



— PRODUCTMIND INTELLIGENCE BRIEF • VOL. 2

An AI Transition Field Guide

A SaaS Builder & Operator's Companion

August 2026 • ~26 pages • A companion technical brief, *The Agent-Ready Data Pipeline*, covers Play 2 in full.

Our first intelligence brief in the series mapped the AI landscape to help SaaS teams map the emerging landscape. This guide answers the question that follows: if you run, advise, or build a product inside a SaaS business today, what do you do about it, and when?

\$285B

**left software, financial & asset-management
stocks in a single day**

Feb 3, 2026, based on an AI tool release (Bloomberg)

–130,180

US customer-service jobs, down 4.8% in a year

May 2024 to May 2025, second straight annual fall (BLS)

18-36

months

before seat revenue reprices at renewal

the contractual delay that hides the actual risk in most P&L

~110%

net retention still reported across the leaders

the lagging indicator many executives mistake as safety

CONTENTS

Six questions an operator has to answer.

01 What is actually changing?

READ THE TERRAIN

02 Where am I exposed, and where do I have an advantage?

KNOW YOUR POSITION

03 What choices do I need to make differently?

CHOOSE YOUR PLAY

04 When do I need to act?

TIME YOUR MOVE

05 What should I measure to know if I'm ahead or behind?

TRACK THE SIGNALS

06 What do we carry forward?

PROTECT WHAT MATTERS

→ Make the Crossing

THE STRATEGIC STANCE

INTRODUCTION

A field guide, not a playbook.

A playbook implies a complete and exhaustive method, and that is not our claim. Our work with clients has taught us that different businesses face different exposure depending on target customers, workflow, contract structure, data rights, margins, and competitive position, and no single sequence of plays works across the board with all of those dynamics to bear.

What this guide intends to do instead is help you recognize the terrain, diagnose your own position, read the signals, weigh the legitimate choices, judge how much time you have, and decide how to respond. Our hope is that you find it actionable without pretending to be exhaustive.

We have organized it around six questions we believe operators actually have to answer, in the order a leader would ask them. The four **plays** we refer to throughout (re-architect the business model, rebuild the product for agentic readiness, realign go-to-market, and restructure the organization to run two engines) are high-conviction actions arising from our work and analysis.

THE OPERATOR'S CORE QUESTION	HOW IT IS ANSWERED	WHAT IT GIVES YOU
1 • What is changing?	The evidence, the shift, the mechanism	A read on signal vs. noise
2 • Where am I exposed?	Five realities, and a seven-question diagnostic	A score and a position
3 • What choices?	Four routes, then plays 1 to 3 within them and the dual engine	A strategy, chosen as a set
4 • When?	The financial coupling arithmetic and four clocks that dictate timelines	A timeline you can defend
5 • What do I measure?	Six internal signals to measure progress	Instrumentation
6 • What do we carry forward?	The asset thesis and the talent contract	What you protect while you move

SECTION 01 · READ THE TERRAIN

What is actually changing?

On February 3rd, 2026, roughly **\$285 billion in market value** left software, financial-services, and asset-management stocks in a single day (Bloomberg). US software fell 6%, its largest one-day drop since the April tariff selloff, on the release of a single AI tool, not on any deterioration in anyone's reported numbers.

In our opinion, the underlying impact matters more than the headline. The worst damage was not to per-seat SaaS. LegalZoom fell 20%, Thomson Reuters 16%, LSEG 13%, FactSet 10.5%, Morningstar 9%, while Salesforce and ServiceNow, the canonical seat-priced names, fell about

7%. The market did not sort on "is this priced per seat." **It sorted on "can the model replace the need for people to do this work."**

The selloff was set off by Anthropic releasing a legal-automation tool, so the professional-services and legal-data names took the sharpest hit, the companies directly in that product's path. But the contagion reached financial data, asset management, and horizontal SaaS the legal tool did not touch. That this spread beyond the immediate target is what marks this as a *capability* sort rather than a same-sector reflex.

For two decades, software **helped** a human do the work (especially knowledge work). What is emerging is software that actually **does** it. We call the second thing **Software-as-a-Worker**, or SaaW, and the distinction is not cosmetic: it changes what you sell, who you sell it to, and what a unit of value is.

— FIGURE 01 • THE REFRAME

From Software-as-a-Service (SaaS) to Software-as-a-Worker (SaaW)

	SaaS · today		SaaW · the destination
The software	Helps humans do the work	→	Performs the work
The operator	The human	→	The AI, increasingly
Value is tied to	User productivity	→	Work delivered, output
Natural unit of sale	The seat	→	The work, the outcome
It competes for	The software budget	→	The labor budget 10× LARGER

HOW TO READ IT

Read the right column as a destination that the transition we are working through is aiming for. Almost no incumbent has gotten there yet. But **the labor budget is an order of magnitude larger than the software budget you compete for now**, so the move rightward can actually be a prize; not the threat it's often perceived as.

Every play we write about is a move from the left column toward the right. Pay attention to the last cell in the last row: the labor budget is an order of magnitude larger than the software budget SaaS companies compete for now, and that is the strategic prize in the transition.

The evidence will show in the labor market before the financials

AI's effect on seat-based software is already visible in one place and absent in another. The work that seats used to pay for is being automated, and the headcount doing that work is falling now. The revenue attached to those seats is not, at least not yet, because enterprise contracts delay the transmission. From the market analysts' lens, the bears expect revenue to be falling already, and it is not, while the bulls read healthy revenue as safety when the mechanism that will erode it is already running underneath all of this. Three facts lay it out.

Fact 1: Seat-heavy headcount is falling. US customer-service-representative employment fell 130,180 jobs, or 4.8%, between May 2024 and May 2025 (BLS Occupational Employment and Wage Statistics). Secretaries and administrative assistants fell 1.8%; wholesale and manufacturing sales reps 2.3%. Across the eighteen most AI-exposed occupations, employment declined while total US employment grew 0.8%. Strip out medical secretaries, whose numbers rise on healthcare demand, and the other seventeen fell for the second consecutive year. BLS projections from July 2026 extend it: CSRs down a further 5.5% by 2034, with AI adoption named as a driver.

Occupation-level decline is not proof of AI causation: offshoring, macro softness, the post-2021 over-hiring correction, and reclassification move these same numbers. Three things separate the AI reading from those alternatives: the decline concentrates in exactly the occupations most exposed to current model capability and existing agent automation, not across the board; it runs against a rising total-employment backdrop, so it is not macro; and the two-year duration fits a mechanism that builds as capability ramps, where a one-year move would read as noise. What none of this establishes on its own is the coupling to revenue, which is why we treat it as a leading signal rather than a closed case.

Fact 2: Seat-based revenue is not falling. Through H1 2026, HubSpot's NRR was 105%, up 1.5 points year over year with customers up 16%. monday.com's NDR was 110% with double-digit seat growth. Atlassian's NRR was above 120%. Workday's gross retention was 97%. Asked

directly about seat compression, HubSpot's management said customers bought more seats across every hub through the year, and monday.com reported no degradation in demand relative to seats.

Fact 3: Multiples repriced, not the revenue. Workday beat on the quarter in February 2026 and fell 10% to a fifty-two-week low. Two different things get called "seat changes," and only one is actually happening. Seats are being *repriced*: vendors shifting from per-seat billing to usage, action, and outcome models: real, happening now, and our read from speaking with sales leaders is that they are doing it to earn more per customer, not because seats are shrinking. Seats are not yet *compressing*. Net retention is flat to rising, and executives deny compression on the record. The falling headcount is the early signal of the second, not the first.

RECAP

- 01** We're seeing human intensive work that can be automated by models affect 'seats' via people being let go. CSRs are the canary in the coal mine.
- 02** We're not yet seeing Seat-based revenue decline in a meaningful way. The way to read that is that the SaaS wave had not saturated demand and is pulling through the last few million seats before there is an agent adoption wave.
- 03** The initial squeeze is coming from market expectations via declines in revenue multiples in both the private and public markets

THE MEASURABLE GAP

The gap between mechanism and financials is the only thing here you can measure rather than guess at, and the work in Section 4 turns that gap into a number that gives you something to act on.

Four tangible shifts to note

Underneath the headline, the SaaS-to-SaaW move shows up as four specific shifts. This is a proprietary framework, synthesized from our client work against current BCG, Battery Ventures, and Bessemer reporting. Each is a classic SaaS advantage that turns into a liability unless you deliberately re-found it.

- **Accessibility to autonomy.** The browser was the SaaS distribution breakthrough; the agent is its AI-era equivalent. You now create the conditions for agents to work, not just for people to log in.
- **Per-seat to value-metric pricing.** Recurring revenue averaged across many customers, easily priced by public markets, now works against you. When one agent carries the load of several seats (which would logically lead to revenue lost), value has to eventually attach to work delivered rather than to human logins.
- **Self-service UI to agent-ready ecosystem.** Interfaces and feature checklists are being commoditized by code generation. The emerging moat is robust APIs, structured data, and the interconnections that let AI agents, yours and your customers', read, write, and execute directly. Agents will still need good UI to command and orchestrate but this UI will be different from the UI previously used by a person to do the work.
- **Best practice to hyper-personalization.** Standardized workflows give way to context-aware software that adapts per customer, per role, per moment, learned from usage.

Previous SaaS advantages were built around the human operator. SaaS advantages that take these four shifts as leverage are built around the work itself. The plays in Section 3 are the recommended response; each individual play is customized, so read on.

SECTION 02 · KNOW YOUR POSITION

Where am I exposed, and where do I have an advantage?

Most writing on this transition is generically bearish on incumbents, which is both wrong and useless. The incumbent's problem is not that it lacks assets. It is that **every asset has a liability welded to it**. Getting specific about which is which is the difference between a winning strategy and a panic.

The five Janus realities

So you do not hold five assets with emerging liabilities. Each thing is more like a Janus; a face that looks forward and looks back. So you hold five Janus realities, each with an asset face and a liability face.

— FIGURE 02 • WHERE YOU ARE EXPOSED

Five Janus realities : every asset has a liability welded to it

ASSET FACE		LIABILITY FACE
Revenue base Cash flow to fund a multi-year transition	→	Old pricing models Licensing approach you must dismantle to move
Go-to-market machine Distribution, renewals, executive access	→	A sales force paid on seats Comp plan that actively fights the new model
Product & data estate Proprietary data, workflow integration	→	Human-mediated architecture You may lack the right to use the data
Trust & credibility Security reviews, compliance, discipline	→	SaaS-optimized org & board Review processes that can starve anything new
The bench • human capital Domain knowledge a challenger cannot hire	→	The one asset that can resign THE EXCEPTION Held only by a soft labor market, not a moat

HOW TO READ IT

The four plays convert each liability while the matching asset is still paying for the conversion. The bench is the exception: your people are held in place by a condition, not a contract, and they can walk out faster than your data migrates.

The revenue base. Asset: cash flow, and the ability to fund a multi-year transition from operations rather than dilution. Liability: it comes from seat-based licensing, which is exactly what Play 1 exists to dismantle. Every point of seat revenue you convert is a point of transition funding

you spend.

The go-to-market machine. Asset: distribution, renewal motions, procurement relationships, and executive access a startup spends years trying to gain. Liability: a sales force compensated to sell seats, calling on departmental budget holders, with a comp plan that will fight the new model harder than any competitor.

The product and data estate. Asset: proprietary data, ranking above features and workflow integration in Brief 1's moat hierarchy. Liability: the architecture holding it was built for human-mediated workflow and dashboards, even old legislation; and you may not have the contractual right to use it the way the flywheel requires (see play 1's permission problem).

Trust, and the industry credibility you have earned. Asset: security reviews passed, compliance standing, an operating discipline that ships. Liability: an org chart, a comp structure, and often a board, all optimized for predictable SaaS metrics, exactly the review process that can starve anything new.

The bench, your human capital. Asset: people who hold the domain knowledge and operating history together, the one pairing an AI-native challenger cannot hire for at any price. Liability: this is the only asset on the list that can walk out the door. The other four are held in place by contracts, architecture, or inertia. This one is held in place by a soft labor market, which is a condition, not a moat, and it can change at any time.

The diagnostic: seven questions to help you understand your exposure

Score each with: 1 (weak), 2 (partial), or 3 (strong). Twenty-one points available.

01 Pricing exposure. What share of revenue depends on per-seat licensing, and what happens to NRR if a customer cuts relevant headcount 40%? Score 1 if over 70% is seat-linked with no variable component; 3 if under 30%, or you already bill on work delivered.

02 Agentic readiness. Can a third-party agent discover and execute your core functions via API today, without a human touching your UI? Score 1 if your API is a reporting API; 3 if an external agent completes your top three jobs-to-be-done end to end.

-
- 03 Data position.** Do you hold the ground truth of the work, the record of what was decided, by whom, on what input, with what result? Score 1 if telemetry only; 3 if you hold the decision record and can attribute results to it.
-
- 04 Data rights.** Do your contracts, DPAs, and BAAs permit cross-tenant training and the retention windows a feedback loop needs, and what share of ARR would need renegotiating data rights? Score 1 if you do not know; 3 if you have the rights or have renegotiated.
-
- 05 Outcome accountability.** If you had to charge only for successful outcomes tomorrow, would customers accept your definition of one, and would your contracts survive it? Score 1 if you have no defensible definition of an outcome; 3 if you have one in a signed contract.
-
- 06 Organizational insulation.** Does a team exist that could ship an AI-native product without being cannibalized by the core roadmap's quarterly demands? Score 1 if no such team exists or it reports into the core P&L; 3 if it reports to the CEO on its own KPI set.
-
- 07 Talent exposure.** If hiring conditions improved sharply next quarter, who leaves, and what walks out with them? Have you asked, and do you know what your strongest people are optimizing for? Score 1 if you rely on the market to hold them and have not asked; 3 if you have a declared talent strategy, a named critical population, and a published route into the new work.
-

Four results and their positions

Exposed (7 to 10). High seat concentration, reporting-grade API, telemetry rather than ground truth, no insulated team. Your revenue is more fragile than your financials show. Start with the dual engine (Play 4): you have no vehicle, and everything else will be absorbed by the core roadmap within two quarters.

Instrumented (11 to 14). You have real data and some programmatic execution, but your pricing and your org are unchanged. This is the most common position in mid-2026, and the most deceptive, because your assets look strong so the urgency reads as low. Stand up the engine first

(Play 4), then price into it (Play 1). Resist the temptation to start with the product rebuild (Play 2): it is where engineering-led teams instinctively go, and the play where the most money buys the least structural change.

Dual-Engine (15 to 18). A working part of the company already exists to take advantage of this moment, as a separate unit with its own metrics or with one variable pricing component live. Run the pricing and product plays in parallel inside that forward AI unit, with the guardrails from Play 4. Your failure mode is not slowness, it is erosion: your AI engineers get pulled back to help the core, or the AI unit shows up in the same dashboards as the legacy business and gets starved.

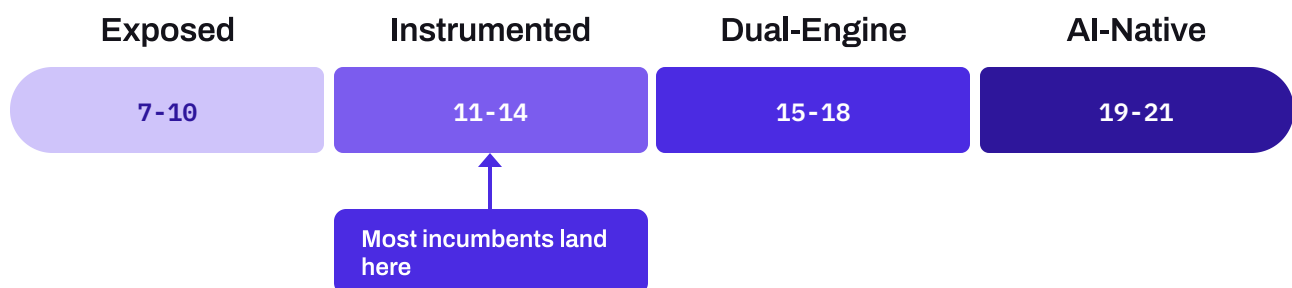
AI-Native (19 to 21). Work-metered or outcome-priced revenue can be material, agents execute core jobs without a human, the feedback loop runs. The focus is to realign go-to-market: the product market fit works, the sales motion just needs to catch up.

— FIGURE 03 • DIAGNOSE YOUR POSITION

Seven questions, twenty-one points, four positions

WHAT YOU SCORE

Score pricing exposure, agentic readiness, data position, data rights, outcome accountability, org insulation, talent exposure



HOW TO READ IT

Instrumented is the most common mid-2026 position and the most deceptive: assets look strong, so urgency reads low. These companies score well on data because a decade of operation produces it, and poorly on architecture, contracts, org, and talent because nobody has been paid to fix them. Whatever your position, answer the timing question this month.

SECTION 03 • CHOOSE YOUR PLAY

What choices do I need to make differently?

There is no ONE answer to this transition. While we focus on the plays in this guide, we know there are fundamentally four legitimate routes to take, and a board should evaluate them as a set before committing to any plays at all. As a leader of a SaaS company, choosing your primary route is the first decision, before executing on the plays below. A route is the path to the plays and there are four of them. Once you settle on the route, the plays make sense. In this brief we will focus on the plays for the first route.

The four routes you need to decide on first

Build your way out

THE PLAYS ARE WRITTEN FOR THIS ROUTE

Fund from operations, run the plays, keep the equity. Dominates when your data estate is genuinely proprietary, your domain trust is hard to buy, and cash flow can carry a multi-year program without cutting into the core.

Buy your way out

Acquire an AI-native challenger and use it as Engine 2. Dominates when your gap is speed and talent rather than data. The failure mode is specific: the acquired team goes under the core P&L, is reviewed on core metrics, and is gone in eighteen months. Acquisition replaces the build, not the structure. You still run Play 4.

Partner your way out

Become the agent-ready data and execution layer under someone else's agent. Dominates when your value genuinely sits in the data and system of record rather than the interface. Undersold as a route, and impossible without the product rebuild: nobody can call a system that is not callable.

Harvest and exit

Run the core for cash, fund no transition, sell while seat revenue still commands a multiple. Dominates when your data is not differentiated, your domain is not defensible, and the read of Section 2 is Exposed with no path to fund the exit. A legitimate answer, and better than a two-year underfunded transformation that arrives at the same place with less cash.

The routes are not mutually exclusive; partner-while-you-build is common. But the choice has a clock on it, because Harvest and Buy both price off a seat-revenue multiple that is compressing. The option you have today is not the option you will have in eight quarters, and that is why route choice cannot wait for certainty.

The plays

PLAY 1

Re-architect the business model

Seat-based licensing puts your revenue in direct conflict with your customer's AI ambitions. Every workflow they automate eventually becomes an invoice line that no longer needs to exist. Structurally, you are betting against your customer's productivity. That is the SaaS-to-SaaS tension expressed as a customer impedance and pricing problem that truly needs to be resolved.

Stage the move; do not switch it. Go to value-based monetization in three stages, metered before outcome-priced. Start with a predictable platform baseline fee, which protects revenue predictability and keeps procurement and your board calm. Add a variable metric tied to the work the agent performs: per action executed, per document processed, per run completed. Then migrate toward delivered outcomes as measurement and attribution mature. Anyone who has worked in advertising and adtech can tell you stories about the risk in the utopian idea of outcome pricing before you have aligned, verifiable attribution.

The public experiments support the staging. Zendesk and Intercom price on resolutions, genuine outcome pricing, and both sit in customer support, the one domain where attribution is single-vendor and event-discrete. Salesforce is the instructive case: Agentforce moved from \$2 per conversation to \$0.10 per autonomous action, then added a flat per-user license (the Agentic Enterprise License Agreement at \$125+ per user per month) in late 2025 after procurement pushed back on consumption pricing. On its FY27 Q1 call it described upgrading existing seats for unlimited AI use, with seven of its top ten deals adding new seats. The largest vendor in the category could not reach a single outcome metric and settled on a portfolio: a) a seat floor procurement will sign, b) a metered middle that scales with the work, and c) outcome pricing where the outcome is clean.

What outcome pricing costs you. When intelligence is a variable cost, a fixed price per outcome transfers unbounded margin risk onto you, and the cases that cost most to serve tend to be the ones the customer values most. Do not price an outcome until you can model its worst case to

serve and the contract caps your exposure: a consumption floor, a volume ceiling, a fair-use boundary, or a stated right to reprice. Metering passes variable cost to the buyer while you learn what the work costs; outcome pricing absorbs it onto your P&L.

WORKED EXAMPLE

Example A: the support-resolution agent. You price a customer-support agent at a flat \$1.50 per resolved ticket. The buyer loves it, because a human-handled contact costs them about \$6. The median ticket the agent closes runs ~3,000 tokens and one tool call, roughly \$0.04 to serve: a 97% margin. It looks like free money.

The distribution is where it bites. About 10% of tickets are gnarly: multi-turn, retrieval-heavy, several tool calls, 80,000 to 150,000 tokens, \$1.80 to \$3.20 to serve. Today, blended, you still make money. But here is the move you didn't price for: the customer's optimal behavior is to route more of the hard tickets to the agent over time, because offloading the hard stuff is the value they bought. As the hard mix drifts from 10% toward 35%, your average cost to serve goes from ~\$0.28 to ~\$0.87, and your margin falls from ~82% to ~42%. At 55% hard mix you're underwater on the flat \$1.50. **The cases that cost you the most to serve are exactly the ones the customer values most, and the contract invites them to send you more of them.**

The fix is to cap exposure without killing the pitch: a fair-use boundary (resolutions past, say, 8 turns or 40k tokens bill at a metered overage), a monthly volume ceiling, or a stated right to reprice at renewal if measured cost-to-serve moves more than X%.

Example B: the legal contract-review agent (value and cost move together). You price at \$200 per contract reviewed, against ~\$600 of associate time. A standard NDA on your own template is ~6,000 tokens, \$0.10 to serve: trivial. But the documents the GC actually loses sleep over, an 80-page bespoke MSA, non-standard indemnity, a foreign-law rider, take multiple passes, a clause-library retrieval, and a human check, \$12 to \$40 to serve, sometimes with a re-run. Still positive at \$200. But if you'd quoted "\$49 a doc" to win the logo, the hard 15% of documents erase the margin on the easy 85%, and the hard ones are precisely what the customer sends you, because that's the work they can't do cheaply themselves. **The discipline: model the 99th-percentile document, not the median, before you name a price.**

The through-line: metering hands the variable cost to the buyer while you're still learning what the work costs; a fixed outcome price absorbs that variance onto your P&L. Don't cross that line until you can model the worst case and the contract caps it.

The P&L consequence. Classic SaaS margins sat in the seventies and eighties because marginal cost was near zero. When inference is a cost of goods sold, model the AI-delivered portion separately or a healthy consolidated number will hide a line of business that does not work. Put the offset in the same conversation: work-metered revenue competes for headcount, BPO, and

consulting budgets, an order of magnitude larger than software budgets. If you are public or PE-owned, disclose the new metric before it is material, with a target mix, a margin profile, and a horizon. You want this seen as the plan, not a deterioration.

WORKED EXAMPLE

How healthy numbers hide a deterioration. Example: a \$250M-ARR public SaaS company. Engine 1 (seats) is \$230M at an 84% gross margin. Engine 2 (agentic, work-metered) is \$20M, but inference and compute are a real cost of goods sold, so it runs at a 38% gross margin.

Consolidated, gross margin is $(230 \times 0.84 + 20 \times 0.38) / 250 = 80.3\%$. It still reads like classic SaaS. The 38% line of business is completely invisible inside the blended number: you can't tell whether it works, and neither can your board.

Now let it succeed, which is the whole point. Six quarters later Engine 2 is \$90M (still ~45% margin as it scales), Engine 1 flat at \$230M. Consolidated margin is now $(230 \times 0.84 + 90 \times 0.45) / 320 = 73\%$. The Street watches blended gross margin fall from ~80% to 73% and prints "margin compression." **You executed the plan perfectly, the AI line is healthy and growing, and the market reads deterioration.**

Two things prevent that. Model the AI-delivered portion separately from day one, so a 38% line never hides inside an 80% average and you can actually see if it works. And put the offset in the same sentence: that \$90M isn't competing for the software budget, it's competing for the labor, BPO, and consulting budget that's roughly 10x larger, so 45% margin on a labor-budget dollar can throw off more absolute gross profit than 84% on a shrinking seat.

Then, if you're public or PE-owned, disclose before it's material: a target mix ("AI-delivered revenue ~30% of total by FY28"), a margin profile ("stabilizing at 50 to 55% as inference costs fall and utilization rises"), and a horizon. The punchline for the board: the number that scares the market, blended margin down seven points, is the same number that proves the strategy is working. The only variable you control is whether they saw it coming.

The permission problem. The data moat has a prerequisite most transition plans skip: the right to use the data. Enterprise contracts, DPAs, BAAs, and public-sector terms routinely prohibit training on customer data, require per-tenant isolation, or cap retention below what a feedback loop needs. Three things, in order. Audit: what share of ARR sits under terms prohibiting cross-tenant training? Segment: separate what tenant-isolated data allows, usually far more than teams assume, from what genuinely requires pooled data. Renegotiate deliberately: fold the terms into renewals rather than opening a separate conversation, and price the permission, because customers will trade data rights for capability and the trade is easier when the capability is visible. Treat it as a twelve-month program with a named owner. It gates the moat.

PLAY 2

Rebuild the product for agentic readiness

This is where SaaS stops being a frame and becomes an engineering program, the strategic shift that changes your product. Traditional SaaS architecture assumes a human mediator: dashboards to read, forms to fill, buttons to click, batch jobs that refresh overnight. Agentic architecture assumes the software does the work: asynchronous, event-driven, surfacing to a human only for judgment and exceptions. Four moves.

- **Deconstruct features into agent-callable tools.** Every core function should exist as a standardized, documented, discoverable capability with structured inputs and outputs, executable by an agent without human interaction. Agents work through tools, so the unit of design is the tool, not the screen. This is the single highest-leverage engineering investment in this guide.
- **Move from scheduled batches to event streams, and audit the new traffic shape.** Human usage is bursty and your customers sleep; agents run continuously, at higher frequency, fetching fresh data rather than tolerating a cached view. Most SaaS infrastructure was not built for that level of interaction and data.
- **Build the memory layer.** Agents need fast semantic recall (what does this contract say?) and structural truth (which customer, which contract, which SLA?), which means pairing vector search with a knowledge graph of your domain's entities and your customer's activity and record data. Done right, your decade of proprietary data becomes a deployed advantage rather than a dormant one.
- **Close the loop.** Embed feedback capture into the core experience so every human correction becomes a training signal. This is the flywheel that separates Type 2 from Type 1 companies (see Brief 1). Clear the permission problem first: using your customers' feedback to train your model is something you negotiate up front.

The board-level takeaway: agentic readiness is a data-infrastructure investment first and a model investment second. Frontier models are converging on capability and competing on price. Your pipeline into them is not. Re-rank your product metrics to match: data completeness, ingestion

frequency, and ground-truth precision become first-class KPIs, above DAU and feature adoption. If your board deck still leads with DAU in 2026, the deck is measuring the last era.

PLAY 3

Realign go-to-market

Here we make the last row of the SaaW table operational. The old SaaS pitch sold marginal productivity to departmental managers: save each user two hours a week. The new pitch sells scalable digital workers to executive budget holders: deliver this outcome, at this cost, with this reliability, and this audit trail. That shift changes who you sell to, what you quantify, and what you compete against. The comparison set for a work-metered agent is not another SaaS tool. It is headcount, BPO contracts, and consulting spend: budgets an order of magnitude larger, carrying an order of magnitude more scrutiny on reliability and accountability.

The positioning we have seen close the deal: our agents and your agents can work together. Ours act on your behalf inside our system, yours reach into our data and execute against it, and both sides hand work back and forth without a person retyping anything. Make the conversation about real work done by agents on both sides, and everyone who is up to speed in the room can follow the pitch and see the value. Three GTM moves:

- **Re-train the sales force** to quantify financial value, risk, and execution speed rather than demo features.
- **Re-align compensation**, because a rep paid on committed seats will fight the new model harder than any competitor.
- **Build the proof apparatus:** attribution dashboards, reliability SLAs, inference cost tools and audit trails are sales collateral now, not just product features.

As of mid-2026, Battery Ventures' June survey puts 49% of enterprises actively deploying agentic AI, up from 33% six months earlier, with 50% scaling across functions. Production is no longer rare. What persists is the evidence problem and the sticker shock of intelligence costs running past

the labor they replace. That is a real incumbent advantage: you have the deployment surface, the data, and the trust to prove outcomes where a startup must ask for faith. Sell against the failed pilot: you have tried the demo, we already run in your system of record.

PLAY 4

Restructure the organization and run two engines

If you are an incumbent or a mid-to-large SaaS company, the hard truth is that the plays above cannot be run inside a single organization optimized to protect quarterly numbers. Ask one org to do both and you get a predictable failure: fragmented AI features that demo well and never reach production, while the genuinely new product starves for talent every time a legacy escalation lands. The answer is separation, which BCG reaches independently and our client work confirms (BCG, *The AI-First SaaS Company*, April 2026).

ENGINE 1

Optimize and Protect

Maximize cash flow, retention, and efficiency of the existing platform; this engine funds the transition. Led by the existing VP Product and VP Sales, measured on churn, NRR, operating margin, and time to value. Its AI mandate is internal efficiency and AI applied to existing features to cut time to value. It does not build the new business. It runs the hybrid pricing bridge on the existing book.

ENGINE 2

Invent and Disrupt

Build agentic, work-metered and outcome-priced revenue lines targeting a minimum 2x expansion in revenue per customer (BCG, April 2026). Reports to the CEO or a dedicated Chief AI Officer, insulated from core quarterly metrics. Staffed with ML engineers, agent architects, PMs fluent in evals, and legal embedded from day one. Measured on model and task execution success rates, production deployments, and new-model revenue, not on legacy KPIs.

Engine 2 will look bad next to Engine 1, and the board should hear this early rather than late. A ten-year flywheel at 110% net retention will beat a new line on every metric a board is used to reading, for years. A board that forgets this will kill the thing it funded while believing it is exercising discipline.

Two things keep the model real rather than theater. **Talent ring-fencing:** Engine 2 engineers cannot be pulled into critical bug fixes or legacy feature requests, because that first exception quickly becomes the policy; planned, chartered rotation between the engines is not only allowed but necessary. **Distinct KPI frameworks, enforced at board level:** if Engine 2 appears against NRR and margin in the same board slide as Engine 1, it will be starved rationally, quarter by quarter, by people doing their jobs well.

How you fund Engine 2 matters as much as where it reports. Fund it as protected, new-growth money, not as a slice carved out of the existing R&D budget. Money folded into core R&D gets judged on core timelines and starved. The market already drifts this way: ICONIQ finds high-growth firms now pointing something like half of R&D at AI (ICONIQ, *State of AI*, January 2026), often out of existing R&D rather than fresh allocation, which is exactly the mistake to warn against, because it will feel like the disciplined choice.

SECTION 04 · TIME YOUR MOVE

When do I need to act?

Most transition advice gets timing wrong in both directions. The people telling you to reprice everything this year are reacting to a stock multiple shifts; they're likely panicing and too early. The people telling you your numbers are fine are reading a lagging indicator; they don't hear the train coming until its too late. Neither is watching the actual clocks, and there is more than one.

Four clocks, running at different speeds

The revenue clock THE SLOW ONE

How long before falling seat usage actually reaches your revenue. The only one you can put a number to, because it is contractual: seat reductions hit revenue only at renewal. For most enterprise books it runs 18 to 36 months, longer than the market is pricing and shorter than your dashboard suggests.

The talent clock RUNS OUT FIRST

How long before the people you need to run this transition start to leave. Your seats are locked in by multi-year contracts; your people are not. A soft hiring market is holding your bench in place the way renewals are holding your revenue in place, masking a number that has already changed underneath. It is the only asset in Section 2 with no contractual protection.

The competitive clock

How long before your customers' work starts flowing through someone else's agent. Every quarter of delay raises the odds, and data is the one asset you cannot re-acquire once it has started flowing elsewhere. You do not control this clock; it runs on your fastest-moving competitor's next product decision.

The capability clock

How long before the economics of outcome pricing turn in your favor. Inference-cost trajectory decides whether pricing on outcomes is survivable, and when. If costs fall far and fast, the margin caution in Play 1 relaxes and the destination arrives sooner. You do not control this one either, but you can watch it (Section 5).

Naming four instead of one gives you permission to sequence your response. The revenue clock gives you room on repricing; you do not have to touch the whole book this year. The talent clock gives you no room at all on the org and the bench, you should rethink org design and motivation very soon. That asymmetry (more time on pricing, less time on people) is the entire argument for doing things in an order rather than all at once.

The revenue clock: how to estimate how much time you have

The lag between headcount moving and revenue moving is not mysterious. It is contractual: seat reductions reach your revenue only at renewal, so the delay is roughly the average remaining term on your book. We use a simple model with clients to turn that intuition into a number they can plan against. The annual drag on gross retention, once it begins, is approximately:

THE MATH

Annual gross-retention drag

$$\approx S \times [1 - (1 - h)^T] \times r$$

S share of ARR attached to seat-heavy job families

h annual headcount decline in those families at your customers

T average contract term in years

r share of the book renewing annually, roughly $1/T$

Worked through: at $S = 40\%$, $h = 5\%$, $T = 3$, $r = 33\%$, a customer arrives at renewal with about 14% fewer relevant heads, and the drag is roughly 1.9 points of gross retention in the first affected year, rising as more of the book renews into the same reality. Against 10 to 15 points of expansion, that is invisible for a while, and then it is not. We use $h = 5\%$ here because it is close to the customer-service decline the BLS data already shows. Substitute your own.

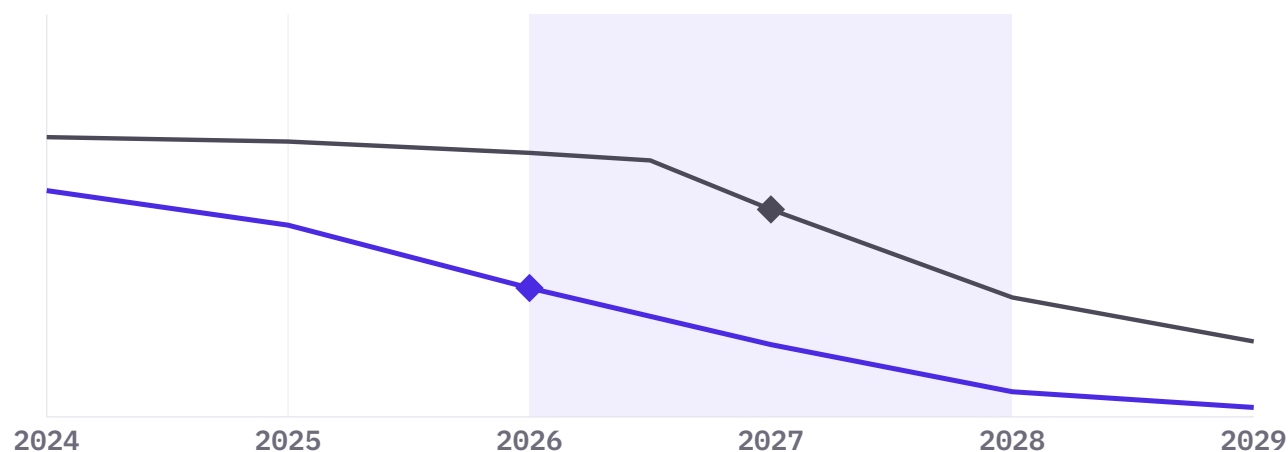
Be clear about what this model does and does not do. It assumes the headcount decline holds steady across the term, that renewals land evenly rather than in lumps, and that expansion elsewhere does not mask the drag. Real books violate all three to some degree. Treat it as a lens for sizing the window, not a promise about any single quarter. Run across the contract terms common in enterprise and SMB SaaS, roughly one to three years, and it lands in that 18-to-36-month range before a seat-heavy book has substantially repriced.

FIGURE 04 • THE SIGNATURE MECHANISM

Labor leads. Revenue lags. The gap is your window.

— **Seat revenue / NRR** revenue only moves at renewal

— **Seat-heavy headcount** already falling (BLS)



THE 18-36 MONTH WINDOW first renewals price in the new reality

AI capability → work automated → headcount falls → seats cut at renewal → **ARR resets**

HOW TO READ IT

The lag is not mysterious, it is contractual: seat reductions reach revenue only at renewal. So you have **more time than the market is pricing, and less time than your dashboard suggests**. The transition is urgent not because revenue is falling, but because revenue is often the last thing to fall.

How much does the engine cost?

Before the sequence, the question every CFO asks first: what does standing up Engine 2 (Invent and Disrupt) actually cost. No incumbent discloses a clean operating budget for a ring-fenced AI unit, so we build the number from cost drivers that are measurable.

The team dominates the cost. The pattern for a new AI unit inside an incumbent is small, autonomous, and CEO-adjacent, closer to a skunk team than a division. A credible minimum viable Engine 2 is roughly ten to eighteen people (depending on the size of our company): a unit lead reporting to the CEO, four to seven applied-AI and agent engineers, one or two ML engineers,

one or two PMs fluent in models, agents and evals, a designer, a forward-deployed engineer who sits with early customers, and legal and data engineering embedded rather than borrowed. Smaller is a pilot, not an engine. Much larger rebuilds the bureaucracy the ring-fence was meant to escape.

Senior AI and ML talent moves everything else. Fully loaded, including equity, benefits, and overhead, budget roughly \$300,000 to \$500,000 per senior technical person, higher in the Bay Area and New York (Carta 2025 startup-compensation data and BLS employer-cost loading). Inference and compute is the second line, and it behaves unlike classic SaaS: per-token prices are falling fast, but agentic workloads consume many times the tokens of a chatbot query, so total spend rises as the product scales. Model inference runs roughly a fifth to a quarter of AI product cost at scaling-stage firms (ICONIQ, 2026). Put together, a ten-to-eighteen-person engine runs on the order of \$6M to \$18M fully loaded in year one; a leaner eight-to-twelve-person launch team, \$3M to \$8M.

We recommend framing the commitment as a share rather than an absolute: roughly **five to ten percent of R&D, or one to three percent of operating cash flow**, for the ring-fenced unit. That sits below the share of R&D companies now point at AI generally, because Engine 2 is a new-growth bet rather than the AI features the core ships, and it echoes the old three-horizons guardrail that reserves about a tenth of resources for new-growth options. Large enough to matter, small enough that the board can carry it through a soft quarter without a special conversation. Whichever structure you choose, stage it against milestones rather than the calendar: gate year one on a shipped agentic capability and first design-partner pilots, gate year two on first paid deployments and a per-query gross margin trending the right way.

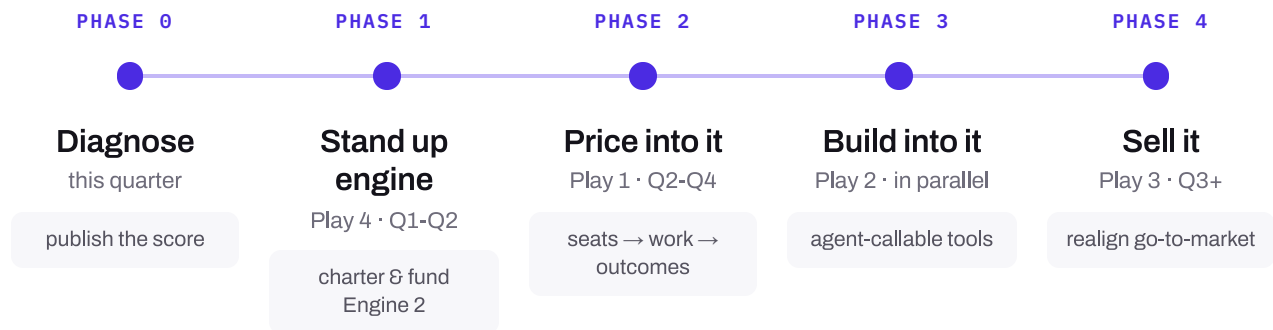
The sequence

When we analyzed the plays, we numbered them in the order that builds an investment argument: model, product, go-to-market, organization. Working with clients, the order for *doing* turned out to be different. Two things kept surfacing as the real constraints, not the pricing or the architecture: the talent risk is the largest and the fastest-moving, and organizational change is far harder than any of the technical work. For that reason we now usually recommend standing up the organization first. Play 4 leads.

So the order is: stand up the engine (the talent structure that produces it), price into it, build into it, then sell what it produces. Section 2 tells you where to enter. Each phase carries a gate: the thing that has to be true before the next phase is worth starting.

— FIGURE 05 · THE SEQUENCE

Stand up the engine, price into it, build into it, then sell it



The plays are numbered for reading. This is the order for doing. Each phase carries a gate: what must be true before the next one is worth starting.

*The failure mode of the transition is not slowness, it is **erosion**: the AI unit shows up in the same dashboards as the legacy business and gets starved because the old metrics do not justify the investment. Sequence, and a ring-fenced engine, are the response.*

Phase 0, this quarter, whatever your position. Run the diagnostic and publish the score (CEO). Audit data rights across the ARR base (General Counsel). Model seat exposure at 20% and 40% headcount decline using the arithmetic above (CFO). Inventory the ground-truth data you hold (CPO/CTO). *Gate*: the exec team agrees on a position, and the transmission window is a number you can take to the board.

Phase 1, quarters 1 to 2. Stand up the engine (Play 4). Charter and fund Engine 2: mandate, reporting line, P&L boundary, its own KPI set, a named team, and a run rate sized against the share target above. Ring-fence talent in writing with a named escalation exception. Embed legal from day one. Publish the talent contract from Section 6 alongside the charter. *Gate*: the board has approved the charter and the distinct metrics, the money is committed for four quarters, and the talent contract has been said out loud to the whole company.

Phase 2, quarters 2 to 4. Price into it (Play 1). Choose the work metric before the outcome metric, with the cost-to-serve model behind it (CPO/CFO). Model worst-case inference cost and set the exposure cap (CFO). Launch platform fee plus work metric in Engine 2, run the hybrid bridge in Engine 1. *Gate:* a contract is signed on the new structure, and you can produce a bill the customer accepts without manual reconciliation.

Phase 3, quarters 2 to 4, in parallel. Build into it (Play 2). Audit agent-executability of your top ten core functions and expose two end to end (CTO). Rebuild data infrastructure for streaming, retrieval, and feedback. Ship one agent-native capability to production, not to demo. Renegotiate data rights on the Phase 0 accounts. *Gate:* an external agent completes two core jobs without a human, and one capability is live with real customers and a measured success rate.

Phase 4, quarters 3 to 4 and beyond. Sell it (Play 3). Move messaging from feature utility to work execution (CMO). Identify executive budget owners and re-align sales incentives (CRO/CFO). Build the proof apparatus (CPO/CTO). *Gate:* the comp plan pays on work-metered revenue without a penalty, and a customer-facing outcome report exists.

SECTION 05 · TRACK THE SIGNALS

What should I measure to know whether I'm ahead or behind?

There is a saying that a goal without a plan is just a wish. The same is true of a destination without a way to measure progress toward it. Most SaaS boards are still measuring the last era: DAU, feature adoption, seat count. Those tell you how the business you are leaving is doing, not how fast the one you are building is arriving.

Two layers matter. The macro dashboard from Brief 1 still runs in the background: outcome-pricing adoption, production deployment rates, regulatory movement on agent liability, and model-cost trajectory. Those tell you where the market is. But the signals that tell you where you sit inside your own business, and there are six of them. Review them quarterly.

The six internal signals

The first four measure how fast Engine 2 is growing.

-
- 01 Executable core.** Share of your core jobs-to-be-done that an agent can execute without a human touching the UI. *Target: above zero and rising every quarter.*
-
- 02 Work-metered revenue** as a share of total. *Target: 2 to 5% and growing.*
-
- 03 Outcome-priced revenue** as a share of total. *Target: above zero in at least one segment.*
Tracked separately from work-metered revenue on purpose, because merged, the number can be hit with metered API pricing that changes nothing structural.
-
- 04 Training-usable data** captured per month under permissive terms. *Target: rising month over month.*
-

The last two measure how fast the assets funding all of it are depleting.

- 05 Latent attrition in the critical population.** Share who have shipped agentic work in the last two quarters, internal applications per Engine 2 posting, and the three-offers answer (Section 6). *Target: measured twice a year at minimum, with a trigger level set in advance.*
-
- 06 Engine 1 erosion.** Seat NRR, seat count, and seat-heavy headcount at your top 20 accounts. *Target: tracked, with a trigger level set in advance.*
-

Four signals tell you how fast the new engine is coming up; two tell you how fast the fuel is running out. Between them, they set your true timeline. A program that watches only its own progress cannot tell the difference between having time and being out of it.

What would prove us wrong

We hold a strong point of view in this guide, and a point of view that cannot be disconfirmed is not worth much. So here is what we are watching that would tell us we are wrong. If you are testing our thesis against your own business, these are the same things you should watch.

- **Seat revenue holds through 2027 while seat-heavy headcount keeps falling.** That would mean enterprises are redeploying freed capacity rather than cutting licenses, and the revenue coupling mechanism is weaker than we think. *Our response: slow Play 1, keep Plays 2 and 4, which pay off under every scenario.*
- **Agent-actions become the dominant new metric, and stay that way.** If by 2028 outcome pricing is still confined to support deflection, Play 1's destination is wrong even though its direction is right. *Our response: stop at the metered stage. It is a complete and valid business model, not just a waypoint.*
- **Inference costs fall far and fast enough that COGS stops being a constraint.** The margin caution in Play 1 goes away. *Our response: accelerate. This is the pleasant version of being wrong.*
- **Ring-fenced units underperform absorbed AI programs.** The play we watch hardest. But ring-fenced units almost never fail because separation was wrong; they fail because it was half-done: judged on the core's clock, cut off from the core's data, weak on talent, or asked for revenue too early. *Our response: rule those out first (see Appendix A). If they are genuinely clean and the unit still lags, a single engine is the honest answer.*
- **Your best people stay through a hiring recovery even without the talent contract.** Then the soft market was not the only thing holding them, and the full contract was insurance you did not need. *Our response: keep the two pieces that serve Engine 2 regardless (the rotation that moves domain knowledge in, and the real Engine 1 mandate), and drop the rest.*
- **Your own top-20 accounts show no seat-heavy headcount decline over four quarters.** Every other test leans on national data; this one runs on your own customers, the clearest read you will get. *Our response: if your customers' headcount is genuinely flat, you have more time than the thesis implies: slow Play 1, keep Plays 2 and 4, and re-run this test every quarter.*

If none of your signals ever move against you, it is very likely you are measuring the wrong things.

APPENDIX A

The dual engine and its risks

Why a dual engine at all. This guide is written mostly for incumbents: mid-size to enterprise SaaS companies, and the larger, well-funded startups, roughly 1,000 employees and up. That scale is the reason the dual engine exists. A company this size cannot simply pivot the core: too many customers on the old model, too much revenue to protect, too many people whose job is to keep the existing business running well. So you run two engines, held apart on purpose because they answer to different economics and different clocks.

That recommendation carries an assumption you should check against your own situation: we assume the competitive threat is real but not yet life-threatening: the mechanism is running underneath you, the window is open, but the building is not on fire this quarter. If your situation is genuinely existential, a competitor is taking your customers now and the core is in freefall, the dual engine is the wrong tool; the whole company moves at once, and you accept the churn from customers and talent. Likewise, below roughly a few hundred people, you may not have enough talent to staff two engines without gutting the core, and a single engine transforming directly, with the CEO owning the quarterly risk, is the honest answer.

When it looks like it is failing, rule out the five ways a separate unit gets botched first. In our work, and in the wider record of corporate innovation units, underperformance almost always traces to one of these, not to separation itself. Ring-fenced units rarely fail because separation was the wrong call; they fail because it was half-done.

01 It was judged on the core's clock. The most common killer. Engine 2 is measured against Engine 1's NRR, margin, and payback, loses every quarter, and gets defunded by people being disciplined. *Mitigation:* a distinct KPI framework, enforced at board level, with a pre-agreed evaluation horizon of eight quarters, written into the charter before the first review.

02 It was starved of connective tissue. Separated so cleanly it lost access to the core's data, customers, and distribution, the very assets meant to be its advantage over a startup. *Mitigation:* separation of incentives, not separation of assets. Guarantee Engine 2 privileged access to Engine 1's data and accounts, in writing.

03 The talent bar was too low, or too borrowed. Staffed with whoever was free in the core rather than people who can actually build agentic systems, or outsiders who do not understand the domain. *Mitigation:* the named-critical-population discipline from Section 6, plus chartered rotation so domain knowledge flows in without the core reclaiming the people.

04 It had no real executive air cover. Reports nominally to the CEO but is run day to day through a divisional leader whose bonus is the core number, so the first hard tradeoff goes against Engine 2. *Mitigation:* genuine CEO or dedicated Chief AI Officer ownership, with the unit lead in the room where capital is allocated.

05 It was asked to show core-scale revenue too early. The eight-quarter reality of first revenue collides with a board expecting a material line in year one. *Mitigation:* stage-gate on capability and design-partner milestones, not revenue, until quarter six or eight (the sequence gates in Section 4).

The genuine failure case, named honestly. After you have ruled out all five, the dual engine can still be wrong for a specific company: too small to staff two engines without gutting the core, or under a threat so immediate that a protected side unit is too slow. In both, a single engine is the better structure, not because separation is a bad idea, but because your scale or your clock does not support it.

The precedents worth studying, in both directions. Amazon's two-pizza-team model and the decision to build AWS deliberately apart from retail are the case for structured separation with real air cover. The long record of corporate innovation labs that produced impressive demos and no business, and absorbed early AI efforts that never escaped the core roadmap, are the case against half-separation. Christensen's finding holds across four decades: incumbents rarely fail because the disruptive unit could not build. They fail because the core reabsorbed its people, its budget, or its attention before the new thing could stand on its own.

SECTION 06 · PROTECT WHAT MATTERS

What do we carry forward?

Most transformation advice tells executives to abandon the old business and run at the new one. Our experience points the other way. The assets that make a successful transition possible are the ones you already own, sitting inside the business you are being told to leave behind. The job is not to escape the old model. It is to take apart its liabilities without breaking the assets attached to them. So the last question is not what to change, it is what to protect while you change everything else. Two things have to hold.

The assets that fund the change have to survive it

Go back to the trap in Section 2: the cash flow paying for this transition comes from the pricing model the transition is meant to dismantle. That is not a clever contradiction. It is a practical constraint, and it sets the pace. You convert each liability while the matching asset is still paying for the work, and never faster than that asset can carry. Reprice the book before you can measure the work, or re-architect before you know which capabilities agents actually call, and you have spent the asset without converting the liability. The five realities from Section 2 are not just a diagnostic score. They are the list of things that still have to be standing when you reach the other side.

Of the five, the one that moves fastest, and the one you cannot buy back, is your data. Every quarter your customers' work runs through your system instead of someone else's agent is a quarter you hold the moat. That is the competitive clock from Section 4, and it is the reason the product rebuild in Play 2 earns its keep under every scenario, including the ones where the pricing thesis turns out wrong.

The bench: the asset that can resign

Four of the five realities are held in place by contracts, architecture, or inertia. Your bench, your people, are held in place by a soft labor market, which is a condition that can change. It is the only asset on the list that can walk out, and it will walk out faster than the data migrates. The ring-fencing rule from Play 4 creates a problem it does not solve: ring-fencing staffs Engine 2 and manufactures

a second problem: capable people left in Engine 1, reading the org chart accurately, concluding that the interesting work has moved somewhere they were not invited. Run the structure without a talent strategy beside it, and you will staff Engine 2 by quietly demoralizing the engine that funds it.

No single study measures the ring-fenced-unit talent drain directly; the mechanism is a synthesis of our experience and three literatures that converge on it. **Organizational ambidexterity** is the structural half: separating an exploratory unit from the core works only when the senior team runs on common-fate incentives anchored on the whole firm rather than the new unit's wins (O'Reilly and Tushman, 2008). **Relative deprivation** is the human half: in a study of a merged company, employees' sense of relative disadvantage predicted intent to leave, fully mediated by weakened identification with the organization (Cho, Lee and Kim, 2014). And **Christensen's account** of why incumbents fail is, underneath, a talent-allocation story: the best people follow the projects that enhance their careers (Christensen, 1997).

The leavers. Your strongest people going to AI-native companies outright. You will not out-compensate a challenger's equity, and you should stop trying. What you have that they do not is the thing your data moat is made of: real domain problems, real customers, real consequences, and a decade of proprietary data to build against. Say that specifically, in the room, or you are competing on the one axis where you lose.

The stayers under duress. Good people who want new problems, are frustrated to be on the core, and are staying because the market is soft, with full knowledge that they will move when it turns. This population does not show up in your regretted-attrition rate, because the soft market is suppressing the departures, not the intent. Gallup put the share of US employees watching or actively seeking at 51% as of May 2024, the highest since 2015, even as actual voluntary turnover had stabilized. This is not a low attrition rate. It is a deferred one, and the deferral ends on a clock you do not control.

The talent contract: five commitments

Declare the critical portions of these commitments to the whole company whenever possible.

-
- 01 A published route into Engine 2.** Selection by tap on the shoulder produces one motivated team and a resentful company. Publish the criteria, run it on a cadence, and build a return path, so a rotation into Engine 2 is not a one-way career bet. Doing this makes it a motivational asset not a liability.
-
- 02 A real Engine 1 mandate, staffed like one.** Engine 1's AI remit is rebuild work, not maintenance: agentic tooling on the core, internal automation, technical-debt reduction at a pace the old staffing never permitted. If that mandate is real, fund it and say so. If it is not, do not pretend; your engineers can tell inside a sprint.
-
- 03 A skills guarantee, with time and money attached.** What your good people are optimizing for is not equity. It is not being obsolete in three years. Hours in the week, a training budget, and production work to apply it to are cheaper than replacing them and more credible than a retention bonus.
-
- 04 The truth about the roles that shrink.** Some functions on both sides will be smaller in three years. People leave faster when they suspect a plan they have not been told about than when they are told a hard one. Name what you know, name what you do not, and commit to a date when you will say more.
-
- 05 A named critical population.** A retention strategy that treats everyone as critical protects nobody. The intersection that matters is people who hold deep domain knowledge and can build agentic systems, the pairing an AI-native challenger cannot hire for and most wants. Identify these people by name, know what each is optimizing for, and review the list quarterly at executive level.
-

The ring-fence of the engine 2 team and a rotation plan are not in conflict. Play 4 bars Engine 2 engineers from being pulled back into Engine 1's escalations; commitment 1 builds a rotation between the engines. The purpose of the ring-fence is to block unplanned, reactive pulls: the 2 a.m.

escalation, the legacy feature request that jumps the queue. The rotation is planned, scheduled, and chartered: a defined tour with a return path, staffed in advance. The rule is "no unplanned movement out of Engine 2," and it belongs in the charter in those words.

If you had three offers next month, would you still be here?

THE LATENT ATTRITION QUESTION. ASK THIS OF YOUR CRITICAL TALENT POPULATION.

→ MAKE THE CROSSING

The most important strategic stance

Brief 1 closed on a simple observation: in every platform shift, the value goes to whoever builds the trains, not the rails. For incumbents, we would add one line. You already own the track, the stations, and the passengers. The only question is whether you put a new train on it before someone else runs one down your line.

The mistake we watch companies make is not moving too slowly on AI features. It is moving fast on features while leaving the business model, the architecture, and the org chart exactly as they were. That path feels like progress and produces a Type 1 company: busy, full of demos, but structurally identical to the thing being disrupted. It is SaaS with better marketing lingo and some good demos. It is not SaaW.

Here is to the bottom-line.: the pressure is probably not in your revenue yet (for some it is). It is in your customers' headcount or seat reductions, and revenue is the lagging indicator. Your advantages are real, but they are perishable, and two of them are running down on clocks you don't control: the data that starts flowing through someone else's agent, and the people who understand that data and can walk out the door in a month. Neither will wait for your strategy to be finished before deciding whether one exists.

The six questions are the whole method, and they are meant to be used in order. Know what is actually changing. Know where you are exposed and where you are strong. Choose your route, then run the plays it calls for. Estimate your clocks and respect that they do not run at the same speed. Instrument the transition so you can tell time remaining from time elapsed. And protect the assets that make the crossing possible, because the incumbent who reaches the far side with its data, its trust, and its bench intact is the one the challengers were never able to become. That, in the end, is the whole advantage: not that you move first, but that you are the only one who could have made this particular crossing at all.

THE WINDOW

So diagnose honestly, and start where you stand.

Stand up the engine, price into it, build into it, then sell what it produces. The window is open now. It will not stay open on your schedule.

If you need some help figuring this out, we can help. Just email hello@productmind.co.

KEY SOURCES & FURTHER READING

Grounded in primary sources.

ANALYST & ADVISORY RESEARCH

BCG · The AI-First SaaS Company: Rethinking the Playbook (Apr 2026)

Battery Ventures · The AI Supercycle Has Arrived (Dec 2025); State of Enterprise Tech Spending (Jun 2026); agentic AI adoption survey (Jun 30, 2026)

Bessemer Venture Partners · Building Vertical AI (Jan 2026); The AI Pricing & Monetization Playbook (Feb 2026)

Gartner · 40% of enterprise SaaS spend shifting to usage-, agent-, or outcome-based pricing by 2030

IDC · Is SaaS Dead? (Dec 2, 2025)

ICONIQ Capital · State of AI: Bi-Annual Snapshot (Jan 2026)

SaaS Capital · AI Assessment Framework (Feb 17, updated Mar 12, 2026)

iMerge Advisors · The AI-Centric Imperative

LABOR MARKET & WORKFORCE DATA

US Bureau of Labor Statistics · Occupational Employment & Wage Statistics (May 2025); AI, IT & Employment 2024-34 (Jul 16, 2026)

McKinsey · analysis of the Understanding Society longitudinal survey

Gallup · State of the Global Workplace (2024); US turnover-intent tracking (May 2024)

Revelio Labs · workforce-attrition modeling

PwC · Global AI Jobs Barometer (2025, 2026 update)

Lightcast · Developing the AI Skills Employers Actually Need (2025)

Josh Bersin Company & AMS · Talent Climate / Internal Hiring (2023)

Carta · State of Startup Compensation (2025)

AllianceBernstein · Software's Big Sell-Off: Structural Risk or Narrative Noise?

COMPANY DISCLOSURES & PRICING

Salesforce · Agentforce pricing (May 15, 2025); FY27 Q1 results (May 27, 2026)

HubSpot · Q4 2025 results (Feb 11, 2026)

monday.com · Q1 2026 results (May 11, 2026)

Atlassian · Q4 FY26 results (Aug 2026)

Workday · Q4 FY26 prepared remarks (Feb 24, 2026)

Zendesk · **Intercom** · resolution-based pricing (accessed Aug 19, 2026)

Bloomberg · software selloff coverage (Feb 3, 2026)

ORGANIZATIONAL & INNOVATION RESEARCH

O'Reilly, C.A. & Tushman, M.L. · "Ambidexterity as a Dynamic Capability," Research in Organizational Behavior (2008)

Cho, B., Lee, D.-G. & Kim, K. · "How Does Relative Deprivation Influence Employee Intention to Leave a Merged Company?," Human Resource Management 53(3) (2014)

Christensen, C.M. · The Innovator's Dilemma (Harvard Business School Press, 1997)

ProductMind · AI Intelligence Brief (Brief 1, May 2026); The Agent-Ready Data Pipeline (companion brief, Aug 2026)