EXECUTIVE INSIGHTS

What B2B SaaS Leaders Can Learn From How AI Labs Price

Business-to-business (B2B) software as a service (SaaS) pricing leaders can borrow several moves from the frontier artificial intelligence (AI) labs, but the greatest opportunity lies in a dimension the labs cannot fully reach: the value the product actually delivers.

A SaaS pricing leader ships an AI feature on Monday. By Friday, five power users have burned more compute than the rest of the customer base combined. Finance wants margin protection, sales wants a story simple enough to sell, product wants to ship the next model next week and no customer wants to reopen a contract. Four demands, one pricing decision. The frontier AI labs have been solving exactly this in public for the past 18 months, so what should other SaaS leaders learn from their success?

The labs' answer is to price a single account across three dimensions at once: the number of users with access, how much they consume and the capability tier they draw on. The price paid and package received are now a function of all three. This is not a quirk of a young category; it is the leading edge of where value-based pricing has been heading for years, forced open by one fact that makes AI different from classic SaaS: The cost to serve is variable and material, not close to zero.

Rather than copying the tactics wholesale, pricing leaders should study the principles underneath. All three of the labs' axes are proxies for value rather than value itself, and a narrower product can price closer to the outcome it delivers. This edition of L.E.K. Consulting's *Executive Insights* examines which of these moves transfer to B2B SaaS, which do not and where the advantage for focused software businesses lies.

1. A common currency lets you launch without repricing

The providers sell models with wildly different costs to serve, yet express them in a single currency: tokens or credits with an internal exchange rate. When a new premium model launches, there is no new negotiation. The customer already holds the currency and simply spends more of it.

Contrast the classic SaaS cycle: Repackage the tiers, reprice, renegotiate, wait two quarters. A consumption currency collapses that. Ship a capability, assign it a credit cost and monetize it on day one. The currency decouples product innovation from pricing friction.

2. Credits are an enterprise trial engine, not just a billing unit

Watch how platforms use user-directed pooled credits (e.g., Microsoft's Copilot Credits, Salesforce's Flex Credits). The buyer commits to a pool; end users spend it on whatever they value, including brand-new features. That does three things at once: It removes procurement friction from trialing a new product, it generates bottom-up willingness-to-pay data as spend reveals what users value and it turns a top-down seat sale into usage-led expansion.

This reframes one of the hardest problems to solve in enterprise go-to-market: getting a new module adopted. Rather than selling each one, you seed a currency, let consumption surface demand and then price against what you learn.

3. Protect the margin tail

“Unlimited” is a margin trap when the cost to serve is real, because a small share of heavy users drives a disproportionate share of costs. The providers have responded with metered limits on unlimited plans (e.g., Claude's weekly limits, Google's compute-based caps) and asymmetric metering that steers behavior toward cheaper, stickier patterns (e.g., OpenAI pricing output tokens above input).

As you embed AI features, your gross margin is now usage sensitive, so pricing needs a cost-defense mechanism alongside its value-capture role. The heaviest 5% of users are a margin question, not just a support one.

4. Make repricing a deliberate posture

Model costs are falling fast, and providers have cut prices to match (Anthropic and Mistral among them). That brings up a question every SaaS business with AI in its cost base will face: When your input costs drop, do you capture the saving as margin or pass it through to defend share? Treat it as a conscious choice rather than a default. There is a quieter lever too: Hold headline prices while adjusting the exchange rate.

5. Reserve capability, not just volume, for the top

Capability-based tiering, where the best models sit in higher tiers, gives customers a reason to upgrade beyond “more of the same,” and a premium power-user tier captures heavy-user value without raising the entry price. Differentiate tiers on what the product can do, not only on how much of it the customers get, and keep the entry point accessible while still capturing the top of the willingness-to-pay curve.

In practice, the tier gate works best when it leaks a little: most labs give lower tiers a metered taste of the flagship (ChatGPT's free plan allows a handful of top-model messages before falling back to a lighter model; Gemini's free tier includes a capped daily allowance of its most capable model). That is product-led growth at work: users who have never seen a stronger model's output struggle to perceive its incremental value, so a capped dose of the best model is often the cheapest way to sell the upgrade while containing cost to serve.

6. Price the outcome — the dimension the labs cannot reach

The three axes the labs price on are all proxies for value. A token carries no inherent outcome; the same unit of consumption might draft a memo or clear a regulatory filing, so the labs cannot peg price to value without knowing what the customer did with it. Their breadth is the constraint.

A narrow product does not have that problem. A collections tool knows the dollars it recovers; a fraud product knows the losses it prevents. That line of sight to the outcome is a pricing dimension the labs structurally lack, and it is where a focused SaaS business can capture value the consumption model leaves on the table. Outcome pricing carries its own frictions (e.g., attribution, measurement and slower sales cycles), which is why most players still lean on proxies. But even a partial tie to outcomes, such as a success fee layered on a consumption base, can convert a narrow focus into pricing power (see Table 1).

Table 1

What to borrow and what not to copy

Lab pricing move	What it solves	SaaS adaptation	Risk to manage
Common currency (tokens or credits)	Monetizes new capability without repricing	One internal consumption currency across features	Opaque exchange rates erode trust
User-directed pooled credits	Drives trial and adoption of new modules	Seeds a credit pool users spend themselves	Revenue harder to forecast
Metered and asymmetric limits	Protects the margin tail from heavy users	Fair-use caps with overage pricing	Backlash if caps feel punitive
Deliberate repricing	Reflects falling input costs	Explicit capture-or-pass-through policy	Margin leakage or price wars
Capability tiering	Captures power-users' willingness to pay	Reserves the best AI features for premium tiers	Hard gating hides the value and stalls upgrades
Outcome pricing (only you can)	Ties price to value delivered	Success fee layered on a consumption base	Attribution and measurement disputes

Note: SaaS=software as a service; AI=artificial intelligence
Source: L.E.K. research and analysis

The limits of the analogy

None of this is a mandate to copy. Credit systems erode trust the moment the exchange rate feels opaque or shifts without warning. Consumption pricing trades revenue predictability for upside, complicating forecasting and the investor story. Competitive dynamics may force pass-through, turning "capture the saving" into wishful thinking. And the labs price from a position of novelty and pricing power most categories do not share, so the tactics travel further than the leverage does.

What to do now

Four questions about your own pricing are worth asking:

1. Does it capture value on more than one axis or is it still just per seat?
2. Would a consumption currency let you ship and monetize faster than your repricing cycle allows?
3. Have you stress-tested your AI-feature margins against your heaviest users?
4. Given your product breadth relative to a frontier model, where could you price to tie to outcomes?

The labs have built a pricing architecture worth borrowing; the durable advantage lies in the dimension they cannot reach. B2B SaaS companies that adopt the labs' discipline on consumption proxies while pricing closer to the outcomes they deliver will be best positioned to capture the value AI creates.

L.E.K.'s B2B SaaS Pricing team advises software businesses on pricing architecture, monetization models and value capture. To discuss how these approaches apply to your business, please **contact us**.

About the Authors

Rob Haslehurst
Rob Haslehurst is a Managing Director in L.E.K. Consulting's Boston office. Rob leads L.E.K.'s global Pricing practice and heads its Software and Technology pricing work. He advises software and technology companies and their investors on pricing and revenue management, drawing on wider experience across growth strategy, go-to-market, digital, and M&A.

Ryan Hoffman
Ryan Hoffman is a Principal in L.E.K. Consulting's Boston office and a member of the Technology, Media and Telecom practice. Ryan brings extensive experience in the GovTech, ISV marketplace, enterprise software and cloud services market spaces. He advises technology companies and their investors on a range of critical issues, including growth strategy, pricing, competitive strategy, channel strategy and partner program design, and M&A.

About L.E.K. Consulting

L.E.K. Consulting is a premier global strategy consulting firm that partners with clients to solve their most critical business challenges and deliver high-impact outcomes. We bring deep industry expertise and rigorous, evidence-based insight to global corporations, growth companies, and private equity investors, helping them turn strategy into action. Founded more than 40 years ago, L.E.K. now operates across 27 offices worldwide. Learn more at **www.lek.com**.