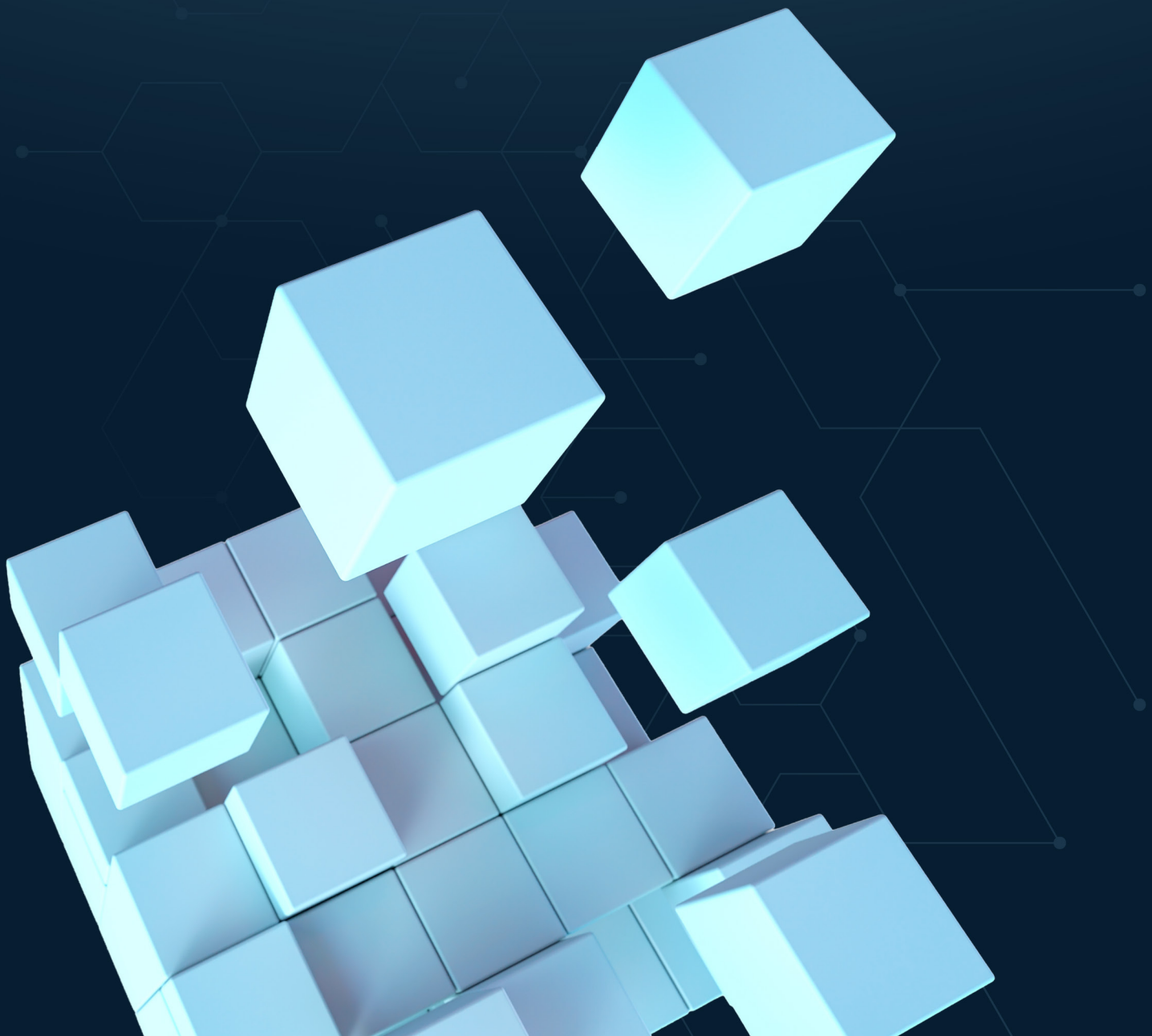
SBI | Grow Intelligently

The Great Unbundling:

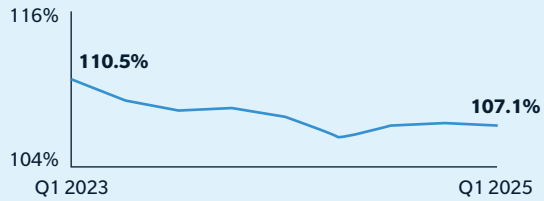
Why Your NRR Is Declining Even as Customers Extract More Value



The SaaS Growth Model Is Breaking

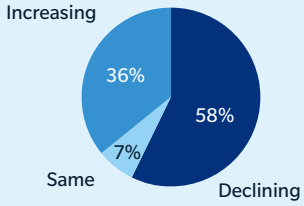
Net Revenue Retention (NRR) in SaaS has declined from 110.5% in early 2023 to 107.1% by 2025. Nearly six in ten SaaS companies now report lower NRR than they did two years ago. For boards valuing companies on NRR strength and revenue leaders who built go-to-market strategies around "land and expand," this erosion represents a fundamental threat to the growth math underpinning SaaS valuations.

NRR is Down and Has Not Recovered
Average NRR by quarter Q1 2023 to Q1 2025



n = 45 Q1 2023 n = 80 Q1 2025
Source: SBI NRR Dataset

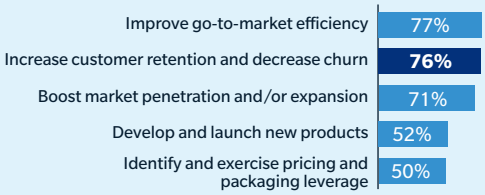
Most Companies' NRR is Lower than Two Years Ago
Share of companies with declining NRR Q1 2023 to Q1 2025



n = 45 Q1 2023 n = 80 Q1 2025
Source: SBI NRR Dataset
Note: Percentages do not equal 100% due to rounding

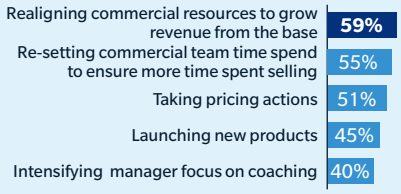
CEOs recognize the urgency. In SBI's latest executive survey, 76% rank "increase customer retention / decrease churn" as a top value creation lever. "Realigning resources to grow revenue from the base" ranks as the top strategic emphasis for 2026. Growth from the base has moved from "nice to have" to "non-negotiable" just as it's getting harder to achieve.

Increasing Retention is The Top Priority
CEOs' top five growth levers to value creation plans



n=118
Source: SBI 2025 Q4 CEO Survey

Growing the Base is the Highest Focus Area
CEOs' top five focus areas to meet revenue/ margin targets



n=118
Source: SBI 2025 Q4 CEO Survey

What CEOs are witnessing is what we call "the great unbundling": the separation of value delivery from seat consumption. For the 80% of SaaS companies that still rely on seats for pricing[1], the great unbundling represents a threat that will continue to erode NRR and put growth targets at risk. The instinct is to respond with product enhancements or AI capabilities. But the root cause runs deeper. Customers are fundamentally changing how they consume value, breaking the economics that made SaaS predictable for over a decade.

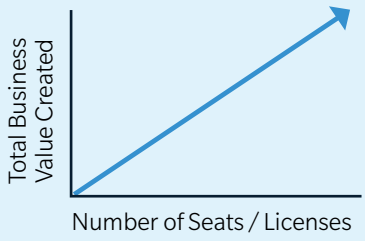
1. Price Intelligently by SBI, 2025 State of SaaS Pricing Report Part 1: 10 Insights for Building Market-Leading Pricing

The Great Unbundling:
Rethinking Value in the AI Era

Traditional Model

Seats Are an Effective Proxy for Value

1 Seat = 1 Unit of Value

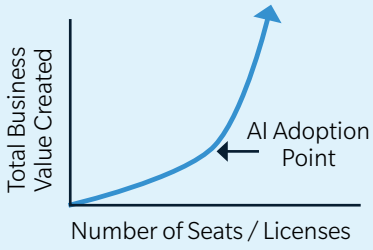


Business Value Scales Linearly with Seat Count

Current Disconnect

AI Productivity Gains Fuel Seat Stagnation

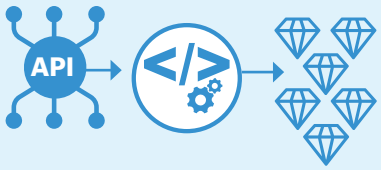
AI Multiplies Human Output



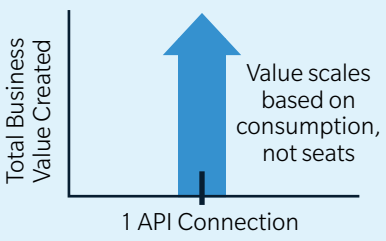
Value Creation Spikes with AI, but Seat Count Stalls

Agentic Future

Differentiation Shifts from UI to API



APIs Scale Value Creation



Business Value Scales via API Consumption, Not Human Capacity

1. The Traditional Model: When Seats Drove Growth

For fifteen years, SaaS operated on an elegant mechanic: seats worked because output was tied to headcount. Take a SaaS product used by software developers. As your customer's company grew, they hired more developers. Each new developer needed a seat. Your NRR increased predictably. The customer's business success translated directly into your revenue growth.

This model created the consistency Wall Street rewarded with premium multiples. The seat was an effective proxy for value delivered.

2. The Current Disconnect: AI Productivity Is Fueling Seat Stagnation

That connection is breaking right now. AI is delivering productivity gains that allow your customers to maintain or increase output without adding headcount. SAP's CFO Dominik Asam states plainly "There are certain tasks which are automated and for the same volume of output we can afford to have less people."[2]

The result: seat counts stagnate or decline even as customers extract more value from your platform. Some are doing more with existing teams. Others are actively reducing headcount while maintaining output. Either way, your seat-based revenue model captures none of the expanding value you deliver.

This productivity-driven seat stagnation is what's showing up in the NRR data. Widespread NRR declines aren't a temporary dip. They're the early signal of a structural shift that will accelerate over the next three years as AI productivity gains compound.

How the Current Disconnect Impacts Different Personas

Persona	Why They're Exposed	Exposure Today
Creator Developer, Copywriter, Designer	Produces tangible outputs—code, copy, designs. AI can now generate these directly.	High
Fixer CS Rep, Security Analyst, QA Engineer	Resolves problems. AI increasingly handles these autonomously.	Medium
Strategist CFO, VP Supply Chain, CRO	Makes judgment calls. AI surfaces insights but can't yet decide.	Emerging

3. The Agentic Future: Differentiation Shifts from UI to API

As AI becomes more capable, employees won't use products through traditional UI interfaces. Autonomous agents will consume value through APIs instead. A customer might need one API license instead of fifty-seat licenses, breaking the connection between seats and revenue completely.

This shift towards autonomous agents threatens the value proposition for companies that built their competitive moat on "ease of use" or "beautiful UI." When agents handle workflows via API calls, the user experience that justified premium pricing no longer matters.

Companies still reliant on seat-based pricing face a compounding problem: not only is their revenue model breaking, but for many, their primary competitive advantage, UI differentiation, is disappearing alongside it.

2. Dominik Asam (SAP CFO), interview with Business Insider, September 2025. Article link: https://finance.yahoo.com/news/cfo-320-billion-software-firm-150201570.html

Three Questions Executives Must Answer

To navigate this transition, executives must answer three questions:

1. **How do you compete when your primary differentiation no longer aligns with customer buying criteria?** Customers prioritize easy-to-integrate APIs over easy-to-use UIs. Your product roadmap and sales messaging were built around the wrong future criteria.
2. **How do you allocate resources when you can't identify which accounts will churn or expand?** Traditional health score metrics like login frequency and seat utilization misclassify risks and opportunities. Teams chase the wrong accounts, based on increasingly obsolete indicators of health, while your best expansion targets go unnoticed.
3. **How do you monetize expanding value when seat counts stagnate?** Seat-based pricing worked when headcount drove consumption. It fails when AI-enabled employees and autonomous agents extract exponential value through a handful of licenses.

What's needed is to rethink how companies detect and respond to growth signals. By applying AI to behavioral and telemetry data, executives can adapt product strategy as buying criteria shift, predict commercial outcomes with 90% accuracy when traditional metrics fail, and evolve pricing to capture value as consumption changes.

How AI Is Rewriting Buying Criteria and Workflows

The product and sales narratives that won deals for the past 15 years are rapidly misaligning with customers' evaluative criteria for SaaS solutions. What closes deals and what drives retention are both shifting.

Today: AI-Augmented Workflows

Buyers evaluate three technical capabilities that weren't purchase criteria three years ago:

- AI integration: Does your product easily connect with the AI tools employees use daily (ChatGPT, Claude, Gemini, private AI systems)? Poor integration creates friction that compounds across every workflow.
- Contextual flexibility: Can employees inject domain knowledge and role-specific requirements to refine AI output? Rigid approaches lose to customizable systems that adapt to how different teams work.
- Data security and rights: Who owns the data flowing through AI workflows? Buyers demand clarity on what information reaches AI providers and explicit contractual protections.

Tomorrow: Agentic Workflows

Within three years, enterprise AI systems will work across multiple vendor platforms simultaneously, orchestrating tasks that currently require human coordination. Buying criteria shift again to pure technical infrastructure:

- Data portability: Can buyers extract their data cleanly, completely, and in real-time? Proprietary formats block AI orchestration.
- API completeness: Does every UI function have an API equivalent? Products built UI-first with APIs as an afterthought can't support agent-driven workflows.
- Response latency: Can your system respond fast enough for AI-driven processes? Acceptable response times for humans become bottlenecks when agents orchestrate multi-step workflows.
- Integration flexibility: How easily does this connect to the customer's orchestration layer? Proprietary integration approaches lose to standards-based connectivity.
- Training rights: Can buyers use data flowing through your system to train their models? Restrictive terms become competitive disadvantages.

These shifts in buying criteria demand immediate action across product, sales, and marketing.

Three Actions Required Now

1. Shift product roadmaps from UI differentiation to API enablement. Prioritize API completeness, data extractability, and integration flexibility over interface polish. Enterprise buyers increasingly evaluate technical integration requirements before user experience. UI becomes table stakes, not differentiation.
2. Prepare sales for technical evaluators. What's old is new again. The past decade saw an explosion in composable business applications, leading to more business-led purchasing than ever before. However, IT and data teams are once again heavily influencing purchase decisions alongside business buyers. Champions need new skills: articulating API depth, demonstrating interoperability, and quantifying productivity gains from AI integration. Demos emphasizing "intuitive interface" no longer close deals.
3. Revise messaging to address AI readiness. Positioning must explicitly communicate how your product enables AI-augmented workflows today and agentic workflows tomorrow. Sales narratives built around ease of use lose relevance when buyers evaluate ease of integration. Marketing content should address data portability, API standards, and orchestration capabilities.

Companies implementing these changes report maintaining competitive win rates in environments where traditional UI-focused competitors see erosion.

Your customer's team is more productive with AI, so they're not adding seats, but they're extracting more value than ever

Using Telemetry to Identify At-Risk and Expansion-Ready Accounts in the AI-Era

GTM leaders face a forecasting crisis. Traditional health score metrics like login frequency and seat utilization systematically misclassify which accounts will churn and which will expand. These metrics were designed when human engagement equaled value consumption. That connection is breaking now and will sever completely in the agentic era.

The problem manifests today in a clear pattern: your customer's team is more productive with AI, so they're not adding seats, but they're extracting more value than ever. Login frequency appears flat while usage intensifies. Health scores flag these accounts as stagnant when they're actually more embedded.

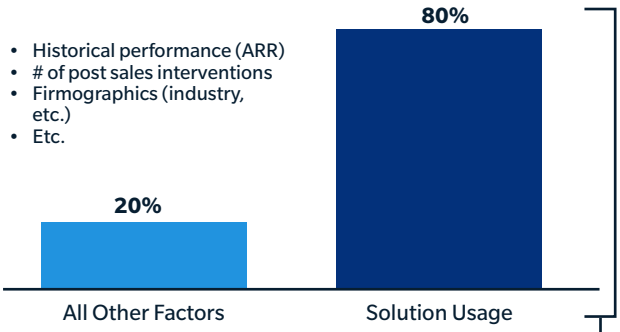
This leads GTM teams to chase the wrong accounts. They run standard renewal plays on expansion-ready customers, leaving revenue on the table. They ignore at-risk accounts that show high activity but have fundamentally misaligned use cases. Resources flow to accounts that cannot be saved while genuine growth opportunities receive formulaic engagement.

Usage Drives 80% of Commercial Outcomes, But Raw Data Creates More Noise Than Signal

To identify what predicts retention and expansion as consumption patterns shift, SBI partnered with the product analytics company QuadSci to analyze 160 billion telemetry datapoints across 9,100 commercial outcomes.

The finding validates what GTM leaders instinctively know: usage matters. It accounts for approximately 80% of renewal and expansion decisions, dwarfing firmographics, historical ARR, and CSM interventions combined. But it also exposes why traditional approaches fail. Raw usage data cannot predict outcomes because the definition of meaningful usage is shifting, and usage volatility creates overwhelming signal-to-noise problems.

Key Drivers of Renewal and Expansion Decisions



Source: QuadSci Retention and Growth Data
n=160 Billion Telemetry Datapoints across 9100 commercial outcomes

Solution usage explains **80%** of a customer's renewal or expansion decision.

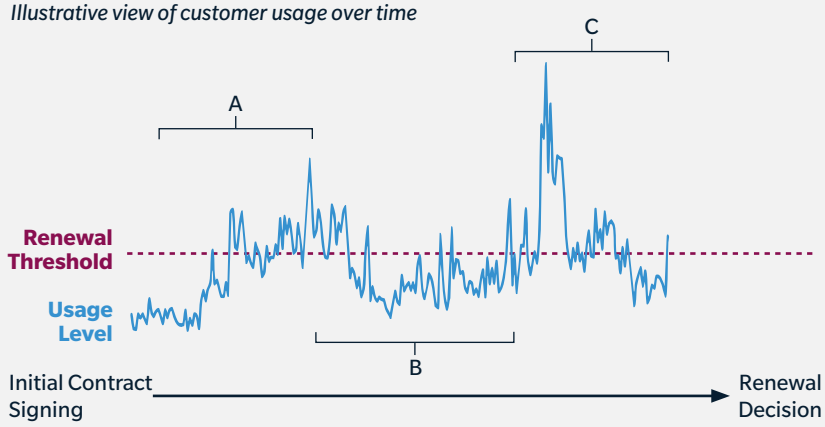
Simply Put: The signal for retention and growth is there. Volatility buries it.

Usage levels naturally fluctuate due to product releases, staffing changes, seasonality, and shifting priorities. The transition from seat-based to API-driven consumption amplifies this challenge dramatically.

Today, AI-enabled workflows mean an account might maintain flat seat counts while per-user productivity doubles. Traditional health scores miss the value expansion entirely. In the API phase ahead, early adopters will generate usage patterns that look nothing like historical benchmarks.

A temporary dip might signal an at-risk account or simply reflect a team vacation. A usage spike might indicate expanding workflows or a one-time project that won't recur. More telemetry doesn't solve this. Accurate forecasting requires distinguishing signal from noise, not accumulating more data.

Illustrative view of customer usage over time



A. ↑ Onboarding complete, champions emerge

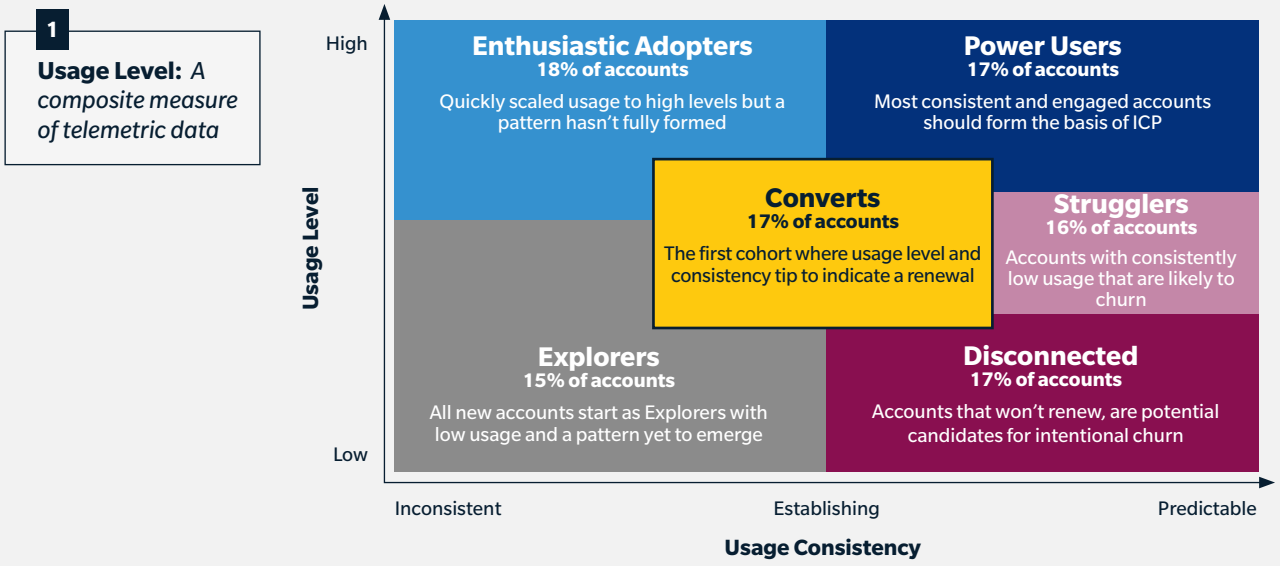
B. ↓ Organizational changes disrupt focus

C. ↑↓ New leadership shifts strategic priorities

Usage Level Plus Consistency Predicts 90% of Outcomes

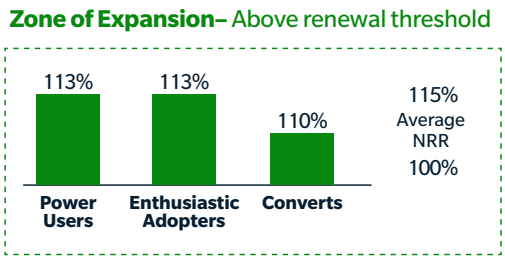
The solution lies in combining two dimensions. Analyzing usage level (volume and intensity) together with usage consistency (predictability of engagement) reveals six distinct behavioral cohorts that predict churn, renewal, or expansion with 90% accuracy up to 12 months in advance.

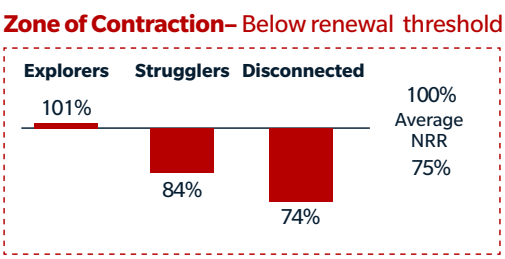
When customers extract consistent and increasing value, whether through UI today or API tomorrow, they renew and expand. When they don't, no amount of outreach changes the trajectory.



Source: QuadSci Retention and Growth Data
n=160 Billion Telemetry Datapoints across 9100 commercial outcomes

Each cohort has a clear NRR profile that enables precise resource allocation. Rather than treating all accounts the same, GTM teams can deploy different strategies based on where accounts sit, improving overall efficiency. Accounts organize into two zones:





Zone of Expansion (Collective NRR > 100%):

- Power Users (17%): High, stable usage. Deeply embedded in daily workflows. Deliver the highest and consistent NRR and serve as engines of expansion.
- Enthusiastic Adopters (18%): High but less stable usage. With targeted enablement, transition into Power Users. Can signal early AI adopters still optimizing workflows.
- Converts (17%): Usage level and consistency recently crossed the renewal threshold. Form the core of reliable renewals and represent early expansion potential.

Zone of Contraction (Collective NRR ≤ 100%):

- Explorers (15%): Low, inconsistent usage. Still testing fit and establishing workflows. Need guidance and effective support to cross the renewal threshold.
- Strugglers (16%): Consistently low usage. Haven't found effective workflows or ROI. Elevated churn risk despite tenure.
- Disconnected (17%): Very low usage. Effectively shelfware. Candidates for intentional churn or targeted win-back experiments.

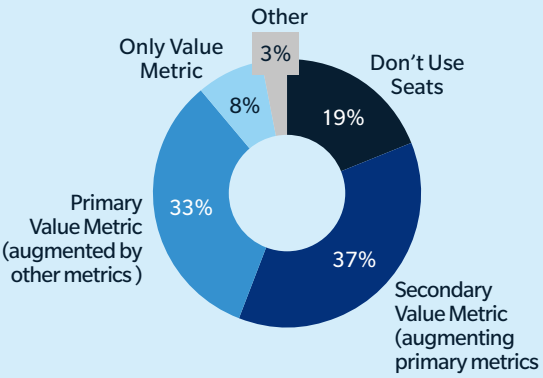
Operationalizing Cohort Intelligence

Identifying at-risk and expansion-ready accounts months in advance only creates value if it changes how teams operate. Three operational shifts are required:

1. Track value consumption, not activity. Cohort placement prioritizes telemetry data that matters to commercial outcomes. In most cases, this means measuring what customers accomplish, not just how often they log in. In the AI era, value equals outputs (contracts executed, code merged), problems solved autonomously (tickets resolved, invoices reconciled), and high-impact decisions triggered by AI insights.
2. Institutionalize cohort-based operating rhythms. Cohorts capture the dynamic nature of accounts. Monthly cohort movement reviews replace traditional pipeline reviews for the base. Accounts slipping downward trigger immediate save plays with executive involvement. Accounts moving upward activate expansion motions. Segmented playbooks align resources with cohort-specific needs rather than generic "high-touch" or "tech-touch" strategies.
3. Deploy role-specific dashboards. CROs track NRR and expansion pipeline by cohort with forecast overlays incorporating cohort-based risk. Chief Customer Officers monitor coverage ratios and intervention success rates by cohort to optimize team deployment. CFOs receive forward-looking NRR projections grounded in cohort distributions, replacing backward-looking averages with predictive models that support accurate guidance.

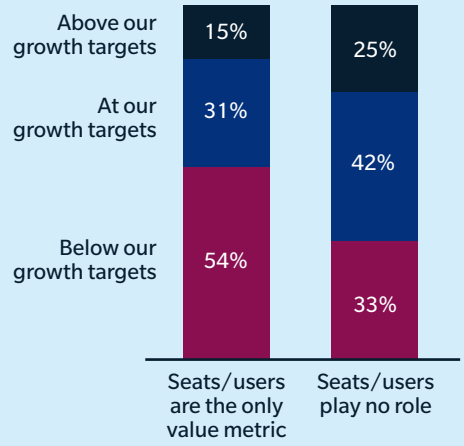
Cohorts directly answer the second question: how do you allocate resources when you can't identify which accounts will churn or expand? Cohort analysis allows executives to understand what usage drives commercial outcomes and predict outcomes with 90% accuracy up to 12 months in advance, enabling teams to match resources to actual potential. Companies operationalizing this approach report 5% NRR improvement while reducing wasted effort on low-probability saves.

Function of Seats in SaaS Pricing Model



N= 321
Source: Price Intelligently by SBI 2025 State of SaaS Pricing Survey
Note: Some "Other" respondents use modified seat pricing, bringing total seat reliance to 80%.

Growth Rate By Seat Importance



N= 95
Source: Price Intelligently by SBI 2025 State of SaaS Pricing Survey

Aligning Pricing to Value in the AI Era

Adapting to new buying criteria and predicting churn risk both assume you can monetize the value you create. As customers progress from individual licenses to API-driven consumption, GTM leaders must transition from seat-reliant to seat-resilient pricing without sacrificing the predictability that enterprises demand and vendors count on.

The stakes are high. According to SBI's State of SaaS Pricing Research, 80% of SaaS companies still rely on seat-based pricing models. Providers that depend on seats as their only value metric are 1.6x more likely to miss growth targets than those that forgo seats for consumption or usage-based models.

Four Ways Seats Are Being Unseated

The transition isn't uniform. Different companies face different unbundling pressures:

Market contraction: Broad headcount reductions as AI makes teams more productive. For example, HCM and intranet platforms see companies maintaining output with fewer employees, directly eroding seat counts.

External AI adoption: Customers augment workflows with external AI tools, extracting more value with fewer users. Developer tool products report 15-30% seat contraction as developers use AI coding assistants to produce more code with smaller teams.

Internal AI adoption: AI-enabled features make customers more efficient per license, cannibalizing seat counts. Customer support and martech platforms face this directly as AI-powered capabilities let one user accomplish what previously required three seats.

New value delivery model: Customers shift from UI to API consumption, extracting value through data access and integration rather than human interaction. CRM platforms increasingly function as databases accessed via API calls rather than applications with licensed users.

Each pattern demands a different pricing response, but all share a common problem: seat-based models fail to capture expanding value as customers extract more while consuming fewer seats.

The Pricing Extremes Both Fail

Pure seat-based pricing breaks for both value and cost reasons. Flat-fee pricing for AI-heavy features destroys margins when power users or agents consume disproportionate compute resources. The AI transition may fundamentally reset SaaS margin expectations from the historically typical 80-90% gross margin range to something closer to 40-50% as compute costs rise and seat counts decline.

Conversely, fully variable outcome-based pricing rarely works commercially. Less than 1% of SaaS providers have successfully implemented true outcomes-based models. The model requires both discrete, measurable outcomes and perfect attribution, criteria most products cannot satisfy. Even when technically feasible, outcome-based pricing creates forecasting problems: CFOs cannot budget when costs fluctuate with business performance rather than usage; procurement refuses to buy without cost certainty.

The Hybrid Model That Works (For Now)

The emerging pattern balances predictability with value capture through hybrid pricing. The structure works because it addresses different buyer priorities: enterprises demand budget certainty while growth-stage companies seek affordability and clear value alignment.

A platform fee (predictable) covers access, data unification, governance, and core workflows. This component remains stable and forecastable, anchoring the commercial relationship and enabling accurate revenue planning for both vendor and customer.

A metered component (bounded) charges for value-linked events: artifacts generated, problems resolved autonomously, or high-impact decisions triggered. This captures variable value consumption as customers shift workloads from human users to AI agents without creating unlimited cost exposure.

The Hybrid Path Forward to Sustainable Growth

The metering must be bounded through tiered pricing that caps exposure at reasonable multiples of the base fee. This structure prevents bill shock that kills renewals while ensuring vendors capture upside as customers extract increasing value.

A Note of Caution on Complexity

Hybrid and consumption-based pricing introduce operational complexity many companies underestimate: tracking multiple value metrics, billing system upgrades, sales compensation aligned to blended models, and customer communication about variable costs. As this complexity accumulates, some portion of the market will likely return to simpler seat-based models with generous fair use caps. The decision point is whether pricing complexity matches the clarity of value delivery.

Case Study:

A Developer Tool Provider Transitions from Seat-Dependent to Seat-Resilient

A long-standing enterprise developer tool provider serving gaming, semiconductor, and software development teams faced the structural challenge outlined earlier. Their flagship product was priced on developer seats, but AI coding assistants were driving 15-30% seat contraction across their customer base. Meanwhile, actual platform usage (code, assets, and data under management) remained robust or grew.

The disconnect was stark. Customers maintained or increased infrastructure load while teams shrank. AI-generated code and assets proved bulkier than human-created equivalents, meaning storage requirements stayed flat or rose even as seat counts declined. Their customers' AI use was breaking the seat-based model.

The Hybrid Solution

Working with SBI, the provider implemented a hybrid pricing model that kept seats as a value metric while adding storage as a bounded secondary lever.

For on-prem deployments, each seat includes a pooled storage allotment at the account level. When aggregate storage exceeds the allotment, overage fees apply. As AI productivity allows customers to reduce headcount, their entitled storage automatically decreases, but their storage needs do not necessarily follow. This pushes heavy users into overage tiers.

For cloud deployments, SBI recommended increasing storage overage pricing by 40% to match competitive rates, introducing tiered storage buckets with volume discounts, and creating feature-lite tiers for smaller teams to expand market coverage.

Results

Scenario modeling showed that in cases with 20% seat reduction, the hybrid model moved from approximately 20% ARR loss under pure seat-based pricing to nearly flat ARR (low single-digit decline) by capturing storage overages. In accounts with heavier storage consumption, the model generated net ARR uplift despite fewer seats.

The Telemetry Imperative

Implementing this model required significant telemetry infrastructure upgrades, particularly for on-premises deployments where storage tracking was inconsistent. The provider invested in real-time storage consumption monitoring across all deployment types, account-level usage aggregation to enable pooled allowances, predictive modeling to help customers forecast overage exposure, and cohort analysis to identify which customer segments would benefit from tiered storage packages versus unlimited plans.

This investment in telemetry capabilities didn't just enable the new pricing model. It provided the foundation for the cohort-based customer success strategies discussed earlier, allowing the provider to predict renewal and expansion outcomes with far greater accuracy.

Key Lesson

The case illustrates how hybrid pricing preserves the simplicity buyers expect (seats remain the primary metric) while ensuring revenue follows where work actually happens in the AI era. Pricing aligns to infrastructure reality: when AI lets customers do more with fewer people, they're still using the platform, and in many cases, using more of it.

Using Telemetry to Define Value-Linked Metered Units

GTM leaders can leverage telemetry and cohort analysis to design metered units that satisfy two commercial criteria:

First, units must represent value the buyer recognizes. Track human-driven outcomes enhanced by AI today and agent-driven outcomes tomorrow. The distinction matters less than whether the unit reflects value customers acknowledge paying for. A CFO cares about invoice exceptions resolved, regardless of whether a human or autonomous agent performed the resolution.

Most successful companies define one primary metered unit tied to their core value driver rather than different metrics for different personas. Persona-specific pricing creates procurement complexity that extends deal cycles and kills momentum.

Second, units must be forecastable with reasonable accuracy. By analyzing historical behavior across cohorts, GTM leaders can model how different customer segments will consume metered units as they scale. Power Users exhibit predictable consumption patterns that enable both vendor and customer to forecast spend with confidence, satisfying finance requirements on both sides.

Companies making this transition successfully use telemetry to identify which capabilities drive disproportionate value for specific customer segments, then build pricing around those capabilities. They capture expanding value as customers shift from UI-based to agentic consumption while maintaining the enterprise predictability that protects renewals.

Conclusion: Engineering Growth in the Agentic Era

The great unbundling of value delivery from seat consumption is not a future threat. It's driving NRR declines today and will accelerate over the next three years. The three questions have clear answers rooted in Growth Intelligence: use telemetry to understand shifting buying criteria and adapt product roadmaps, leverage cohort analysis to predict churn and expansion with 90% accuracy, and evolve pricing to capture expanding value without sacrificing predictability.

Traditional differentiation based on UI polish erodes when autonomous agents mediate interactions. The new competitive moat is understanding which customers will succeed, why they'll succeed, and how to price accordingly. Companies that master telemetry-driven insights and pricing evolution will engineer predictable growth through this transition. Those continuing to optimize for the seat-based era will find their growth models breaking faster than they can adapt.

Why SBI?

Driven by insights and delivered from experience, SBI continues to help clients grow their revenue, margin, and enterprise value in ways never before possible.

Working with us, go-to-market leaders can expect confidence and trust with experienced partners every step of the way. We engage and support our clients as an extension of their team, both guiding and working side-by-side to deliver relatable, practical strategies that work for today and tomorrow.

Connect with SBI today and talk to us about how we can help you on your growth journey.

550 Reserve Street
Suite 190
Southlake, Texas, 76092
www.sbigrowth.com