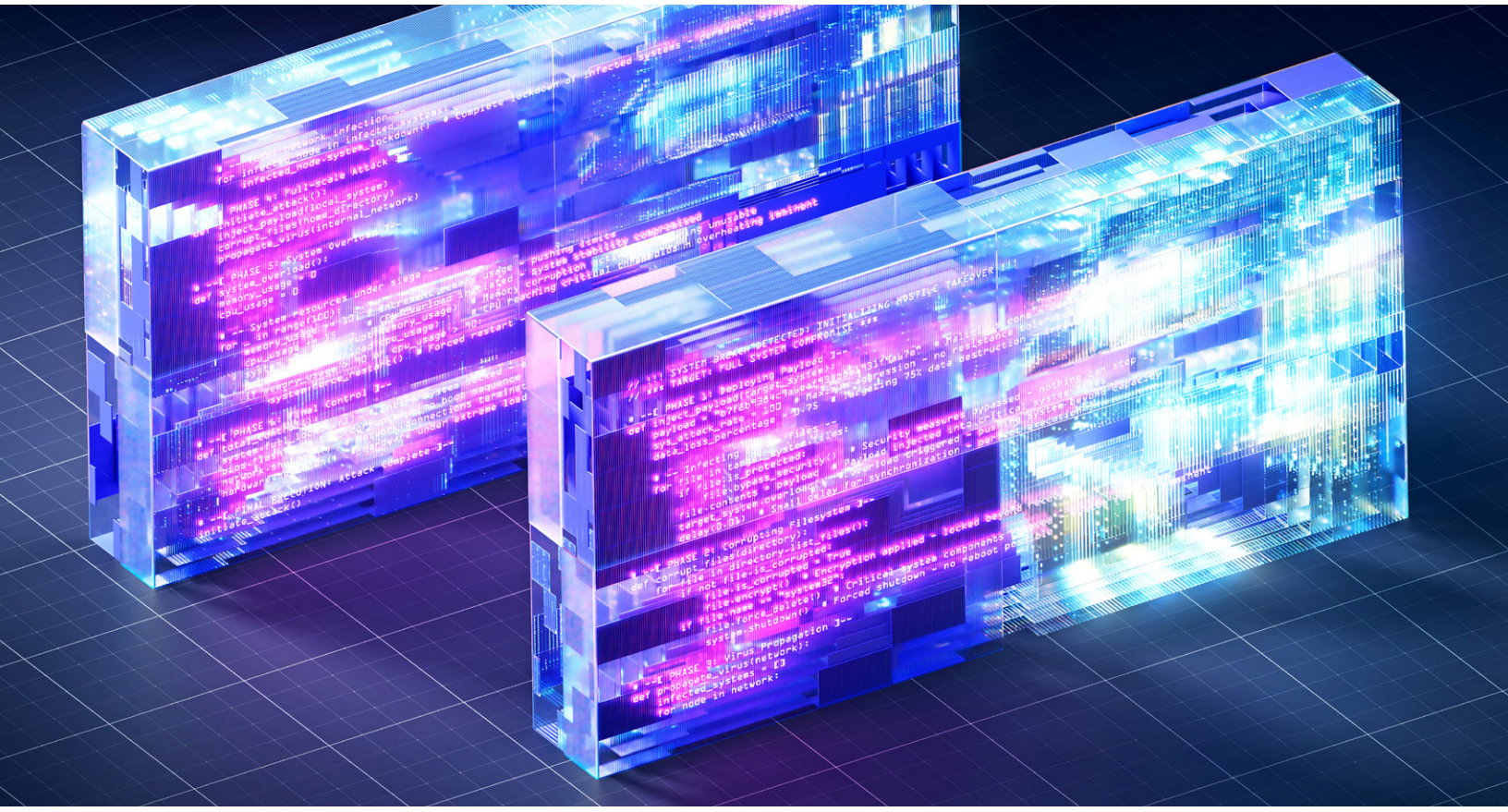
McKinsey & Company

Technology, Media & Telecommunications Practice

Upgrading software business models to thrive in the AI era

AI is transforming software from a tool that enables work into a platform that actively performs and orchestrates work. How should industry players evolve their business models to lead in this new era?

by Mohit Khanna and Nimish Mittal
with Ashley Wu and Marc Castillo



September 2025

Roughly three years since gen AI emerged as the most disruptive force in enterprise software since the dawn of software-as-a-service (SaaS), companies are racing to launch agentic use cases and embed AI in every workflow. Whereas previous shifts in the industry have involved clear technological transformations, this burgeoning era is more combinatorial; AI+SaaS adds the intelligence and automation potential of AI with the scalability and accessibility of cloud-based software. But while many software players are bringing AI capabilities to existing products and rapidly devising new ones, the most suitable business model is still in question.

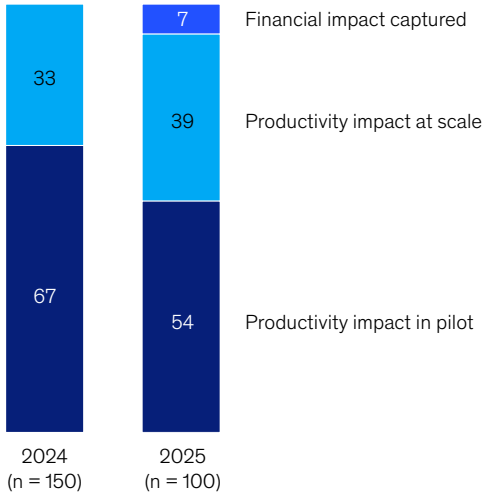
There is little doubt that the potential opportunity is massive: Previous McKinsey research indicates that as much as $4.4 trillion of incremental economic potential could be generated from AI-driven increased productivity.[1] The total addressable market for AI+SaaS has expanded beyond IT budgets to include labor—and momentum appears to be building. One year ago, 33 percent of companies surveyed saw productivity impact at scale or were already capturing financial impact from AI. In 2025, 46 percent of companies are (Exhibit 1).

Even as customer traction accelerates, a critical question remains: How will leaders evolve their business and go-to-market (GTM) models to thrive in this new AI+SaaS era, as they have needed to in previous shifts?

Exhibit 1

Many companies that have experienced productivity impact in AI pilots are now seeing it across the organization, with some capturing financial impact.

Enterprise AI implementation impact, by type, 2024–25[1]
% of companies



[1]For companies with annual revenues exceeding $1 billion.
Source: McKinsey Enterprise Tech Buyer Survey, North America and Europe, Feb 2024; McKinsey Enterprise Tech Buyer Survey, North America, May 2025

McKinsey & Company

[1] "The economic potential of generative AI: The next productivity frontier," McKinsey, June 14, 2023.

This article aims to answer that question, based on an analysis of pricing models and success metrics from 150 global vendors (both incumbent SaaS players and AI natives), and conversations with more than 50 companies launching AI products. It examines monetization challenges that AI providers have already encountered and the actions and strategies required to capture the full growth potential of this emerging technology.

Lessons from initial AI monetization efforts

Based on earnings reports from several SaaS leaders, there have been some early signs of traction with AI monetization, such as the nearly two million paid users GitHub has reported for its Copilot.[2] But while global enterprise spending on AI applications has increased eightfold over the last year to close to $5 billion, it still only represents less than 1 percent of total software application spending.[3]

Three consistent challenges underlie the slower-than-anticipated growth in AI software monetization:

— *Value communication and realization:* Many companies highlight potential use cases for AI, but only 30 percent have published quantifiable ROI in dollar terms from real customer deployments. Companies that do this well, such as Salesforce's Agentforce ROI Calculator—which demonstrates how AI-agent-led customer service inquiry handling has a quantifiable cost savings versus human agents—leave less to the imagination to close a deal or need to be proved as part of a pilot. At the same time, many companies are experiencing rises in IT costs without yet being able to make corresponding decreases in labor costs. For instance, enabling AI across the full customer service tech stack of a typical organization could result in a 60 to 80 percent increase in list prices. "All of these copilots are supposed to make work more efficient with fewer people, but my business leaders are also saying they can't reduce head count yet," said one HR executive at a Fortune 100 company. Other companies are finding that low-cost labor still undercuts the financial case for AI.

— *Price predictability:* Customers want to understand how AI costs will scale with usage, but many current pricing models are complex and opaque. One CFO of a Fortune 500 company described the problem: "It is frustrating that I have no idea what we're going to spend on AI this quarter. My business units have no forecast of what they are going to use, and it is spread across tens of software vendors. By comparison, my spend on cloud computing is also usage-based, but it's predictable because of the structure of our negotiated buys."

— *Sustained adoption (postpilot):* Even when pilots go live, many fail to scale due to underinvestment in change management. Our experience from successfully scaled pilots suggests that a good rule of thumb is that for every $1 organizations spend on model development, they should expect to have to spend $3 on change management (such as forward-deployed engineering, employee user training and reinforcement, and standing up performance monitoring).[4]

[2] Satya Nadella, Microsoft Annual Report Shareholder Letter 2024, October 18, 2024.
[3] Tim Tully, Joff Redfern, and Derek Xiao, "2024: The state of generative AI in the enterprise," Menlo ventures, November 20, 2024; Chandra Gnanasambandam, Ari Libarikian, and Cem Turkeli, "The SaaS factor: Six ways to drive growth by building new SaaS businesses," McKinsey, July 19, 2022; "S&P Global Market Intelligence foresees rapid expansion of generative AI software market by 2028 to $52.2 billion," S&P Global, June 6, 2024.
[4] "Moving past gen AI's honeymoon phase: Seven hard truths for CIOs to get from pilot to scale," McKinsey, May 13, 2024.

Adapting SaaS for the AI+SaaS era

As AI+SaaS products increasingly perform instead of merely support work, the new era calls for a business model that aligns customer value with units of work completed. Consumption-based models are a natural fit, a flexible and seemingly fair way to monetize the wide range of customer value that AI generates. This is especially important as the growth in human users (and their associated "subscription user seats") may slow and AI takes on a greater share of the work.

The other force for change is customer expectations: well-funded AI-native start-ups are resetting the bar with dynamic, usage-aligned pricing models that offer buyers greater control, transparency, and scalability—putting pressure on incumbents to follow suit.

As a result, leaders in monetizing AI capabilities typically start this transition by focusing on key pricing design choices:

— *Pricing model:* How should software companies evolve the pricing model to incorporate the range of possible AI actions in a portfolio?

— *Pricing meter:* What is the "pricing unit" that organizations should use to best reflect the effort and value of each AI action?

— *Scaling:* How should providers scale customer spend across more volume and across the portfolio?

— *Price levels:* How should vendors set price levels when the cost of inferencing is dropping rapidly? How should they balance value capture with scaling adoption?

Then, as the software provider pilots changes with a small number of products and customers, they can begin to address cross-functional execution challenges that revolve around three main areas: GTM (what happens to the traditional sales-success-renewals motion?); product, IT, and billing (is the tech stack ready to deliver consumption-based value?); and finance and investor relations (how should they track results and explain them—and strategy—to investors?).

Asking and answering pricing design questions

As software providers start to capture the massive potential from commercializing AI+SaaS offerings, they should examine several fundamental questions about how to optimally monetize the new technology.

Pricing model: How should software companies evolve the pricing model to incorporate the range of possible AI actions in a portfolio?

It's unlikely that the traditional per-user monthly subscription model will disappear entirely in this new era, since it remains a simple and effective way to package multiple capabilities and align with familiar procurement workflows. But as AI capabilities scale, software incumbents will likely need to incorporate some form of consumption-based pricing into the business model mix.

In the near term, consumption-based pricing raises the bar for providers because unlocking monetization requires clear value delivery and customer usage. However, in the long term, it can create the potential for substantial upside. If monetization of new AI products can scale with usage, companies will create a "land and expand" organic growth motion. As consumption

models become more popular in the marketplace, customers' tolerance for paying for unused user seats, or "shelfware," will also decrease.

Many companies are starting with hybrid models. When adding a consumption element to a subscription model, they must first decide how to treat "additional" consumption that goes beyond a capacity cap. With AI functionalities becoming table stakes in the business, turning to a consumption model for additional usage allows companies to satisfy customer expectations and monetize power users. "Additional" consumption can be treated in one of two ways.

'Buckets' for additional usage: Companies focused on their power users tend to use this model. HubSpot, for instance, favors this approach, including AI-powered features such as lead scoring within most of its subscription plans. A certain number of HubSpot credits, ranging from 500 to 5,000, is included, depending on the tier. Once they exceed their allotted amount, customers can purchase additional credits in 1,000-unit increments as their usage scales.

Metered throughput: As used in ChatGPT Enterprise/Business plans, this approach employs dollar-per-user pricing, but limits the number of tokens processed daily, weekly or monthly for certain models. Users who exceed their quota can still access the product, but are temporarily downgraded to lower-powered models.

Even subscription models are evolving in this new era. For many modern SaaS products, where work can be done through AI, automation, or integrations of the two, value is not highly correlated to how many people log in. In these cases, platform fees untethered to customer user counts can feel more linked to value, and they are becoming more common (Exhibit 2), especially among AI natives. For example, Clay, an automation platform for lead enrichment and prospecting, charges enterprise-wide subscription flat fees regardless of the number of users.

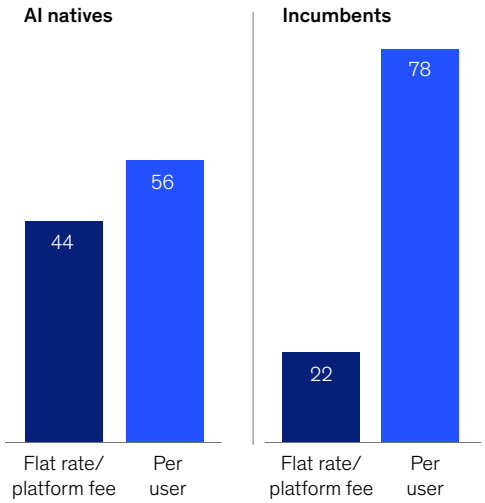
Given the rapid pace of AI innovation, companies with hybrid models will need to revisit their choices frequently. Capabilities that are cutting-edge at launch can quickly become standard issue. For example, Zoom's AI transcripts and summaries features were first monetized as separate add-ons or only part of higher-tier plans. These features are now included in all paid plans at no additional cost.

For AI+SaaS portfolios that enable a broad range of tasks with varying degrees of workload intensity, a more consumption-centric approach may be a better fit. By maintaining a subscription model for the legacy core portfolio but monetizing differentiating AI features on a consumption basis, vendors can achieve a more scalable model that is better aligned to incremental value.

Exhibit 2

Led by AI-native vendors, software companies are increasingly moving away from subscription fees based on the number of users.

Predominant subscription pricing model for software vendor, by company type, 2025,
% of companies



McKinsey & Company

For instance, ServiceNow offers a separate product, Now Assist, that is priced based on volume of "Assist" credits; Salesforce's Agentforce is similarly a stand-alone SKU, available in multiple pricing models: pay-per-action, bank of Flex Credits, or per-user licensing. As AI agents automate increasingly complex tasks, customers' usage (and thus their derived value) can vary significantly and grow rapidly. Consumption-based approaches can seem "fairer" to customers when the number of potential activities AI conducts is vast—for example, when the work required to dub a 30-minute video is more than the work required to remove the background from an image. Likewise, the range of marginal costs incurred can be too large for a flat-fee structure.

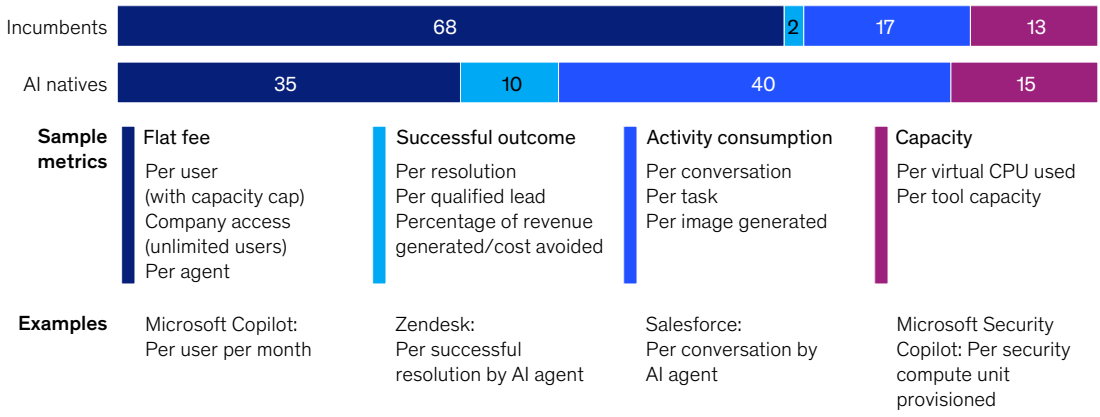
Pricing meter: What is the 'pricing unit' that organizations should use to best reflect the effort and value of each AI action?
Choosing to switch to a consumption pricing model, in whole or in part, is just the start of the process for a software provider embracing AI+SaaS. Picking the right pricing metric—the one most aligned to customer value—is just as critical, and complicated, a decision (Exhibit 3).

Activity-based metrics are the most common and practical to start with. The best activity-based metrics are those that can be linked to business value generated rather than resources expended. Compare two AI sales development representative (SDR) agents: one monetizes based on effort expended (per thousand emails sent) and another charges per sales development workflow completed (leads identified). The latter is closer to the business value that matters to the sales organization. True outcome-based pricing would be one step further, adding a layer of quality or success of the leads generated, or per qualified lead.

Exhibit 3

In the emerging, AI-driven software era, providers have a wide range of consumption pricing metrics to build their businesses around.

Software pricing meter type and sample metrics, by company type, %



McKinsey & Company

While outcome-based pricing can sound (and may ultimately be) ideal, it is often challenging to execute. Few companies have yet to adopt this approach. Making it work at scale requires several key factors: the outcome being monetized must be highly correlated to the work the product does; the lag time between the workflow execution and the outcome should be low; customers should value the outcome roughly equally; and it must be possible to confirm success automatically, without human assistance or judgment.

One AI-native SDR company offered customers the option of pricing models that are outcome-based (per qualified lead) and activity-based (per prospective customer outreach), and found that 90 percent of customers opted for activity-based. Forecasting spend and negotiating the definition of "qualified lead" were different in each customer context, leading to a cumbersome negotiation process for both the company and its customers.

Successfully commercializing an outcome-based metric takes no small effort. For example, Zendesk, which charges $1.50 per successfully resolved customer interaction, tested for edge cases in monthslong trials, built customer- and internal-facing telemetry, and developed a seven-step flowchart to determine whether each customer interaction could be successfully resolved.

Companies figuring out the most suitable pricing metric have other important decisions to make: Which specific parts of the agentic process, including AI-agent consumption, should be monetized—put another way, what is the monetizable unit? In the era of agentic AI, agents feature new capabilities and use existing system-to-system tools (such as APIs and data transfer). For example, an agent that can book business travel can consider customer preferences and constraints, which is a new capability. But it also uses existing hotel and flight booking APIs to execute the workflow. Software providers need to decide whether to monetize separately all the

tools agents call upon to execute workflows. Generally, when the workflow of the agent is highly customizable, companies charge separately for agent execution and the tools the agents use. When the workflows are standard, there is typically a charge only for agent execution.

Companies should also consider portfolio-wide coherence of the pricing metric. For company portfolios with a large range of possible AI actions, it is important to define and communicate one coherent metric across the portfolio. ServiceNow has Assists, Zendesk has resolutions, and Salesforce has Flex Credits based on actions. To reflect that not all AI actions are equal in terms of value and effort, the number of credits decremented for each action can be adjusted depending on cost intensity and business value generated.

Scaling: How should providers scale customer spend across more volume, and across the portfolio?

Beyond the pricing metric, companies offering a consumption model need to decide how they want to scale customer spend alongside usage. In a linear model, each incremental unit of usage costs as much as the prior. In a consumption-based tier, each "tranche" of usage is associated with a differential rate (Exhibit 4).

These decisions aren't just mechanical. They shape revenue predictability, customer upsell, and product usage behavior. Broadly, companies with an assisted sales motion tend to use volume discounts, which align well with negotiated enterprise deals and provide incentives for large, up-front commitments. Consumption-based tiers are more common in self-serve or product-led growth motions, where pricing needs to be transparent, progressive, and flexible enough to encourage organic adoption.

Regardless of the scaling model a company chooses, it also has to make an important structural choice regarding how to set up buying programs to encourage greater usage. Key elements are flexible payment terms, fungibility across products, and fair overage treatments.

— *Payment model:* Nearly all software players with consumption pricing models offer both pre-commit (paid up-front for a certain usage level) and flexible commitment structures (a pre-agreed-upon level of usage, paid as the usage is incurred). Differing discount rates depending on commitment level can help balance customer choice and maintain revenue predictability.

— *Fungibility:* In a recent survey of purchasing decision-makers, 65 percent of buyers said that exchanging usage or spending commitments from one product to another was very or extremely important.[5] To manage growth priorities while satisfying this customer expectation, some companies offer conditional fungibility (limited to certain product families, for instance).

— *Overages:* Leaders are shifting from charging overages immediately to offering "true forward" mechanisms, adjusting future commitments based on actual consumption in the prior period.[6] At the same time, enterprise license agreements that provide customers with discounted and predictable long-term pricing should also be reevaluated more frequently (every six to 12 months, for instance, rather than every three to five years) to adjust based on actual realized usage.

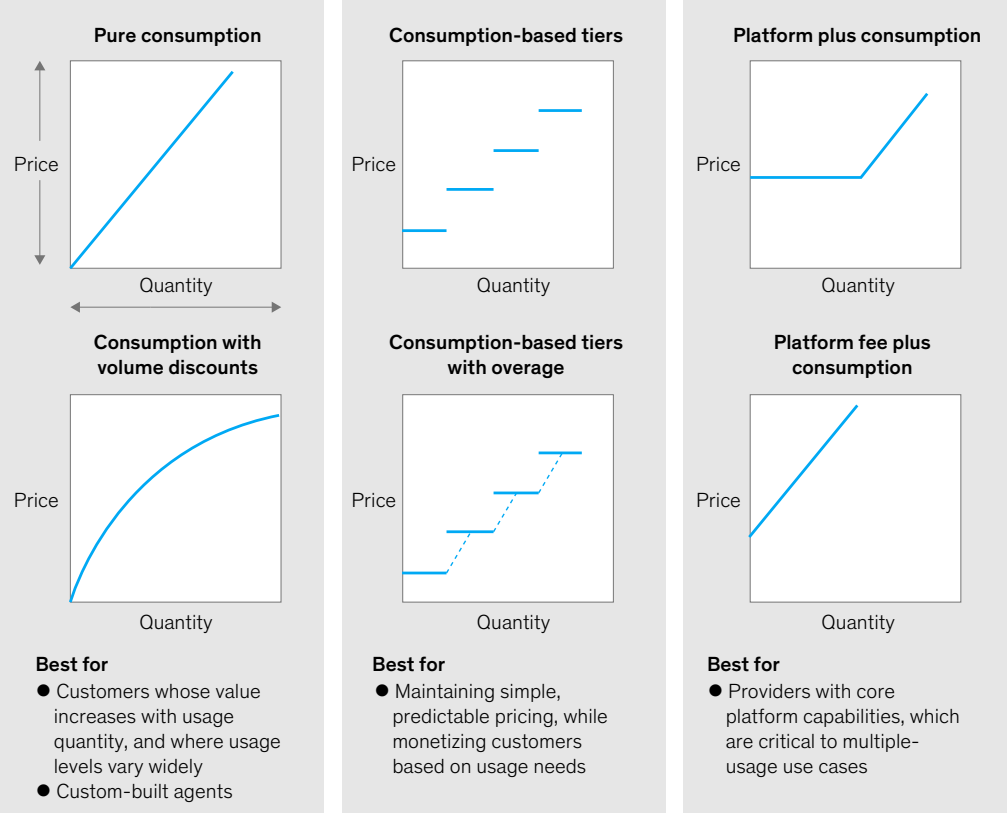
[5] McKinsey Enterprise LOB and IT Software Buyer Survey, October 2024, n = 150.

[6] In a true-forward structure, if committed spend was $100 in the period and actual usage was $105, the customer still pays $100, but committed spend for the next period is adjusted to $105.

Exhibit 4

Figuring out the best consumption pricing model to monetize customers as their usage grows is a critical decision in the new AI software era.

AI+software-as-a-service (SaaS) consumption pricing model type, by best fit (illustrative)



McKinsey & Company

Price levels: How should vendors set price levels when the cost of inferencing is dropping rapidly? How should they balance value capture with scaling adoption?
Figuring out the appropriate price levels is one of the most challenging, uncertain aspects of the AI+SaaS business model. The cost of goods sold in SaaS has typically been de minimis. In this new era, those direct, variable expenses can be quite substantial, meaning companies will have to balance answer and workflow fidelity with the cost of compute consumption. As part of these efforts, they will continue experimenting with how best to deal with countervailing trends on declining cost of computational delivery versus the increasing complexity and quality of AI actions.

Declining cost of delivery: Amid declining costs of large language model (LLM) delivery (more than 80 percent per year over the past two years), solutions where the LLM is the primary cost driver could see price levels come down or capacity increase. Take Jasper AI: in 2022, it capped usage and monetized additional text generation at $10 per 5,000 words. In 2023, users now receive unlimited access to text generation as part of the standard plan.

Increasing complexity of agent actions: Conversely, AI agents are now executing more complex, multistep workflows that often involve external API calls, database queries, and interactions with third-party services. Price levels for such offerings may go up, as use cases move toward multistep autonomous workflows.

Balancing value capture and adoption: As in the prior era of cloud migration, many AI+SaaS companies are finding that they need to lower the barriers to trial to increase adoption and long-term value capture. Two strategies are proving effective. One is to subsidize or offer free initial usage allocations for AI capabilities. For one IT SaaS player, customers with traditional enterprise agreements receive enough free credit allocation to activate a few lighthouse use cases. Once customers adopt and see value, the company works to upsell to a higher allocation for additional use cases.

Another approach is to start by including AI capabilities in core products, then move toward a stand-alone, monetized AI offering as capabilities mature. Adobe's journey is an example of an evolution that many companies may also experience. At first, it embedded capabilities for free within the existing subscription, providing blocks of usage ($5 per additional 100 credits) for customers that exceeded their free allotment. Then, by the start of this year, as customer traction grew, it shifted to monetizing its AI capability as a separate SKU, adding video generation and translation capabilities and marketing a range of capacity-capped tiers ($10 to $200 for 2,000 to 50,000 generative credits) after a customer had exhausted an initial free 1,000 credits.

Adobe exited the first quarter of 2025 with $125 million in revenue from stand-alone AI products. While still a very small share of the company's record $5.7 billion in total revenue that quarter, the company expects AI-fueled business to double by the end of the fiscal year. The broader market suggests that making this journey quickly is rewarded. Although only 16 percent of SaaS incumbents have commercialized AI applications as stand-alone products, those that have report two to three times higher customer traction and revenue.

As in the prior era of cloud migration, many AI+SaaS companies are finding that they need to lower the barriers to trial to increase adoption and long-term value capture.

Cross-functional execution keys in the new era

The shift to monetize AI and incorporate consumption-based business models involves the entire organization, far beyond just pricing. Go-to-market, product, IT/billing, and finance and investor relations (IR) are a few of the most highly impacted functions.

GTM: What happens to the traditional sales-success-renewals motion?
Traditional hand-offs between sales, customer success (CS), and renewals are dissolving. Many companies now treat their account executives as the "quarterback" to own the full customer journey—selling, onboarding, and managing usage. Account executives in these models are responsible for both the initial land and the subsequent expansion, supported by specialized onboarding or solution teams. Others maintain a more traditional role separation but add commercial goals to their CS teams. Unlike traditional subscription SaaS, where annual contracts smooth revenue predictability, consumption models require ongoing customer engagement and value delivery to drive continuing spending, which necessitates role changes and compensation KPIs.

Sales compensation is moving away from large up-front payouts at deal close. Some vendors pay on pre-commits with clawbacks for underuse, while others pay as customers consume. For the second model in particular, the sales organization's mindset needs to shift from closing big deals (often late in the quarter) to enabling ongoing customer success, including early-quarter activation, to allow time for usage to accrue and to boost in-account expansion.

Under subscription models, customers only had to worry about how many user seats they needed, and companies had a relatively simple job of forecasting future revenue. In this new era, *roles such as implementation or forward-deployed engineers and sales engineers* that work on-site with customers are more critical than ever. They not only have to help guide customers to the right products and use cases, but they are also essential for predicting account-level usage growth, both for customer expectations and internal forecasting.

The buyer itself is also changing. As AI starts performing and not just supporting work, even more purchasing decisions are shifting from IT to business units. Line-of-business leaders are increasingly making budget trade-offs between head count investment and AI deployment, and they expect sellers to engage them on value and outcomes, not just features.

Product, IT, and billing: Is the tech stack ready to deliver consumption-based value?
Consumption models require new infrastructure—especially around telemetry and billing. Many companies must build customer-facing usage tracking from scratch, including real-time visibility (with clear translation of usage to cost, and spend prediction tools). Financial operations best practices (such as cloud usage dashboards and usage calculators) offer strong templates. Internally, account teams can use granular usage data to spot trends or warning signs and intervene early if necessary. The same systems can help instrument sales compensation based on usage. On the back end, billing systems must support usage mediation, cost translation, invoicing, and revenue recognition—capabilities often missing from traditional SaaS stacks.

Finance and IR: How to track results and explain them (and strategy) to investors
Finance: Whereas subscription models can easily calculate forward-looking metrics such as annual recurring revenue and annual contract value, companies need to identify different consumption indicators that reflect a positive revenue trajectory (including cohort revenue growth and active customer count growth). Accounting must also ensure that pricing changes comply with revenue recognition rules.

Investor relations: A clear investor narrative is essential when starting to incorporate consumption models. Market confidence hinges on handling two questions: What short-term revenue dip is expected, and what leading indicators suggest long-term growth will materialize (for instance, lower churn, more multiproduct adoption)? And how directly do usage metrics tie to revenue?

AI+SaaS is fast emerging as the next software "super cycle," but how (and how successfully) providers will monetize the technology is less certain. As companies wrestle with the key questions and shifting models of this new era, players that can quickly make initial design choices, test them with real customers, and iterate just as fast should thrive. Those that align pricing with customer value and go-to-market with usage activation will be best positioned to capture an outsize share of the potentially massive opportunity.

Mohit Khanna is a partner in McKinsey's Bay Area office, where **Marc Castillo** is an associate partner; **Nimish Mittal** is a partner in the New York office, and **Ashley Wu** is an associate partner in the Denver office.

The authors wish to thank Begum Erdogan, Brendan Gaffey, Brian Elliott, Illán García, Jeremy Schneider, Jonathan Shulman, Martin Harrysson, Naveen Sastry, Siddharth Shekhar, Tejas Shah, and Zachary John for their contributions to this article.

This article was edited by Daniel Eisenberg, an executive editor in the New York office.

Designed by McKinsey Global Publishing