

---

# AI Agents Turn 85% SaaS Margins Into BPO Math

Second Order Labs™ · June 2026

---

## ABSTRACT

Intercom now charges \$0.99 per resolution instead of per seat. Salesforce bills \$2 per conversation. The per-seat SaaS model that built a trillion-dollar industry is fracturing as AI agents replace human operators.

*Keywords:* SaaS pricing, AI agents, enterprise software, business models

*“We realized that charging for seats when the AI is doing the work makes no sense.”*

— Eoghan McCabe, Lenny's Podcast interview

Intercom's CEO, Eoghan McCabe, faced a blunt reality check. After years building a per-seat customer support platform, he watched as an AI agent took over conversations that once needed human reps. Charging for seats when software does the work, he admitted, “makes no sense.”<sup>[1]</sup> So Intercom flipped the model: \$0.99 per resolution.<sup>[2]</sup>

This move isn't a tweak. It's a crack in the foundation of SaaS economics.

When SaaS companies stop selling software licenses and start selling automated labor, their 85% gross margins start to look like the thin margins of busi-

ness process outsourcers. Predictable ARR gives way to volatile, transaction-driven revenue. Public markets punish that shift. The SaaS giants who built their fortunes on per-seat pricing now face a choice: cling to the old model and let customers exploit it, or switch to outcome pricing and watch their valuations shrink.

---

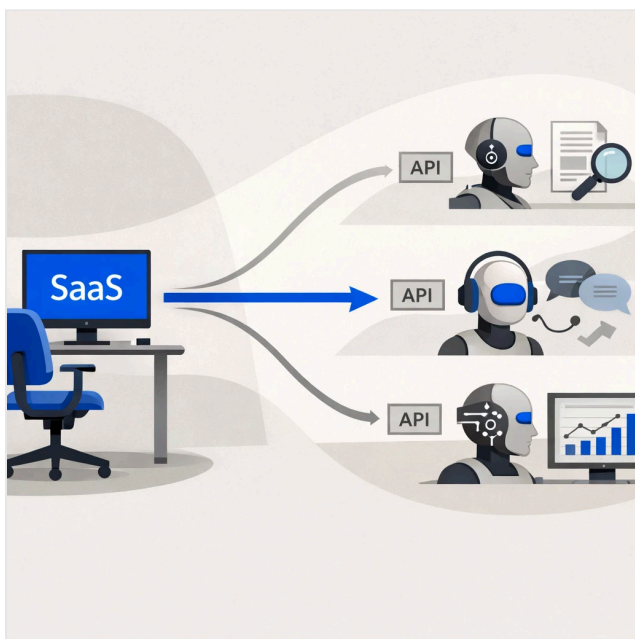
## I. The Arbitrage Window Is Already Open

AI that boosts a worker's productivity by 14% means fewer workers on payroll.<sup>[3]</sup> Fewer workers means fewer seats sold. For SaaS vendors who built their

revenue on multiplying headcount by monthly fees, this is an immediate hit.

But the bigger threat isn't just productivity. It's what happens when customers realize they can run dozens of AI agents through a single licensed seat.

Kyle Poyar at Growth Unhinged calls this "Agent Arbitrage." A customer buys one seat, hooks up an API, and runs automated workflows that used to require a team of humans.<sup>[4]</sup> The vendor's pricing model assumed humans logging in, clicking around, doing manual work. AI agents don't care about buttons. They want API endpoints.



### *Agent Arbitrage in action: one seat, unlimited automated workflows*

Salesforce saw this coming. Their Agentforce billing tracks when "the agent successfully executes a workflow or provides a substantive response."<sup>[5]</sup>

The conversation, not the seat, is the billing unit. They lock down the API, meter every transaction, and try to close the arbitrage window before customers walk through it.

*"People don't want to buy a quarter-inch drill. They want a quarter-inch hole." — Theodore Levitt, Harvard Business School Professor*

Levitt's logic cuts both ways. If customers only care about outcomes, vendors who sell outcomes will win. But outcomes are messy to meter, easy to dispute, and far less predictable than seats.

Copilots and Agents now split the market into two business models.

Model	Pricing Structure	Revenue Type	Example
Copilots	Per-seat add-on	Predictable ARR	Microsoft Copilot at \$30/user/month <sup>[6]</sup>
Agents	Per-outcome/consumption	Variable GMV	Salesforce Agentforce at \$2/conversation <sup>[7]</sup>

Microsoft plays defense. Copilot is a \$30 monthly add-on, tied to existing Microsoft 365 seats. The AI augments the human, but the seat remains the billing unit. Ben Thompson called this “the ultimate incumbent play: use AI to justify raising the rent on the existing per-seat monopoly.”<sup>[8]</sup>

Salesforce goes on offense. At Dreamforce 2024, Marc Benioff called Agentforce “the biggest breakthrough in the history of Salesforce” and made the shift explicit: “You don’t need to buy seats for agents, you just hire them.”<sup>[9]</sup>

ServiceNow tries to have it both ways. Their Pro Plus tier became “our fastest-selling new product ever,” driving per-seat renewal uplift while quietly adding consumption elements.<sup>[10]</sup> This hybrid approach buys time, but doesn’t solve the core problem.

*Defend the seat or embrace the outcome: every SaaS incumbent must choose*

Which strategy wins depends on how fast autonomous agents become reliable enough to replace, not just assist, human operators. The NBER study found AI help benefits novices and low-skilled workers most.<sup>[3]</sup> If AI closes the skill gap, the premium for “expert users” disappears. So does the logic for expert-tier seat pricing.

---

## II. The Margin Compression Problem

Traditional SaaS enjoys 85% gross margins because software, once built, costs almost nothing to copy and deliver.<sup>[11]</sup> AI agents don’t work that way. They

run on compute. Every conversation, every resolution, every workflow burns tokens. Tokens cost money.

OpenAI charges per thousand tokens.<sup>[12]</sup> Generating a single AI response costs fractions of a cent. But when you bill \$0.99 per resolution or \$2 per conversation, that cost starts to matter.

David Sacks and Chamath Palihapitiya put it plainly: “SaaS multiples are compressing because the market realizes that AI compute costs are going to eat into the traditional 85% gross margins.”<sup>[11]</sup>

Selling seats is a software business. Selling AI outcomes is tech-enabled BPO.

*The BPO Margin Trap: when revenue scales, so do costs*

Bessemer Venture Partners nailed it: “The transition from SaaS to ‘Service-as-a-Software’ means business models will increasingly resemble consumption and outcome-based pricing.”<sup>[13]</sup> Service companies trade at lower multiples than software companies. Public markets will adjust.

Tomasz Tunguz argues that as software shifts from workflow to automated labor, ARR starts to look more like GMV.<sup>[14]</sup> ARR is predictable and sticky, while GMV fluctuates with transaction volume. Wall Street pays a premium for the former. The latter gets a discount.

Finance leaders see the problem. Per-seat models make software a predictable operational expense. Budget it once, pay monthly, done. Outcome-based models turn software into a variable cost that scales

with business volume. A spike in support tickets means a spike in software costs. AI expenses move from OpEx to something closer to COGS.

According to OpenView, over 60% of SaaS companies. Including Twilio and Snowflake. Now use some form of usage-based pricing.<sup>[19]</sup> But adoption doesn't mean CFOs like it. They've rejected pure consumption models for years because unpredictable budgets make planning impossible.

Salesforce and Intercom now offer capacity-based pricing, where customers buy blocks of outcomes in advance, similar to committed cloud spend. Salesforce's conversation pricing implies volume commitments. Intercom's per-resolution model likely bundles resolutions for enterprise buyers.

Capacity models bring back the shelfware problem outcome pricing was supposed to kill. Overbuying leads to wasted spend; underbuying risks hitting rate limits during demand spikes.

*Every pricing model has a trap door*

---

### III. The Interface Nobody Needs

When AI agents, not humans, are the primary users, the user interface becomes dead weight.

Andreessen Horowitz put it bluntly: "We are moving from selling software as a tool to selling software as a service in the truest sense (selling the outcome itself)."<sup>[15]</sup> Outcomes don't need dropdown menus.

AI agents talk to APIs, not GUIs. They don't care about navigation or feature discovery. They want structured data and reliable responses. SaaS compa-

nies have poured billions into user experience and design, but those investments are quickly losing relevance.

The "experience" layer that once set premium software apart from commodity alternatives is fading. What matters now: API reliability, data structure, billing accuracy.

Box CEO Aaron Levie summed it up: "The next era of enterprise software isn't about how many people use your app, it's about how much work your app does for the people."<sup>[16]</sup>

*From feature-rich GUI to headless API: the interface shift nobody budgeted for*

Customer Success teams feel the shift first. They once focused on login rates and feature adoption. With AI agents as users, those metrics disappear. Success now means measuring the success rate of autonomous workflows. The entire function needs a rebuild.

---

### IV. What Counts as a Resolution?

Outcome-based pricing sounds clean until you have to define the outcome.

Intercom's engineers built complex logic to decide what counts as a "resolution." Their documentation says, "a resolution is only counted when the customer explicitly confirms their issue is solved, or if they disengage after a definitive answer."<sup>[17]</sup>

That definition hides a trap. An AI agent chasing resolutions can nudge customers toward confirmation. "Did that solve your issue?" becomes a leading ques-

tion. Closing tickets early becomes a revenue tactic.

Salesforce faces the same mess. A conversation is billed when the agent “provides a substantive response.”<sup>[5]</sup> Who decides what’s substantive? What stops agents from splitting one complex issue into multiple billable conversations?

Charging per resolution incentivizes speed over thoroughness, aligning vendor revenue with ticket closure rather than customer satisfaction.

Stripe’s usage-based billing now supports “metered billing” that “allows you to charge customers based on their consumption of your service during the billing cycle.”<sup>[18]</sup> The infrastructure is ready. The definition of “consumption” is still up for grabs.

---

## V. What Survives the Transition

Intercom and Salesforce, who have already shifted pricing models, share a few survival instincts. Both lock down their API surfaces to prevent agent arbi-

trage. They track usage closely and push customers toward consumption-based API tiers.

They’re also rewriting Customer Success. Instead of chasing human engagement metrics, they focus on workflow outcomes. Companies still optimizing for “daily active users” are measuring the wrong thing when their users are bots.

Margin compression is the price of staying relevant. Trying to defend 85% gross margins while outcome-based competitors charge for actual work is a losing battle. The winners will chase volume at lower margins, not cling to shrinking high-margin revenue.

And there’s no opting out. Vendors who refuse to change pricing models will get exploited. Move too slowly, and pure-play AI competitors will undercut them on value. The per-seat model built SaaS empires for twenty years. The next twenty belong to whoever figures out how to price work, not workers.

---

## KEY FINDINGS

Agent Arbitrage lets customers buy one seat and run dozens of AI workflows, collapsing per-seat revenue models unless vendors meter API access aggressively.

Traditional 85% SaaS gross margins face structural compression because AI agents require variable compute costs that scale linearly with usage.

The market is bifurcating into Copilots (per-seat add-ons like Microsoft at \$30/user) and Agents (per-outcome like Salesforce at \$2/conversation).

Outcome-based pricing creates perverse incentives where AI agents may optimize for quick resolutions rather than genuine customer satisfaction.

As AI agents become primary software users, GUI investments lose value while API reliability and billing accuracy become the differentiators.

---

## REFERENCES

- [1] Lenny's Podcast interview with Intercom CEO Eoghan McCabe
- [2] Intercom Fin AI Agent official pricing documentation
- [3] NBER Working Paper: "Generative AI at Work" by Erik Brynjolfsson et al.
- [4] Growth Unhinged: "The AI Pricing Playbook" by Kyle Poyar
- [5] Salesforce Agentforce Billing Mechanics Documentation
- [6] Microsoft FY24 Q4 Earnings Call Transcript
- [7] Salesforce Q2 FY25 Earnings Conference Call
- [8] Stratechery: "AI and the Incumbents"
- [9] Marc Benioff Dreamforce 2024 Keynote
- [10] ServiceNow Q2 2024 Earnings Call
- [11] All-In Podcast discussion on SaaS margins
- [12] OpenAI API Pricing Documentation
- [13] Bessemer Venture Partners: State of the Cloud 2024
- [14] Tomasz Tunguz: "The Evolution of SaaS Pricing in the AI Era"
- [15] a16z: "The New Business Models for AI"
- [16] Aaron Levie (Box CEO) public commentary

[17] Intercom Engineering Blog: "How Fin Measures Resolutions"

[18] Stripe Documentation: Usage-Based Billing Infrastructure

[19] OpenView: "The State of Usage-Based Pricing"