

A STRATEGY BOOK

After Code

*The Repricing of
the Software
Agency*

Mo Fadaly

WHAT'S STILL
WORTH PAYING FOR



A STRATEGY BOOK

After Code

The Repricing of the Software Agency

MO FADALY

WHAT'S STILL WORTH PAYING FOR

After Code: The Repricing of the Software Agency

Every named quotation is from a public source (see A Note on Sources and Method).
Illustrative dollar figures are hypothetical worked examples, not reports of real
transactions.

This is operating guidance, not legal, accounting, or financial advice.

Contents

Introduction

1 · The Repricing Engine

2 · What Still Gets Paid For

3 · Sell the Slice

4 · Sign What You Can Carry

5 · Show Your Work

6 · Survive the Gap

7 · Keep the Room

Appendix A · The Operating Kit

Appendix B · The Case Files

Appendix C · The Repricing Index

Appendix D · The Lexicon

A Note on Sources and Method

The Repricing of the Software Agency

The work your agency does is still good — maybe the best it has ever been. What changed is the price, and where in the work the value now sits. But value moving is not the same as margin moving: the buyer can pocket the AI savings and still pay you less. This book is about that shift, and the harder problem of getting paid for the half that holds.

IN THIS CHAPTER

- Why the math can feel like it's bending even when revenue and pipeline look fine
- The split between what you make and what you answer for — and why the margin doesn't follow the value on its own
- The five moves, and why each is harder than it sounds

You run a good agency. The team is sharp, the clients are real, the portfolio holds up in any room you walk it into. And yet something underneath the business has started to feel wrong in a way the dashboards don't quite explain.

The proposal that won last year comes back with more questions and a harder number attached. A client who never asked before wants to know exactly what your people did that their own tools couldn't have done without you. A deal you would once have closed on reputation stalls in a review you didn't know existed. Revenue is fine. The pipeline

is fine. And still, the numbers are starting to work against you — quietly, before it shows up anywhere you'd flag it.

You're not the only one who has noticed. In late 2025, the chief executive of Endava said it to investors. Endava is a publicly traded engineering firm with thousands of people, and he told the market something most agency owners only mutter in private. As the firm delivered much higher productivity, he said, that productivity "is having an erosion on the revenue that's coming through the business." The better his people got at the work, the less the work was worth. What he was describing wasn't a recession but a repricing — and he was saying it out loud, on the record, the way the people running the largest firms in the industry have started to.

The work isn't ending — it's being repriced

The headline version of the story is that AI is coming for the agencies: the machines write the code and design the screens, and the firms in the middle simply disappear. That version mistakes a change in price for the end of the work. The work isn't ending so much as being repriced. For thirty years a software, design, or product agency sold one thing above all others — the making of the thing. The code written, the interface designed, the deck built, the feature shipped. That is the half of the business AI compresses. The visible artifact got cheap enough that buyers no longer trust the old production price. We are, in the plain sense of the phrase, *after code*: past the era when writing it was where the value lived.

The value didn't vanish, though. It moved to the parts of the work the model can't do alone. That means deciding what to build and integrating it with the tangle that already exists. It means catching what the machine got confidently wrong, and standing behind the result when a client is betting their quarter on it. Judgment, integration,

verification, accountability. The single dollar your buyer used to hand you for production is splitting in two — a shrinking half for the making, and a growing half for the judgment around the making.

The split: what you make, and what you answer for

The split is plain enough to see on any project you deliver. On one side is the work you merely perform — the code, the screens, the decks, the production the model now does for almost nothing. On the other is the work you **carry**: the judgment about what to build, the integration with what already exists, the verification of what the machine got wrong, and the accountability for the result. If a shorthand helps, the line between the two sides is the carry line, and the carry is the half clients still pay for. It is not a framework to adopt — just the part of the work the buyer can't get from the machine.

On the production side, the model takes the margin. On the carry side, the buyer still pays — but only for the carry it can trust, and only if you can prove it. That last clause is where most of this book lives.

What changed, and why now, sits in your buyer's shop rather than yours. Eighteen months ago your client's own team couldn't credibly claim to do your work without you. Now it has the same tools you have, and a chief financial officer watching the line. So the buyer no longer pays a premium for the making, which is the thing they believe they can now do themselves.

And here is where the conventional framing gets it backwards. The story is usually that buyers haven't learned to name the value of judgment yet. But they have. Procurement understands this work well. It already buys risk-transfer, service-level agreements, indemnity, assurance, managed services — it has decades of machinery for paying someone to carry a consequence. The gap is not a missing word for the

carry. The gap is that agencies have never *packaged* the carry into the forms procurement already knows how to buy, so it gets folded into an hourly rate and given away. Your price is being set in that gap.

Which sets up the hardest part of the whole book, worth saying before the buyer does. The savings from AI are the *buyer's* savings, not yours. Nothing about cheaper production hands that money to the firm that used to charge for the labor. The buyer can take the whole gain, keep paying you the modest fee for the judgment it was already getting for free, and ask you to carry *more* — more accountability, more risk, more proof — for *less*. Judgment becoming valuable is not the same as judgment getting paid for. Whether the carry ends up as your margin or the buyer's free expectation is the whole fight, and it is the fight the rest of this book is about.

This is a repricing, not a downturn

This is a repricing rather than a downturn to wait out, and the hard part is that being excellent at the work no longer settles the question. Everyone already agrees judgment is valuable; agreeing it is valuable is not the same as getting paid for it. The firms that don't make it through this transition won't, for the most part, be the ones with the worst craft. They'll be the ones that kept selling the production half at a premium while the carry — the half that actually holds its value — went unnamed, unpriced, and unproven on their own invoices. They delivered the carry. They just never made the buyer pay for it as its own thing.

When the cost of making something collapses, the money moves to whoever carries the consequence of the result. That much is an old mechanism, and not in dispute. What is new is only the speed, the fact that this time it runs through software, design, and product — and that

the buyer is now sophisticated enough, and equipped enough, to keep the savings unless you make a deliberate case otherwise.

Who this is for, and the five moves

This book is written for the person who runs one of those firms. I mean the fifty-to-three-hundred-person software, design, or product agency whose clients are big banks, Fortune 500s, and major media. It's the firm whose name lands on the same procurement table as the global consultancies. It's neither a forecast nor a eulogy, but a repricing manual for an unusually confusing moment, and it makes one argument in five moves.

- **Sell the slice** — sell the carry as a bounded result the buyer can see, name, and pay for, packaged in the language of risk procurement already approves, instead of folding it free into the production rate that no longer holds. - **Sign what you can carry** — take only the contract terms whose risk your firm can actually support, and refuse the ones that hand you carry you can't. - **Show your work** — make your carry auditable, because carry the buyer cannot verify is carry the buyer will not pay for, and proof is the one thing a competitor can't copy off your website. - **Survive the gap** — cross the cash-flow trough as you move from labor margin to carry margin, the trough that has quietly closed more good agencies than any model ever will. - **Keep the room** — hold the people who carry the judgment the whole new model runs on, and lead them through a crisis they can read as clearly as you can.

Each is a chapter, and each is harder than it sounds, because the forces pulling against them are sitting in your buyer's procurement office right now, doing math on your invoice. Together they answer the one question asserting that judgment is valuable never does: how a firm turns value the buyer admires in the abstract into margin the buyer is

contractually buying. So the book begins where the pressure is most visible. Not with a prediction about the future of artificial intelligence, but with a single real company, on a single real day, doing the exact thing this whole book is about. Going through the public record, the failures rarely turned on the future. They turned on a price that had already changed on the invoice — and on value the firm had delivered for years but never made the buyer pay for while it still held the pen. The open question is whether you can price the right half, in terms your buyer's own systems already approve, before the benchmark arrives to do it for you.

CHAPTER ONE

The Repricing Engine

The dollar a Fortune 500 pays your firm in 2026 isn't shrinking — it's splitting into two halves that ride different curves. The work of the year is knowing which half of each invoice sits on which curve, account by account.

IN THIS CHAPTER

- Why composition on an invoice changes before the total does
- The three forces splitting the dollar — from above, from the side, from below
- The gap between them that actually kills firms, and the four shapes that take it differently

In October 2025, Mike Cannon-Brookes went on the 20VC podcast and made three claims in one breath. Five years from now, he said, Atlassian would employ more engineers than it does today. They would be more efficient. And technology creation is not output-bound.

On March 11, 2026, the same company eliminated 1,600 jobs — 10% of its workforce — and split its CTO role in two.

The financial press read the layoff as the retraction. Cannon-Brookes had promised more engineers; here he was cutting them; the AI era had caught another executive in the gap between vision and balance sheet. I don't think that reading holds. The layoff wasn't the contradiction of the October promise. It was the first visible move toward keeping it.

The three claims only look like a paradox

Go back to what he actually said. More engineers in five years. More efficient. Technology creation is not output-bound. They look like a paradox only if you assume headcount is the thing that matters, and it isn't. The first claim is about a total. The second is about a coefficient. The third names the real constraint.

And the real constraint on building technology was never how fast anyone types. It's judgment about what to build, integration with what already exists, verification of what the system actually does in production, accountability for the binary that ships at the end. The 1,600 were the roles whose work AI now reproduces. The rebuild underneath them is the roles whose work AI now exposes the need for. Add it up across five years and the headcount rises while the inside of that number looks nothing like it did.

So: AI compresses the keystroke, not the judgment. The rest of this book is that line applied to an invoice, and it explains something happening right now on an invoice that has nothing to do with Atlassian. Composition changes first. The total moves last, if it moves at all.

One question is waiting on the agency invoice

Bring that model to the agency invoice and a single question is sitting there. It is being asked, in writing, across nearly every multi-year MSA renewal now in motion. An MSA is a master services agreement — the multi-year contract that governs the relationship. The parties are a 50-to-300-person enterprise software, digital, design, or product agency on one side and a Fortune 500 buyer on the other. *Everyone is using AI now. Programming hours have to be down at least 50% by now. Why is your rate unchanged and your hours unchanged?*

It doesn't arrive as an accusation. It arrives as a form. A Vendor Management Office questionnaire with a field that wants a number.

A Third-Party Risk disclosure request. A CFO benchmark deck forwarded down to procurement with a deadline attached. In private-equity-owned accounts it arrives as a portfolio-wide cost study from the sponsor's operating partner, against which the agency's rate is now a line. It is the same invoice the firm sent eighteen months ago — same scope, same buyer, same procurement table — coming back with a different burden of proof.

And notice what the buyer is not asking. The buyer is not asking whether the agency's people are any good. The buyer is asking what the agency now owns that the buyer cannot produce, audit, or safely absorb on its own.

The honest answer: the rate held and the hours held because the composition of what those hours buy has changed faster than the invoice can show it. Some of those hours produce work AI really has made cheaper. Other hours produce work AI has made harder. The buyer's own AI is now generating more code, more designs, more first drafts — more confidently wrong output that someone has to integrate, verify, secure, and put a name behind.

The dollar isn't shrinking — it's splitting

The plain version, then. The dollar paid to one of these agencies in 2026 isn't shrinking so much as splitting in two.

That is **The Bifurcation**, the principle the rest of this book hangs from — and it's a lens, not a law. Not every dollar splits cleanly. Some of what an agency sells is access, urgency, trust, taste, political cover — the plain value of being a throat to choke that the buyer's own org chart can't supply. And some production is so fused with judgment you can't pull the two apart. But the dominant line on a 2026 invoice does divide, and it divides hardest exactly where the buyer can benchmark it.

On one side sits the production the buyer now believes AI made cheap: UI scaffolding, boilerplate code, first-draft copy, layout, the component variants that used to eat three weeks and now seem to take a Friday afternoon. On the other side sits the judgment, integration, safety, verification, and accountability the buyer still can't supply alone. The two halves aren't moving at the same speed. The production half is bleeding pricing power wherever the buyer can substitute. The judgment half is gaining surface wherever AI deployment exposes a problem it doesn't solve. For now the invoice total holds, because the two motions roughly cancel at the line. The composition has changed before the total has. That lag — composition shifting while the total sits still — is **The Repricing Engine**. Less a metaphor than a sequence, and the sequence has a predictable victim.

First the distinction the whole argument turns on, because two lazy stories are competing for the CEO's attention and both are wrong. The story that AI is killing the agency is overstated. The story that nothing has really changed is a fantasy. The true version is more demanding than either. The dollar is splitting into two parts that ride different curves, and the work of 2026 is to know which half of each invoice sits on which curve — account by account, line by line. Get that wrong in either direction and the buyer reprices the firm before the firm has finished arguing about whether repricing is real.

And there is a second trap underneath the first, one this chapter only names and the next one has to solve. Knowing which half holds its value is not the same as keeping the margin. When production gets cheaper, that saving is the *buyer's* first — it belongs to whoever has the leverage to keep it, and left alone that is the buyer, not the firm that used to bill the labor. The judgment half can gain all the value in the world and still not gain a dollar of price, because value the buyer admires is not the same as value the buyer is contractually paying for.

Composition is the diagnosis. Whether the firm captures the half that holds, or merely delivers it for free, is the fight the rest of the book is about.

The defensible half isn't permanently defensible

One caveat rides with the Bifurcation for the rest of the book: the defensible half is not permanently defensible. The same models compressing production are climbing into judgment. Agentic systems now draft architecture, wire integrations, and run their own verification, and what looks safe in 2026 will look less safe in 2028. So the moat was never the task. It's the privileged context underneath the task: the buyer's systems, the buyer's failure history, the years of scar tissue no model can read off a repository. Tasks compress. Context compounds. The firm that mistakes the task for the moat is defending a wall the next model release walks straight through.

Three forces are doing the splitting. They come from three directions, and not one of them is hypothetical — each sits in a filing, an earnings call, or a published study from the last twelve months. One arrives from above. One from the side. One from below.

The pressure from above: a channel over your head

The pressure from above is the one the financial pages were watching the week it happened. In a single week in May 2026, the supply of AI engineering quietly acquired a distribution channel that runs over the agency's head.

On May 4, Anthropic, Blackstone, Hellman & Friedman, and Goldman Sachs announced a venture the Wall Street Journal put at roughly \$1.5 billion, built to ship Anthropic engineers directly into the operating companies inside Blackstone's and Hellman & Friedman's

portfolios. Goldman in the cap table. Anthropic supplying the engineers. The PE side supplying the demand. Seven days later, on May 11, OpenAI launched what it calls the Deployment Company — more than \$4 billion of initial investment from nineteen partners. TPG led; Advent, Bain Capital, and Brookfield co-led; and Bain & Company sat in the cap table holding priority access into PE-owned portfolios. Reuters' pre-launch coverage valued it at \$10 billion pre-money; Axios put the post-money valuation at \$14 billion.

The exact multiple is a footnote. The mechanism is the headline. Both vehicles ship straight into private-equity portfolios, and both route around the agency to reach buyers the sponsor already owns.

The threat here is not that AI writes code — that was already true in 2024, and the agencies survived it. The threat is the channel. AI deployment may now reach the prospect through the board seat, the operating partner, the portfolio-CEO call, or the capital stack, and arrive before the agency is ever invited to pitch. The buyer's chairman takes a call from Bain. The buyer's CEO gets an introduction to a deployment team. The buyer's CFO opens a portfolio-wide rate card that was approved a level above the procurement office. By the time the agency walks into the room, the deal that put the AI engineers inside the building was signed somewhere the agency cannot see.

This is the lever private equity has been hunting since 2019. Six years of programmatic value-creation programs — cybersecurity, procurement, talent, ERP — most of them quietly underdelivering against the deal memo, because programmatic value creation across a fragmented portfolio is genuinely hard. The labs offer something cleaner: one tool stack, one engineering bench, one contract structure hung off the GP's relationship instead of the operating company's procurement. For TPG, for Blackstone, for Hellman & Friedman, this is the lever the

industry has been trying to build in-house for a decade. They are not inclined to lose it to a 180-person shop.

The captive may not scale — and it doesn't matter

Now the objection, because the argument is worth more if it survives its own best one. PE captives have a long history of overpromising scale. Data fragmentation, legacy systems, budget ownership in the wrong place, compliance, culture, the last mile of change management — the same drag that slowed every operating-partner program of the 2010s applies here in full. The OpenAI Deployment Company and the Anthropic venture may both discover that embedding engineering into sixty portfolio companies with sixty different stacks is mostly services work, not platform work. And services scale worse than the press release suggests. Call that Scenario 2. It is real, and Chapter 7 returns to it.

It does not blunt the pressure, though, and the reason is subtler than the captive eating your lunch. Most captives won't: the CFO has watched GP-mandated shared services stall at the last mile of legacy integration for fifteen years, and has little reason to think this wave clears it. But the repricing does not require the captive to be a credible substitute. It requires only that it become a visible reference point. The moment a Bain operating partner names the Deployment Company in a portfolio-CEO meeting, the agency's proposal stops being measured against last year's invoice. It starts being measured against a number approved a level above procurement — and an anchor you can't see is still an anchor. The captive doesn't have to win the work, or even be capable of it. It only has to sit on the same slide long enough to reset what counts as a reasonable price.

That is the pressure from above. Capital is the moat, portfolio access is the demand contract, the filings name the players, and the mechanism is the channel rather than the tool. The agency that misreads this thinks

it is losing on craft, when it is really being routed around at an altitude it can't see from inside procurement. That's one direction; two to go.

The pressure from the side: the bundle beats the craft

The second force comes in from the side, at fifty times the agency's size, and it has spent three years getting ready. The cleanest way to see it is two holding companies that started the cycle in roughly the same place and ended it in radically different ones.

Publicis grew 5.6% organic in FY2025 at an 18.2% operating margin. Those two numbers belong in the same breath. Mid-single-digit organic growth at almost 20% margin is what holding-company shareholders waited two decades to see and rarely got. Publicis got there by bundling: software, data, procurement trust, compliance posture, audit infrastructure, delivery teams, and the Sapient engineering bench, all sold under one wrapper to one procurement officer. The 2015 Sapient deal the whole industry remembers as a long, ugly integration is finally working the way the memo promised. Worth noting against the AI-pure-play story: Publicis's technology line is still only 14% of net revenue, and roughly flat year over year. The bundle is the story. The AI label is not.

In the very same year, WPP's net cash from operations fell from £1.408 billion to £724 million, adjusted free cash flow collapsed from £738 million to £202 million, and the company dropped out of the FTSE 100. Some of that is governance and account-portfolio damage rather than AI — a CEO transition, a Pfizer loss, CPG share erosion. It would be dishonest to pin it all on the machine. But the two firms began from comparable ground. One bundled. One didn't. The market noticed which was which.

A slide is not a SKU

The rest of the field corroborates the same move at smaller scale. Globant reported \$32.8 million in AI Pods annualized revenue across 40% of its top 20 accounts as of March 2026 — about 1.3% of a \$2-billion-plus base, which is an early packaging signal, not a finished shift. But the dollars are not the threat. The packaging is. Globant sells AI Pods to enterprise procurement as a structured SKU: defined inputs, defined outputs, defined service levels, defined audit posture. The mid-sized agency’s “we do AI work” lives on one slide of a deck. Procurement knows the difference between a slide and a SKU.

Accenture makes the same point at a scale that ends the argument: FY2025 revenue of \$69.7 billion, bookings of \$80.6 billion, \$5.9 billion of that in generative-AI bookings and \$2.7 billion of generative-AI revenue already recognized. All of it sold inside contracts already calibrated to the buyer’s enterprise risk posture. So the buyer adds the AI work without standing up a new vendor governance regime. The regime is already running.

The single most useful number in the sideways story is none of those. It is CI&T’s. In Q1 2026, CI&T booked \$136.6 million in revenue, up 23.2% year over year, with 20% of new sales on new pricing models. It is a firm sized like the upper end of this book’s readership — proving in a public disclosure that buyers will sign new commercial models when the model is offered clearly. The cleanest evidence the market will move is not Accenture’s \$5.9 billion. It is CI&T’s 20%.

Here is what unites them, and it’s the uncomfortable part if craft is what you sell. None of these firms is winning on craft. Globant is not better at React than the mid-market product specialist. Accenture is not better at brand systems than the brand-led shop. Endava is not better at a given bank’s legacy integration than the firm embedded in it for years. The competition has moved up a level — to the buyer’s Vendor

Management Office, the buyer's Third-Party Risk function, and, in regulated industries, the buyer's Model Risk Management function. All of them run parallel evaluations. The mid-sized agency has to satisfy four functions with four red lines, and often finds the fourth one only after the deal is already in trouble. The bundlers have been calibrated for all four since the 2000s.

That is the pressure from the side. The bundler does not win on craft. It wins because it looks safer — and in 2026 procurement is choosing safer more often than it used to.

The pressure from below: the buyer's own AI

The third force is the quietest, which is why it gets underrated. The captives announce themselves. The bundlers announce themselves. The buyer's internal team running Claude Code on its own backlog does not. It just shows up at the next quarterly business review with a list of work it now believes it can do without the agency.

The tools are all public — Cursor, Claude Code, GitHub Copilot, Gemini, Lovable, v0, internal copilots, the AI baked into the CRM and the analytics stack and the design tool. Add a generation of generative-design tools producing passable first-draft layout and copy with no agency seat in the room. The buyer's PM runs Cursor at his desk. The buyer's engineers run Claude Code on the backlog. The buyer's CMO has v0 spinning out landing-page variants between meetings. And the buyer's CFO compares the agency's bill rate against all three.

The CFO is one enforcer, not the only one. The Vendor Management Office now benchmarks vendors against an AI-augmented internal-cost number that didn't exist eighteen months ago. Third-Party Risk asks new questions about model use, data handling, and the indemnity boundary on AI-generated content. And in regulated industries the Model Risk Management function asks a harder question still — one

a great many agency CEOs have never once had to satisfy. Are the AI tools the agency uses in delivery themselves subject to the institution's model governance? Four functions, four red lines, all four live on every renewal.

What the productivity numbers actually say

Then the productivity numbers arrive, and they don't say what the people leaning on them assume. They are contested, and they are moving. DORA's 2024 research tied rising AI adoption to lower throughput and lower delivery stability. Its 2025 research reversed the first half — AI now tracks with higher throughput — while still finding delivery instability climbing as adoption rises. Taken together, the honest verdict is neither "AI makes teams faster" nor "AI makes teams slower." It is that AI buys speed and charges for it in stability: teams ship more, and a larger share of what they ship breaks. The METR study published in July 2025 caught the human half of it. Experienced developers using AI tools ran 19% slower on real tasks while believing they were faster — a result METR itself later flagged for selection limits, which only sharpens the point. No one has clean numbers yet.

Two things fall out of that, and they point in opposite directions.

The first is that the buyer's confidence in its internal productivity gain is running well ahead of the evidence. The real gain is smaller, and far less settled, than the CFO believes. The repricing pressure applied on a *we can do this faster ourselves* basis is calibrated against a number the research itself keeps revising. Often it's whatever the CFO saw in a consulting deck, not anything his own team measured. Eventually the gap between the claim and the measurement gets discovered. Not immediately. Procurement pressure arrives before measurement catches up to it, and that lag is where the bad renewal gets signed.

The second is the part the doom narrative misses entirely. The buyer's internal AI is generating more code, more designs, more drafts — and

a meaningful share of that output is technical debt, broken integration, hallucinated logic, drift, rework. The instability DORA now measures rising alongside AI adoption reads less as a footnote than as a demand signal. Someone has to debug, refactor, integrate, secure, and verify what the internal team shipped, and in a meaningful number of cases that someone is the incumbent agency. So the buyer's own AI does two things at once. It compresses the perceived value of agency work that looks like raw production-without-proof. And it opens a new demand surface for architectural rescue, integration governance, and verification of whatever the internal team just shoved into production. The dollar splits because those two effects net out to a smaller production half and a larger judgment-integration-verification half.

The obvious rejoinder is that the buyer, now armed with the same AI, simply hires that judgment in-house and stops paying for it. Sometimes it does — and a firm that has not made its judgment worth buying has no answer to that. But the structural reasons it tends to stay outside are worth stating plainly, because they are also the reasons it can be priced. A buyer pulls the judgment in-house only when being wrong about it is cheap. The point of paying an outside firm is to move the consequence of being wrong onto someone else's balance sheet. That buys accountability the buyer's own team cannot personally own, on systems and across vendors no single internal group sees end to end, where the cost of shipping the wrong thing into production is high. None of that argues procurement is naive about the value. It is the opposite: the Fortune 500 procurement office has bought risk-transfer, indemnity, and managed services for decades and knows exactly what it is paying for. The error in the optimistic story is assuming the buyer hasn't noticed the judgment is valuable. It has. The unsolved problem is that agencies keep handing it over folded into an hourly rate the buyer

is now benchmarking down — which is the next chapter's subject, not this one's.

That is the pressure from below. All three are now named — top-down through the captive channel, sideways through the platform incumbents, bottom-up through the buyer's own AI. They are not three stories. They are three forces converging on one invoice from three sides. And the firm sitting on that invoice feels them arrive not as a thesis but as a tightening.

The gap between the forces is what kills firms

Which brings us to what actually kills firms, and it is none of the three pressures. It is the gap between them.

Endava's Q3 FY2026 is the failure mode written into an annual report. Hold four numbers together: £364.6 million of goodwill impaired, adjusted PBT margin falling from 12.6% to 1.8%, revenue down 8.4% year over year — and AI-driven revenue climbing from 5% to 15% in the very same period. The AI line tripled. The legacy base contracted faster than the AI line could fill it. Cash conversion deteriorated in the space between. The goodwill impairment is the accounting echo of that: management's forecast through FY2031 came down, the auditors recalculated, the write-down was the consequence. The auditors named outcome-based contracts as part of the miss. Outcome work takes longer to execute, cash lands later, working capital tightens — and the firm gets caught between an old commercial model collapsing and a new one not yet ramped.

▲ THE TRANSITION TRAP KILLS IN SEQUENCE.

This is the trap that deserves to be stated as a sequence, because that's how it works:

- The new line grows on schedule.
 - The old line shrinks faster than the new one can replace it.
 - Cash falls into the gap. Margin falls into the gap.
 - The auditors revalue the firm to reflect the gap.
- The press calls it an AI loss. It is a sequencing loss.

And the pattern is not one company's bad quarter. S4 Capital reported FY2025 revenue down 10.8%, its own disclosure naming client budget shifting toward AI infrastructure as a driver. Dentsu booked ¥310.1 billion of goodwill impairment across the Americas and EMEA. DXC guided FY27 organic decline at minus 3% to minus 5%. Stack them with Endava and the holding companies and you have a row of public firms reporting versions of the same dynamic inside four quarters. The ones that bundle are growing. The ones that don't are shrinking. And the ones caught halfway are getting their cash flow squeezed and their multiples reset.

The uncomfortable part is that Endava's leadership is not incompetent, and was not blindsided. Two quarters before the impairment, on an autumn 2025 earnings call, its chief executive, John Cotterell, had already named the mechanism in plain words. The firm's much higher AI-driven productivity was "having an erosion on the revenue that's coming through the business." He was not describing a risk. He was describing a result already arriving. The strategy is roughly right. The execution is roughly right. The math of the transition produced the impairment anyway, because the calendar on which the old contracts collapse does not line up with the calendar on which the new ones mature.

And the 50-to-300-person firm is more exposed to that sequencing problem, not less. It does not have £364.6 million of goodwill to impair before the cash conversation reaches the bank covenant. It has ninety days of runway and a top-five client concentration that can't absorb the loss of a single contract. This is why this chapter runs in the order it does — bifurcation, then pressures, then the trap, and only then the decomposition and the exits. The firm that picks an exit before naming the pressure on its own book lands exactly where Endava landed, with smaller numbers and no goodwill to cushion the fall.

Why a giant's weather is yours

A fair objection has been building, and it deserves a straight answer. Every company named so far is a giant — Accenture at seventy billion dollars, Publicis and WPP at holding-company scale, Endava with more than eleven thousand people on the payroll. You run a hundred. Why should their weather be yours?

Because you sell to their buyer. The forces in this chapter don't travel by vendor size; they travel through the procurement office. And the Fortune 500 procurement office runs the same lifecycle on a hundred-person vendor that it runs on Accenture. The same Vendor Management Office questionnaire. The same Third-Party Risk disclosure. The same CFO benchmark deck, now pointed at your invoice. The public companies are not the subject of this chapter; they are the instrument panel. They are the one place the pressure shows up in audited numbers, because they are required to report what you are not. They are not a representative sample and they are not your destiny. They are the weather report. Endava's impairment is your next bad quarter with three more zeros and a press release attached.

But the symmetry stops at the threat, and missing that is its own trap. A hundred-person firm cannot answer the way a holding company

answers. It can't bundle software and data and audit infrastructure under one wrapper the way Publicis does, or carry a balance sheet deep enough to underwrite enterprise risk transfer the way Accenture does. That asymmetry is the entire reason this book's answer reads *sell the slice* rather than *become Publicis*. The slice is the one scaled-down posture a small firm can actually hold, precisely because it doesn't require a holding company's balance sheet to deliver. The giants show you the pressure; they can't show you the move.

A few firms are genuinely insulated, for now. A shop built on a founder's personal relationships, or buried so deep in a single niche that no procurement office has a benchmark to swing at it, gets a softer version of all of this. The question arrives later and lands lighter. *For now* is the operative phrase. The benchmark is coming for the niche too. It is only a matter of which renewal.

Four shapes, each on its own clock

So before any talk of exits, the dollar has to come apart on the page. The four firm shapes that dominate this segment each take the Bifurcation on a different clock — and one of them is yours.

There is the **project shop**, running fixed-price SOWs at \$1 million to \$3 million deal sizes for PE-portfolio and growth-stage buyers. The production-heavy lines compress 20% to 30% per renewal cycle, while architecture, integration, security review, and launch accountability hold. The clock runs fast — six to twelve months — because every new SOW is a new signature, and the clock resets at each one.

There is the **retainer firm**, billing monthly for brand, growth, and product work. Production hours compress as the buyer's internal team picks up first-draft layout and copy, while strategic-operator hours and brand-system stewardship hold. The clock is slower here — twelve-to-

eighteen-month — with the question landing at quarterly or annual review and the pressure compounding quietly in between.

There is the **staff-augmentation MSA shop**, billing engineer-months on multi-year masters to Fortune 500s and large banks. The bill rate compresses 4% to 7% on renewal, while growth migrates toward architectural rescue of internal AI-generated code the buyer doesn't yet know how to buy. This is the slowest clock of all — eighteen to thirty-six months — patient and unavoidable.

And there is the **hybrid premium enterprise firm**, carrying a mix of project, retainer, and MSA work across Fortune 500, federal, big tech, and financial-services buyers. All three clocks run at once, and the simultaneity is the thesis: top-down, sideways, and bottom-up pressures landing on different lines, on different timelines, inside a single fiscal year.

Those four shapes get established once here and are only referenced after, since what matters isn't the taxonomy but the line items underneath it. And the decomposition isn't theoretical — it happens on the invoice.

The teardown: a \$2 million SOW, line by line

► A \$2 MILLION FIXED-PRICE SOW, DECOMPOSED.

Take the project shop's signature event: a \$2 million fixed-price SOW to rebuild a mobile app for a PE-owned mid-market SaaS company. It breaks into seven lines, and the buyer believes AI made some of them cheaper. The buyer is right about some and wrong about others.

- **\$400,000 UI and front-end build** draws a *we can do this in Cursor for \$150,000* — partly true for drafts, false for polish, accessibility, and integration. Hold the price; sharpen the acceptance criteria; ship it as a fixed deliverable with an evidence pack.
- **\$300,000 commodity back-end build** draws *AI writes this* — mostly true for boilerplate, false for non-trivial domain logic. Compress it 20% to 30%; rebadge it as commodity; defend it on speed and stability.
- **\$500,000 architecture and integration, \$200,000 security review, \$250,000 QA and verification, \$200,000 product judgment and PM, and \$150,000 launch accountability and warranty** are the lines the buyer can't reliably substitute. Those hold or rise.

Of the \$2 million, \$700,000 sits in lines the buyer believes AI made cheaper: \$1.3 million sits where substitution fails. Do nothing and the next bake-off compresses the total toward \$1.5 million. Repackage — name the architecture and judgment lines as the value spine, ship the evidence — and the firm holds close to \$1.8 million with stronger margin on the lines that matter. The dollar isn't smaller. The composition has changed.

That teardown is the whole chapter in one invoice. The numbers are a model, not a market quote — the floor and the recovery land differently on every real deal; the shape is the point, not the decimals. Of a \$2 million deal, \$700,000 was always going to get repriced. The only open question was whether the firm did the repricing or the buyer did. The firm that names the \$1.3 million as the spine, out loud, on its own paper, walks away with \$1.8 million. The firm that says nothing watches the same deal drift to \$1.5 million and calls it an AI loss.

The same move, the other three shapes

The same decomposition runs through the other shapes, and the moves rhyme without repeating.

For the **retainer firm**, an \$80,000-a-month engagement eleven quarters deep does not get defended by opening the books to line-item attribution. That hands procurement the knife, and attribution is a fight a firm can win and still lose. It gets defended in aggregate: continuity of judgment across eleven quarters, institutional memory the buyer's own team has lost to turnover, the standing ability to decide what *not* to do with the AI output. The compressible production hours drift if left alone. The brand-system stewardship and strategic-operator hours and the quarterly proof hold as an eligibility floor — possibly at the cost of a 5% to 8% adjustment. That is the price of keeping the master closed. The **MSA shop** is where the mechanics get unforgiving, and where an expensive mistake hides inside a plausible-sounding fix. The case is a \$25,000 engineer-month at a top-five US bank, \$24 million a year of time-and-materials billing, renewal in eleven months. The bill rate is not the unit being repriced; the workstream is. The bank benchmarks against Accenture, Cognizant, TCS, Wipro, and now its own AI-augmented internal team. Net-new feature work is partly compressible. Legacy maintenance holds, because knowledge of the monolith is privileged context. The regulatory and compliance workstream holds or rises, because Model Risk Management is its advocate.

▲ **DON'T CONVERT THE MSA TO MILESTONE BILLING AS A RENEWAL STUNT.**

At a top-five bank, a mid-stream commercial change reopens the third-party-risk review the relationship already passed. Claw-back is not how a regulated buyer cures a vendor problem. And procurement approves contractor headcount against a budget cycle that milestone billing scrambles. Milestone billing breaks the bank's accounting before it breaks the firm's.

The move that works instead is a **Slice Order rider** hung off the existing MSA — a named workstream with a **Performance Rebate** (variable consideration netted against price up front), not an at-risk fee — that signals repricing readiness without forcing the master open.

The bank absorbs a 4% to 7% compression on renewal; the firm's growth comes from the new category: architectural rescue of internal AI-generated code the bank doesn't yet know how to procure and the firm is positioned to teach it.

And the **hybrid firm** — the most common shape in the segment, and the one the doom narrative misreads worst — simply runs all three of those clocks at once. The project line reprices first and fastest, because every engagement is a new signature. The retainer line reprices on its slower cadence, each quiet quarter adding to the buyer's belief that the retainer overpays. The MSA line reprices slowest but carries the largest single-account exposure, where a 6% compression on \$5 million is \$300,000 of margin gone. The instinct to fix the whole firm at one offsite is theater. The math says sequence the clocks — project line first, retainer second, MSA third — and let them integrate into one operating system over twenty-four months.

One pattern runs under all four: the defensible half has to carry the price, and the compressible half has to stop pretending to be premium work. But carrying the price is not automatic. The same firms that deliver plenty of defensible work usually find, on the first honest pass, that they never charged for it as a separate thing. It was folded into a

blended rate and defended as part of the relationship. So two numbers come out of the pass, not one. Roughly 30% to 45% of invoiced work sits in the compressible half and is being sold on judgment-line pricing — that gap is the compression the procurement office is already calibrating. And the defensible half, the part that should hold, is mostly being delivered for free. The first is the value the firm is overcharging for. The second is the value the firm is giving away. Both have to be fixed, and the second is the harder one.

Three ways out, and a fourth way to stay

Once the dollar is decomposed, the squeezed middle has three ways out — and a fourth way to stay where it is, with discipline. None is a menu item; each is a math problem with a specific answer for a specific book of business, and which one fits depends on which pressure dominates and which substrate the firm can actually build.

Exit Up — eligibility, not premium

The firm can go up. **Exit Up** means building verification, governance, security review, and AI-use disclosure. It means SBOM and AIBOM readiness — a software bill of materials and an AI bill of materials, the ingredient lists a buyer’s audit now wants. And it means contract discipline and delivery evidence strong enough to stand on the same procurement form as Globant and Accenture. But the honest framing matters more than the ambition: Exit Up buys eligibility to bid, not pricing power. The CISA and G7 SBOM-for-AI minimum elements, published May 12, 2026, will not stay a differentiator — they become a checklist every credible vendor satisfies.

The firm that builds audit-grade infrastructure in 2026 is not earning a 15% premium for it in 2027 — only the right to bid against Accenture, which is a different and lesser thing. What defends margin is the privileged context the audit logs prove is real. The firm holding eleven

quarters of a specific bank's legacy code prices 8% to 12% above the scaled vendor on context. The audit infrastructure is merely what gets it past procurement to have the context conversation. Build the audit grade without the context and you become a slower, smaller Globant; spending \$400,000 on audit infrastructure for buyers who never asked is margin destruction with better documentation. Globant is the public case — five years building the substrate, \$32.8 million of AI Pods ARR that is small in itself but underwrites the next \$300 million landing inside existing MSAs without renegotiating risk.

Exit Down — the only consistent premium

The firm can go down. **Exit Down** means owning one vertical, one workflow, one regulated surface, one platform ecosystem more deeply than any broader firm can. Vertical depth is the only substrate that consistently earns a 15% to 25% premium in 2026 procurement testing. It is the strongest pricing-power exit, and the hardest to execute. The firm that earns the premium has committed to one vertical in a way that excludes others.

The shop pitching healthcare and fintech and edtech is a generalist with a vertical-themed deck. The shop that turns down the edtech work to go deeper on healthcare is the one earning the premium. That means losing revenue on purpose in the short term to build category depth in the medium term. The firm looks smaller before it looks stronger, and most firms will not make that trade. CI&T did. Its 23.2% growth at \$136.6 million sits on a concentrated financial-services and consumer book. The 20% of new sales on new pricing models is the proof: buyers signed new commercial terms because the vertical depth was credible enough to absorb the risk transfer.

Exit Sideways — the accountable layer

The firm can go sideways. **Exit Sideways** means becoming the trusted, accountable layer. It is the firm that knows the buyer's systems,

translates AI output into operating progress, and takes bounded responsibility for what the buyer can't safely verify. The form is quarterly evidence packs and named acceptance criteria. Fee-at-risk on 10% to 20% of delivery KPIs, but only where the firm can credibly underwrite the KPI — because attribution-fuzzy domains are where outcome retainers go to die.

Publicis is the case at scale. It does not pitch itself as best at any single craft. It pitches itself as the accountable layer between the buyer's CMO and the buyer's procurement office — the bundle as differentiator and the cash conversion as proof. For the mid-market, Exit Sideways is the retainer posture made commercial. The retainer doesn't transform; it just produces the proof artifact every quarter, so the CMO who got her CFO's question has a document to hand back.

Stay Center — the architectural rescuer

Or the firm can refuse all three and **Stay Center** with discipline. This is the architectural-rescuer — billing time-and-materials for elite senior engineering that cleans up the AI-generated mess the buyers' internal teams now produce at speed. The demand surface is real. The instability that climbs with AI adoption — the stability cost in DORA's latest research — is rescue work. Industry estimates put custom-software development on a 22% CAGR through 2034. IT outsourcing is projected past \$750 billion by 2031.

But **The Fourth Path** demands three hard things. A senior bench deep enough that there are no juniors to displace — an inverted pyramid that is structural rather than aspirational. Internal AI adoption that is brutally real rather than performative — because the delivery economics depend on the firm using the tools at least as well as the client's own team. And a fixer brand rather than a transformation-partner brand. That is a different sales motion, with different references and a different deck. The failure mode is staying center because the

firm can't stomach changing — the squeezed middle wearing a strategy costume. Staying center because the rescue market is underserved and the firm can credibly serve it is a real strategy. The only difference is whether the internal AI adoption is real, or whether “AI-augmented” went on the website and got called a plan.

The case for Jevons — and why it doesn't save you

There is a serious case against all of this, and it earns the right to be heard on its own terms.

The case is Jevons — the paradox that when a resource gets cheaper, total spend on it can rise rather than fall, because demand grows faster than the price drops. Applied to software: as the cost of producing code falls, demand rises faster, and total spend on engineering services grows. The doom narrative is then wrong, because the pie is expanding faster than the unit cost is compressing. The evidence is real and should be honored. Accenture booked \$5.9 billion of generative-AI work against \$2.7 billion recognized — bookings outrunning recognition, meaning the pipeline is growing faster than delivery can keep up. NewsCorp signed a \$50-million-a-year licensing deal with OpenAI: one publisher, one deal, a revenue line that didn't exist three years ago. CI&T's 23.2% is not contraction. A custom-software market on a 22% CAGR is not a shrinking pie. If Jevons holds in aggregate, the CEO can argue the Bifurcation is just a composition shift inside a growing total, and growth absorbs the composition risk. That is the steelman, taken whole.

It still doesn't blunt the chapter, for one reason. Composition change happens whether the total grows or shrinks. The firm that doesn't name which half of its invoice sits on which curve gets repriced inside a growing pie, not protected by it. Accenture's \$5.9 billion lands on Accenture's P&L, not on the mid-market generalist's.

And the strongest form of the objection — that demand is outstripping delivery capacity, so even the overflow has to land somewhere — is true and still doesn't save the undifferentiated firm. Overflow flows to whoever the buyer can hand the judgment half to without a fight, not to whoever has idle production hours. Spare capacity in the compressible half is not a moat; it is inventory the buyer can now make for itself. The generalist gets the procurement question and the bill-rate compression regardless of what the aggregate market is doing. Growth is the cover under which the composition shift becomes either an opportunity or an extinction event, account by account. Which one lands on a given firm comes down to timing: does the firm name the Bifurcation early enough to repackage on its own terms, or late enough to be repackaged on the buyer's?

From the industry view to the account list

The industry answer stops mattering here. The account list starts. Inside thirty days, five things belong on the desk — and the work below is directive on purpose, because this is the part you actually do.

→ **INSIDE 30 DAYS**

1. **A revenue heat-map.** Score every active engagement against the three pressures and the Fourth Path — red where multiple pressures are loud and the posture is failing, yellow where one is loud and the defense is partial, green where the pressures are quiet, the substrate is intact, and the relationship can compound. The proportions matter less than the honesty of the labels; the first pass almost always surfaces two accounts mislabeled green that are really yellow turning red. Those come first, because the next signature on them is the wrong one unless the posture changes before it.
2. **An exit map.** Assign each account to Up, Down, Sideways, or Stay Center — a hybrid firm legitimately running three at once if they map to different revenue lines. Then the hard test: does the firm's actual substrate — verification infrastructure, vertical depth, privileged context, senior bench — support the exit assigned? Where exit and substrate don't match, the account is lost, sold to a peer, or rebuilt inside twelve months.
3. **An artifact gap list.** The Slice Order primitive, the AI Provenance Schedule, the Engagement Boundary Memo, the Acceptance and Cure Protocol, the Mutual Audit Cooperation Clause, the Liability-and-Insurance Worksheet from Chapter 3 — mark what exists, what exists in rough draft, what has never been written. The artifacts that don't exist are precisely the conversations the firm has been having on the buyer's paper, which is why those conversations have not gone the firm's way.
4. **A 90-day distressed triage — but only if** cash sits within six months of a line-of-credit covenant or client concentration tops 30%. Most readers do not need this immediately; the ones who do already know, and the math is in Chapter 5.
5. **A trust-artifact draft for the next Fortune 500 RFP** — AI-BOM language, model-version pinning, human-review attestation, audit-cooperation framework, indemnity boundary — drafted even if no RFP is on the desk today, because the draft is the forcing function for the substrate. Write it honestly and the substrate is there. Fail to write it and it isn't, and that failure is the first hiring brief.

The COO owns the heat-map and the exit map. The CFO owns the artifact gap list and the trust-artifact draft. The CEO owns the next forty-five days with the top-five accounts. The heat-map will show that two of those five are repricing inside ninety days — and it is better to start that conversation than to receive it.

What this all comes to

That is The Repricing Engine. The dollar splits because three pressures push on it from three directions. The first move is to know which half of each invoice sits on which curve. But knowing is only the diagnosis. The agency that doesn't know is being repriced by the buyer who does. The agency that does know still has to do the harder thing: capture the half that holds before the buyer captures the saving. And it has to do that on its own terms — account by account, line by line, signature by signature.

Atlassian fires the engineers AI now reproduces. Atlassian hires the engineers AI now exposes the need for. The total goes up over five years; the composition already changed. The same motion is running through every invoice in this segment right now. The firms that name it are repackaging the production half so the compression is anticipated, and the judgment half so the price holds. The firms that don't are watching their totals stay flat while margin erodes underneath, until the cash lands somewhere the covenant or the partner draw cannot absorb.

Which raises the question the next chapter takes up. If the dollar splits, and the defensible half has to hold the price, where exactly does the split fall — and which side of it is your firm actually being paid for? There is a plain line running through every project you ship, dividing the work you merely perform from the work you answer for. The next chapter shows how to find it on your own invoices. It also confronts

the harder problem this one only flagged: the value moving to the defensible half does not, on its own, move the margin there too. The buyer can pocket the savings and still pay you less. Seeing the split is the easy part. Getting paid for the half that holds is the work.

THE REPRICING ENGINE IN 20 SECONDS

- The dollar a Fortune 500 pays your firm in 2026 isn't shrinking — it is *splitting* into a compressible production half and a defensible judgment half. The total holds while the composition moves.
- Three forces do the splitting: a captive *channel* from above (PE + the labs), platform *bundlers* from the side, and the buyer's *own AI* from below.
- The killer isn't any one pressure — it's the *gap* between an old model collapsing and a new one ramping. That sequencing loss is the Transition Trap.
- The giants are the instrument panel, not the destiny: same procurement office, same forces, three more zeros.
- Four firm shapes take it on four clocks. Decompose the invoice, find the compressible third, and price the defensible half out loud — on your own paper.
- Seeing the split is the easy half. The saving belongs to the buyer first, so the defensible half holds its price only if the firm *captures* it — not if it keeps delivering it for free.

Sell the slice. Sign what you can carry. Show your work. Survive the gap. Keep the room.

CHAPTER TWO

What Still Gets Paid For

Production got cheap, so the value in a project moved to the judgment around it — what to build, how to integrate it, whether it's right, who answers when it breaks. But value moving is not the same as margin moving. The buyer can pocket the AI savings and still pay you less. This chapter is about seeing the split on your own invoices, and the harder problem of getting paid for the half that holds.

IN THIS CHAPTER

- Where the value goes when production gets cheap — and why the margin doesn't follow on its own
- The four things the buyer still can't get from the machine: judgment, integration, verification, accountability
- Why the buyer won't simply insource them, and what it takes to make them paid

Sir Martin Sorrell spent three decades assembling the largest advertising company the world had ever seen, and he built it on the simplest unit there is: time. Bodies, hours, day rates, marked up. So it means something that the man who built WPP now goes on camera and asks, out loud, the question every agency owner is working not to ask. “Do you charge on the basis of time or output, value added?” he said in a 2025 interview. He was walking through what happens to a business like his when the cost of making the work falls by most of its old number. “So we have to change our model.”

When the architect of the old model says the model has to change, it is worth asking what replaces it. And worth being honest: “value added” is the easy half of his sentence to say, and the hard half to bank.

When production gets cheap, the value moves up

Start with the part that is not in dispute. When the cost of producing something collapses, the value in the work does not vanish. It moves — toward whoever carries the consequence of the result rather than the labor of making it.

The visible artifact got cheap enough that buyers no longer trust the old production price. The deck, the build, the first draft — a competent team produces these now in a fraction of the hours it used to bill, and the buyer knows it. So the buyer stops paying the old number for the making. What it will still pay for is the part it cannot get from the machine. That is the judgment about what to make, the integration into the mess that already exists, the catch on what the machine got wrong, and the name that stands behind the result.

That much is widely agreed. The disagreement — the one that decides whether any of this is good news for a firm — is whether the *margin* moves with the value.

Value moving is not margin moving

Here is the trap underneath the optimism, and it is worth naming before anything else, because most of the advice on this subject quietly assumes it away.

Say production used to cost the buyer a million and the judgment around it was bundled in free. Now the production costs two hundred thousand. The eight hundred thousand of savings is real — but it is the *buyer's* savings, not the firm's. Nothing about cheaper production hands that money to the firm that used to charge for the labor. The buyer can take the whole gain, keep paying the firm the same modest fee for the

judgment it was already getting for nothing, and feel it got a bargain. Often that is exactly what happens.

So value migrating to judgment does not guarantee that margin migrates with it. The savings belong to whoever has the leverage to keep them. Left alone, that is the buyer.

This is the real work of the rest of the book, and it is not a pep talk about how valuable judgment is. Everyone already agrees judgment is valuable. The question is how a firm makes its judgment *paid* — turns a thing the buyer values in the abstract into a thing the buyer is contractually buying. That takes four levers, and they are the spine of every chapter after this one. **Scarcity:** the carry has to be hard to get elsewhere. **Proof:** the buyer has to be able to see it was real. **Packaging:** it has to be sold as a bounded thing, not folded into a rate. **Contractual discipline:** the terms have to make the firm the one holding the consequence, on purpose, for a price.

Miss those, and you can be completely right that the value moved up and still watch your fee fall every year. Value is the easy claim. Margin is the one you have to engineer.

The split, on a single invoice

The split itself is plain enough to see on any project a firm delivers. Run a line down the middle of one.

On one side is the work you merely perform. The code typed, the screen pushed from a brief, the deck built, the test pass — the production the model now does in an afternoon for a small fraction of its old cost.

On the other side is the work you answer for. The judgment about what should be built. The integration into the tangle that already exists. The verification of what the machine got confidently wrong. And the accountability for the result, when a client has bet their quarter on it.

If a shorthand helps, the line between the two sides is the carry line, and the work above it is the carry. It is not a framework to adopt; it is just the half of the work the buyer is still, genuinely, paying for — when you make it pay. The making is racing to the cost of the tools. The carry is the only part with a price the buyer can't benchmark away in a single procurement cycle — *if* you have made it scarce, proven, packaged, and put it in the contract.

The carry is four specific things, and each can be priced

“Move up the value chain” was always too vague to act on. The carry is not one thing but four, and each can be sold as its own line of value.

Judgment — deciding what to build

The first is **judgment**: deciding what to build and what not to. The model has no stake in being wrong, so it cannot do this for the buyer. It will generate four plausible directions in a minute. Choosing the one that is right for *this* client's market, and killing the three that are merely competent, is judgment. It is worth most when being wrong is expensive — which is precisely when the buyer will pay to have someone else own the choice.

Integration — making it work inside the mess that exists

The second is **integration**: making the thing work inside the decade of accumulated systems, exceptions, and politics the client already lives in. The model can write a clean function. It cannot know that the payments service it calls was patched in 2019 by an engineer who has since left, and that the patch will fail under the new load. Carrying that risk — and being the one accountable when the integration is the thing that breaks — is worth paying for.

Verification — catching what the machine got confidently wrong

The third is **verification**: catching what the machine got confidently wrong. The new production is fast, fluent, and occasionally, invisibly false. The model does not flag the line it hallucinated — it renders it in the same confident voice as the lines it got right. Someone has to be accountable for the difference between *looks done* and *is done*. The faster and more fluent the production gets, the more that someone is worth.

Accountability — standing behind the result

The fourth is **accountability**: standing behind the result. Being the name on the contract, the throat to choke, the firm that does not get to say “the AI did it” when the system fails in production at 2 a.m. This is the part the buyer most wants and can least manufacture internally. It is also the one that survives every model release, because no model will ever sign an indemnity — a promise to cover the loss if it goes wrong. Judgment, integration, verification, accountability. It is exactly what your buyer’s own AI cannot supply for itself. Which is the *reason* it stays valuable — but not yet the reason the buyer pays you well for it. Value is the opening. Price is a separate fight.

Why procurement won’t just take it for free — and why it sometimes does

There is an obvious objection here, and it is the right one: if the carry is so valuable, why doesn’t the buyer simply insource it now that AI is the army of doers underneath? The Fortune-500 CTO has the AI. Why keep paying an outside firm for the judgment on top?

Sometimes the CTO does insource it, and a firm that has not made its carry paid has no answer. But the structural reasons the carry tends to stay outside are worth stating plainly, because they are also the reasons it can be priced.

A buyer insources the judgment only when the judgment is cheap to be wrong about. The whole point of paying an outside firm is to put the consequence of being wrong onto someone else's balance sheet. It buys accountability the buyer's own team cannot personally own, on a system the buyer's own team did not build. Add the cross-vendor integration scope that no single internal group sees end to end. Add the plain fact that being wrong in production is expensive. The outside firm has a real seat — *if* it is willing to sit in it and price the seat.

And this is where most of the conventional framing gets the problem backwards. The story is usually that buyers “haven't learned to name” the value of judgment yet. But they have. Procurement understands this work well. It already buys risk-transfer, service-level agreements, indemnity, warranties, assurance, managed services. It has decades of machinery for paying someone to carry a consequence. So the gap is not a missing word for the carry. The gap is that agencies have not *packaged* the carry into the forms procurement already knows how to buy. The buyer is fluent in paying for risk-transfer. The seller keeps offering it as a free courtesy folded into an hourly rate.

The move, stated once: a firm makes the carry paid by selling it in the language of risk procurement already approves. That means a bounded result, a warranty, a named liability the firm accepts for a price — while the production underneath costs what production now costs. Asserting that the carry is valuable does none of that work.

Why the four hold even as the tools improve

The four have one thing in common, and it is the source of their durability. Each is a place where the model's great weakness — that it has no stake in the outcome — becomes the entire job.

Judgment requires caring whether you are wrong; the model does not. Integration requires remembering one specific client's undocumented

history; the model knows the world in general and your client not at all. Verification requires distrusting fluent output; the model produced the fluent output and cannot see its own confident errors. Accountability requires being a *someone* the buyer can hold responsible; the model is a tool, and a tool cannot be sued, fired, or trusted with a quarter.

The next release will write better code and catch more of its own mistakes. It will move the split, too — taking more of the production with it. It will not cross into the carry, because the carry rests on the trait the model structurally lacks: skin in the game. The tools will change; the split keeps landing in the same place. What the tools cannot do is make the buyer pay you for the carry. Only packaging and terms do that.

The people running the largest firms describe the same split

None of this is a forecast. The people running the largest firms in the business are describing it in real time. What struck me, reading the public record, was how completely they agree on the diagnosis — and how much quieter they get on the cure, which is the margin.

The engineering-services firms are watching the production half deflate on their own income statements. Endava’s chief executive, John Cotterell, told investors in late 2025 that the firm was getting “significant step-ups in productivity” from AI. But the gain, in his own words, was “having an erosion on the revenue that’s coming through the business.” There is the value-versus-margin problem stated by a public company: the productivity is real and the revenue is falling anyway, because the savings are flowing to the buyer. Then he named the response. Outcome-based work had risen to roughly a quarter of revenue and was still climbing, with the other three-quarters still billed by time-and-materials. That is a firm trying, in public, to move its *pricing*

— not just its value — onto the outcome it carries instead of the hours it spends.

The unit of billing is following the same logic. Globant’s Martín Migoya put the labor inversion in a sentence: “The initial version of AI was humans accelerated by AI. The next generation is AI supervised by humans.” And his pricing tracks it — away from the hour and toward what he calls *supervised tokens*: a charge for the supervision, which is carry, rather than the keystrokes, which are production. That is the attempt to make the carry paid as an explicit unit, not a bundled freebie.

The venture capitalists financing the disruption say the quiet part directly. a16z’s Alex Rampell sums up the whole services-as-software wave as one pitch: “We’re not giving you software — we’re going to do a job for you.” The bet is that buyers will pay for the job done — the outcome — and stop paying for the tool that does it. Note the structure: it is a bet about *capturing* the value, not just creating it.

The creative side says it too. David Droga, who runs Accenture Song, told a Cannes audience that the staffing model of agencies — the rows of people doing the production — “is going to disappear.” As everyone gets the same AI-driven best practices, he added, “when everybody is doing best practices, nobody is.” The commodity middle hollows out; what is left has to be scarce, or it is just a slower way to be a commodity. Five firms, five seats at the table, one diagnosis. The argument in the industry has moved past whether the value is moving and onto the harder question none of them has fully solved: how to keep the margin once it does.

Finding your share as a number

The split is most useful as a number, not a metaphor — the way you would find your runway. The instrument is simple enough to run in a week, though it tends to be an uncomfortable one.

► **THE CARRY AUDIT.**

Pull your last five statements of work. Print them. Go down each one, line by line, and mark every deliverable with a letter:

- *P* if it is **production** — the making the model now does cheaply.
- *C* if it is **carry** — judgment, integration, verification, accountability the model cannot supply.

Be honest: “senior oversight,” billed as forty hours of a principal’s time, is usually three hours of carry wrapped around thirty-seven of production. Now total the fees against the letters. The share of the fee that sits in the carry is your **carry ratio**.

Then run it again with a harder question against each *C*: is the buyer actually being charged for this as a named thing — or is it bundled in, given away, defended as “what good partners do”? That second pass is the one that hurts, because it usually shows the carry is real and the *price* for it is zero.

Here is what that looks like in practice. Consider a \$240,000 fixed-price build a mid-size product agency shipped last quarter: a customer portal for a regional bank. On the SOW it is one number. Run the line through it.

The React front end, the component library, the boilerplate API wiring, the first-pass copy, the test scripts — call it \$150,000 of the fee, nearly all of it production. A competent team with current tools now produces it in a fraction of the hours it used to bill, and the bank’s procurement team can price that for itself.

What sits in the carry is quieter and smaller on the page. There were the three weeks spent deciding which of the bank’s twelve legacy authentication flows the portal had to honor — judgment. There was the integration against a core-banking system that failed in undocumented ways under load. There was the security pass that caught a token-scoping bug the model had introduced with total confidence — verification. And there was the firm’s name on the indemnity when it shipped — accountability. That is maybe \$90,000 of the fee, and

it is the only part of the project a competitor cannot reproduce off a prompt.

The agency's carry ratio on its flagship engagement was 37%. That number is not the good news it looks like, though. The firm had never named those four items as things the bank was buying. They were bundled into a blended rate and defended as part of the relationship. So the 37% was value the firm delivered, not margin it had protected — and the entire pitch had gone into defending the hourly rate on the \$150,000 that AI was already busy halving.

That is what the audit tends to reveal, and the pattern repeats from firm to firm: the work has been priced backwards. The hour gets billed for the production that is racing toward the cost of the tools, and the carry — the part with no benchmark — is given away. The carry ratio comes out as *delivered* value, not *captured* value. The firm does the carry. It just never made the carry pay.

Why it hides on a good month

The trouble is hard to see because it doesn't show up on a good month. A firm competes, and wins, on the half of its work that is collapsing in price — so the wins get cheaper every cycle while the logo still feels like a victory. Trace it forward and it compounds.

▲ EACH STEP LOOKS SURVIVABLE. THE SUM IS A SLOW BLEED.

- This year the buyer benchmarks your production against its own AI-augmented team and trims your rate 8%. You hold the account by absorbing it.
- Next year the model is better, the trim is twelve, and the junior roles whose hours you marked up have less to do.
- The year after, a competitor bids the production at cost to win the logo — and the only thing that could have justified your premium, the carry, was never priced as a separate thing the buyer agreed to pay for. There is nothing on paper to charge for.

Revenue looks fine most of the way down. The carry ratio — and whether any of it was ever actually *billed* as carry — is the gauge that was falling, and almost no one watches it.

The sum is a firm doing the same excellent work for a falling number, having trained its best client, year over year, to expect the most valuable thing it does for free. The value was always there. The firm simply never converted it into margin while it still had the leverage to.

Two firms, same work, different fate

Two firms make the point. Both build digital products for regional banks; both shipped roughly the same portal last year for roughly the same fee.

The first did what most do. It bid the whole engagement as a blended rate, won on price and relationship, and never priced the carry as its own thing. Its renewal this spring came back benchmarked against the bank's new internal AI team. The production line items were cut to the bone. There was nothing else on the contract to defend, because the judgment and the integration had always been folded into the rate the bank now considered inflated.

The second firm had run the audit and acted on the second pass — the one about whether the carry was actually being charged for. It split

the engagement on the page. There was a fixed, frankly cheap price for the build. And there was a separately named, separately priced commitment — written in the language procurement already buys — for the architecture decisions, the integration risk, and the security accountability it would carry, with a warranty attached. When the same benchmark hit the same production line, that priced commitment was untouched. The bank had been paying for it on purpose for a year and could see precisely what it would be giving up.

Same work, same client type, same AI. One firm's price is collapsing toward the cost of production; the other's is anchored to a carry it made the buyer agree to pay for. The difference was not who delivered more carry. Both delivered about the same. The difference was who turned it into a priced, defensible line before the benchmark arrived.

? “BUYERS WILL NEVER PAY FOR ‘JUDGMENT’ AS A LINE ITEM. PROCUREMENT APPROVES DELIVERABLES, NOT ADJECTIVES.”

Correct — and procurement is right to. It is the reason the move is never to write “judgment: \$90,000” on the SOW.

It is to sell the bounded *result* the judgment produces — the working, verified, integrated portal you put your name behind, with a warranty and a defined liability you accept — as a named thing the buyer can purchase. Procurement already knows how to buy exactly that: it buys risk-transfer, SLAs, and indemnity every day. The job is to package the carry into those forms, not to invent a new one and hope the buyer values it.

You are not asking the buyer to admire your carry in the abstract. You are selling the outcome only your carry can guarantee, priced as the risk you are taking on. That is what makes value into margin. How you package it is the next chapter. That the production half can no longer carry the price — and that value alone won't carry it either — is this one.

The move is the rest of the book

The move is the rest of this book. It is not complicated to state, only hard to do. Stop selling the production at a premium. Start pricing, proving, and protecting the carry — so the value that moved up actually shows up as margin.

That means packaging the carry as a bounded result the buyer is purchasing, not a courtesy — which is **selling the slice**. It means taking only the carry your firm can actually stand behind, and refusing the carry you can't, which is the contractual discipline of **signing what you can carry**. It means making the carry auditable, because carry the buyer cannot verify is carry the buyer will not pay for — which is the proof in **showing your work**. It means financing the months between the old labor margin shrinking and the new carry margin arriving — which is **surviving the gap**. And it means holding the people in whose heads the carry actually lives, and leading them through a change they can read as clearly as you can — which is **keeping the room**.

Sorrell is right that the model has to change. The change is not a new way to bill the old work, and being correct that the value moved is not enough on its own. The change is the discipline of capturing that value before the buyer does. The buyer is going to draw this line through your firm in the next eighteen months whether you draw it or not. It will arrive as a procurement form, a benchmark, or a renewal. It will strip out everything on the production side and ask, pointedly, what is left — and what you can prove you should still be paid for. The question that matters is whether you priced the right half, in terms the buyer's own systems already approve, while you still held the pen.

THE SPLIT IN 20 SECONDS

- When production gets cheap, the value in the work moves to whoever *carries the consequence* of the result — but the savings belong to the buyer first.
- Value moving up is not margin moving up. The buyer can take the AI savings and still pay you less, unless you make the carry paid.
- The carry is four things: *judgment, integration, verification, accountability* — the four the model can't supply, because it has no stake in the outcome.
- Procurement isn't naive. It already buys risk-transfer, SLAs, and indemnity. The gap is packaging the carry into those forms, not naming it.
- Your **carry ratio** is the share of the fee in the carry — but the harder number is how much of it you actually *bill* for. Usually near zero.
- The whole book is one move: stop selling production at a premium; price, prove, and protect the carry so value becomes margin.

→ ON MONDAY

1. Pull the last five SOWs and run the carry audit — mark each deliverable *C* (carry) or *P* (production).
2. Total the fees against the letters, write down the carry ratio — then mark which *C* items the buyer was actually charged for as a named thing. That second number is the real one.
3. For your flagship account, name the four carries — judgment, integration, verification, accountability — and the consequence the firm is taking on for each.
4. On the next proposal, separate the cheap, bounded build from a named, separately priced carry written as risk-transfer: a result, a warranty, a defined liability.

Sell the slice. Sign what you can carry. Show your work. Survive the gap. Keep the room.

CHAPTER THREE

Sell the Slice

Value moving to the judgment around the work does not make it margin. Margin comes from packaging that judgment into the risk-transfer forms procurement already buys — a bounded result, a warranty, a named liability — so the buyer pays for it on purpose. This chapter is the packaging: why a slice holds a price that a blended promise can't.

IN THIS CHAPTER

- Why the procurement room is four sophisticated buyers, not one naive client
- The Five Parts that package the carry into a slice procurement can sign
- The four billing models, and the contract shape each one needs

Two firms want the same engagement. Six weeks apart, they walk into the same enterprise procurement office and pitch the same buyer on the same scope. The first firm promises everything — it will own the build, the launch, the outcome, and it will sign whatever the contract asks, because confidence closes deals. The second firm does something that looks, in the moment, like a tactical error. Partway through the meeting it slides a single page across the table: here is the list of things we will not be responsible for.

The second firm wins. It wins the work, and at a higher number than the first firm quoted.

It runs against the instinct the trade is built on — that you win by promising more, signing more, standing behind more. In the public

record, the opposite holds consistently. The rest of this chapter is the machinery of why.

Packaging is what turns value into margin

The last chapter ended on an uncomfortable fact: the value moved to the judgment around the making, but the savings belong to the buyer first, and value moving up is not the same as margin moving up. This chapter is the part of the answer that does the converting. It is packaging — taking the carry the buyer can't get from the model and selling it in the forms procurement already buys: a bounded result, a warranty, a named liability the firm accepts for a price. Assert that judgment is valuable and the buyer agrees, then keeps the savings. Package the judgment as risk the buyer is purchasing, and the value finally shows up as margin.

The short version fits on a card. Sell the slice. Own the risk you control. Prove the outcome. The slice is how you package the carry — the half the buyer still pays for — instead of burying it in a blend that procurement knows how to crush. The firm that holds the most money is the one that gives away the most risk out loud.

Sit in that procurement room a moment longer, because it is where the dollar lives or dies. The buyer isn't one person; it is four people who don't report to each other and don't trust the same things. One is counting cost against a number that didn't exist eighteen months ago. One is counting concentration — what happens to the buyer if this vendor disappears. One is counting which models touched the data, and whether the copyright passes through clean. One is counting whether anybody validated the thing before it shipped.

The vendor who promised everything walked in with one answer. The room had four questions. So procurement doesn't reject that vendor. It does something worse. It agrees, takes the all-in promise at the all-

in price, and then quietly discounts the whole invoice to the cost of its cheapest component. There is nothing inside the promise it can hold separately. You can't defend a price for the part you never named.

The firm with the one-page refusal survives the room because it gave the room something to grip. That is what a slice is — not a smaller deal, a graspable one.

*

**

A slice is bounded, not thinner

The distinction is easy to get backward, and it is worth getting right before anything else. A **slice** is not a smaller project. Told to package their services, most firms carve the same blended engagement into thinner blended engagements. A \$2 million build becomes four \$500,000 phases, each still folding the production work the buyer believes AI made cheap into the judgment work the buyer can't supply alone. Thinner is not bounded. A blend cut into quarters is still a blend, and procurement compresses a blend at any size. The moment an invoice mixes a benchmarkable line with a defensible one, the benchmark sets the price for both.

A slice is a bounded operational result a firm can name, price, deliver, verify, and defend. It has edges a procurement officer can trace with a finger. It carries a price that survives contact with the buyer's internal-cost number. It throws off evidence strong enough that finance, legal, and security each sign their own line without taking the firm's word for anything. And it draws an explicit border around the risk the firm will and won't carry. Strip any one of those and the slice falls back into the blend that procurement already knows how to grind down.

There are **Five Parts** to it — a buyer, a boundary, a proof standard, a risk surface, and a commercial wrapper — and they are less a checklist than a single load-bearing structure. Taken out of order they make no

sense. Taken together they make a thing that can be sold next quarter. The place to begin is the part most firms never even identify.

The buyer is four functions, not a client

The first edge is the buyer, and the trap is the word *client*. At the scale this book is about — a top-five US bank, a global media holding company, a name on the Fortune 100 — there is no client. There is no single signature. There are the four functions from the procurement room, each evaluating a different universe, and any one of them can kill the deal alone.

The chief financial officer is benchmarking the price against the cost of doing the work in-house with a junior team and a frontier model. That number genuinely did not exist two years ago, and it compresses the production half of any invoice on contact. The Vendor Management Office — the team that approves suppliers — isn't pricing the work at all; it is pricing the firm. Concentration, continuity, whether this vendor is a single point of failure, whether the relationship can be exited without a fire. It reads every offer through a lifecycle the regulators handed it. The 2023 interagency guidance from the Federal Reserve, the FDIC, and the OCC treats a third-party relationship as five stages, cradle to grave: planning, due diligence, contracting, monitoring, and termination. The VMO approves vendors, not outcomes.

Then there is Third-Party Risk. It didn't scrutinize creative and engineering vendors this way three years ago, and does now, on every signature cycle. Which models, whose training data, does the copyright pass through, does the indemnity survive an AI tool sitting in the loop. And Model Risk Management — the function that has long governed a bank's own models and its vendors' alike — sits in a revealing bind with generative AI. The 2026 refresh of US model-risk guidance, which replaced the 2011 framework, holds that generative and agentic AI fall

outside formal model-risk scope. Yet it tells banks to govern those out-of-scope tools by the same principles: materiality, validation, ongoing monitoring, effective challenge.

So at the most regulated buyers the discipline reaches vendor-supplied AI anyway — by principle rather than by letter. The development controls, the monitoring, the governance, and a hard answer to the question of what happens when the model is wrong. That gap, between “outside formal scope” and “still has to be governed,” is exactly where a vendor gets caught. None of this is a creative-review checklist; it is a governance machine.

It is worth being plain about what that machine implies, because it overturns the comfortable version of the story. This buyer is not naive about the value of judgment, waiting for a vendor to teach it a word for the carry. It buys risk-transfer, service-level agreements, indemnity, and assurance every quarter; it has decades of machinery for paying someone to hold a consequence. The four functions are the proof of that sophistication, not the absence of it. So the seller’s problem was never that the buyer couldn’t see the carry. It is that the seller kept handing the carry over as a free courtesy folded into a rate, in a form the machine had no line to pay it on. Packaging is the fix: putting the carry into the risk-transfer shapes the machine already approves.

Name the slice once per buyer

So the slice names which of the four signs, and — this is the move — it names itself differently for each. The work might be the same six people doing the same thing. The offer can’t be. Sold to the CFO, a checkout-flow engagement becomes **Checkout Conversion Instrumentation and Release Readiness**, priced against the internal-AI benchmark precisely on the ground that benchmark can’t reach. Sold to Third-Party Risk, the identical work is **AI-Assisted Personalization Boundary and Evidence Pack** — model use,

data exposure, indemnity posture, all named on the face of it. Sold to Model Risk Management, it is **Human-Reviewed Recommendation Workflow with Drift Monitoring Exclusions**. There, *exclusions* does the heavy lifting: the fence around what the firm won't warrant is what makes the rest credible.

Aim the same slice at the wrong function and it dies of irrelevance. So naming the buyer isn't manners; it is the first cut of the border.

*
**

The boundary is the spine

The other four parts hang off the boundary, and the reason is mechanical, not poetic. The buyer signs a boundary. The proof certifies delivery inside a boundary. The risk surface is the boundary's outer edge, drawn in the open. The wrapper is the contract shape that carries the boundary into paper the buyer already has. Get the boundary wrong and everything else is good scaffolding around the wrong building.

A boundary is not an old-style scope document — a list of what gets built. It is a risk-tagged scope, where every line wears one of four tags: agency-controlled, client-controlled, shared, or excluded. The architecture decision is agency-controlled. The decade-old legacy system the new work has to plug into is client-controlled. The data migration is shared. The behavior of a third-party language-model API the design leans on is excluded. The tags aren't decoration. They are the map of who answers the phone when something breaks.

Most agency scopes are written as promises. A boundary is written as a border. The difference only shows up the first time a project goes sideways — and then it is the whole game. The firm holding a promise argues about what everyone meant. The firm holding a border points at a tag. One of those arguments happens in the cold light of the contracting room, when both sides can still think clearly. The other

happens in the heat, when neither side can. Draw the border early and you have already won the argument you haven't had yet. Procurement doesn't hate boundaries. It hates surprises, and a boundary is the opposite of one.

*
**

The proof standard makes it signable, not expensive

A slice you can't prove is a slice you can't defend, which makes the proof standard the part that turns a good story into a signature. It is the evidence pack — the artifacts that let each of the four functions sign its own line. The pack has a floor of three things. Acceptance criteria written before the work, defining *done* in terms the operating owner agreed to. A disclosure of what models were used and where and under what controls. And — wherever a model produced or shaped an artifact — a provenance record, a trace of what made it.

Engineering already has names for this. The NTIA published the minimum elements of a software bill of materials — the SBOM, an ingredient list for code — in 2021. CISA's 2025 work pushed supply-chain transparency from a courtesy into a procurement expectation. And the model-inclusive successor, the AIBOM, is forming now around model, dataset, lineage, and configuration. The buyer may not say the acronym. The buyer says the demand behind it: show me what is inside the thing.

Creative and brand work has its own version, and it deserves the same rigor — not a portfolio and a warm reference. The name for it is **Generative Provenance**. That means the record of which model versions made or assisted the work, a training-data attestation, and the copyright pass-through chain. It also means the human-review logs proving a person was accountable at the points that mattered. The

C2PA standard — a content-authenticity spec — reaches into AI and machine-learning provenance, and points straight at this. Authenticity built into the model and the output, not bolted onto the file afterward. So the holding companies — your WPPs and Publicis-scale shops — are discovering that their Third-Party Risk reviewers no longer sign a creative deliverable on the strength of a reel. They sign on the strength of a provenance record — the same way the buyer signs an Endava or a Globant on the strength of an SBOM.

Proof is the ticket; context sets the number

Proof is the ticket, not the price. The evidence pack gets the slice through the four-function review — it makes the thing signable. It does nothing, by itself, to make the thing expensive. What holds the price is the one input a competitor can't reconstruct from the artifacts. That input is the firm's privileged knowledge of this buyer's systems, this buyer's failure history, this buyer's actual constraints. Anyone can copy your evidence pack. No one can copy three years inside the buyer's integration layer. Proof gets you in the door. Context sets the number on the way out.

*
**

Own the risk you control — and refuse the rest, just as loudly

Now the part where firms either give away the company or refuse to sign anything. Both failures come from the same error: treating risk as one undifferentiated mass to be swallowed whole or spat out whole. It is not one thing. The discipline that splits it has a name: **Own the Risk You Control.**

The firm accepts — explicitly, in writing, without flinching — exactly the risks it controls. Scope: it built what the boundary said. Quality: it met the acceptance criteria. Integration: the slice connects to the

systems it was contracted to connect to. Continuity: the work survives a key person walking out the door. Security process: the controls were followed and logged. Verification: the proof exists. Measurable progress: the milestones landed on the agreed dates. These are the firm's to own because they are the firm's to control, and owning them loudly is the entire source of the premium. A firm that will put its name on its own quality standard is worth more than one that hedges it into fog.

And then the firm refuses, with equal volume, the risks it doesn't control. Upstream model drift, when the frontier model the design depended on changes its behavior in a release the firm didn't write. Third-party LLM API changes — an endpoint moved, a rate limit cut, a pricing tier reshuffled. Hallucination failures it took every reasonable control against and still couldn't eliminate, because the technology can't yet be made to. Attribution-fuzzy KPI gates in domains where a result can't be cleanly traced to the work. And the buyer's own risk: its delays, its approval cycles, the regulatory shifts that move the target after the ink dries.

That line between the two lists isn't a negotiating posture to be softened later — it is the carry line drawn on the contract where everyone can see it. The carry you'll hold on one side, packaged as a priced liability; the risk you refuse on the other.

▲ ACCEPT A RISK YOU CAN'T HONOR AND YOU'VE WRITTEN THE BUYER A FREE OPTION.

A firm that accepts a hallucination warranty it can't honor hasn't won anything — it has handed the buyer a free option it will exercise the first time a model misbehaves. A firm that refuses to stand behind its own quality hasn't protected itself — it has announced to procurement that there is nothing here worth buying.

Which returns us to the two firms in the opening room. The one that promised everything sold an option against itself. The one that handed over the page of refusals wasn't being modest. It was demonstrating, in the only language procurement believes, that the things it did promise were promises it could keep. The refusal page is what makes the rest of the promise credible.

The accepted risks, for the record, resolve in the contract into five categories: performance, IP indemnity, confidentiality and breach, consequential, and liquidated damages or SLA credits. Each is separately priced, capped, allocated, and checked against the insurance that will actually respond. The mechanics of each are the next chapter's work. Named here, they are simply the lines the border draws.

Legibility earns the hearing; context holds the price

None of this makes a firm immune to price. A competitor can copy the refusal page — show up just as legible, just as disciplined about what it won't sign, and undercut the number anyway. That firm exists, and on a bare rate sheet it sometimes wins. What it can't copy is the input from a few pages back: the privileged read on this buyer's systems and failure history that no artifact carries out the door. Legibility is what earns the premium a hearing; context is what holds it.

So this is a tendency, not a law — the structure doesn't guarantee a higher number on any single deal. What it does is move the fight. Off the production line, where the firm loses to anyone holding the same tools, and onto the judgment lines, where the cheap-and-legible competitor has nothing the buyer can benchmark. None of that promises victory; it is just the only ground on which the firm can win one.

*

**

Globant productized AI to clear procurement

If all of that sounds like theory, it helps to watch a \$2 billion firm do it in public and lose money on purpose.

In March 2026, Globant reported a number small enough to skip. Its AI Pods — the company’s name for packaged units of AI delivery — had reached \$32.8 million in annualized recurring revenue, sitting inside 40% of its top 20 accounts. Against a first-quarter revenue around \$607.1 million, annualized, that is 1.3% of the run-rate base. A rounding error on a \$2-billion-plus business. The kind of line you could cut from the earnings deck and lose nothing.

The number is small; the packaging is the point. An AI Pod is a SKU, not a project or a team you rent. Defined inputs, defined outputs, written service levels, an audit posture fixed before the first invoice, priced as a thing that recurs. And recurrence is the tell, because you can only put a subscription line on something the buyer can describe without you in the room.

So ask the only question that matters: why would a firm that size bother for 1.3%? Not to grow revenue — to survive the four-function room. A bounded, legible, procurement-grade unit is the single shape that clears the gauntlet where the blended dollar gets crushed. Globant didn’t productize AI to add a point of growth. It productized AI because the Pod is a packaged unit priced to clear procurement. And a 180-person agency doesn’t need a \$2 billion balance sheet to copy the grammar, only a narrower wrapper.

That narrower wrapper is the fifth part, and the one that decides whether any of this can actually be sold before the holidays.

*

**

The wrapper has to ride the paper that already exists

The firms in this book aren't selling into empty fields. They are selling into existing relationships, governed by master agreements negotiated over months. Those agreements were signed by people who have since moved on, and guarded by procurement teams whose entire job is to keep them shut. A repricing strategy that requires reopening the master isn't really a strategy; it is a two-year legal project the client has no reason to begin.

So the commercial wrapper has to let the slice ride inside the paper that already exists — the same Five Parts every time, four different shells. The wrapper is the only part of a slice that bends to fit the buyer's existing contract. The shells differ by revenue model. The slice gets built once; the wrapper gets built once per model.

The project shop: concede the cheap work, rename the rest

The project shop sells discrete builds. Its instinct under pressure is to plant a flag on the total — the \$2 million build is still worth \$2 million, hold the line, dig in. That is the losing move, because the buyer has already taken the invoice apart, and the buyer is right about half of it. The seven-line teardown that opened this book showed the anatomy; what it did not show is the move on the page — the same engagement billed two ways, and where the margin survives the second time. That is worth re-cutting here, because the rename stays abstract until it lands on an actual invoice.

► **ONE ENGAGEMENT, TWO INVOICES — WHERE THE PRICE HOLDS.** (ILLUSTRATIVE; FIGURES ARE THE WORKED \$2 MILLION BUILD, NOT A REAL DEAL.)**

The old blended bill. A single line: *Mobile app rebuild* — \$2,000,000. The judgment is in there — the architecture call, the integration against the client’s legacy core, the security pass, the launch warranty — but none of it is named. It rides inside the number for free.

The buyer’s objection, AI-driven. Procurement runs the build against a junior team and a frontier model and answers with a number that didn’t exist eighteen months ago: *we can produce this for roughly \$700,000 in-house.* Against one blended line, that benchmark sets the price for the whole thing, and the engagement sags toward \$1.5 million — the blend punished for the sins of its cheapest line.

The revised offer — the carry packaged as risk the buyer buys. The same work, re-cut into lines procurement already knows how to pay for:

- **Production build — fixed, frankly cheap.** Conceded to the benchmark, priced to lose the argument the firm can’t win. This is the ~\$700,000 the buyer was right about.
- **Integration-liability boundary.** A named liability the firm accepts for the slice working inside the client’s legacy systems — and an explicit refusal to warrant the behavior of those systems themselves.
- **Verification and release readiness.** Priced as a separate result: the proof that what shipped is *done*, not merely *looks done*, with the evidence pack attached.
- **Warranty and launch accountability.** A bounded warranty on the working outcome — the throat to choke at 2 a.m. — priced as the risk it is, not given away as relationship.
- **Generative provenance.** The record of which models touched the work, the data attestation, the human-review logs — sold as the artifact Third-Party Risk now requires to sign.

The margin math. The production line compresses to the cost of the tools, as it was always going to. But the four packaged lines — the \$1.3 million defensible half — carry a price the benchmark can’t reach, because procurement is buying risk-transfer there, not keystrokes. The engagement holds near \$1.8 million. The firm did not save the \$700,000. It stopped defending it, and moved its margin onto the lines a competitor can’t reproduce off a prompt.

The two numbers that matter survive the re-cut. Roughly \$700,000 of that build is production — the work a competent internal team with frontier tools can now largely reproduce. Fighting for it is fighting the benchmark, and the benchmark wins. The other \$1.3 million is the defensible half.

A firm that defends the blended total and refuses to decompose hands procurement a single mass it can't pull apart. The production discount gets applied to the whole thing, and the engagement sags toward \$1.5 million as the blend gets punished for the sins of its cheapest line. So the move is to concede the cheap work first and let the \$700,000 compress. Then hold the rest with something that is not a discount and not a discipline but a rename.

The slice is not “mobile app rebuild” — that is the container. The slice is **Cross-System Launch Readiness and Accountability**, and the change of name changes the conversation the buyer is having. “Mobile app rebuild” competes with AI-assisted production — and loses, because the buyer now has a cheaper way to produce. “Cross-System Launch Readiness and Accountability” competes with the cost of the app failing on launch day. And there is no cheaper way to buy a throat to choke when revenue is on the line. The first name walks the firm into a rate comparison. The second walks it into a risk conversation.

▲ **THE RENAME IS A REFRAME, NOT A RELABEL.**

A firm can print Cross-System Launch Readiness and Accountability on a cover sheet and still lose the room if it can't defend the boundary beneath it — procurement has read enough grand names to discount the ones that are only names. It works because it changes the cost the buyer measures against, from the price of production to the price of a launch that fails. The words are how the firm holds that frame. The boundary is what makes them true.

Underneath, the slice does exactly what the Five Parts require. Its buyer is the operating sponsor and the CFO. Its boundary tags the legacy systems client-controlled, the architecture and integration agency-controlled, the migration shared, the app-store policy call excluded. Its proof is the integration suite, the architecture decision record, and the release evidence. Its risk surface owns integration and quality. And it refuses two things: the behavior of the client's own legacy systems, and the downstream revenue no honest firm can underwrite.

Repackaged this way, the engagement holds near \$1.8 million — \$300,000 above the do-nothing \$1.5 million, collected on exactly the lines the buyer couldn't benchmark. The firm didn't hold \$1.8 million by holding the line. It held \$1.8 million by drawing it, and by conceding the \$700,000 first. That is the paradox again, with a price tag on it: you keep the larger number by surrendering the smaller one out loud.

The retainer firm: the memo, not the gate

The retainer firm has the more dangerous temptation, because its bad idea sounds like courage. Under pressure, it wants to prove its worth by tying the fee to outcomes — convert the flat monthly into a KPI-gated deal. *Pay us on results.* In attribution-fuzzy domains — brand, growth, demand — that conversion is the costly mistake.

► **HOW AN OUTCOME GATE BECOMES A STANDING DISPUTE.**

An \$80,000-a-month growth retainer converts to an outcome structure gated on a KPI — pipeline, qualified leads, some revenue-influenced number.

- In Q2, the KPI hits, and the agency expects to be paid.
- The buyer’s CFO, doing precisely the job a CFO is paid to do, re-attributes the win: the number moved because a platform algorithm shifted in the agency’s favor that quarter, and because a competitor went out of stock and the traffic redistributed. Neither was the agency’s doing.
- The gate technically triggered, but the attribution is now contested — and contested attribution in a CFO’s hands is just a discount with extra steps.

The \$80,000 settles to about \$75,000 effective. Worse than the \$5,000, it installs a standing quarterly attribution dispute the agency can never win, because the domain genuinely doesn’t permit clean attribution.

The firm didn’t sell an outcome. It sold itself a recurring trial in which it is permanently the defendant.

What defends the retainer is not a gate but the **Engagement Boundary Memo**. The slice it carries is not “growth marketing” — a phrase abused past meaning — but Quarterly Demand-System Improvement and Evidence Pack. The memo is one page, signed every quarter, and it does four things. It states what was in and out of scope. It states what was delivered. It states what the firm stands behind, and what it does not. That fourth line is the credibility. A firm willing to write down the limits of its own accountability is a firm whose other claims you can believe.

The asymmetry is the design. The EBM carries a voluntary fee adjustment on demonstrated outsized contribution. When the quarter’s work truly drove a result, the buyer can pay up — by choice, on evidence, never by formula. What it doesn’t carry is a claw-back or an SLA credit. There is no mechanism to dock the fee because a KPI in a non-

deterministic domain came in soft for reasons no one can untangle. Upside is shared on evidence; downside is not gated on attribution that can't be established.

? “ISN'T THE MEMO JUST A PLACE TO HIDE FROM ACCOUNTABILITY?”

It has it exactly inverted. The claw-back retainer is the one pretending to an accountability it can't deliver, staking the fee on a number the domain will never let anyone honestly attribute.

The EBM defends the aggregate work — named and stood behind — without betting it on a single fuzzy gate. It doesn't say *trust us, marketing is hard*. It says: here is the part we controlled, here is the evidence, here is the part no honest vendor can underwrite. That is a stronger sentence than a brave invoice, and it doesn't spend next quarter being re-litigated.

The staff-augmentation shop: a rider, not a conversion

The staff-augmentation shop bills time-and-materials against a master agreement. Its instinct is to move “up the value chain” by converting that T&M relationship into milestone or outcome billing — to look modern. At an enterprise buyer, and the sharpest case is a top-five US bank, that conversion is usually a bad trade. Nothing forbids it. The trouble is that it runs into the buyer's own governance and spend accounting before it ever reaches the agency's economics. So it stalls in the Vendor Management Office for reasons the agency is never told.

The bank doesn't lack imagination. It governs every vendor through the lifecycle the regulators handed it, the one that runs the whole relationship — planning, due diligence, contracting, monitoring, termination. Inside that machine, a mid-stream change to the commercial terms is rarely a quiet amendment. Convert a settled time-and-materials arrangement into milestone or claw-back billing, and you can reopen the third-party-risk file on a relationship that was already approved. A change to payment terms, remedies, and risk allocation is

exactly the kind the guidance says to reassess, not rubber-stamp. So a commercial tweak can become a governance event before it becomes a better deal.

Then look at how the bank actually buys bodies. Contractor headcount is approved against a budget and a forecast, on a vendor-governance cycle, with the spend expected to land when the plan says it will. A deferred or milestone-gated fee complicates that accounting — the money now arrives on the slice's clock, not the plan's. And a claw-back, a routine commercial remedy elsewhere, buys less than it looks like here. At a regulated bank, a refund alone does not resolve a vendor's underperformance. The bank works it through its own governance and supervisory channels, on a timeline no vendor-issued credit can stand in for.

None of this makes the conversion impossible — banks execute milestone riders every day. It makes it a poor trade. The agency thinks it is proposing a more sophisticated structure. The bank hears a request to reopen a settled governance file and disturb its own spend forecast, for an outcome it can reach a cleaner way. The better economics rarely get a clean hearing.

The wrapper that works is the **Slice Order**. It is a rider on the existing T&M master that names a workstream as a bounded slice, without converting the master at all. The base team stays time-and-materials. The slice sits beside it, a named bounded result with the Five Parts. You call it what it is to the people who must sign it: Legacy Claims Workflow Evidence and Release Readiness. Its buyer is the VMO plus the operating executive. Its boundary gives the agency workstream delivery, integration notes, test evidence, defect triage, and release-readiness reporting; the bank keeps environment access, policy calls, data approvals, and release windows. Its proof is a monthly evidence

pack. The master terms never reopen; the rider rides the buyer's normal contract-governance path.

Two structures that look like twins

Inside the Slice Order, where the survival actually lives, is a choice between two structures that look like twins and behave like strangers. The right one is a **Performance Rebate** — what the accountants call variable consideration, revenue you expect to give back. It is estimated at the contract's inception and netted against the price up front. So the firm books only the revenue it expects to keep. Like any variable-consideration estimate, that figure has to be reassessed as the facts develop. The wrong one is an **At-Risk Fee**: the same dollars, but booked as earned and then placed in jeopardy, which leaves a contingent receivable that may never be collected.

Here is the correction to the tidy version of this advice, the one in which the rebate is free. It is not. A lender funds payroll against the receivables pool, and any concession dilutes that pool — a rebate shrinks it by a known amount. But an at-risk fee is worse, because a contingent or disputed receivable is not merely smaller; it is typically ineligible. Most credit agreements strike contingent or disputed receivables from the base until the condition clears, and the uncertainty drags down the advance rate on the receivables that remain.

So the rebate costs working capital predictably, on the firm's terms. The at-risk fee costs it unpredictably, on the buyer's. And the at-risk fee carries one hazard more. A fee refunded because the work disappointed starts to read like a warranty. But an errors-and-omissions policy underwrites negligent work, not a commercial bet on a fee. So the firm can find itself uninsured on exactly the dollars it put in the fire. Two structures meant to share upside. One is a known cost the firm controls. The other is a quiet, uneven bet against the firm's own credit

line and its own insurance. The Slice Order signals repricing readiness inside paper the bank can sign. No conversion. No reopened master.

The hybrid firm: one grammar across three modes

The hybrid premium firm runs all three motions at once — project builds, retainer tails, embedded teams inside enterprise clients. Its failure is to let them drift into three commercial dialects spoken to the same procurement office, which reads the inconsistency as a vendor that doesn't know what it is. A fixed build in one voice, a retainer in another, an embed in a third — and the buyer has been trained to see three suppliers where there is one.

The fix is one slice grammar across three billing modes, carried on the contract shape the enterprise already trusts: the exhibit-based MSA, standard since the early 2000s. That is a master that sets the terms once and hangs an exhibit on each engagement. The hybrid firm writes the same Five Parts vocabulary into every exhibit. That holds whether the exhibit is a fixed-SOW build, a quarterly EBM retainer, or a Slice Order on a T&M base. Same definition of a slice. Same boundary tags. Same proof standard. Same risk language. Three billing modes, one grammar.

And the payoff compounds. The same procurement office meets the firm's paper across a build, a retainer, and an embed — and finds one coherent posture. A firm that knows precisely what it sells, and precisely which risks it signs. That coherence is itself a slice the firm sells without ever naming it: the legibility premium. It is worth more than any single project on its own. The firms collecting it are the ones that look most like Accenture in their discipline, and least like it in their size.

*
**

“Isn’t a slice just a SOW in new clothes?”

There is an honest objection here, and it deserves an honest answer, because part of it is simply right. Exhibit-based MSAs are old. Fixed-scope SOWs are old. Statements of work with acceptance criteria and risk allocation have existed for two decades, and a sharp reader is right to ask what is new besides the vocabulary. If a slice is just a SOW wearing a fresh name, this chapter is repackaging dressed as doctrine.

? “WHAT’S ACTUALLY NEW BESIDES THE WORDS?”

The contract shape is old. The burden it carries is new. Four things changed, and together they reload the old instrument.

- The **four-function review** — CFO, VMO, Third-Party Risk, Model Risk Management each signing alone — didn’t gate creative and engineering vendors a few years ago, and does now.
- The **AI-augmented internal-cost benchmark**, the number the CFO measures the production half against, didn’t exist eighteen months ago, because the tools that generate it didn’t.
- The **verification artifacts** — SBOM and AIBOM in code, Generative Provenance in creative — are demanded today on signature cycles where a portfolio and a reference call once carried the day.
- The **bifurcation** itself — the dollar splitting between a production half the buyer believes AI made cheap and a judgment half the buyer still can’t supply alone — is what made the blended invoice newly indefensible, and what made packaging the judgment as priced risk the only defense.

None of those existed in this form when the exhibit-based MSA became standard. The old SOW protected delivery. The slice protects price.

The market is already paying for it, which is the part no thesis can argue with. CI&T reported \$136.6 million in revenue for the first quarter of 2026, up 23.2% year over year from \$110.9 million, with a fifth of its new sales written on new pricing models. That last figure is the one worth holding. Twenty percent of new business, booked on shapes other than the old time-and-materials default — the bounded and

packaged and provable unit clearing procurement and commanding a number. CI&T didn't just say AI made delivery faster; it connected AI to the shape of the invoice. So the slice is less a thesis waiting on evidence than a sale already closing on the firms that built it.

*
**

Every winning move is a concession made out loud

So return one last time to the two firms in the procurement room. The one that promised everything is gone — discounted to the floor or refused outright, because an all-in promise gives a four-headed buyer nothing to hold and every reason to compress. The one that handed over a single page of refusals walked out with the work and the premium. It understood the thing this whole chapter has been circling. In a room where the dollar gets crushed, the firm that names what it won't stand behind is worth more than the firm that swears it will stand behind it all.

Globant gave away the appearance of growth to keep the substance of survival. The project shop gave away \$700,000 to keep \$1.8 million. The retainer firm gave away the brave invoice to keep the durable fee. The bank engagement gave away the milestone to keep the master closed. Every winning move in this chapter is a concession made out loud, in writing, on purpose — because legibility is the whole game. What is legible can be priced and proven and held. What is blended gets ground down by a procurement office that has every tool it needs to find the cheapest reading of an invoice it can't break apart.

But naming a slice is not signing one. A boundary the firm can't defend in the contract is a boundary the buyer's lawyers will redraw. A risk surface that sounds disciplined on the page still has to survive two readers: a carrier who has never heard the word *slice*, and a

lender whose covenant doesn't care. The five categories named here — performance, IP indemnity, breach, consequential, liquidated damages — have to become real clauses. That means caps the insurance will actually honor, and structures the borrowing base can actually carry. That is the arsenal, and the next chapter builds it. The Slice Order as a contract primitive. The Redline Field Manual for the clauses procurement will try to move. The boundary instruments that hold under pressure. And the Liability-and-Insurance Worksheet that prices, to the dollar, how much risk a firm can sign before it is quietly betting against its own carrier and its own lender.

Name the slice in this chapter. Sign it in the next.

SELL THE SLICE IN 20 SECONDS

- A *slice* is a bounded result you can name, price, deliver, verify, and defend — not a thinner blend. A blend cut into quarters is still a blend, and procurement compresses it at any size.
- The buyer is *four functions* — CFO, Vendor Management Office, Third-Party Risk, Model Risk Management — each signing alone, each able to kill the deal.
- The *Five Parts*: a buyer, a boundary, a proof standard, a risk surface, a commercial wrapper. The boundary is the spine; the others hang off it.
- *Own the Risk You Control* — accept what you control out loud, refuse what you don't just as loudly. That refusal line is the carry line drawn on the contract, the carry packaged as priced liability.
- Proof is the ticket, not the price. Context — your privileged read on the buyer's systems — sets the number.
- One slice, four wrappers: a renamed fixed build, the Engagement Boundary Memo, the Slice Order rider, one grammar across all three.

→ **ON MONDAY**

1. Take one live engagement and split it: which lines can the buyer's own AI reproduce, and which can't. Concede the first set; name the second as the slice.
2. Rename that slice to the cost it actually defends against — a launch that fails, not “a rebuild” — then write the boundary that makes the name true.
3. Tag every scope line agency-controlled, client-controlled, shared, or excluded — and write the one-line refusal of the risks you don't control.
4. Pick the wrapper that fits the existing paper — renamed SOW, Engagement Boundary Memo, or Slice Order rider — so nothing forces the master open.

Sell the slice. Sign what you can carry. Show your work. Survive the gap. Keep the room.

CHAPTER FOUR

Sign What You Can Carry

Selling the carry only matters if the contract makes it paid. This is the lever that does it: writing the carry down as risk-transfer the buyer already knows how to buy, so valued judgment becomes purchased obligation instead of a courtesy folded in the rate. This chapter is about the three clauses that decide whether the price holds, and the one piece of arithmetic that tells you whether you can actually carry what you just signed.

IN THIS CHAPTER

- Why the carry only becomes margin when it's priced as risk-transfer in the contract
- The three clauses that decide the deal — and the part of each that founders fight on by mistake
- The worksheet that tells you, before you sign, whether to push back or walk

The slice comes back bleeding.

You sent a clean one-page rider for the \$1.8 million build — a bounded result, four signatures, the line between the risk you would carry and the risk you would not, all argued onto a single page. It is back on your screen now. The buyer's outside counsel has been through it in tracked changes, and there is more red on the page than black.

One clause carries the whole story. Your rider capped the firm's liability at the fees paid — \$1.8 million, the most you could ever lose. The redline strikes the cap and leaves a margin note in a lawyer's flat courtesy: *Cap not acceptable for indemnified claims — see §9.* You open §9. The

indemnity now reaches third-party intellectual-property claims, data breach, and “any claim arising from Vendor’s use of artificial intelligence.” Uncapped. Surviving termination. Payable on demand. The cap you wrote still sits in the contract, untouched and now meaningless — because everything that could actually take the firm down has just been lifted out from under it.

That one clause can cost more than the whole deal pays. And it has almost nothing to do with how well you built the slice.

This is where strategy meets a person whose job is to undo it

The redline is where the strategy meets a person paid to undo it. Everything until now is the pitch — the split, the carry named in four parts, the slice, the risk you refused out loud. Now it meets a person whose entire job is to move your risk onto your balance sheet. He has never heard the word *slice* and would not care if he had. What he does understand, fluently, is risk-transfer — caps, indemnities, warranties, SLAs. That is the whole of his trade.

He is not your enemy, and he is not naive. He is doing his job exactly as well as you did yours when you drew the boundary, and he reads these clauses for a living. The difference is that his job is to widen every definition in the document until “the risk you control” quietly comes to mean everything. Yours is to stop him without losing the work — and to know, before you sign, precisely how much you have agreed to carry, and for what price.

So it helps to recall the split from two chapters back, in *What Still Gets Paid For*. Below the line sits production — the keystrokes, the boilerplate, the generated draft — and the cost of that work has fallen toward zero, which the buyer knows better than you do. Above it sits

the carry: judgment, integration, verification, accountability. That is what the buyer still values, and the only thing you are actually selling. But that chapter ended on a harder point, and this is the chapter where it comes due. Value moving up to the carry is not the same as margin moving with it. The buyer can pocket the AI savings and still pay you less. Making the carry *paid* — turning a thing the buyer values into a thing the buyer is contractually buying — takes four levers: scarcity, proof, packaging, and contractual discipline. The first three are the chapters around this one. This is the chapter on the fourth.

Which changes what a contract is. When you sit down to a software deal in the post-code era, you are no longer negotiating the price of labor. You are pricing the carry — and writing it down in a form the buyer's own systems already pay for. Every clause fight — every cap, every indemnity, every service credit — is where the carry either becomes a paid, contracted obligation or leaks back into a courtesy folded in the rate. Contractual discipline is the lever that converts valued judgment into purchased risk-transfer. The firm holds the consequence, on purpose, for a price the buyer agreed to. The expensive mistake is the inverse: signing for carry your balance sheet, your team, and your proof system cannot hold, for no price at all.

There is an industry ready to sell you a binder of agency-favorable templates for this moment — one airtight clause for every fight. The binder is the wrong tool. (What follows is how operators think about these clauses, not legal advice for your jurisdiction; you still need your own counsel.) You do not hold the line with a binder. You hold it with three moves, a discipline borrowed from a much older profession, and one piece of arithmetic most firms skip.

*
**

Don't bring your own paper

Start with the move firms get backward before they ever reach the redline: they bring their own paper.

It feels like strength — proof you are a real shop with real terms, that you have done enough of these to have a position. To the enterprise across the table it reads as the opposite. It looks like a vendor expecting to set terms for an institution that sets terms for a living, behind a procurement policy written precisely so that vendors do not.

At this scale the contract that governs is almost always the buyer's. The master agreement underneath it is a quarter-long fight you will not win. It was negotiated over months by people who have since moved on, and reopening it summons legal, risk, and procurement — any one of whom can stall it forever, and none of whom gain anything by hurrying. So the move is not to reopen it. The move is to hang the slice off the paper that already governs. Use a short order — the Slice Order — that names the bounded result and sets the handful of terms the slice actually needs. For everything else, it points back to the existing master. The base relationship does not move. Only the new, bounded thing gets new, bounded terms.

A rider that varies three things inside a structure the buyer already blessed is something a general counsel can sign in a week. You will not out-paper a bank. You can out-structure one — by making the thing you ask it to sign small enough that yes is easy, while insulating the carry you are actually taking on.

*

**

Three clauses decide the deal

Riding the buyer's paper does not save you from the clauses. It narrows them to the few that actually decide the slice. There are three. The rest your counsel can hold alone; these three earn you your seat at the table.

Each is a place where the principle from the last chapter — own the risk you control, refuse the risk you do not — either becomes priced carry or evaporates into a lawyer’s definitions. These three clauses are not legal housekeeping. They are the instruments that turn the carry into risk-transfer the buyer pays for — the same ones procurement uses to buy assurance from anyone else.

The cap, and the part of it founders fight by mistake

The first is the **limitation of liability** — the cap on what you can lose. Founders fixate on exactly the wrong part of it. They argue the number: one times the fees, two times, a hard dollar ceiling. Fight that fight if you like, but the number is not where the danger lives.

The danger lives in the carve-outs — the categories of claim the contract lifts out from under the cap. For those, exposure is not limited to a multiple of fees; it runs to the full size of whatever went wrong. Every cap has them, and the buyer’s counsel will work to lengthen the list: intellectual property, confidentiality, data breach, “gross negligence and willful misconduct,” and now anything that touches AI.

The effect on the math is stark. A two-times cap on a \$1.8 million order looks like a \$3.6 million ceiling — a number you could survive. Carve IP and breach and AI out from under it, and the ceiling is a fiction. On the only claims likely to be catastrophic — the ones that run to eight figures in enterprise damages — your exposure is unbounded.

So the leverage does not go on the multiple. It goes on the carve-outs, fought on the principle that drew the slice. Accept uncapped exposure only for harms you genuinely control and could have prevented — your own willful misconduct, your own breach of someone’s confidence. Refuse it for harms you cannot. That means the behavior of a third-party system you did not build, or the buyer’s own data flowing through your work. It means a regulatory line that moves after the ink dries, and the hallucinated edge cases of a model whose weights you do not own.

A clean-looking cap with the dangerous claims carved out from under it offers no real protection at all. It is a number that exists to make the firm feel protected, while the only exposures big enough to be fatal get signed away underneath it.

The indemnity, where the usual advice backfires

The second clause is the one where the usual advice gets you killed, so it is worth slowing all the way down. The IP indemnity is your promise that if the work you hand over infringes someone else's intellectual property, you will cover the buyer for it. In 2026 the buyer's version arrives with AI named on its face, because that is the fear of the moment — and the buyer is right to have it.

Somewhere out there is a copyright claim waiting for the first enterprise that shipped a foundation model's output into a product. The buyer would very much like that claim to be your problem, not theirs. They are asking you to carry the model.

The instinct, when the demand lands, is to cap it. *We will indemnify you, but only up to the fees you paid us.* It sounds disciplined and responsible. It is also the move much of the agency-contract advice online will hand you, and it usually fails — on the single clause where failing costs the most. A serious general counsel will tell you, flatly and correctly, that an indemnity capped at fees is barely an indemnity. If \$200,000 of fees has to stand in front of a multi-million-dollar infringement claim, the buyer is uninsured for the entire gap. No enterprise pays a vendor to deliver clean intellectual property and then accepts being uninsured on it. Lead with the cap and you have told the sharpest person in the room that you do not understand what the clause is for. You lose the negotiation that matters most before it begins.

▲ THE MOVE IS THE TRIGGER, NOT THE CAP.

The cap is *how much* the indemnity pays. The trigger is *what it covers in the first place* — and that is the part you can actually defend.

Narrow the trigger to the IP risk you create and control: your own original work, the code your people wrote, the design and architecture choices they made and could have made differently.

Then carve out, in plain and specific language, the inputs you did not create and cannot stand behind:

- The buyer's own materials and instructions — if they handed you the brand, the data, the spec, you do not warrant where those came from.
- The third-party components the buyer told you to use — the library they standardized on, the vendor they required.
- The raw output of a foundation model whose training data you did not assemble, and whose provenance no human can warrant.

That last carve-out is the honest one, and it is worth being honest about why. You cannot open a frontier model and check whether it reproduced a paragraph, a melody, or a few lines of someone's copyrighted code from the ocean of material it learned on. Neither can the company that built it — which is exactly why their own indemnities arrive hedged, conditioned, and carved.

So you do not promise what no one on earth can verify. You will not win this carve-out clean, and the fight is not easy. The buyer is paying for output it can use without fear, and a good GC will push back hard. Handing the model's IP risk back to the customer is precisely what the customer is trying to avoid. You will lose ground in it.

What you can hold, even when you cannot hold all of it, is the principle underneath. You warrant your *process* — that a human reviewed the work, that you kept a record of which models touched it and when, that you ran the controls you said you would. And you scope the indemnity to what that process actually covers, rather than to a guarantee about a model's training data that was never yours to give. The cap is a number

you will lose. The trigger is a position you can defend — and where the defense gives way, you have at least stopped writing a check against a fact you can never check.

The credit stack, the quiet one your counsel waves through

The third clause is the quiet one — the one your counsel is most likely to wave through, because each piece of it looks so reasonable. These are the **service-level credits**. A few points of the fee back if you miss a milestone date. A few more if the defect rate runs above the agreed line. A little more for an availability miss once the slice is live. Each is modest. Each is fair, taken alone.

But they do not arrive alone and they do not stay small, because credits stack. Miss the date, breach the defect rate, and slip on availability in the same rough quarter — not implausible on a hard build. Now the 5% and the 5% and the 10% become a fifth or a quarter of the fee handed back. Almost no contract counts those credits against the liability cap you just fought so hard to set. They live outside it: a second ceiling you never saw, an uncapped penalty wearing the costume of a service level.

This clause is turning more dangerous as the pricing model itself moves. The whole industry is sliding off the billed hour and onto the outcome — charging for the carry now, not the keystrokes — and that shift is exactly what loads the service level. John Cotterell, the CEO of Endava, put a number on how far it has traveled, on his company's November 2025 earnings call: "This quarter, we were 24% outcome-based, which is still rising. But it does mean that 76% of our revenue is coming in a T and M basis."

The contracts follow the pricing. As you move from that 76% toward that 24%, the buyer will expect you to carry the outcome — and the

service-level agreement is where carrying the outcome gets written down.

“ TWO MOVES CLOSE THE TRAP.

1. **Cap the credits in the aggregate, and tie that aggregate to the liability cap** — so the agreement has one ceiling, not a hidden one beneath the floor.
2. **Refuse any credit gated on something you do not control** — a third-party API’s uptime, the buyer’s own approval cycles, or an outcome no one can attribute cleanly. A promise of conversion-rate lift is the classic trap, when the buyer owns the pricing, the media spend, and the inventory.

That is the refuse-list again — written into the paper this time, instead of asserted across a table.

Win all three and you have done what almost every careful firm does. It is still not enough.

*

**

An illustrative case: signing carry you can’t hold

Consider an unnamed, mid-sized product agency — eighty people, a deep design bench, the kind of shop that wins on taste. They took an illustrative \$2.5 million build for a global retail brand, hungry for the logo in a thin year. They negotiated hard. But they conceded two things they had quietly filed under *remote*. They accepted an uncapped IP indemnity that swept in AI-generated assets, because they trusted their creative directors to rework everything a model produced. And, moving with the market toward outcome pricing, they accepted an SLA that tied a fifth of the fee to conversion-rate lift on the new checkout flow. They signed the carry.

Four months into the deployment, the remote risk arrived. A hero asset on the checkout page — heavily iterated through a frontier image model — turned out to carry a structural resemblance to a competitor’s

copyrighted global campaign. A cease-and-desist landed. The retail brand panicked and paused the entire \$10 million media buy pointed at the page. With the traffic gone and the layout reverted to a legacy fallback, the conversion rate fell through the floor.

The brand did not just ask for a fix. It invoked the clauses. It demanded indemnity for the legal costs of the copyright dispute. It demanded damages for the paused media spend as a direct consequence of the breach. And it demanded the SLA credit back, because conversion had dropped below the baseline. Three claims, from three clauses, in one letter.

The agency turned to its errors-and-omissions carrier — the insurer that covers professional mistakes. The carrier read the claim and pointed at the contractual-liability exclusion. The policy covered the firm's negligence — work performed below the professional standard of care. It did not cover a fee the firm chose to put at risk through an SLA. And it did not cover a liability assumed by contract that the firm would not have owed at law. The broad AI indemnity was exactly that kind of assumed liability. The carrier declined.

The agency then turned to its bank, where a large receivable tied to the engagement had slipped past due during the dispute. It fell out of the borrowing base — the pool of invoices the lender uses to size the credit line. The line tightened against its covenant; the bank called to ask what was going on. The indemnity — uncapped and, it turned out, uninsurable — came due in cash, on demand, in the same few weeks the carrier walked and the credit line froze. The founders personally guaranteed payroll to make it through the month. They settled for a number that wiped out their operating cash, laid off a third of the firm, and lost their equity in a distressed recapitalization.

They did not die because they were bad at software. They died because they signed carry they could not hold.

*

**

The older professions already solved this

The mature professions worked this out a long time ago, and the cleanest model sits one trade over, in structural engineering.

A structural practice designs a forty-story tower. It charges a premium for its judgment and its stamp, and it draws a ruthless, non-negotiable boundary around the carry. The engineer signs off on the load calculations. He carries the integration of the math, and if the building fails because the math was wrong, his policy pays. But he does not carry the steel mill that rolled substandard rebar. He does not carry the contractor who poured the concrete in the rain. And he absolutely does not carry the developer's risk that the tower never reaches 90% occupancy. If the steel was cheap, the supplier answers for it. If the building sits empty, the developer eats it. He stamps the load, not the pro forma — the developer's profit projection.

This is the discipline a software firm now has to adopt toward AI. You carry the integration. You carry the architectural math. You carry the verification — the part where a human stood behind the work. You do not carry the foundation model's training data, which is the steel you did not smelt. You do not carry the client's shifting market, which is the occupancy rate. Every clause fight in the last section was a single instance of holding that one line: charge for the load you actually engineered, and refuse to carry the conditions you cannot.

The analogy has a hard limit, and it is worth naming rather than leaning on. The engineer's carve-outs are not something he argues for deal by deal; they are codified — written into building codes, professional-liability statutes, and a century of precedent that already says the developer eats the occupancy and the mill answers for the rebar. A software firm has none of that scaffolding. Nobody has written the

code that says a vendor does not warrant a frontier model's training data. So the discipline is the same, but the protection is not inherited — you have to write it into the buyer's paper yourself, one clause at a time, against counsel paid to leave it out. The structural engineer shows you the shape of the boundary. The contract is where you have to build it.

*

**

The arithmetic that the failed agency skipped

You enforce the line by doing the arithmetic the failed agency never did. It is simple to describe and hard to act on.

Before you sign, total the worst case across every clause in the deal — not the likely case, the worst one, because the worst one is the only case a bad week will care about. Then set that total against the two numbers that decide whether a company lives through a bad engagement.

The first is the real sublimit on your E&O policy — the most the insurer will actually pay on this kind of claim. Not the reassuring number on the binder, the \$2 million aggregate, but the far smaller figure the carrier will pay on the *kind* of claim a contract like this creates. That can be a quarter of a million or less, and it excludes, by its own plain terms, the contractual promises and SLAs you volunteered to take on. The second is the headroom on your line of credit — the cushion the covenant leaves before one slipped receivable trips a default and hands the lender the keys to the conversation.

If the worst-case exposure you signed exceeds what your carrier will pay *and* what your lender will tolerate, you did not transfer that risk anywhere. You are self-insuring the gap, with money you do not have, and you did not decide to. The contract decided it for you, one reasonable-looking clause at a time, while you were busy winning each clause instead of summing them.

If the math does not fit inside both numbers, you do not have a sales problem or a pricing problem. You have one of two honest options: a clause to renegotiate, or a deal to walk.

And walking is the hardest move on this list. Walking away from a nearly \$2 million engagement in a thin year, with the team watching the number and the competitor down the street happy to take it, is the call most firms won't make. No framework makes that easy.

But the choice is not really between brave and timid. A firm that signs past what it can carry is not braver than one that walks; it is just placing a bet — that no client, on no engagement, ever has an outcome bad enough to call the exposure due. Across enough signatures, that bet pays exactly once, against you, on the deal you could least afford to lose. The worksheet that runs this math is an afternoon in a spreadsheet: three columns, your maximum exposure, your real insurance, your real credit headroom. The math takes an afternoon; acting on a bad answer is the hard part.

Write the boundary down before the redlines start

What holds that nerve when the pipeline is thin is taking the judgment out of the heat of the deal — fixing it on paper before the redlines start. Call it the Do-Not-Sign Carry Checklist — the boundary made literal, agreed by sales, delivery, and counsel in advance. It is the same discipline read from the other direction: each line names a carry the firm will only hold for a price, and refuses to hold for free. Each line is a hard stop. A clause that trips one goes back across the table as a commercial decision — a question of what the buyer will pay to transfer that risk — not down the hall to legal as a wording problem.

“ **THE DO-NOT-SIGN CARRY CHECKLIST.**

1. **Uncapped AI IP indemnity.** You cannot underwrite the copyright hygiene of a model you did not train; it is uninsurable carry. Indemnify your own original code and assets, hold the line on process, and disclaim warranty of non-infringement for raw third-party model output, which rides the provider’s terms.
2. **Uncapped data-breach liability on the buyer’s systems.** If you are working inside the client’s cloud and their data, their security posture is their carry. Cap breach liability at a defined multiple of fees and condition it on your failure to follow the agreed controls — not strict liability for any breach.
3. **Outcome SLAs on metrics you do not control.** Conversion lift, sales volume, retention — tying your fee to these when you do not own the pricing, the marketing, or the market is carrying the client’s business risk for free. Tie SLAs only to deterministic technical metrics: uptime, latency, defect-resolution time, delivery dates.
4. **Open-ended consequential damages.** A delayed launch that costs the client a season can run to many times your fee, and your E&O will not touch it. Insist on a mutual waiver of indirect, special, and consequential damages, lost profits and lost revenue named explicitly.
5. **Fitness-for-purpose on the buyer’s data.** If the system runs on the client’s data lake, the output is only as good as that data. Disclaim the implied warranty of fitness for a particular purpose, and state plainly that accuracy is contingent on the quality and legality of what the client hands you.

Print it. Make it the gate every order clears before signature.

When a client pushes back on a line, what they are really asking is for you to be their uncompensated insurer. So move the conversation from the legal column to the price column. *If you want us to carry that, the number is not \$1.8 million. It is far higher.* The conversation usually ends there, which tells you it was a pricing question all along.

*

**

What makes the redline safe

So go back, one last time, to the redline still bleeding on your screen. It is not dangerous because the buyer is hostile, and it is not safe because you named the slice well in the chapter before this one. It becomes safe when you have done three unglamorous things. You hung the slice off paper that already governs, so you are not fighting a war over the master you were always going to lose. You held “own the risk you control” through every clause a lawyer is paid to widen — pricing the carry you took and refusing the carry you would not, in the language of risk-transfer the buyer already buys. You won the carve-outs instead of the number, narrowed the indemnity’s trigger instead of capping it, and put one ceiling over the credit stack. And you weighed the whole of what you signed against what your carrier and your lender can actually carry — so that no run of ordinary bad luck can quietly add up to the end of the firm.

This is the lever doing its work. The carry was valuable the moment production got cheap; that much the last two chapters settled. It becomes *paid* only here — when the contract turns the judgment you carry into an obligation the buyer purchased on purpose, priced as the risk you agreed to hold. Value alone never moved a dollar. The clause is where it moves.

The market has this backward. The mood this year treats the threat as the technology — the model coming for the work, the margin, the future. But the firm most likely to die in the next two years will not be killed by a model at all. It will be killed by a clause. An indemnity it did not narrow, a credit stack it did not cap, a risk envelope it never totaled — signed in the warm relief of winning the work. The deal looked like salvation, right up until it became the thing that took the company down.

And there is one demand still sitting in that redline that no clause can settle, waiting directly underneath the AI indemnity you just fought to narrow. It is the buyer's quiet, reasonable, entirely uncontractable follow-up to everything you refused to warrant. Prove it. Prove which models touched this work. Prove a human stood behind the parts that mattered. Prove that the thing you carved out of your indemnity is genuinely outside your control.

That proof is not a contract term, and no redline will ever produce it. It is an artifact: something you build, and keep, and hand over. Learning to build it before the buyer demands it — and long before a court does — is the work of the next chapter.

SIGN WHAT YOU CAN CARRY IN 20 SECONDS

- The contract is the lever that makes the carry *paid*. It prices the carry as risk-transfer the buyer already buys — every clause a decision about how much risk you are signing for, and for what price.
- Don't bring your own paper. Hang the slice off a short *Slice Order* on the master the buyer already blessed.
- Three clauses decide it: the *cap* (fight the carve-outs, not the number), the *IP indemnity* (narrow the trigger, don't cap it), and *service-level credits* (one ceiling, not a hidden one).
- The illustrative failure didn't die of bad work. It died of signing carry it couldn't hold — uninsurable indemnity plus an outcome SLA.
- Like a structural engineer: stamp the load you engineered, refuse the steel you didn't smelt and the occupancy you don't control.
- Before signing, total the worst case against your real E&O sublimit and your real credit headroom. If it doesn't fit, renegotiate or walk.

→ **ON MONDAY**

1. Draft the *Slice Order* template — bounded result, the three terms the slice needs, everything else pointing back to the master.
2. On your live redlines, move the fight off the cap *number* and onto the carve-outs, and narrow the IP indemnity's trigger to your own original work.
3. Write the *Do-Not-Sign Carry Checklist* and get sales, delivery, and counsel to agree the five hard stops in advance.
4. Run the three-column worksheet on your largest open deal: worst-case exposure, real E&O sublimit, real credit headroom.

Sell the slice. Sign what you can carry. Then prove what you signed.

CHAPTER FIVE

Show Your Work

“We use AI” is a claim, and a claim sits in the proposal until someone asks you to back it up. Proof is how you back it up — and proof is one of the few levers that turn carry the buyer merely values into carry the buyer actually pays for. Carry the buyer cannot verify is carry the buyer will not pay for. This chapter is about the record that converts the claim into proof, and the proof into a defensible price.

IN THIS CHAPTER

- Why “we use AI” stopped being a differentiator the moment it went universal
- The five-field folder that answers the vendor questionnaire in an afternoon
- The same folder doing three jobs: defending the indemnity, clearing the audit, surviving a dispute

There is a sentence on the agency homepage now, and the firm cannot prove it. *AI-powered. AI-native. We use AI to ship faster.* It went up about eighteen months ago, because a competitor put it up first. It cost nothing to write and it won meetings. It is also a claim — a promise about how the work gets done. And the bill on a claim comes due the first day someone asks you to back it up.

That day is closer than it used to be, and it does not arrive as a lawsuit. It arrives as a line in a vendor security questionnaire. As a follow-up from the buyer’s risk team. As a clause in a renewal asking you to *describe your AI governance*. The trouble is not that the sentence is a lie. Assume it is perfectly true. The trouble is that a claim does not evaporate when the

pitch ends. It goes into the proposal, sometimes into the contract, and it waits. The day someone has reason to check, it stops being marketing and becomes a thing you have to prove. And the firm that wrote it usually cannot.

The two carries you can't see

Chapter 2 split the agency dollar in two: a production half AI is dragging toward zero, and the carry you answer for — judgment, integration, verification, accountability — the half the buyer still pays a premium for, because it still cannot get that anywhere cheaper. It also named the catch: the buyer can value the carry and still keep the savings, so the carry becomes margin only when you make it *paid*. Chapter 2 set out four levers for making it pay. Proof is the last of them. This chapter is the proof lever. It matters more than it looks, because two of the four carries cannot be seen without it.

Verification and accountability are peculiar that way. They are not visible in the work the way judgment and integration are. A buyer can see your judgment in the thing you made. He can feel your integration in the way it slots into his systems. He cannot see, just by looking, that a named person checked the model's output on the third of March and changed four things. He cannot see that you can account for every line of what shipped.

So verification and accountability are claims until you can show them. And carry the buyer cannot verify is carry the buyer will not pay for. Until you can show them, you are charging a premium for two of your four carries on good faith alone. You are hoping for margin you have given the buyer no record to grant.

Chapter 3 said sell the slice and own the risk you control. Chapter 4 said sign only what you can carry. It ended on the one demand a redline leaves behind that no clause can answer: *prove it*. This chapter is

how you answer. And why the answer turns out to be worth more than the claim ever was. “We use AI” is a statement about the production half, the half going to zero. “Here is exactly what our AI did, what our people did, and how we know” is something else. It is proof of carry — the lever that converts valued judgment into billable, defensible carry. It is the only form in which verification and accountability can be seen at all.

*
**

“We use AI” stopped being a moat

“We use AI” was a moat for about a year. It isn’t now, because everyone says it.

David Droga, who runs Accenture Song, put the general law plainly at Cannes in June 2024: “when everybody is doing best practices, nobody is doing it”. The moment a capability becomes universal it stops setting anyone apart, and the edge slides to the next scarce thing. The capacity to use AI went universal sometime in the last eighteen months. What is still scarce — what almost no one can do on demand — is prove what their AI did. The differentiator already moved.

The four-function room from the earlier chapters shows what the claim actually does to each seat. The chief financial officer does not pay a premium because you use AI. He benchmarks you harder, because the same tools are why his own team believes it can do the work without you. Your claim is evidence for his cost-cutting, not yours. The Vendor Management Office hears it and adds a row to the questionnaire. Third-Party Risk reaches for the disclosure form — which models, trained on whose data, and does the indemnity survive an AI tool sitting in the workflow. And Model Risk Management, at the regulated buyer, does not hear a feature at all. It hears a governance event, a new thing that has to be assessed before anyone can sign.

So the claim meant to mark you as modern instead marks you, to all four of them, as a fresh category of risk to evaluate. That is the opposite of standing out. It is closer to volunteering for an audit.

And these are not naive readers. Procurement and risk are the most sophisticated buyers of carry in the building. They have spent decades paying outside firms to absorb consequence — buying indemnity, service-level guarantees, assurance reports, managed-service warranties. They know precisely how to pay for a carried risk; it is most of what they do. So they do not need the value of your verification explained to them. They need it evidenced, in the forms they already purchase. The firm that treats them as people who “haven’t learned to name it yet” has the relationship backwards. They named it long ago. They are simply waiting for a seller fluent enough to hand them the proof.

The claim outruns the evidence

That is the harmless half. The half that ends companies is that a claim about how you work has a long memory, and almost no agency keeps the evidence to back it.

You said a human reviewed the model’s output — which human, on which day, reviewing what, and changing what? You said the model drafted and your people supplied the judgment — can you draw that line after the fact, on a specific deliverable, a year later? You said the work was original — can you prove the model did not hand you, buried in a thousand plausible lines, a stranger’s copyrighted code or a paragraph lifted whole from its training data?

For most firms the honest answer to all three is no. Not because they lied — because they never built the habit of writing it down as it happened. The claim outran the evidence. The gap between them is a liability with no line item on the balance sheet, invisible right up until the day it is the only thing anyone wants to discuss.

▲ YOU DO NOT HAVE TO COMMIT FRAUD FOR THIS TO HURT.

You only have to be unable to show your work the day someone asks. The carry was real — your people did review it, judgment did happen. But a carry you cannot evidence is, to the people who matter at renewal, indistinguishable from a carry that never happened. Work above the line is worth nothing at the table if it cannot be shown above the line.

**

Builder.ai: a claim no one could produce

Builder.ai is the cautionary version, and the details matter.

There are two stories about Builder.ai. The one the internet believed: that hundreds of engineers secretly hand-wrote code while the company pretended a machine did it. That one is disputed — careful reporting pushed back hard on it — so treat it as unproven. The story that holds up: Builder.ai sold itself for years on an AI product called Natasha, which by the best reporting was a real AI system. It also ran a large network of outsourced human engineers, hundreds of contractors across India, Vietnam, Romania, Ukraine, and Poland. In May 2025 it entered insolvency after its lender, Viola Credit, swept roughly \$37 million from its accounts. Bloomberg reported it had allegedly inflated sales through round-tripping with an Indian firm, VerSe Innovation — which VerSe flatly denies — and federal prosecutors in Manhattan opened a criminal investigation into its financial reporting.

Set the fraud allegations aside; they are Builder.ai's own, and not your concern. The detail that is your concern is what happened to the *AI claim* once scrutiny arrived: no one could cleanly say how much of the AI was real. Not the critics, not the defenders, not, by all appearances, the company itself. The line between what the machine did and what

the people did had never been written down in any form that could survive a hostile reading.

There may well have been a defensible answer in there. Real AI doing real work alongside real humans is not a scandal — it is an ordinary Tuesday at a modern shop. But a true answer you cannot produce on demand is, under hostile reading, no answer at all. The claim could not defend itself. With no record, the loudest interpretation won.

That is the mechanism, and it has nothing to do with fraud. A claim about how the work got done, with no contemporaneous record underneath it, does not survive scrutiny even when it is true. Once the reading turns hostile, the absence of evidence reads exactly like the absence of the thing. It takes no fraud probe to trigger that. It takes a buyer's risk team, a soured engagement, or a discovery request — and a folder that turns out to be empty.

*

**

The smaller version that happens to a firm your size

The likelier version has no fraud and no headlines, and it lands on a firm your size.

A hundred-person agency wins a build at a mid-market company that happens to be owned by a regulated parent — a bank, an insurer, a health system. The work goes well. Everyone is happy. Then the renewal comes up, and with it something that was not there two years ago: a vendor security review. It is the same four-function scrutiny the big institutions run, now trickling down to the mid-market because the parent's risk team made it policy for every supplier.

The agency's day-to-day contact — the product owner who loved the work — is nowhere near the room for this part. It has gone to a procurement officer the agency has never met, and a third-party risk

analyst who has never seen the product. A demo would not move her if she had. Her job is not to like the work. Her job is to check boxes she can defend to an auditor above her, and an empty box is a finding whether or not the underlying work was excellent.

The questionnaire is unremarkable and merciless. Attach your AI usage policy. List the models you use and the versions. Provide evidence of human review on AI-assisted deliverables. Attest to how client data is handled and whether any of it touches a training process. Each line wants an artifact, not an assurance — a document, a log, a dated record. “Yes, a person always reviews it” is not an answer to “provide evidence of human review.” It is the thing the evidence is supposed to support, offered in its place.

▲ REAL-AND-UNRECORDED AND NEVER-HAPPENED PRODUCE THE IDENTICAL PDF.

The competitor down the street has *AI-native* across its homepage in a font twice the size of yours. Behind the slogan there is nothing: no policy, no inventory, no logs, because they never built any.

Their carry may even have been real — they may have reviewed every line. But when the analyst opens the folder, real-and-unrecorded and never-happened look the same, which is to say they produce no PDF at all. The review stalls.

So the competitor’s review stalls. The procurement officer sends the questionnaire back with half the fields flagged. They scramble to reconstruct, after the fact, who reviewed what across a year of work. And reconstruction is exactly the thing a risk analyst is trained to distrust. A record built the week of the audit proves the week of the audit and nothing before it. The legal back-and-forth drags into a second month. Eventually the risk officer writes the three words that end it — *unable to evidence controls* — and the renewal quietly routes elsewhere.

The agency that kept the records answers the same questionnaire in an afternoon. It attaches a folder it has been filling the whole time. Inside: a one-page AI usage policy. A model-and-version inventory that updates itself out of the project tracker. A sampling of dated human-review entries pulled straight from the pull-request history. A data-handling attestation it wrote once and reuses. The analyst opens it and finds a box for every line on her form. She signs off — not because she was charmed, but because she now has something for her own file when her own auditor asks why she approved the vendor.

That is the quiet thing the folder does. It does not just answer the buyer’s risk team; it arms them. It gives the one person whose approval you need the cover she needs to give it. The agency signs for another year. No catastrophe, no tabloid — just a deal won on paper that the other firm lost on the same paper. In that review the word “AI” was never spoken aloud as praise or accusation. It came up only as a thing to be evidenced, which one firm could and the other could not.

*
**

The folder is five fields, not a governance program

That folder is the whole chapter, and it is not a six-month governance program or a compliance hire. The minimum viable version is five fields, logged as the work happens, in the tools already open on your screen. Nothing here is a new system. Each field is a home for a fact you are already generating and currently throwing away.

1. Which model and version, and when

Which model and version touched each deliverable, and when. This lives as a column in the project tracker — the same board where the ticket already moves from “in progress” to “done.” Not “we use AI” but “Claude 4.x drafted this module on March 3; the layout came from

an internal tool on the eleventh.” Specific enough that a stranger could reconstruct it. The person who owns it is the person who moves the ticket. It is one field, filled at the moment they would have updated the status anyway.

2. What went in

The data and inputs, with a one-line attestation of what you are and are not permitted to use. This lives in the delivery doc, one line near the top, written once per engagement and rarely changed. The buyer’s biggest fear is that their data trained somebody’s model. One honest line retires it: *no client data was used to train any model; inputs were limited to the materials provided under the MSA*. The account lead owns it. It tracks what the contract already permits.

3. Who reviewed it, and what they changed

The human in the loop, named and dated. This lives where review already happens — the pull-request approval in a code shop, the sign-off field in the delivery workflow in a creative one. You are not adding a review; you are recording the one you already do. The reviewer owns it by the act of approving.

This is the field that converts “a person was accountable” from a claim into a record. Accountability, recall, is one of the two carries that is otherwise invisible. This is where it becomes visible.

4. What you asked and what you kept

The prompt-and-decision trail: what the model produced, what you rejected, what you shipped. This lives in a short section of the pull-request template or the delivery doc — three lines, not a transcript. It is the evidence that judgment happened, not just generation. The maker owns it, at the moment they decide what to keep. Anyone can generate; the kept-and-rejected trail is the residue of someone choosing, which is what puts the work above the line.

5. The license and copyright pass-through

Where each component and asset came from, and what rights came with it, so the buyer knows exactly what they are allowed to do with what you delivered. This lives in the delivery doc as a manifest line per third-party component or generated asset. The maker owns it on the deliverable that ships.

Own it where the work already lives — the engineer who commits the code, the designer who delivers the file — not in a new department that meets on Thursdays. A team that meets on Thursdays produces minutes; the engineer who fills one field at commit produces evidence. The record is made at the same instant as the work and cannot drift from it. The firms that get this right start this week, in the tool already in front of them, on the next deliverable that ships. Not a backfill of the last year, which fools no one — the next thing out the door, logged as it goes. One deliverable, five fields, the tools already open. That is the entire ask. And it is the whole distance between answering the questionnaire in an afternoon and losing the renewal you thought you had won.

? “A LEAN SHOP HAS NO ROOM FOR OVERHEAD, AND THIS LOOKS LIKE OVERHEAD.”

It is overhead exactly backwards. The firm too small to absorb a stalled two-month renewal — or to spare the senior people who would otherwise spend a week reconstructing a year of history after a dispute — is precisely the firm that cannot afford to skip the record.

The real bloat is the alternative: the policy deck nobody reads, the governance committee that meets after the work is done and produces minutes instead of evidence. This is the opposite of that. It is the smallest record that preserves the truth, made at the moment the truth is still free to capture. After a couple of weeks it’s just part of shipping.

*

**

Creative shops are not exempt

This is not only a problem for code shops. If anything the exposure is sharper for a creative agency. The copyright question on a generated image or a block of AI-assisted copy is hotter and less settled than it is on a function.

For a brand, design, or content shop the same folder has a name the earlier chapters already gave it — **Generative Provenance**. The fields rhyme, and live in the same tools a creative shop already has open. Which generative model made or assisted each asset, logged on the asset card in the project tool where the work is tracked. An attestation about the data and the licensing behind that model, written once in the statement of work. The copyright chain that tells the client whether they can safely run the work in a paid campaign, kept as a manifest line beside each delivered asset. And the human-review log — the art director's sign-off, dated — that shows where the art direction and the editing actually happened, recorded at the approval step that already exists in the studio's workflow.

The code shop proves its build. The creative shop proves its originality. Both are answering the same question from the same nervous buyer — *can I use this without it coming back on me* — and both answer it with a record or not at all.

The standards are converging on exactly this

The clock behind the buyer's question is real, not a someday. And a light word of caution, because the moment we touch this material we are near the law. What follows is strategic orientation, not legal or compliance advice. The person who signs your attestations should see counsel.

In code, the discipline is hardening into a standard. The software bill of materials — the SBOM — is growing its model-inclusive successor, the

AIBOM. And CISA has published minimum elements for an SBOM-for-AI, built on shared work among the major economies.

In Europe, the AI Act's heavy obligations on high-risk systems — the documentation, the records, the demonstrable governance — were first written for August 2026. Then, in the simplification drive Brussels called the Digital Omnibus, the EU agreed to defer them, pushing the core high-risk requirements into 2027 and 2028. The lesson outlasts the date. The regulator moved its own deadline, and the buyer's questionnaire did not move with it. Brussels can grant itself an extension; the procurement officer cannot, and is asking now.

There is, alongside it, a certifiable international standard for managing AI — ISO/IEC 42001. A sophisticated buyer can require it the way it already requires its security cousin, and check the box rather than read your essay.

All three have the same thing in common. They do not ask whether you use AI. They ask you to document what it did. The standards are converging on exactly the discipline the five fields produce. So the firm that starts the folder now answers more than this year's questionnaire. It is pre-building the artifact next year's standard will demand. The question is not going away; it is going to multiply, and arrive earlier in the deal each year.

*

**

One folder, three returns

The folder is worth far more than the afternoons it costs, because the same artifact does three jobs — and you only ever build it for one of them. One discipline, three returns.

It makes the indemnity defensible

The job nobody expects comes first, because it reaches back into the last chapter and quietly doubles the value of work you already did.

It is the indemnity you narrowed in Chapter 4. There you refused to warrant the raw output of a foundation model you could not vouch for, and held the line that you would stand behind your *process* instead. That carve-out was a good move on paper. But on the day it is tested, it is worth exactly as much as your ability to prove the boundary was real and not a convenient fiction.

A carve-out you cannot evidence looks, to opposing counsel, precisely like a firm inventing a boundary after the fact to duck a warranty it always meant to give. Provenance is the difference. It shows, with dated records, precisely where your controllable work ended and the model's uncontrollable output began. The prompt you wrote, the output you got, the human who checked it, the version that shipped. That is the difference between a principled boundary a court will respect and a lawyer's dodge it will tear apart in an afternoon. The clause you signed in Chapter 4 is only ever as strong as the folder you build in Chapter 5. They were never two problems; they were the same problem, addressed at two desks.

It clears the audit

The second job is the one you just watched win a renewal: it clears the room. When Third-Party Risk sends the questionnaire and Model Risk Management asks what happens when the model is wrong, you do not compose a careful essay and hope. You hand over the folder. The four-function gauntlet that stalls your competitor becomes the thing that advances you. The audit you used to dread becomes the audit you quietly want, because it is the one part of the bake-off your sloppier rivals cannot fake. Judgment they can claim. Integration they can claim. But verification and accountability they either logged as they worked or they did not, and the folder is the only place the difference shows.

It is your defense if you are ever tested

The third job you hope never to need and cannot afford to be without. It is your defense the day the claim itself is read back to you in the worst possible light — a dispute, a discovery request, a deal gone bad and lawyered. The firm that can show its work is the firm whose claims survive the reading. “Here is exactly what ran, who reviewed it, and when” is a sentence that ends an argument. “No one can really say” is the sentence that ended Builder.ai’s.

You build it once, for the questionnaire. It then makes the indemnity you already signed defensible, clears the audit that stalls your rivals, and stands as your evidence if you are ever tested. Three returns on one discipline — and you would have paid the full cost of the folder for any one of them alone. That is why the math is not close.

And somewhere in the doing of it, the sentence on the homepage quietly turns inside out. “We use AI” was a claim: cheap, universal, indistinguishable from every competitor’s, and a liability the instant anyone leaned on it. “Here is precisely what our AI did, what our people did, and how we know” is not a claim at all but evidence. And it is the only part of the whole pitch a competitor cannot copy off your website overnight. It’s evidence of a year of actual work, which is why it can’t be copied.

*

**

Show your work

So the sentence stays on the homepage. But in a market where the same line sits on every rival’s site, the *using* of AI has stopped being the thing that sets you apart. Everyone uses it; everyone says so; the slide is the same slide. The differentiator has moved — the way it always moves — once a capability becomes common, the way Droga said it moves — from having the thing to proving the thing.

And proving it is not a marketing act so much as the act of making two of your four carries visible — and a carry the buyer cannot see is a carry the buyer will not pay for. Judgment and integration show up in the work. Verification and accountability do not — they live or die in the record. Proof is the lever that turns them from value you assert into carry the buyer will fund. The folder is where they become legible to the only people who decide whether you get paid to carry them again next year. The agencies that win the next two years will not be the ones with the boldest claim on the site. They'll be the ones who can produce the record on request.

That is the arc of this book, compressed into a single working week. You sold a slice the buyer could grasp. You signed only what you could carry. And you can prove what you did — which is what lets you defend the price, honor the indemnity, and win the audit, all from one set of records kept as the work happened. Be legible. Own what you control. Show your work.

There is one thing proof cannot do, and it is the thing that decides whether the rest of it matters: it cannot pay you next quarter. A moat is not a paycheck. The firm doing every bit of this correctly — selling the slice, signing only its carry, logging every field — is still living in a gap. The old model is shrinking; the new one has not yet scaled. That gap is not measured in strategy or craft or provenance; it is measured in cash, on a calendar, against a payroll. Surviving that gap is the next chapter.

SHOW YOUR WORK IN 20 SECONDS

- “We use AI” is universal now, so it differentiates no one — and it sits in the proposal as a claim you may be asked to prove.
- Verification and accountability are two of your four carries, and they are invisible until you record them — and carry the buyer can’t verify is carry the buyer won’t pay for.
- Proof is one of the levers that makes carry paid; procurement already buys carried risk fluently, so the job is to evidence it in the forms they purchase, not to explain its value.
- A claim with no contemporaneous record does not survive a hostile reading, even when it’s true — that is what collapsed under Builder.ai.
- The whole answer is five fields, logged as the work happens, in the tools already open: model-and-version, inputs, reviewer, prompt-and-decision trail, license chain.
- The same folder does three jobs — defends the indemnity, clears the audit, survives a dispute — and you only build it once.
- Creative shops aren’t exempt: same folder, named Generative Provenance.

→ ON MONDAY

1. Pick the next deliverable shipping this week — not a backfill — and add the five fields to the tool it already lives in.
2. Write the one-line data attestation once, and reuse it: no client data used to train any model, inputs limited to materials provided under the MSA.
3. Put model-and-version logging in the project tracker, owned by whoever moves the ticket.
4. Pull a sample of human-review entries from your pull-request history, and confirm you could attach them to a questionnaire tomorrow.

Sell the slice. Sign what you can carry. Show your work.

CHAPTER SIX

Survive the Gap

You can get the whole strategy right — reprice for the carry, package it, sell the slice — and still run out of cash before the margin lands. This chapter is about that gap: the cash cost of moving to carry-based pricing, where it comes from, and how to spot it in time. You survive it only if the carry is actually getting paid on the far side.

IN THIS CHAPTER

- Why a full pipeline can feel like safety when it isn't
- The three shapes the cash gap takes, and how to tell which one fits
- The move that tends to come too late: the lender call, before the breach

Most of this book is about getting the repricing right. This chapter is about a quieter risk: getting it right and running out of cash before the new model pays off.

It happens because repricing changes when you get paid, not what you spend. You staff and pay the team now. You collect when the outcome lands, a quarter or two later. The monthly dashboard doesn't show the squeeze, because it averages — and the danger sits in a single week, when payroll clears before the receivable meant to cover it.

Going through the public record, what struck me was how few of the failures were really about strategy. Most were about timing — good firms, doing the right things, a quarter too early for the cash. So even a book about strategy has to stop on the cash for a chapter.

The rest of it holds up, as far as I can tell. The dollar is bifurcating; the slice is how you reprice it; the contract is how you hold the price; the provenance is how you prove the work. None of that changes the sequencing underneath: the new way of working gets paid later than the old way stops getting paid. That distance, measured in weeks of payroll, is the gap.

Where the gap comes from

This is the repricing squeeze from Chapter 1, seen from the inside. Not a line on a chart — a specific Friday, when the payroll run is larger than the bank balance and the money to cover it doesn't land until the quarter after next.

The gap is the cash-shadow of repricing toward the carry. Chapter 2 split the work in two — below the line, the production a model now does for almost nothing; above it, the carry the buyer still pays for: judgment, integration, verification, accountability — and warned that the buyer can keep the savings unless the carry is made *paid*, packaged and proven and written into the contract. That move is right. The gap is its cash cost. Shift the billing from the hour to the outcome and the spending doesn't change; only the timing of the collection does. Payroll stays on the fifteenth. The revenue slides into next quarter — if it arrives at all. The strategy and the cash trough are the same act, seen from two sides. You fund the trough on a bet: that the carry pays. If you never made it pay, you funded the trough for nothing.

▲ EACH RIGHT MOVE PUSHES THE MONEY OUT.

Sell the slice, and revenue waits on a bounded outcome instead of a monthly hour. Sign only what you can carry, and the bank may balk at the new shape of the receivable. Show your work, and the proof that wins the renewal still collects on the renewal's calendar, not this Friday's.

A full pipeline can feel like safety

The instinct under that pressure is to look at the pipeline for reassurance, and it is the one place that won't help. A full pipeline is future cash; it makes none of this month's payroll. It can even make things worse, because every signed deal is more delivered work to fund before any of it pays. The healthier the bookings look, the deeper the trough being funded to deliver them.

Here is roughly how it arrives. It is a Thursday; payroll runs tomorrow. The pipeline is the best it has been in two years, the team is busy, the work is good, and a deal in late-stage legal is already treated as closed. The founder opens the banking app out of habit — and the balance is under the payroll file. Not by a catastrophe; by a sliver. A \$300,000 invoice that should have cleared Tuesday is stuck in a client's accounts-payable queue, behind an approval step nobody mentioned and a controller who is out until Monday. The money is owed and on its way. It just isn't in the account on payroll day, which is the only date that matters.

The monthly reports weren't wrong, just monthly. They averaged, and the squeeze lived in a single week.

It happens to public companies too

Chapter 1 met Endava on the income statement: a £364.6 million impairment, revenue down 8.4%, an adjusted pre-tax margin falling from 12.6% to 1.8% — in the same quarter its AI-driven revenue tripled to 15% of the business. From the cash side, those numbers are this whole chapter. The new line was growing exactly as the strategy promised, and the firm was squeezed nearly flat anyway. Management told the market why: its large, complex outcome-based contracts were taking longer to execute than planned, and cash was converting more slowly than planned.

That is the mechanism, in the careful language of an earnings call. Paid for an outcome instead of an hour, you staff the team and run payroll now, and you collect when the outcome arrives — quarters later. The migration the whole industry is being urged to make is the same one that stretches the distance between spending and collecting. Endava has a public balance sheet to absorb that distance. A hundred-person shop has ninety days of runway and a payroll on the fifteenth.

When the lender moves first

Madwell, a Brooklyn creative agency that had been around since 2010, turned off the lights at 11:59 p.m. on April 30, 2025. The CEO emailed staff that evening; roughly eighty people lost their jobs; vendors went unpaid. It didn't die of a bad pivot or a lost bake-off — there is no rival in the story who made better work for less. It died of a loan. It had defaulted on a Bank of America facility — by court documents reviewed by Adweek, it owed more than \$4 million. The bank had moved to seize its assets about three weeks before the lights went off.

The order matters. The seizure came first, the shutdown second, and the market never got a vote. That is the real shape of trouble in the gap: not a competitor who beat you on the work, but a covenant breached while the work was still good. The creditor who held that loan stopped being a partner the moment the breach landed, and became a liquidator.

Three shapes of the same hole — and one will fit

The gap doesn't arrive the same way for everyone. Each billing model fails in cash in its own predictable shape, and three of them cover most agencies. Knowing which one fits tells you what to model, what breaks first, and what to ask for in the call this chapter ends on.

1. The milestone trough

With engineer-months billed against a master agreement, revenue arrives on the first, as reliably as rent. Then comes the move from “What Still Gets Paid For” — pricing the carry instead of the hour — and the conversion to milestone billing. The revenue isn’t lost; it just moves later, while the costs stay where they were — every two weeks, per head. It comes back bigger only if the carry was packaged to be paid. Otherwise the timing slips and no margin arrives to justify the wait.

► COUNT THE PAYDAYS IN THE GAP.

Lay the real payroll dates on a calendar. Mark the last old-model collection and the first new-model one.

- Last monthly invoice clears: early April
- First milestone signs off: late June
- Fortnightly payrolls in between: five

Each of those five funds work already delivered and not yet billable — drawn against a reserve sized for a business that got paid on the first.

That business no longer exists. Five payrolls times the fortnightly burn is the trough — the number to carry into the lender call.

2. The acceptance-window subtraction

For fixed-price work, the gap hides inside a clause that rarely gets negotiated hard. A \$2 million build with a thirty-day acceptance term looks like a thirty-day receivable. Let acceptance stretch to seventy-five days — a stakeholder on vacation, a reorg across the table, one more round of “just a few tweaks.” Now the receivable ages past ninety days, the line where the lender stops counting it.

► **HOW ONE SLOW INVOICE SHRINKS THE WHOLE CREDIT LINE.**

A credit line secured by invoices only lends against the *eligible* ones — invoices under ninety days old. That pool is the **borrowing base**.

- Past ninety days, the invoice drops out of the base.
- **Cross-aging:** once enough of one client's balance is overdue, the bank can disqualify that client's *other* invoices too — including the fresh, current ones earmarked for payroll.
- So one stretched window does not just freeze the \$2 million it delayed. It can knock that client's entire balance out of the base — and the credit shrinks at the exact moment it's needed.

That can trip a loan covenant no one remembered, which turns a slow invoice into a technical default on the whole relationship.

The way to see it coming is to age the invoices forward, not backward. Start with the three or four largest open invoices; push each to its worst plausible acceptance date; see what crosses ninety days. Then add every other invoice from those same clients, because cross-aging takes them too. What falls out of the base is the real exposure.

3. The outcome-retainer leak

For retainers tied to an outcome — a KPI, a pipeline target, a share of revenue — the failure mode is the dispute that never quite closes. The attribution fight from the earlier chapters isn't a one-time haircut; it is closer to a permanent tax on collection. The \$80,000-a-month that settles to \$75,000, after the client's CFO re-attributes the quarter's win to a platform change, is the small version. The structural version is worse: you book revenue every month that you can't reliably collect, in a domain that never lets the number be proven cleanly. Forecast and bank balance drift further apart each quarter.

► FIND THE LEAK.

Take the last four quarters. Overlay booked outcome payments against what actually cleared the bank. The delta is the leak. Project it forward at the same rate.

This one isn't a cliff. It is a slow leak that is easy to mistake for a rounding error — until the year the roundings add up to a quarter of payroll.

? “I WATCH THE AGING REPORT, AND THE CLIENT BASE IS SPREAD. ISN'T THAT COVERED?”

Not in time, unfortunately.

An aging report is a rear-view mirror. By the day an invoice shows ninety-one days, it has already dropped out of the borrowing base; it reports a wound, not a risk.

And concentration isn't something you can diversify in the quarter the gap opens. The anchor client who is 40% of the book took four years to land and will take longer than the runway to replace.

Both are real disciplines. They just run on the wrong clock — they prevent next year's gap, not this one's.

How distressed operators find the date

There is a way to find this date that tends to get reached for too late — the one distressed operators use. It is a thirteen-week cash-flow forecast: not revenue and margin, but cash in and cash out, week by week. Thirteen weeks because that is the horizon you can still steer inside; weekly because a monthly view averages away the exact Fridays that decide it. The Thursday founder wasn't careless. He was looking at a cadence that hid the only column that mattered.

► **THE THIRTEEN-WEEK FORECAST — FIVE COLUMNS.**

- **Week-ending date.**
- **Cash in** — every receipt at its *realistic clearance date*, not its invoice date.
- **Cash out** — every payroll (flag the months with three Fridays), plus rent, taxes, benefits, insurance, vendors, interest.
- **Net change** for the week.
- **Running balance** — colored red the first week it dips below next Friday's payroll.
That red week is the date.

▲ **THE FIRST VERSION IS ALMOST ALWAYS TOO KIND — WHICH IS MORE DANGEROUS THAN TOO HARSH.**

It is easy to forget the months with three Fridays, so two payrolls quietly become three. Easy to leave out the quarterly tax remittance and the annual insurance renewal. Easy to date the big receivable to its due date rather than to the fifteen-days-late the client has actually paid every quarter for three years — because the contract is easier to believe than the pattern.

Each of those is a week of runway that was never really there. The fix is to date every line to behavior, not to terms.

Most of what feels like bridging is theater

Once the shape is visible, the work is to bridge it — and a lot of what feels like bridging is theater. The first reflex under cash pressure is to cut, which is half right and easy to get wrong. Cutting the senior bench funds this quarter by selling next year. That bench is the rescue-and-judgment talent — the whole carry this book has been building toward — so losing it balances the forecast by dismantling the reason anyone paid a premium. The second reflex is worse: chasing a big, desperate deal to fill the hole. If that deal is a milestone build with a long acceptance window, it doesn't bridge the gap so much as deepen it, while

looking like a fix in the all-hands. Growth that consumes cash isn't a bridge; it is the gap, accelerating.

The call that tends to come too late

What actually buys time is unglamorous, and it tends to come too late: the lender call, before the covenant breaks rather than after. The lender call usually gets treated as the last call, the confession. It works better as the first — because it is really two different relationships wearing the same logo. A lender called at ninety days of runway, with a forecast in hand, is holding a performing loan from a borrower with a plan. A lender surprised by the breach is holding a defaulted loan with nothing left to protect but its collateral — and a reason to grab it first. That is the bank that moved on Madwell. Same situation, different week; the only thing that changed is who called first.

“ THE LENDER CALL — IN ORDER.

1. **Send the forecast two days ahead**, with a one-page note: exactly what's needed, and why the numbers close.
2. **Name the shape of the gap plainly** — *“a milestone trough of five payrolls,”* or *“an acceptance subtraction of this much dropping out of the base on these dates,”* or *“a retainer leak running at this rate.”* A lender can underwrite a named, bounded problem. It can't underwrite a vague one.
3. **Ask for the smallest intervention that buys the most time**, matched to the shape:
 - a covenant waiver through the dry months;
 - a temporary overadvance against the specific milestone that clears in June;
 - a borrowing-base re-cut that steps down with the milestones instead of off a ninety-day cliff;
 - tighter weekly reporting, traded for breathing room.

None of that is charity. It is a lender protecting a performing loan — worth more than the desks at auction.

The rest is mechanical and mostly familiar: invoicing the day a milestone is met rather than at month-end, billing in smaller and more frequent increments, taking deposits, tightening the acceptance language, cutting the overhead that isn't the future while protecting the bench that is. All of it helps. But the call comes first, because it is the one move whose value collapses the longer it waits.

If the arithmetic still won't close

If the numbers won't close no matter how the cash gets pulled forward, the remaining choices are still real, and still far better made early than late. A bridge is expensive money that buys a quarter to finish what was started. A sale or recapitalization to a stronger firm works while there is still a book worth buying, rather than a fire sale of desks and a client list that scatters the day the news leaks. An orderly wind-down at least pays the people and the vendors on the way out — the difference between Madwell's vendors and a version of that night where they were made whole. Six months of runway keeps all three on the table. Six weeks leaves only the last, taken on someone else's terms.

None of this is a clever instrument or a restructuring trick. Mostly it is deciding to look at the run-out date early enough to still have options. *(This is operating guidance, not legal or financial advice — but the decision is yours regardless of who you hire to execute it.)*

Be right on a calendar you can survive

In the end it comes down to one number — not revenue or pipeline, but the date. The rest of the book runs backward into it: sell the slice so the revenue is legible, sign only what you can carry, show your work, and watch the calendar so the gap doesn't close overhead before you cross it. The gap isn't a sign the strategy was wrong, so much as the price of getting it right early — the cash-shadow any firm throws when it reprices for the carry instead of the hour. You cross it only if the

carry is paid on the other side. Bet that value alone becomes margin, and the trough has no far edge.

The agencies that don't make it through this transition mostly won't be the ones with the worst strategy. They will be the ones with the right strategy and the wrong runway — a calendar that ran out before the correct things could pay. Right strategy on the wrong runway still fails. Endava had a balance sheet and still bled into the gap; Madwell had fifteen years and a loan, and the loan won.

There is a non-financial side too. A runway is also a room full of people who can read it — they see the same short calendar you see, they are deciding every week whether to stay, and their leaving widens the gap and worsens the work at the worst possible moment. The senior bench you protected from the layoffs can still walk out on its own. So crossing the gap turns out to be as much about who stays — and how they are led through a crisis everyone in the building can see — as about the cash. That is the last thing this book has to work through.

THE GAP IN 20 SECONDS

- The number that matters most here is the *date the cash runs out* — not revenue, pipeline, or margin.
- The gap is the cash cost of repricing toward the carry: you spend now and collect later, so every right move widens it before it closes. You cross it only if the carry actually gets paid on the far side.
- A full pipeline is future cash; it makes none of this month's payroll.
- Three shapes — the milestone *trough*, the acceptance-window *subtraction*, the retainer *leak*. One will fit. Model it on a calendar, not a spreadsheet.
- The thirteen-week cash-flow forecast turns the dread into a date: the first red week.
- The lender call lands far better *before* the breach. Same bank, two different weeks.

→ **ON MONDAY**

1. Build the thirteen-week forecast — five columns, dated to *behavior*, not terms. Find the red week.
2. Name the gap's shape — trough, subtraction, or leak — and put a number on it.
3. Move the acceptance clause to a hard fifteen-business-day window, with automatic acceptance if no written objection lands.
4. Put the lender call on the calendar at ninety days of runway, not nine.

Sell the slice. Sign what you can carry. Show your work. Survive the gap.

CHAPTER SEVEN

Keep the Room

You can sell the slice, sign what you can carry, show your work, and survive the gap — and still lose, because the capability that does all four lives in a few people who can leave. This chapter is about keeping them.

IN THIS CHAPTER

- Why the public record is so incoherent — and what your best people do with it
- The reshape: four roles the new work splits into, and the one that quietly rots
- Leading and selling through it — true/plan/ask/by-when, and the four-function room

Everything else in this book has been about the work. This last chapter is about the people who carry it — because that turns out to be the same thing.

There is a scene that is playing out on weeknights all over the industry. Your best person is reading the same headlines you are. Not as an employee, but as someone with a mortgage, a reputation, a résumé, and a set of skills worth more in one version of the future and less in another.

The headlines do not agree with each other. In one tab, the chief executive of a company many times your size says replacing junior people with AI is roughly one of the dumbest things he has ever heard. In the next, another chief executive says — with something close to pride — that his AI agents have taken over the work of about four thousand support staff, and that he simply needs fewer heads. A third

company cuts fourteen thousand jobs; its head of HR ties the decision to AI, and then, days later, its own CEO says it was not really about AI at all. It was about culture. A Stanford paper reports that early-career workers in the most exposed occupations have already watched their employment fall by about 16%. And the founder of the most famous AI company in the world, who spent something like two years warning of a white-collar apocalypse, now says he is delighted to have been wrong. Your best person reads all of it and lands on the only conclusion the evidence supports: nobody knows. Not the people with the most data. Not the people with the most to gain from sounding certain. So this person — sober, good, the reason three of your hardest accounts have not churned — does the sensible thing a smart adult does in front of a genuinely uncertain future. They keep their options open. They take the recruiter's call they would have sent to voicemail two years ago. They open the résumé they have not touched since the morning they signed with you, and spend twenty quiet minutes making it current — rewriting the bullets to emphasize judgment over output. Nothing dramatic. No decision made. Just a door, unlocked, in case.

Those twenty minutes are the subject of this chapter, and the thing the whole book has been walking toward.

You can do everything else right — sell the slice so the work is legible, sign only what you can carry, show your work, survive the gap so the calendar stays open — and still lose, completely. The capability that does all four of those things is not a process or a deck or a methodology; it is a few people. On some Tuesday you will not see coming, one of them walks out carrying judgment that can't be rehired in a quarter, and the new model leaves with them. That is the loss the title is about: not a resource, but a room.

The firms that cross to the other side rarely win on the cleverest strategy slide. They tend to be the ones whose best people, having done

their own math, choose to stay. The repricing is, in the end, a human problem.

Why the public record is so incoherent

That incoherence is worth sitting with, because it is the actual thing your people are metabolizing on those weeknights — and it is not going to resolve in time to help you.

The chief executives contradict each other, and they contradict themselves, and the reason is rarely that any of them are lying. They are describing genuinely different things and calling all of them by the same two-word phrase. When the head of AWS says it would be foolish to replace your juniors with AI, he is making a capability argument. Who, exactly, learns judgment if no one is ever made to do the junior work that teaches it? When the head of Salesforce says his agents absorbed the work of around four thousand support staff and he needs fewer heads, he is describing one narrow, deeply automatable task. That task is tier-one support, and it genuinely does compress toward zero. And when Amazon cuts fourteen thousand roles, its HR chief reaches for AI as the reason while its CEO reaches for culture. Both are grabbing the more comfortable story available for the same painful decision.

So “AI and jobs” is really a dozen questions wearing a single headline, not one question at all. Your best people are sophisticated enough to feel that — to sense the headlines are not about the same thing, which is exactly why no amount of outside reading will settle them. There is no authoritative source out there that resolves it, because the incoherence is real, not a media artifact. So they will turn, whether they say so or not, to the one person whose read they can test against the firm they work inside every day. They will turn to you.

The true thing is more useful than the comforting one: AI is not emptying your firm so much as changing what the work is.

The production layer is the part that compresses, because production is exactly what the model is good at. Those are the hands that turned a spec into a screen, a brief into a deck, a ticket into a working function — the making Chapter 2 called production, below the carry line it drew. What does not compress is the carry: the work you answer for. Deciding what is worth building. Catching the thing the model produced confidently and wrongly. Integrating four machine outputs into one coherent thing a client can stake a quarter on. Standing behind the result when that quarter is on the table. Judgment, integration, verification, accountability. The buyer still pays for those — but only when the firm has made them paid, sold as a bounded thing instead of folded into a rate. That is the whole strategy of this book, and it lives in the carry.

The reshape gives the line a human shape

The work above the line has four layers, and they are worth naming, because the org chart gets redrawn around them and the compensation has to follow.

There is the **Editor** — the senior judgment that decides what good looks like, holds the standard, and is accountable to the client for the outcome. The Editor does not merely review. The Editor frames the problem before the machine touches it, and stops a beautiful wrong answer from entering the room.

There is the **Verifier** — the rescue-and-judgment talent who checks the machine's confident output against reality. The Verifier finds the hidden dependency, the hallucinated fact, the workflow that looks plausible until a real user touches it.

There is the **Operator** — who runs the models and the pipeline. The new craft skill, fluent in the tools the way a great production hand used to be fluent in the software.

And there is the **Apprentice** — the early-career person who, in the old model, did the production work that is now evaporating. Somehow this person has to acquire judgment without the years of grunt work that used to teach it.

The rung that quietly rots

That last one is the problem the AWS chief executive was actually pointing at. It is real and slow and easy to ignore until it is too late to fix. If the Apprentice rung rots out — if you stop growing early-career people because the model does what they used to do — the Editor bench never replenishes. A decade from now you have a firm full of senior judgment with no one rising to replace it. The present gets optimized by quietly mortgaging the future, and the bill comes due exactly when it is hardest to pay. If AI removes the drudgery, the learning that drudgery used to smuggle in still has to be replaced. That is the succession plan for the judgment layer.

The carries live in people

This is the hinge of the whole thing. The mid-level designer who, two years ago, spent her week pushing pixels from a brief now spends it differently. She decides which of four machine-generated directions is actually on brand. She catches the one that subtly misreads the client's market — the kind of miss that would have been embarrassing in front of the board. She kills the direction that tests well and sells wrong. She walks into the room and defends the choice to a client who is nervous and paying for confidence. Her hands do less; her judgment does far more.

She has been promoted in everything but title and pay. She is doing Editor work for an Operator's wage. And for that exact reason she is the single most extractable person in the building. Manage her as a production designer, speak to her as if the machine made her less important, and she will believe the market before she believes you. The market's story is simple: your firm depends on her judgment and has not noticed what it is worth.

That is the thread that can't be lost hold of: the carries are not abstractions. They live in people. The Editor is judgment and accountability, embodied. The Verifier is verification and integration, embodied. When this book says the firm's value moved to the carry, what it means, physically, is that the value moved into a handful of heads — and heads can read a recruiter's email on a weeknight.

This is also where the hard economics of the whole book come to rest. Production raced to the cost of the tools; the carry is the only part of the work with a price the buyer can't benchmark away. But that price holds for one reason — the carry is scarce, and it is scarce because it lives in specific people who are hard to replace. That makes those people the firm's single asset whose margin cannot be benchmarked away in a procurement cycle. Everything else on the invoice has a market rate the buyer can look up. They do not. Lose them, and the scarcity that justified the price is gone with them — and the buyer, benchmarking what is left, pays for production at the cost of production. So keeping the room is not an HR concern running alongside the strategy; it is the strategy itself, the only way to protect the one margin a competitor cannot reproduce off a prompt.

Which makes the reshape the same problem as the runway, not a separate one. The talent risk and the transition risk are not two risks on two tracks. They are one, because the capability that lets you survive the gap is carried in the skulls of the very people most able to leave

during it. Keeping the room is keeping the carry. Lose the people and the captured margin walks out with them — there is no version where you lose the room and keep the value.

**▲ THE MOST EXPENSIVE REFLEX IN THIS CHAPTER:
WELCOMING QUIET ATTRITION.**

The spreadsheet seems to bless it. AI means fewer people are needed, the logic runs, so a little attrition is a gift you don't even pay severance for.

It is wrong twice over. You don't need fewer people. You need *different* people — weighted toward the judgment layers, away from the production layer that is compressing. And the attrition is not falling evenly. It concentrates in exactly the layer you can't rebuild in a year.

The spreadsheet sees a saved salary. It does not see that the person who resigned was the one who remembered why the client's last transformation cratered, and knew which stakeholder kills good work and how to route around her. That isn't a salary saved. It is the new business model walking out the front door. The grunt work can go to the machine. The judgment walks on two legs.

Leadership is deciding how you talk to people who already know

The capability you most need can read the situation as clearly as you can. That is what makes it valuable. So the central act of leadership in this transition is deciding how you talk to people who already know something is happening.

You do not get to choose whether they see it. They see it in the pipeline, in the kind of work being sold, in the hiring freeze nobody has called a hiring freeze, in the senior leaders asking strange new questions about utilization. You only get to choose whether they hear it from you, framed so they can act — or infer it from your silence and fill the vacuum with the worst version their imagination can build. It will build the worst version, because a smart person under uncertainty always

does. The absence of information never reads as neutral; it reads as the bad news you weren't brave enough to say.

There are two ways leaders get this wrong, and they are mirror images.

False reassurance

The first is **false reassurance** — the all-hands where everything is fine, the future is bright, nothing fundamental is changing, the deck heavy with adjectives and the hard questions answered with phrases. It buys one quiet month. Then reality contradicts it — a client pauses, a role goes unbackfilled, a senior person leaves — and now it is not merely a hard spot. You are the person who stood up and said it wasn't. Every sharp person in the room runs the same recalculation in the same half-second: if that was managed rather than told, what else is being managed? The one currency you cannot reissue is the belief that what you say is true, and that is the one you just spent.

The panic-dump

The second is the **panic-dump**. The leader, in a spasm of what he tells himself is authenticity, hands the whole team the full unprocessed weight of the runway and the uncertainty and the 2 a.m. fear, and calls it transparency. It is not transparency but offloading, and it does to morale what the first mistake does to trust — faster, because it broadcasts the one signal guaranteed to make your best people leave: there is no plan. A short runway with a plan is a problem. A short runway without one is a sinking ship, and people tend to leave sinking ships.

True, plan, ask, by-when

The discipline between those two failures is narrow, and it has a shape that holds in your head. Tell them what is true, including the hard part you would rather soften. Tell them the plan — what you are doing about it, specifically, not in the abstract. Tell them the ask — what you need from each of them, concretely enough that they could

start tomorrow. And tell them by when they will know more, so the uncertainty has an edge instead of stretching out forever.

True, plan, ask, by-when. The true part earns the right to be believed; without it the other three are noise. The plan part is the entire difference between leadership and confession. The ask part converts spectators into participants — and a participant is far less likely to leave than a spectator. The by-when part lets a person decide to stay through one more milestone rather than leaving on open-ended ambiguity. Smart people do not flee bad news. They flee the absence of a plan, and the suspicion that they are being handled rather than addressed.

► **WHAT IT SOUNDS LIKE WHEN IT LANDS.**

A founder stands up and says: *we have nine months at the current burn; here is the specific plan to close that gap; here is the part of it I need from each of you, by name; and you will have the next real update in three weeks.*

Nobody bolts; mostly there is relief that the number is finally on the table.

The senior engineer who had quietly taken the recruiter's call the week before — who had told himself it was just due diligence — stays. Not because the news was good; nine months is not good. He stays because, for the first time in months, the uncertainty has edges and he has a job to do inside them. The ambiguity had been the exhausting part. A hard number with a real plan wrapped around it was, strangely, a relief — a hard number with a plan around it was easier to work with than the open-ended version.

Retention is the interesting problem, not the counteroffer

Underneath the words, what the plan has to contain is the reshape made real and personal. Roles redrawn around judgment. The people you most need made owners of the new thing rather than survivors of the old one.

And it helps to be clear about what retention is and is not in a transition like this. A counteroffer war is the version that fails. You will lose it to firms with deeper pockets, and every counteroffer you win teaches the whole firm the same corrosive lesson: that leaving is how you get a raise here. That is a slow auction of your best people, run on your own time — not retention.

What actually holds the Editor and the Verifier is being handed the interesting problem. Give the Editor the authority to define what good looks like across the work, not just on her own accounts — her judgment becomes the firm’s standard. Give the Verifier ownership of the rescue function, the hardest and most respected work in the new model, with the headcount and mandate to build it into something real instead of heroics performed alone at midnight. Tell both, in plain language, that they are not being asked to ride out a decline. They are being asked to build the thing that replaces it, and to own a genuine piece of what they build.

As Martín Migoya of Globant put it, “The initial version of AI was humans accelerated by AI. The next generation is AI supervised by humans.” The supervision is the job now. A firm that makes its best people the supervisors, with the title and the stake to match, has stopped asking them to stay out of loyalty and started giving them a reason that survives the next recruiter’s call. Anything less is extraction with better language, and the people worth keeping can tell the difference quickly.

? “WON’T NAMING A SHORT RUNWAY OUT LOUD SET OFF THE STAMPEDE I’M TRYING TO PREVENT?”

It is backwards, and seeing that it is backwards matters. The stampede is caused by ambiguity, not by news. The people who leave are fleeing the vacuum the silence created — not the hard truth finally said out loud.

The choice was never between calm and alarm. It is between a plan people can act on and a vacuum they will fill. The way frightened people fill a vacuum is by leaving.

The room you sell to is a committee, not a person

There is a second audience reading you in this transition, and it is not the one most pictures when imagining the sale. When you pitch the repriced firm — the one that sells the slice, signs only what it can carry, shows its work — it feels like persuading a buyer, a single person. What you are really persuading is a committee, and the members do not share a single fear among them. The public position that wins the new work has to answer each of them in their own language. A story built for only one will die, quietly, in the room where the other three are sitting.

There are four of them — the same four the first chapter named, now turned around to read you — and four entirely different red lines.

The **CFO** is doing the math this whole book has been about: is the price legible, is the margin real or hidden subsidy, is this a number she can carry upward and defend without flinching? She is the audience the pitch is naturally built for, which is exactly the trap.

The **Vendor Management Office** does not care about the vision and is faintly annoyed by it. It cares about delivery and continuity — will you still be here in eighteen months, does the work survive a key person leaving? That last question is the precise risk this chapter is about, and the **VMO** asks it in writing. A sole-founder origin story, the

one that charms in a first meeting, reads to the VMO as a single point of failure waiting to happen.

The **Third-Party Risk** function cares about none of that and everything about security posture, data handling, the SBOM and the AIBOM — the questions Chapter 5 was built to answer. It can kill a deal the CFO loved without the CFO ever finding out why.

And in regulated industries there is a fourth seat: **Model Risk Management**, which wants to know how the AI inside your delivery is governed, validated, and bounded. It will quietly sink a deal the other three were excited about if it cannot get a straight answer to a single question: when your model is wrong, what catches it before it reaches us?

The position is a matrix, not a message

The default mistake is to pitch all four as though they were the CFO. The deck that dazzles on price and vision lands in front of the TPR reviewer as an unanswered security questionnaire and dies there, silently, weeks later, with no feedback to learn from.

So the public position works less like a message than like a matrix. Pitch and pricing speak to the CFO. References and the continuity story — the documented bench, the no-single-point-of-failure answer — speak to the VMO. The evidence pack and security posture speak to the TPR function, in its own native language, before it has to ask. Volunteering the answer is itself the signal that you are the kind of firm that knows the question matters. And the governance story — how the model is bounded, who verifies it, what catches it when it is confidently wrong — speaks to the MRM seat. In regulated rooms it is a first-class part of the pitch, not an awkward appendix produced when cornered. Four audiences, four fears, four pieces of evidence, deliberately aimed.

Here is the part that ties the public position all the way back to the people, and the reason these were ever one chapter rather than two. The external story only holds if the internal belief is real. You cannot send a salesperson to convince a Third-Party Risk officer that the firm runs rigorous verification while the actual Verifiers are demoralized, underpaid, and updating their résumés on weeknights. A committee can smell a position the firm does not believe. It leaks through the reference call that is a half-beat too careful. Through the security review that comes back half-staffed, because the person who owned it left in the spring. Through the answer that does not quite cohere, because the people who would have made it cohere are gone.

The firm that wins the four-function room is the one whose people are genuinely, demonstrably doing the thing the position claims. The position is not marketing laid over the firm so much as the firm itself, described accurately — to four different kinds of fear. Which is why keeping the room and winning the room are the same act: the Editor and the Verifier you fought to keep are the living proof the committee is checking for. Lose them, and the position becomes a story about a firm that no longer exists.

? “ISN’T THIS OVERTHINKING A SALES DECK?”

It is the opposite of overthinking. It is finally noticing that the deal you were certain you lost on price, you actually lost in a security review you never saw — or on a continuity question a sole-founder story could not answer, or on a governance question treated as an afterthought.

The four functions are already evaluating you, whether or not you address them. Declining to position for all four doesn’t simplify the sale. It just guarantees you lose the ones whose red line you ignored — and you never find out which one it was.

Where the book closes: the conviction and the cut

This book opened on Atlassian, a company confident enough in the new era to be building hard toward it. It is worth letting it close there too.

In October of 2025, the company's co-founder sat for an interview. He said that five years out, his firm would have more engineers than it had that day. His reasoning: technology creation is not bound by output, so better tools do not mean fewer people. They mean more ambition, a larger surface of things worth building, more hands wanted, not fewer. It was a genuine conviction, stated plainly, by someone with every reason to know. Five months later, in March of 2026, the same company cut about sixteen hundred jobs — roughly a tenth of its workforce — and said the savings would help self-fund its push into AI. The same announcement noted that its chief technology officer was leaving, the role split in two.

The contradiction does not resolve, and pretending otherwise is the false-reassurance mistake performed at the scale of an entire industry. More engineers in five years, and sixteen hundred fewer jobs this quarter. Both true. From the same leadership. Five months apart. That is not hypocrisy, and reading it as hypocrisy teaches nothing. It is what a transition actually looks like from inside the lives of the people living through it. A real, defensible belief in a larger future, held at the very same time as a present-tense decision about who is in the building to build it. The role split in two, rather than handed whole to a successor, is the reshape from this chapter happening at the top of one of the most sophisticated software companies on earth — judgment redistributed, not simply replaced.

Every founder this book was written for will live some version of those five months. The conviction and the cut, held at once, in front of people who can read both as clearly as you can — who are doing their own

math at their own kitchen tables on their own weeknights. There is no clean answer to that. The firms that pretend there is one lose their best people to the firms that don't insult them with it. What there is instead is the discipline this whole book has laid out, now carried all the way down to its human floor.

Sell the slice, so the work is legible. Sign what you can carry, so the contract holds. Show your work, so the proof is there. Survive the gap, so the calendar stays open. And keep the room — keep the people in whom the carry physically lives. The Editors and Verifiers hold the judgment and the verification the whole new model runs on. They can read the same uncertain future you can. And they will decide, on some quiet weeknight, whether to build the next thing with you or with someone else.

The repricing of the software agency is real, and it is already under way. In the end, though, it is less an economic event that happens to firms than a human one. It is carried out by people who get to choose where they do their best work — and who are choosing right now, whether or not they have been given a reason. The firms that make it across are the ones that give them one — and build the firm they choose.

KEEP THE ROOM IN 20 SECONDS

- The capability that makes the whole strategy work lives in a few people — and people can leave on a weeknight.
- The public record is incoherent because “AI and jobs” is a dozen questions wearing one headline. Your people turn to *you* to read it.
- AI is not emptying the firm; it is moving the value into the carry — into the *Editor* and the *Verifier*. The *Apprentice* rung is the one that quietly rots.
- Welcoming quiet attrition is the expensive reflex: it cuts the exact layer you can’t rebuild in a year.
- Lead with *true, plan, ask, by-when*. Smart people flee the absence of a plan, not the bad news.
- The buyer is a committee of four — CFO, VMO, Third-Party Risk, Model Risk Management — with four red lines. The position is a matrix, and it only holds if the people behind it are real.

→ ON MONDAY

1. Name your Editors and Verifiers — the people in whom the carry physically lives — and check what they’re actually paid against what they actually do.
2. Map the Apprentice rung: who is learning judgment, and how, now that the grunt work is gone?
3. Draft the *true, plan, ask, by-when* version of where the firm stands — including the hard part you’d rather soften.
4. Build the four-function matrix for your next pitch: one piece of evidence aimed at each of CFO, VMO, Third-Party Risk, and Model Risk Management.

Sell the slice. Sign what you can carry. Show your work. Survive the gap. Keep the room.

APPENDIX A

The Operating Kit

The book made an argument for running an agency through the repricing. An argument is not a tool, so here is the tool set — nine artifacts, each tied to the chapter that explains why it works, each built to be filled in rather than admired. The first time through, they tend to make most sense in order. After that, four of them earn a permanent place: the carry audit (quarterly), the provenance folder (per deliverable), the thirteen-week forecast (weekly), and the Do-Not-Sign checklist (every order before signature).

1. The Carry Audit (*Chapter 2*)

Purpose: measure the share of your fee the buyer pays for the carry rather than the production — a number most dashboards never break out.

Do this: pull your last five statements of work. For each, list every deliverable. Mark each one with a letter:

- **B — below the line (production):** the keystrokes, the screens-from-a-brief, the boilerplate, the first-pass copy, the test scripts — work a competent team with current tools now does in a fraction of the hours it used to bill. - **A — above the line (carry):** judgment (deciding what to build and what to kill), integration (making it work inside the client's decade of accumulated systems), verification (catching what the model got confidently wrong), accountability (the name on the indemnity).

Be honest about the split. “Senior oversight,” billed as forty hours of a principal, is usually three hours of carry wrapped around thirty-seven of production — and the two belong in different columns.

Compute: total the fee against the letters. Your **carry ratio** = (fee on A-lines) ÷ (total fee).

Deliverable	Fee	A / B
<i>e.g.</i> React front end + component library	\$150,000	B
<i>e.g.</i> legacy-auth-flow decision (which of 12 to honor)	\$40,000	A
<i>e.g.</i> core-banking integration (failed under load)	\$30,000	A
<i>e.g.</i> security pass (caught the model’s token-scoping bug)	\$20,000	A
Carry ratio		A ÷ total

Read it: the first run usually returns a carry ratio lower than expected — not because the firm does little carry, but because it never priced the carry it already does. The number collapsing in the business is the B-fee. The number worth learning to name and price is the A-fee. Everything else in this kit is how.

2. The Slice — the Five Parts (*Chapter 3*)

Purpose: convert a blended engagement into a bounded result procurement can grasp, price, and sign — instead of grinding it down to the cost of its cheapest line.

A slice is **not** a smaller project. It is a *bounded operational result* with five load-bearing parts. Strip any one and it collapses back into a blend.

1. **The buyer** — name which of the four functions signs (see Tool 9). The work may be the same six people; the *offer* is named differently for each. (“Checkout Conversion Instrumentation” to the CFO; “AI-Assisted Personalization Boundary & Evidence Pack” to Third-Party Risk.)
2. **The boundary** — a *risk-tagged* scope, every line

tagged **agency-controlled / client-controlled / shared / excluded**. Not a list of what gets built; a map of who answers the phone when it breaks. 3. **The proof standard** — the evidence pack (Tool 6) that lets finance, legal, and security each sign their own line without taking your word for anything. 4. **The risk surface** — the border between the risk you accept (Tool 4) and the risk you refuse, drawn in the open. 5. **The commercial wrapper** — the contract shell that carries the slice inside paper the buyer already has (Slice Order, EBM, fixed SOW, or MSA exhibit).

The rename test: the slice is not “mobile app rebuild,” which competes with cheap production and loses. It is “Cross-System Launch Readiness and Accountability,” which competes with the cost of the launch failing — and there is no cheaper option for that. The rename is a *reframe*, not a *relabel*: it only holds if the boundary beneath it is real.

3. The Slice Order (*Chapters 3–4*)

Purpose: reprice inside an existing master agreement without reopening it. Reopening the master is a quarter-long fight, and usually a losing one.

A Slice Order is a short rider on the existing T&M or master agreement that:

- **names** the bounded result (e.g., “Legacy Claims Workflow Evidence and Release Readiness”); - **varies** only the handful of terms the slice actually needs (the three clauses in Tool 4); - **points back** to the existing master for everything else; - **leaves the base relationship untouched** — no vendor re-onboarding, no reopened third-party-risk file.

Rule of thumb: if what a general counsel is being asked to sign is small enough that *yes* is easy, it is a Slice Order. If it triggers legal,

risk, and procurement to reconvene, the master has been reopened — which is the signal to stop.

4. The Three Clause Fights + the Do-Not-Sign Carry Checklist (*Chapter 4*)

Purpose: sign only the carry your balance sheet can actually hold.

The three clauses that decide the slice (the rest your counsel can hold alone):

1. **Limitation of liability — fight the carve-outs, not the multiple.** A 2× cap on a \$1.8 million order looks like a \$3.6 million ceiling. Carve IP, breach, and AI out from under it and the ceiling is a fiction. Accept uncapped exposure only for harms you control and could have prevented (your own willful misconduct, your own breach of confidence). Refuse it for harms you cannot (a third-party API, the buyer’s own data, a model’s hallucinations).
2. **IP indemnity — narrow the trigger, not the cap.** An indemnity capped at fees is barely an indemnity; leading with the cap tells the sharpest person in the room you do not understand the clause. Narrow *what it covers*: your own original work, yes; the raw output of a foundation model whose training data you did not assemble, no. Warrant your **process** (a human reviewed it; you logged which models touched it; you ran the controls), not the model’s training data.
3. **Service-level credits — one ceiling, not two.** Credits stack (5% + 5% + 10% in one bad quarter = a fifth of the fee) and almost never count against the liability cap. Cap them in the aggregate and tie that aggregate to the same liability cap. Refuse any credit gated on something you do not control (a third-party API’s uptime, the buyer’s approval cycles, “conversion lift”).

THE DO-NOT-SIGN CARRY CHECKLIST — print it; make it the gate every order clears before signature. Each line is a hard stop; a clause that trips one goes back to the business as a *pricing* decision, not down the hall to legal as a wording problem.

#	Clause	Why it is uninsurable carry	The redline
1	Uncapped AI IP indemnity	You cannot underwrite the copyright hygiene of a model you did not train.	Indemnify your own original code/assets only; disclaim warranty of non-infringement on raw third-party model output (it rides the provider's terms).
2	Uncapped data-breach liability on the buyer's systems	Their security posture is their carry; you cannot insure their infrastructure.	Cap at a defined multiple of fees; condition on <i>your</i> failure to follow agreed controls — not strict liability for any breach.
3	Outcome SLAs on metrics you don't control	Tying fee to conversion/sales/retention when you own none of the pricing, media, or market is carrying the client's business risk for free.	Tie SLAs only to deterministic technical metrics: uptime, latency, defect-resolution time, delivery dates.
4	Open-ended consequential damages	A missed launch can run to many times the fee; E&O will not touch it.	Mutual waiver of indirect, special, and consequential damages — lost profits and lost revenue named explicitly.

5	Fitness-for-purpose on the buyer’s data	Output is only as good as the data lake it runs on, which you did not build.	Disclaim the implied warranty of fitness for a particular purpose; state accuracy is contingent on the quality and legality of client materials.
---	--	--	--

A client who pushes back on one of these lines is asking you to be their uncompensated insurer. The move is to shift the conversation from the legal column to the price column: *if you want us to carry that, the number is not X — it is far higher.*

5. The Worst-Case Carry Worksheet (Chapter 4)

Purpose: know, before you sign, whether the risk has actually transferred — or whether you are quietly self-insuring with money you do not have.

Before signing any order, fill three columns:

Column	What goes in it
A — Maximum exposure	The <i>worst-case</i> total across every clause: full carve-out exposure, stacked SLA credits, consequential damages — not the likely case, the worst one.
B — Real E&O sublimit	Not the binder’s headline aggregate. The far smaller figure the carrier will actually pay on the <i>kind</i> of claim this contract creates — and which excludes contractual liabilities and SLAs you volunteered. (Ask your broker the sublimit for this claim type, in writing.)
C — Real credit headroom	The cushion your covenant leaves before one slipped receivable trips a default and hands the lender the conversation.

The rule: if $A > B$ and $A > C$, the risk hasn't transferred anywhere. You are self-insuring the gap. That leaves two honest options: a clause to renegotiate, or a deal to walk. The arithmetic is an afternoon in a spreadsheet; the nerve to act on it is the harder half.

6. The Provenance Folder (*Chapter 5*)

Purpose: turn “we use AI” — an unprovable claim, and a liability — into “here is exactly what our AI did, what our people did, and how we know.” That is evidence, and the one thing a competitor cannot copy off your website. It makes the narrowed indemnity (Tool 4) defensible, clears the audit, and holds up if the claim is ever read back in a dispute. Three returns on one discipline.

Five fields, logged as the work happens, in the tools already open — owned by the person closest to the deliverable, not a Thursday governance committee:

Field	Lives in	Owner	Example entry
1. Model & version touched each deliverable, and when	project-tracker column / PR-template line	the person who moves the ticket	“Model X drafted this module Mar 3; layout from internal tool the 11th”
2. What went in — inputs + a one-line data-permission attestation	delivery doc, top	account lead	“No client data used to train any model; inputs limited to materials provided under the MSA”

<p>3. Who reviewed it, and what they changed — named, dated</p>	<p>PR approval / sign-off field</p>	<p>the reviewer</p>	<p>“Reviewed by lead eng Mar 5; rewrote error handling, removed 2 hallucinated edge cases”</p>
<p>4. What you asked and what you kept — prompt/ decision trail</p>	<p>PR-template section / delivery doc</p>	<p>the maker</p>	<p>“Generated 3 directions; kept 1; rejected 2 for brand mismatch”</p>
<p>5. License & copyright pass-through per component/ asset</p>	<p>delivery manifest line</p>	<p>the maker</p>	<p>“All third-party assets MIT/Apache; no training-data conflicts identified”</p>

Creative shops — Generative Provenance: the same folder, fields rhyming — which generative model made or assisted each asset; a data-and-licensing attestation; the copyright chain that tells the client whether they can run it in a paid campaign; the art-director sign-off log. The code shop proves its build; the creative shop proves its originality. Both answer the same buyer question: *can I use this without it coming back on me?*

The version that works starts this week, on the next deliverable that ships — not a backfill of last year, which fools no one. One deliverable, five fields, the tools already open.

7. The Thirteen-Week Cash-Flow Forecast + the Lender Call (*Chapter 6*)

Purpose: find the date the cash runs out while it is still far enough away to do something about it. Pricing the carry moves revenue *later* while payroll stays where it is — the gap is the cash-shadow of the strategy, measured in weeks of payroll, not points of margin.

Build it (weekly, thirteen weeks, cash not margin):

Col	Contents
A	Week-ending date
B	Cash in — every receipt at its <i>realistic clearance date</i> (when the client actually pays, not the invoice date)
C	Cash out — every payroll (flag the months with three Fridays), rent, taxes, benefits, insurance, vendors, interest
D	Net weekly change (B – C)
E	Running balance — color red the first week it dips below next Friday's payroll

The red week is the date. The first version tends to read kinder than the truth — it is easy to miss the three-Friday months, the quarterly tax remittance, the receivable that always clears fifteen days late. The fix is to date every line to *behavior*, not to terms.

Know your failure shape: a *milestone trough* (revenue moved later by converting to milestone billing — count the payrolls in the gap); an *acceptance-window subtraction* (a stretched acceptance term ages a receivable past 90 days, and cross-aging drops the client's *other* invoices out of your borrowing base too); or an *outcome-retainer leak* (booked-vs-collected drift that is easy to mistake for a rounding error).

THE LENDER CALL — the move that works before the covenant breaks, not after. Its value collapses the longer it waits. The call lands when it walks in with three things, in order: 1. **The honest forecast**, sent two days ahead with a one-page note: exactly

what is needed and why the numbers close. 2. **The shape of the gap, named** — “a milestone trough of five payrolls” / “an acceptance subtraction of \$X dropping out of the base on these dates.” A lender can underwrite a bounded problem, not a vague one. 3. **The smallest ask that buys the most time, matched to the shape** — a covenant waiver through the dry months; a temporary overadvance against the specific milestone that clears next quarter; a borrowing-base re-cut that steps down with your milestones instead of off a cliff. A lender called early holds a *performing loan from a borrower with a plan*. A lender surprised by the breach holds a *defaulted loan with only collateral to protect* — and a reason to grab it first. Same bank, two different Tuesdays; the only variable is who called first.

8. The Reshape Map (*Chapter 7*)

Purpose: redraw the org chart, and the pay, around the four layers of the new work — by *responsibility*, not title.

Layer	What they own	Carry
Editor	decides what good looks like; frames the problem before the machine touches it; accountable to the client	judgment + accountability
Verifier	checks the machine’s confident output against reality; owns the rescue	verification + integration
Operator	runs the models and the pipeline; turns intent into output at speed	the new production craft
Apprentice	the early-career person who must learn judgment without the grunt work that used to teach it	the succession plan for the Editor bench

The trap: welcoming attrition because “AI means fewer heads.” The need is not fewer people; it is *different* people, weighted toward Editor/Verifier. The senior person quietly doing Editor work for an Operator’s wage is the most extractable person in the building — and the one to find, and pay and title for the carry she already carries, before a recruiter does.

9. The Leadership Script + the Four-Function Position Matrix (Chapter 7)

Purpose A — talk to people who already know something is happening. The stampede comes from ambiguity, not from news. Four beats, every time:

1. **True** — what is true, including the hard part. (*Earns the right to be believed.*)
2. **Plan** — what you are doing about it, specifically. (*Separates leadership from confession.*)
3. **Ask** — what you need from each of them, concretely. (*Turns spectators into participants.*)
4. **By-when** — when they will know more. (*Gives the uncertainty an edge.*)

“We have nine months at the current burn. Here is the plan to close the gap. Here is the part I need from each of you, by name. You’ll have the next update in three weeks.” — the room does not panic; the room exhales.

Purpose B — pitch the committee, not the buyer. Four functions, four red lines, four pieces of evidence — deliberately aimed:

Function	Its fear	What you aim at it
CFO	Is the price legible and defensible upward?	the pitch + the slice’s per-line pricing
VMO (Vendor Mgmt)	Continuity — does the work survive a key person leaving?	references + documented bench + no-single-point-of-failure story

TPR (Third-Party Risk)	Security, data, model exposure	the evidence pack / SBOM / AIBOM (Tool 6), <i>before</i> they ask
MRM (Model Risk)	When the model is wrong, what catches it?	the governance + verification story, first-class in regulated rooms

The external position only holds when the internal belief is real. It is hard to convince a TPR officer the verification is rigorous while the Verifiers are updating their résumés. Keeping the room (Tool 8) and winning the room (this matrix) turn out to be the same act.

Monday Moves

If you do nothing else this week:

1. **Run the carry audit** on one recent SOW. Find your carry ratio. (Tool 1)
2. **Rename one live engagement** from a production noun to a carry/outcome noun, and test it on the next buyer conversation. (Tool 2)
3. **Add the five provenance fields** to your PR template / delivery doc, and fill them on the next deliverable that ships. (Tool 6)
4. **Build the thirteen-week forecast.** Find your date. (Tool 7)
5. **Print the Do-Not-Sign checklist** and make it the gate on the next order before signature. (Tool 4)
6. **Name, for your three most valuable people, which layer they actually work in** — and whether you are paying them for it. (Tool 8)

Sell the slice. Sign what you can carry. Show your work. Survive the gap. Keep the room.

APPENDIX B

The Case Files

The evidence behind the chapters, written up as twelve short case files — the strongest cases for the argument, the strongest against it, and the failures that show what the repricing does to a firm that ignores it.

IN THIS CHAPTER

- Each file is tagged Verified (named source) or Illustrative (a flagged composite)
- The discipline is one anchor, two corroborators, and one honest counter per major claim — so the counter-cases are here too, not buried
- Verified facts carry their source; Illustrative cases show a mechanism only, never a named firm's fate

These are the case files behind the chapters. Each is tagged by status.

Verified means it is drawn from a published earnings transcript, court filing, or on-the-record public statement, with the source named.

Illustrative means it is a composite built from the verified mechanics.

It is flagged so no reader mistakes it for a real named firm.

The book's discipline is one anchor, two corroborators, and one honest counter per major claim. So the counter-cases sit here too, not buried.

The discipline

A strategy book is only as good as its willingness to show the case that argues against it. So these files turn on the book's central claim. *The repricing is real, and it forces a change in how agencies price and what they sign.* Each claim gets the strongest evidence for it and the strongest against

it. The failure cases then show what the repricing does to a firm that ignores it.

Verified facts carry their source. Illustrative cases are labeled as such, and used only to show a mechanism. They never assert that a specific named firm met a specific named fate.

I. The repricing is real

Case 1 — Endava: the deflation confession (*Verified — Q1 FY2026 earnings call, Nov 11 2025*)

The anchor. On its Q1 FY2026 earnings call, Endava’s CEO John Cotterell said, on the record, that “as we’re delivering a much higher productivity, that is having an erosion on the revenue that’s coming through the business” — and put numbers to the response. That was 24% of revenue outcome-based and rising, 76% still time-and-materials. Productivity step-ups ran “in the five to 10 times type range” — and over 70% of services were now AI-related.

From the income-statement side, the same quarter tells the other half. It carried a £364.6 million impairment and revenue down 8.4%, with an adjusted pre-tax margin collapsing from 12.6% to 1.8% — even as AI-driven revenue tripled to roughly 15% of the business.

What I take from this file: the new revenue line was growing exactly as the strategy promised, and the firm was squeezed nearly flat anyway. Outcome-based contracts convert to cash more slowly than the hours they replace. Here is a public company narrating its own migration toward the carry, and the gap that migration opened. It is the strongest single piece of evidence in the book, because it is HIGH-confidence and self-incriminating.

Case 2 — Globant: the AI Pod as survival mechanism
(Verified — Globant IR, reported March 2026; Q1 2026 revenue per Globant IR press release, May 14 2026)

Corroborator. Globant reported its “AI Pods” — packaged, subscription-priced units of AI delivery — at \$32.8 million in annualized recurring revenue. They sit inside 40% of its top-20 accounts, against a Q1 2026 revenue base around \$607.1 million ($\approx 1.3\%$ of run-rate). The number is a rounding error; the *packaging* is the point. An AI Pod is a SKU — defined inputs and outputs, written service levels, an audit posture fixed before the first invoice, priced to recur. CEO Martín Migoya has described the underlying pricing shift as moving “from ... [an] hourly model and fixed-price models ... to ... supervised tokens.” The reading here: a \$2-billion-plus firm productized AI not to add a point of growth, but to clear the four-function procurement room. The bounded, legible unit is the shape that survives where the blended dollar gets crushed. A 180-person agency does not need the balance sheet to copy the grammar.

Case 3 — CI&T: the market is already paying *(Verified — CI&T Q1 2026 results)*

Corroborator. CI&T reported \$136.6 million in Q1 2026 revenue, up 23.2% year-over-year from \$110.9 million — with about a fifth of its new sales written on *new pricing models* rather than the time-and-materials default.

The reading: the bounded, packaged, provable unit is already a sale, closing in volume — not a thesis waiting on evidence. CI&T connected AI to the *shape of the invoice* rather than merely to delivery speed, and the new-pricing line is the result.

Case 4 — Sorrell / S4 Capital: the repricing question*(Verified — recorded public interviews, 2025)*

Corroborator. “Do you charge on the basis of time or output, value added? ... So we have to change our model,” said Sir Martin Sorrell — founder of S4 Capital and, before it, builder of the modern holding company at WPP. He framed the problem as time compresses; in his example, a commercial goes from four months to four weeks. He also bet, contrarian for a scale-builder, that “in a super ... AI age, scale is not necessarily the most important [thing] — [it] is brain[power].”

The note I’d make: the person who built the time-based agency model is the one declaring it broken. (*MEDIUM confidence — auto-caption. The dollar figures in the example are flagged for audio-verification in Appendix C.*)

Case 5 — Droga / Accenture Song: creativity bifurcates*(Verified — Cannes Lions, June 2024)*

Corroborator (creative side). “Not all creativity is worth saving ... the majority of advertising is not creative — it’s written by something far more dangerous than AI, which is research,” said David Droga, and “when everybody is doing best practices, nobody is doing it.” He also conceded the staffing model: “a lot of the roles ... the need to have the staffing the same way ... is going to disappear.”

What this points to: the production half of creative work commoditizes, while a premium accrues to the genuinely original top. The staffing pyramid that served the commodity middle shrinks with it. That bifurcation lands on the agency, not just on the coder. (*MEDIUM confidence.*)

II. The honest counter**Case 6 — Persistent Systems: “the deflation hasn’t hit us”** *(Verified — earnings interviews, 2024–2026)*

The counter. Pressed directly by an ET Now interviewer on “AI deflationary revenue ... taking the entire surplus out,” Persistent’s

CEO Sandeep Kalra pushed back. Being ahead of the curve “doesn’t mean that we will see the deflation come in first,” and the firm was still growing — roughly 17% year-on-year, after a four-year CAGR north of 25%. His optimistic case: AI lets a firm profitably take on work that was previously uneconomic — “legacy ... mainframe-based ... COBOL-based” systems nobody would touch — “open[ing] a newer addressable market.”

Why it is here: the repricing is real, but it is not a uniform death sentence. For some firms, in some years, growth and an expanding addressable market genuinely offset per-unit price erosion. The honest version of the thesis is that the deflation is coming, not that it has already arrived everywhere. And timing — the runway, Chapter 6 — is exactly what decides who survives being right.

Case 7 — a16z / Rampell: the outside view (*Verified — “Software is Eating Labor,” Oct 2025*)

The counter from above. Alex Rampell frames the whole shift from the investor’s side. Software is no longer selling tools; it is going after the labor market (“\$300 billion ... SaaS ... 13 trillion ... labor”), and the pitch becomes “we’re not giving you software — we’re going to do a job for you.” Seat- and hour-based pricing dies of its own success: “how many seats do you need if every one of your agents is 9,000 times more productive? ... Zero.”

Why it is here: it is the bull case for the disruptor and the bear case for the incumbent agency — the same force the book describes, seen by the people funding it to happen *to* services firms. It corroborates the mechanism. It also warns that the agency is on the wrong side of that force unless it repositions. (*MEDIUM confidence; stage figures flagged for audio-verification.*)

III. The failure cases — how a firm actually dies

Case 8 — Madwell: killed by the lender, not the market

(Verified — reported by Adweek from court documents)

The covenant, not the competitor. At 11:59 p.m. on April 30, 2025, Madwell — a Brooklyn creative agency in business since 2010 — turned off the lights. The CEO emailed staff that evening; roughly eighty people lost their jobs; vendors went unpaid. It did not lose a bake-off. It had defaulted on a Bank of America facility — by court documents reviewed by Adweek, it owed more than \$4 million — and the bank had moved to seize its assets about three weeks before the end. The reading (Chapter 6): death in the gap traces to a covenant breached while the work was still good, not to a competitor who beat you on the work. Once the firm breached, the lender behaved like a lender — it moved to seize the assets. The seizure came first, the shutdown second; the market never voted.

Case 9 — Builder.ai: the claim and the scrutiny *(Verified — with care: see status)*

Status: one version of this story is **disputed** treat it as unproven. The internet's version — that hundreds of engineers secretly hand-wrote code while the company pretended a machine did it — was pushed back on hard by careful reporting. The version that holds up is narrower. Builder.ai sold for years on an AI product called Natasha, which by the best reporting was a real AI system. It also ran a large network of outsourced human engineers — hundreds of contractors across India, Vietnam, Romania, Ukraine, and Poland.

In May 2025 it entered insolvency after its lender, Viola Credit, swept roughly \$37 million from its accounts. Bloomberg reported it had *allegedly* inflated sales through round-tripping with an Indian firm, VerSe Innovation — **which VerSe flatly denies** — and US federal prose-

cutors in Manhattan opened a criminal investigation into its financial reporting.

The reading (Chapter 5): set the fraud allegations aside — they are Builder.ai’s own. The detail that bears on an agency is what happened to the AI claim once scrutiny arrived. No one could cleanly say how much of the AI was real. The line between what the machine did and what the people did had never been written down in a form that could survive a hostile reading. If you can’t evidence it under hostile review, it won’t survive — true or not. The dynamic doesn’t need a fraud probe to do damage — only a risk team, a soured engagement, and a folder that turns out to be empty.

Case 10 — The signed-carry death (*Illustrative composite* — Chapter 4)

Illustrative, not a named firm. An eighty-person product agency takes a \$2.5 million build for a global retail brand. Hungry for the logo, it concedes two “remote” risks: an uncapped IP indemnity that sweeps in AI-generated assets, and an SLA tying a fifth of the fee to conversion-rate lift. A hero asset — heavily iterated through a frontier image model — turns out to resemble a competitor’s copyrighted campaign. A cease-and-desist lands; the brand pauses a \$10 million media buy; conversion collapses; the brand invokes all three clauses at once. The E&O carrier declines (contractual-liability exclusion); the bank tightens the line as a receivable ages out of the borrowing base; the uncapped indemnity comes due in cash. The founders personally guarantee payroll, settle for a number that wipes out operating cash, lay off a third of the firm, and lose their equity in a distressed recapitalization.

The reading: they did not die because they were bad at software. They died because they signed carry they could not hold. (*Figures illustrative.*)

The mechanics — contractual-liability exclusion, borrowing-base cross-aging, uncapped indemnity — are real and standard.)

Case 11 — The unprovable-renewal loss (*Illustrative composite — Chapter 5*)

Illustrative. A hundred-person agency wins a build at a mid-market company owned by a regulated parent. The work goes well. Then the renewal arrives with a vendor-security review, now policy for every supplier. The demands: *attach your AI usage policy; list models and versions; provide evidence of human review; attest to data handling*. The competitor down the street, “AI-native” in a font twice the size, has no policy, no inventory, no logs. Its review stalls, and the risk officer writes the three words that end it: *unable to evidence controls*. The firm that kept the five-field folder answers in an afternoon and signs for another year.

The reading: in a market where everyone says “we use AI,” the differentiator moved from *having* the capability to *proving* it. The audit is the one part of the bake-off a sloppier rival cannot fake. (*Illustrative. The regulatory drivers are real: SBOM/AIBOM, ISO/IEC 42001, and the EU AI Act’s high-risk obligations. Those obligations were since deferred to 2027–2028 under the Digital Omnibus.*)

IV. The bookend

Case 12 — Atlassian: the conviction and the cut (*Verified — public interview Oct 2025; company announcement March 2026*)

Both true, five months apart. In October 2025, Atlassian’s co-founder said in an interview that five years out the firm would have *more* engineers than it did that day. His reasoning: technology creation is not bound by output, so better tools mean more ambition, not fewer people. In March 2026, the same company cut about sixteen hundred jobs — roughly a tenth of its workforce. It said the savings would help

self-fund its push into AI, and noted that its chief technology officer was leaving, the role split in two.

The reading (Chapters 1 and 7): both can hold without resolving. More engineers in five years, and sixteen hundred fewer jobs this quarter — both true, from the same leadership. It is less hypocrisy than what a transition looks like from inside the people living it: a genuine belief in a larger future, and a present-tense decision about who is in the building to build it. The CTO role *split in two* rather than handed whole to a successor is the reshape (Chapter 7), happening at the top of one of the most sophisticated software companies on earth. Judgment was redistributed, not simply replaced. Most firms going through this will see some version of those five months.

Source & confidence note

The earnings-call and court-document cases — Endava, Madwell, and the financial figures for Globant, CI&T, and Atlassian — are **Verified** from published transcripts, filings, or company statements. The recorded-interview cases are Sorrell, Droga, Kalra, Rampell, and Migoya’s pricing quotes. They are **Verified as on-the-record statements** but **MEDIUM confidence** on exact numbers. Those numbers are flagged in Appendix C for audio-verification before any figure is quoted precisely.

The Builder.ai file is reproduced with its hedges intact. The sensational version is **disputed**/unproven; only the load-bearing, defensible facts are asserted. Cases 10 and 11 are **Illustrative** composites built from real, standard mechanics — labeled so no reader mistakes them for a named firm’s documented fate. Full citations live in Appendix C and the working research files.

APPENDIX C

The Repricing Index

Why this exists

A strategy book about a shift this fresh runs into one evidence problem: the people living the shift haven't written it down yet. The usual workarounds are thin. An anonymous survey can't be attributed. A handful of off-the-record interviews can't be verified. So I took a third path.

The executives of the firms living the repricing are already saying it in public, on the record. S4 Capital, Globant, Accenture Song, Persistent, Endava, and the venture investors funding the disruption talk about it on earnings calls, at Cannes, on podcasts, in keynotes. The Repricing Index is what came out of mining that public record: pulling the transcripts, coding each statement against the book's thesis, and citing it to its source and timestamp.

The standard behind it is a narrow one: a statement is included only if it actually appears in the cited recording. Each one is tagged for confidence:

- **HIGH** — verbatim from a published earnings transcript or human-reviewed captions; wording reliable. - **MEDIUM** — from auto-captions, which occasionally mis-hear proper nouns and numbers. The *substance* is reliable; any dollar figure, percentage, or multiple is flagged to be confirmed against the source audio before it is quoted as an exact figure.

Nothing here is fabricated, and nothing is paraphrased into a quotation. Where the public record had a gap, I named the gap rather than fill it.

The source ledger

S1 · Sir Martin Sorrell — Founder & Exec Chairman, S4 Capital (founder, ex-WPP). Recorded interview, “AI, agencies, and the future of advertising”, 2025. *auto (MED)*.

S2 · Sir Martin Sorrell — Founder & Exec Chairman, S4 Capital. “Marketing Connected — Agency Agenda”, 2025. *auto (MED)*.

S3 · Martín Migoya — Chairman & CEO, Globant. “How Enterprises Actually Get ROI From AI”, 2025. *auto (MED)*.

S4 · David Droga — CEO & Creative Chairman, Accenture Song. Cannes Lions, in conversation w/ Mira Murati, June 2024. *auto (MED)*.

S5 · Sandeep Kalra — CEO, Persistent Systems. ET Now earnings interview, Jan 2026 (Q3 FY26). *auto (MED)*.

S6 · Sandeep Kalra — CEO, Persistent Systems. Earnings interview, July 2024 (Q1 FY25). *auto (MED)*.

S7 · Alex Rampell — General Partner, Andreessen Horowitz (a16z). LP Summit, “Software is Eating Labor”, Oct 2025. *auto (MED)*.

S8 · John Cotterell — CEO, Endava. Q1 FY2026 earnings call, Nov 11, 2025. *published transcript (HIGH)*.

(Full URLs and approximate timestamps are held in the working research file; the Endava transcript is public on the earnings-transcript record.)

The findings, by thesis

The statements sort into five claims. The first three are the book’s spine; the last two are its consequences.

A · Production work is compressing toward zero

“As we’re delivering a much higher productivity, that is having an erosion on the revenue that’s coming through the business.” — **John Cotterell, Endava** (Q1 FY2026 earnings call). **HIGH.** *A public-company services CEO, on an earnings call, saying AI productivity erodes revenue under the current model — the deflation thesis confirmed from inside.*

“[We’re] getting significant step-ups in productivity — in the five to 10 times type range.” — **Cotterell, Endava.** **HIGH.** *A named, quantified productivity multiple for software delivery.*

“Not all creativity is worth saving. ... the majority of advertising is not creative — it’s written by something far more dangerous than AI, which is research ... so many things that we deem as creative ... are formulaic already.” — **David Droga, Accenture Song** (Cannes, 2024). **MED.** *Most “creative” output is formulaic, and therefore automatable; only top-tier originality is defensible.*

“Anything that is repeatable, anything that can be commoditized — application support, infrastructure support, business processes ... with our AI progress, automation ... we are being brought into those deals.” — **Sandeep Kalra, Persistent Systems** (Q3 FY26). **MED.** *Names precisely which work commoditizes.*

B · What holds — and grows — in value

“The initial version of AI was humans accelerated by AI. The next generation is AI supervised by humans.” — **Martín Migoya, Globant** (2025). **MED.** *The cleanest one-line account of the labor inversion, and where human value relocates: supervision, verification, accountability. This is the carry in a sentence.*

“When everybody is doing best practices, nobody is doing it.” — **David Droga, Accenture Song** (Cannes, 2024). **MED.**

Once a capability goes universal it stops differentiating; value bifurcates — a premium for the top, commodity for the middle.

“In a super digital age, or super AI age, scale is not necessarily the most important [thing] — [it] is brain[power]. ... I’m not sold on [scale].” — **Sir Martin Sorrell, S4 Capital** (2025). **MED.** *A contrarian repricing thesis from a man who built two holding companies on scale.*

C · The pricing model is shifting — hours to outcomes

This is the best-evidenced claim in the book: three independent, named confirmations of the same mechanism, plus the macro frame.

“Do you charge on the basis of time or output, value added? ... So we have to change our model.” Context: *“we sell time ... if the time to make a commercial can be compressed — which it can — from 4 months to 4 weeks, and from \$2.5 million of cost to 500,000 ... what do you do?”* — **Sir Martin Sorrell, S4 Capital** (2025). **MED** (verify the dollar figures against audio). *The repricing question, put by the architect of the modern holding company.*

“This quarter, we were 24% outcome-based, which is still rising. But it does mean that 76% of our revenue is coming in a T and M [time-and-materials] basis.” — **John Cotterell, Endava** (Q1 FY2026 earnings call). **HIGH.** *A hard number on the pace of the hours → outcomes migration at a public engineering-services firm.*

“We’re changing from ... [an] hourly model and fixed-price models ... to ... the same way that [OpenAI] charges you ... per tokens, per consumption ... in this case it would be like supervised tokens ... a monthly subscription that includes a certain amount of supervised tokens.” — **Martín Migoya, Globant** (2025). **MED.** *A named CEO describing the concrete new pricing primitive: hours → “supervised tokens.”*

“We’re not giving you software — we’re going to do a job for you. And that’s the sales pitch.” Frame: *“The worldwide SaaS market is about \$300 billion per year. ... The labor market in the US alone is 13 trillion. ... the prize that [software] is going after is the labor market.”* — **Alex Rampell, a16z** (2025). **MED** (round figures stated on stage; verify against audio). *The services-as-software pivot in one line, with the market frame behind it.*

D · The headcount and the pyramid reshape

“We’re ... continu[ing] to see us training and bringing in AI-native people and losing people who are not going to make that shift into the future.” — **John Cotterell, Endava** (Q1 FY2026). **HIGH**. *Workforce churn toward AI-native people, attrition of those who don’t convert.*

“A lot of the roles that we have in agencies and design businesses ... the need to have the staffing the same way ... is going to disappear.” — **David Droga, Accenture Song** (Cannes, 2024). **MED**. *A named agency CEO conceding the staffing model shrinks.*

“[We must] wean away from large numbers of people and even larger amounts of time to smaller numbers of people who are more efficient, effective and use technology and data.” — **Sir Martin Sorrell, S4 Capital** (2025). **MED**. *Headcount-and-hours compression stated as strategy, not fear.*

E · What they are actually doing about it

“We are writing more outcome-based deals with clients which locks in the opportunity to deliver greater benefit to the client using AI, but also to improve our margins.” — **John Cotterell, Endava** (Q1 FY2026). **HIGH**. *The strategic response to T&M erosion: convert to outcomes to recapture the AI-productivity margin.*

“Our revenue per employee should move up, our profit per employee should move up, because we are doing a platform and services ... rather than being a human-centric ... services delivery arm.” — Sandeep Kalra, Persistent Systems (Q1 FY25). MED. *Decoupling revenue from headcount: from bodies-for-hire to platform-plus-people.*

“The platform should ... help us disrupt our own business and ... the business of our competition.” — Sandeep Kalra, Persistent Systems (Q1 FY25). MED. *Explicit willingness to cannibalize the firm’s own labor-based model.*

The complete coded dataset

The by-thesis section above surfaces the strongest statements. For completeness, the full set of coded data points follows, by source — the working evidence base behind the chapters. Each is verbatim from the cited recording; MEDIUM unless marked HIGH; numbers flagged for audio-verification per the integrity note below.

S1 · Sir Martin Sorrell, S4 Capital (2025 interview) - *On the AI savings pitch*: “When we pitch AI we say you will save at least 30% from your cost.” (*MED* — *verify 30%*.) Quantifies the deflationary pressure S4 itself markets. - *On the production test*: a “normally produced \$2.5 million [commercial that] take[s] 4 months” against “a[n] [agentic] commercial ... about 4 weeks and cost half a [million].” (*MED*.) Concrete production-compression experiment. - *On the org*: “they will flatten it hugely ... the way that you control a silo is by controlling information ... in an era of AI that’s [going away].” Information-democratization flattens the agency pyramid. - *The one shift he’d bet on*: “Democratization of knowledge. ... It makes everybody in the company as aware of

what’s going on as anybody else.” Value moves to whoever deploys collective knowledge.

S2 · Sir Martin Sorrell, S4 Capital (2025, “Marketing Connected”) - *On differentiation*: “if everybody’s personalizing at scale ... does that mean ... Coca-Cola’s advertising will be the same as everybody else’s? ... that puts a premium on strategic and creative.” When AI commoditizes output, differentiation gets the premium. - *On cannibalization*: “We’re not frightened to do ourselves out of a job because ... media planning ... in many cases ... would be better in-house than out[sourced].” The agency-of-record model openly questioned by a holdco founder. - *On the outsourcing reversal*: “Too much stuff was outsourced ... when CFOs were saying ... lessen the headcount, and in a ... first-party-data tech world there are functions that are better in-house.” Pressure on the services revenue pool.

S3 · Martín Migoya, Globant (2025) - *On verification as the product*: “AI agents on the back will produce that software ... in order for [it] to be enterprise class ... we need ... human supervision checking ... the software ... is correct.” Production automates; verification is sold. - *On the durable wrapper*: “all the enterprise-class things ... security ... accessibility, accountability, traceability, cost control.” Enumerates the value layer around raw model output. - *On token transparency*: “You have a monthly limit of tokens that represents ... the effort that ... took to be created. You have full transparency on which ... assets [were] created, how many tokens each ... required.” Pricing re-based on machine effort, not labor hours.

S4 · David Droga, Accenture Song (Cannes, June 2024) - *On buyer repricing pressure*: “a lot of clients ... just look at it for ‘oh, I’ll never have to pay a cop[ywriter] ... a photographer’ ... our job ... is to find the counterbalance between what is still worth paying for.” Buyer-side production-repricing is already explicit. - *On the non-automatable residue*:

“eventually the clients are going to have to pay us ... [for] what we add to it ... we still bring to the table intent and sincerity.” What humans still get paid for: intent, taste, sincerity.

S5 · Sandeep Kalra, Persistent Systems (Q3 FY26, Jan 2026) -

The deflation framing, put by the interviewer (ET Now): “there’s an AI deflationary revenue which is ... taking the entire surplus out ... you will see the headwinds first with respect to revenue deflation.” *(Attributed to the interviewer, not Kalra.)* - *On today’s AI-revenue share:* “the ... true AI revenues ... are high single-digit at max ... or low double-digit ... down the line I don’t think anyone will talk about AI revenues and other revenues separately.” AI dissolves into all services.

S6 · Sandeep Kalra, Persistent Systems (Q1 FY25, July 2024)

- *On demand-side cost pressure:* “those [large] deals have been led by cost-optimization initiatives from our customer side ... vendor consolidation.” Clients use the moment to cut services spend.

S7 · Alex Rampell, a16z (Oct 2025, “Software is Eating Labor”)

- *On the new production function:* “the capital ... we give to companies ... They buy GPUs ... hire engineers ... and then out pops software that does the job of labor. ... the new $E=mc^2$.” *(MED — figures on stage.)* - *On why prior software waves didn’t reprice services:* “almost every software company has ... turned [a filing cabinet] into a database ... nothing has ... gotten that much more efficient because the ... records are read by humans ... the whole business model of software has to change.” This wave removes the human-in-the-loop. - *On the live outcome-pricing jump:* a support vendor “piloting outcome-based pricing ... maybe [it] could charge \$5 million a year ... Don’t pay us 1.4 million. Pay us 5 million. Save [\$]75.” *(MED — verify figures.)* A seat-vendor leaping to outcomes — the destination the agency model is pushed toward.

S8 · John Cotterell, Endava (Q1 FY2026 earnings call, Nov 11 2025) — HIGH - *On the AI-native transition speed*: “that greater than half has moved to over 70% of our services [being] AI related.” HIGH.

Integrity note

Of the eight sources, one (Endava) is HIGH-confidence, drawn from a published earnings transcript. It supplies the hardest numbers in the book: the 24%/76% outcome-vs-T&M split, the 5–10× productivity range, and the on-the-record statement that productivity is *eroding* revenue. The other seven are MEDIUM. The wording of the argument is reliable. But every dollar figure, percentage, and multiple drawn from auto-captions — Sorrell’s “\$2.5 million → 500,000” and “30%,” Rampell’s market figures, Kalra’s growth rates — should be confirmed against the source audio before it is printed as an exact pull-quote. Auto-captions occasionally mis-transcribe numbers and proper nouns. No quotation here is paraphrased or invented; each one exists in the cited recording.

Gaps, named rather than filled. Rich public material exists from WPP (Mark Read; Cindy Rose) and EPAM (Balázs Fejes), but I couldn’t secure a clean, verifiable A/V or transcript source for a verbatim pull, so none is asserted here. Several further targets — Thoughtworks, Capgemini, Code & Theory, Media.Monks, and others — are identified for a future pass, not yet mined to the same standard. The Index is a living instrument; it grows as the record does. It isn’t complete, and doesn’t claim to be. What it claims is narrower: every line in it is real, attributed, and checkable — which is the kind of evidence I’d want before pricing a strategy on it.

APPENDIX D

The Lexicon

The Repricing — the event this book is about. AI collapses the cost of *producing* the artifact. So the money in software and creative services moves to whoever carries the *consequence* of the result. It is a permanent change in what a dollar of agency work buys, not a downturn. (*Chapter 1*)

The Bifurcation — the splitting of the agency dollar into two halves that now behave differently. One is a **production** half racing toward zero. The other is a **judgment** half that holds or grows. The invoice that blends them is the one procurement crushes. (*Chapter 1*)

The carry line — a plain name for the split on any engagement, not a framework to adopt. On one side is the work you merely perform: the keystrokes, the screens-from-a-brief, the boilerplate — what the model now does cheaply. On the other is the half the buyer still pays for: judgment, integration, verification, accountability — the work you **carry**. (*Chapter 2, “What Still Gets Paid For”*)

Carry / the carry (*noun*) — the work above the line that the buyer still pays a premium for, *if* you can prove it. It is the model-proof part. It is defined by what the model lacks: a stake in being right. (*Chapter 2*)

The four carries — what the carry is made of. There is **judgment**, deciding what to build and what to kill. There is **integration**, making it work inside the client’s existing tangle. There is **verification**, catching what the model got confidently wrong. And there is **accountability**, the name on the indemnity. (*Chapter 2*)

Carry ratio — the share of a fee that sits in the carry. It is probably the most important number about a firm, and it shows up on no

dashboard. Audited honestly, the ratio often comes in lower than a firm had assumed. (*Chapter 2; Appendix A, Tool 1*)

The slice — a bounded operational result a firm can name, price, deliver, verify, and defend, with edges a procurement officer can trace. It is a *graspable* project, not a smaller one. It is the first move to get paid for the carry. (*Chapter 3*)

The Five Parts — what every slice needs, or it collapses back into a blend: a **buyer**, a **boundary**, a **proof standard**, a **risk surface**, and a **commercial wrapper**. (*Chapter 3*)

The four functions — the procurement committee, which is not one person. The **CFO** asks whether the price is legible. The **VMO / Vendor Management Office** asks about continuity: does it survive a key person leaving? **Third-Party Risk / TPR** weighs security, data, and model exposure. **Model Risk Management / MRM** asks what catches the model when it is wrong. Any one of them can kill a deal on its own. (*Chapter 1, Chapter 7*)

Own the risk you control — the discipline that splits risk instead of swallowing or refusing it whole. You accept, in writing, exactly the risks you control. You refuse, just as clearly, the ones you don't. The border between the two lists is the carry you will stand behind, drawn on the contract. (*Chapter 3, Chapter 4*)

The Slice Order — a short rider that hangs a bounded slice off an existing master agreement. It varies only the few terms the slice needs and points back to the master for the rest. So you reprice without reopening — and losing — the master. (*Chapter 3, Chapter 4*)

The Do-Not-Sign Carry Checklist — the clauses that hand a firm carry its balance sheet cannot hold. The list: uncapped AI indemnity, uncapped breach liability on the buyer's systems, outcome SLAs on metrics you don't control, open-ended consequential damages, and fitness-for-purpose on the buyer's data. It is a hard stop, returned to the

buyer as a *pricing* decision rather than a legal one. (*Chapter 4; Appendix A, Tool 4*)

The provenance folder / the five fields — the minimum record, logged as the work happens, that turns “we use AI” from an unprovable claim into evidence. The five fields: model & version, inputs & permission, who reviewed it and what they changed, prompt-and-decision trail, and license pass-through. The one discipline pays off three ways. It makes the narrowed indemnity defensible, it clears the audit, and it defends you in a dispute. (*Chapter 5; Appendix A, Tool 6*)

Generative Provenance — the creative-shop version of the folder. It records which generative model made or assisted each asset, the data/licensing attestation, the copyright pass-through chain, and the human-review log. The code shop uses it to prove its build, and the creative shop uses it to prove its originality. (*Chapter 5*)

The gap / the Transition Trap — the cash-shadow of pricing the carry. The new model gets paid *later* than the old model stops getting paid, while payroll stays where it is. It is measured in weeks of payroll, not margin. That is why a firm can be right about the strategy and still not survive it. (*Chapter 1, Chapter 6*)

The thirteen-week cash-flow forecast — the distressed-operator’s tool. It tracks cash in and cash out, week by week, with every receivable dated to when it *actually* clears. It finds the exact date the cash runs out, while that date is still far enough away to act on. (*Chapter 6; Appendix A, Tool 7*)

The reshape / the four layers — the redrawn org chart of the work above the line, sorted by responsibility rather than title. The **Editor** holds senior judgment, accountable for what good looks like. The **Verifier** checks the machine against reality and owns the rescue. The **Operator** runs the models and the pipeline — the new production craft. The **Apprentice** is the early-career person who has to learn

judgment without the grunt work that used to teach it, and who is the succession plan for the Editor bench. (*Chapter 7*)

Keep the room — the final move. You retain the people in whom the carry physically lives, the Editors and Verifiers. The capability that lets a firm survive the repricing sits in a handful of heads. They can read the same uncertain future you can, and they can also walk out the door. (*Chapter 7*)

The Repricing Index / the Quote Bank — the book's evidence base. It collects attributed, on-the-record statements by named executives, mined from the public record: earnings calls, keynotes, interviews. Each one is coded against the thesis and confidence-tagged. (*Appendix C*)

The refrain — the doctrine in five imperatives, in order: **Sell the slice. Sign what you can carry. Show your work. Survive the gap. Keep the room.**

A NOTE ON SOURCES AND METHOD

Where the Facts Come From

This is where the facts come from, how they were checked, and what was left out. A book that says you should prove your claims owes the reader the same.

IN THIS CHAPTER

- What's on the public record, and how confidence was tiered
- Which figures are real disclosures and which are illustrative
- What was deliberately left out, and why

This book makes a sharp claim about a fast-moving moment. A reader is right to ask how much of it is proof and how much is just my say-so. The honest answer: the two were kept on separate shelves and labeled. This note explains the labels.

Everything attributed to a named person is on the public record

There's no anonymous survey behind this book. No made-up "industry executive." No off-the-record chat rebuilt from memory. Six people are quoted: Sir Martin Sorrell, Martín Migoya, David Droga, Sandeep Kalra, Alex Rampell, and John Cotterell. Each quote is copied word for word from a recording anyone can find — an earnings call, a conference keynote, a recorded talk. The full list — speaker, role, firm, venue, date — is **Appendix C, the Repricing Index**.

That's the same rule the book asks of an agency: keep a provenance folder. A claim you can't source tends to fall apart the moment someone leans on it. (Provenance just means a paper trail back to the source.)

Confidence is tiered, and the tiers are visible

Lines drawn from a published earnings transcript — chiefly Endava's — are marked HIGH. Those carry the book's hardest numbers: the 24%-outcome-versus-76%-time-and-materials split, the five-to-ten-times productivity range, and the on-the-record point that productivity is *eroding* revenue.

Lines drawn from auto-captioned video are marked MEDIUM. The wording of the point holds up. But auto-captions can mis-hear a name or a number. So each dollar figure, percent, and multiple from those sources is flagged to be checked against the audio first. A figure from a MEDIUM source is given as the speaker's stated number — not as an audited fact.

The corporate financials are from primary disclosures

A handful of figures come straight from the companies. Globant's AI Pods figures and quarterly revenue. CI&T's quarterly revenue and growth. Endava's impairment and margins. Atlassian's headcount cut. Madwell's debt and shutdown. These come from earnings materials, company notices, and — for Madwell — court documents in the trade press. They are the spine of the case files in **Appendix B**.

The dollar figures inside the playbook are illustrative, and say so

The book walks through a \$1.8 million build, a \$2.5 million engagement, a \$300,000 invoice stuck in accounts payable, and a thirteen-week forecast. Those numbers are worked examples, not reports of real

deals. They're there to show how a thing works at a scale the reader can feel. How a cap's carve-outs play out in the math. How a milestone conversion opens a cash trough.

They're labeled illustrative wherever they appear. But the moving parts they show are real and standard. Each one was checked against how these tools really behave: borrowing-base cross-aging, the E&O contractual-liability exclusion, the way service-level credits stack outside a liability cap. (E&O is errors-and-omissions insurance.)

The cautionary cases got extra care

Cautionary cases invite carelessness, so these got a harder look. The Builder.ai story is the clearest example. The splashy version is that the company faked its AI with hidden human labor. That account is *disputed*, and the book treats it as unproven, because careful reporting pushed back on it.

What the book uses instead is the narrower, safer account. There was a real AI product and a real network of human contractors. Then came an insolvency, an *alleged* revenue irregularity the named counterparty flatly denies, and an open probe. The lesson here doesn't rest on the fraud claim at all. It rests on one fact no one disputes. Once scrutiny arrived, no one could cleanly say how much of the AI was real. Where a claim is contested, the book says so in the same sentence.

Two cases are fictional, and flagged every time

The signed-carry death in Chapter 4 and the lost-renewal in Chapter 5 are composites. They're built from real mechanics but tied to no real firm. Both are marked as illustrative in the text and in Appendix B. The other option would be dressing a made-up story as a reported one. That is the exact dishonesty this book argues against.

The frameworks are the author's, and the vocabulary is new on purpose

Several terms are the book's own coinages: the slice and its Five Parts, the four functions, the four carries, and the four layers of the reshape. They're defined in **Appendix D, the Lexicon**. They aren't standard industry terms, and the book doesn't pretend they are. They earn their place only if they help a reader act. That's the test the Lexicon holds them to.

What was left out is named, not hidden

Some firms had rich public material — WPP, EPAM, and others. They were checked, but didn't yield a clean, checkable quote to the standard above. So no quote from them is used. The gap is logged in Appendix C rather than papered over.

The rules moved while the book was being written. The biggest shift: the European Union deferred the AI Act's high-risk obligations from August 2026 into 2027–2028, under its “Digital Omnibus” simplification. The text reflects where things stood as it was written. The caveat is explicit: the buyer's procurement questionnaire doesn't wait for the regulator's calendar.

None of this is professional advice

None of this is legal, accounting, tax, or financial advice. It's how operators reason before they hire the people who give that advice. The contract mechanics, insurance behavior, lending covenants, and accounting treatments here are real. As the author understands them, they're correctly described. But each firm's case differs. The worksheets in Appendix A are starting points for a talk with your own counsel, broker, and lender. They aren't stand-ins for one.

The aim here was simple. It matches what the book asks of the work.
A claim you can check beats one you can't.