The Next Computing Bill

When AI takes over the work, whose bill shrinks and who gets paid?

An essay by Tarry Singh



FIVE CONTRACTS / ILLUSTRATIVE BANK RENEWAL

Where the $1 million goes

Existing suppliers	AI supplier fees	Transition costs	Kept by the bank
$828,000	$79,000	$21,000	$72,000

40,000 narrative words / 38 exhibits

THE READING EDITION

Contents

Prologue	The second signature	4
Chapter 01	What the float bought	7
Chapter 02	The income statement underneath Claude	13
Chapter 03	When the computer sells the work	20
Chapter 04	How large is the bill today?	26
Chapter 05	The cloud sends two invoices	31
Chapter 06	Two routes into the client	38
Chapter 07	The consulting partner’s arithmetic	44
Chapter 08	What becomes of a seat?	51
Chapter 09	The database collects its toll	57
Chapter 10	A bank that keeps its mainframe	63
Chapter 11	Five industries rewrite a purchase order	70
Chapter 12	The wage bill beyond the CIO	78
Chapter 13	Who keeps the saving?	84
Chapter 14	A company larger than the cloud	91
Chapter 15	The case for the incumbent	97
Chapter 16	The terms of the next renewal	101
Coda	Whose signature counts	105
Notes	Sources and reading notes	108
Index	The 38 exhibits	118

THE ARGUMENT IN BRIEF

The investment case at renewal

Start with the buyer

A cloud supplier sells capacity. A software company sells access. A consultant sells time or a deliverable. An AI supplier that promises to close a bank's case can bid for work on all three contracts. It must still pay for the machines, the records and the people who repair its mistakes.

Price the counteroffer

Chapter 11 gives one bank five contracts worth $1 million a year. Ordinary renewal cuts the bill to $952,000. Three winning AI bids take it to $928,000. The AI supplier collects $79,000 in fees. The additional benefit to its customer is $24,000. The buyer, offers and cash flows in this exercise are illustrative.

Read the shareholder's accounts

The buyer's saving tells a shareholder little about the cost of earning the next fee. The prospectus records the lab's sales, expenses and contracted capacity. The closing notice distinguishes the shares Anthropic sold from shares sold by existing holders. The five-market exercise traces cash the public filings cannot identify.

Set a date for the prediction

The essay asks whether an AI computing company can overtake a major cloud segment on trailing four-quarter recognised revenue by 2031. Comparable renewals, reopened cases and the costs carried by each supplier give the 2028 review something to test before then.

Reporting through 20 November 2026. Read the prospectus, the closing notice and the five-contract comparison. The web edition contains the controls for all 38 exhibits.

PROLOGUE

The second signature

Two disputed charges for a food order lead to a six-month correspondence. An investor in AI has an interest in finding out what it costs to close that file.

An online food order left two charges on a Metro Bank debit card, one for £38.78 and one for £40.58. The customer believed he had been charged twice. He approached the merchant, then the bank. That was in March 2024. By September the deadline for raising a chargeback had passed. The bank said it had asked for evidence it had never received. The customer said he had sent everything he had. Two amounts under forty-one pounds apiece had acquired a six-month correspondence and, eventually, an entry in the public record. [117]

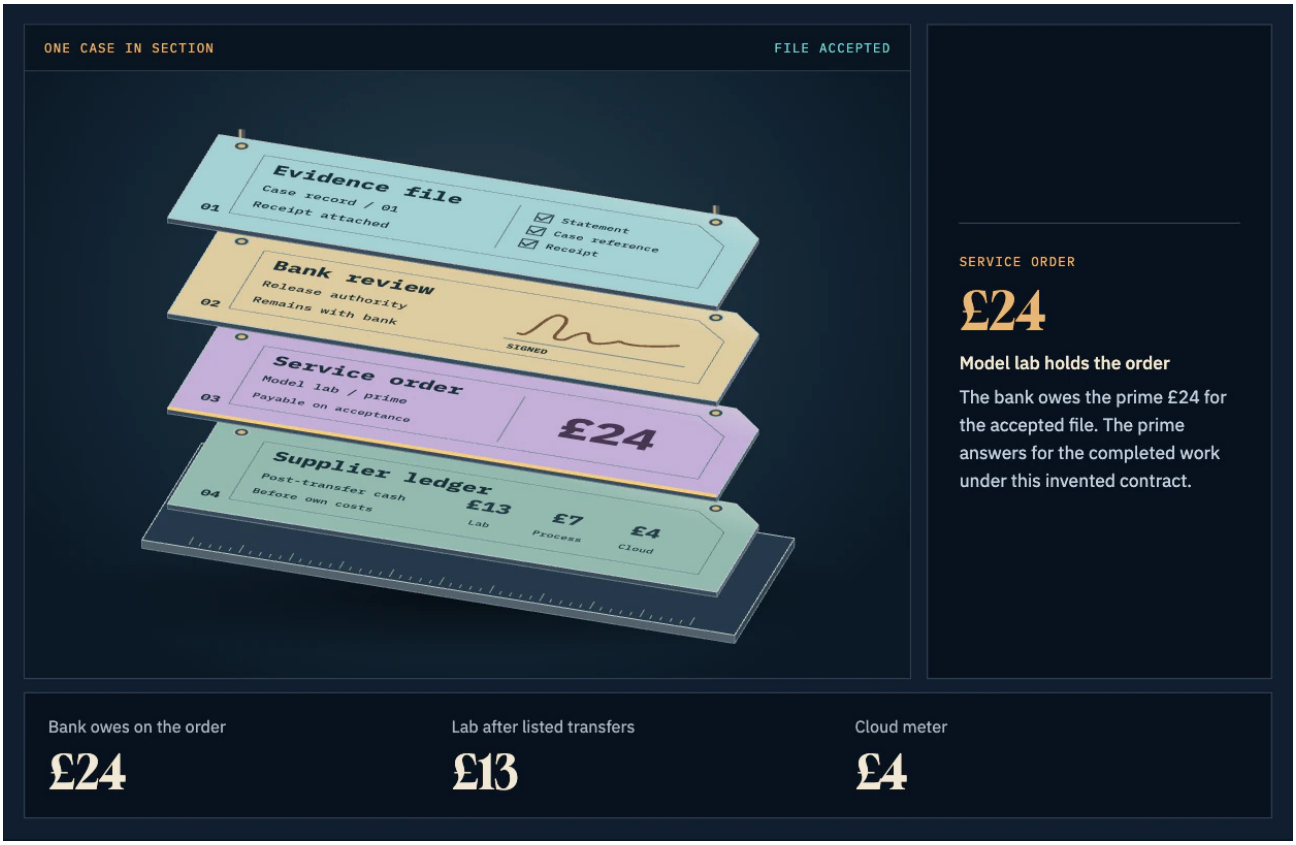
The Financial Ombudsman Service decision makes careful reading. The ombudsman found that Metro had reasonably asked for further information and had acted reasonably in declining to pursue the chargebacks. Metro had also acknowledged shortcomings in its handling of the complaint. The ombudsman decided that the bank should pay the £100 it had offered. That sum recognised the service the customer received. It tells us nothing about whether the merchant collected twice for a single order, and remarkably little about the cost of answering the question. Someone had to read the correspondence, retrieve the transactions and establish what Metro could still do. [117]

If you were going to build an AI computing company, which part of that file would you offer to buy back from the bank? You could sell the investigator a better search tool. You could supply the infrastructure on which the search runs. You could offer to assemble the evidence, identify what is missing and deliver a checked file for a fixed fee. The bank might accept the third proposal while keeping its database, its complaints team and the authority to pay the customer. A software supplier would then have acquired a piece of the bank’s operating business without replacing a single core application.

On Thursday, 19 November 2026, a price appeared on Nasdaq for one company capable of making such a bid. Anthropic had offered its shares at $83. The first trade was $111.50 and the close was $97.17. The following day the company announced completion of the offering and a concurrent placement with seven technology companies. Several have businesses selling the infrastructure on which model developers depend. Shareholders paid money into a company that may use some of it to pay those very suppliers. A buyer considering an AI contract faces a smaller version of that circularity. A familiar cloud company can appear on both sides of the negotiation. [01] [02] [03]

EXHIBIT 01 / AN ILLUSTRATIVE SERVICE CONTRACT

One bill through the enterprise



Four separated sheets trace an accepted bank file from evidence and release authority to a £24 service order. The lab retains £13 after the listed transfers.

An illustrative service contract in pounds. One completed computation costs the lab £4. Both prime bids pay only when the bank accepts the file. The £24 fee excludes compensation or a refund due to the customer. The service prices and contracting parties in this instrument are fictional. The opening Metro Bank dispute is a separate published case. Ombudsman decision

Open exhibit 01 in the interactive edition

The file in the illustration is an invented purchase, with its own prices. A bank promises £24 for an accepted case. Before acceptance the lab has already incurred £4 in computing charges. A missing receipt would be enough to put the fee on hold while leaving the bill for that first run intact. Give the order to the lab and it buys help from the processor. Give the same order to the processor and it buys intelligence from the lab. The bank gets the same work for the same money. Investors in the two suppliers would have quite different accounts to read. The fee buys a file in a condition to be decided. Any redress due to the account holder requires another entry in the bank's customer accounts.

My prediction is that a company which bids for completed work can grow larger, on recognised annual revenue, than a major cloud business. I expect one to pass at least one of two named businesses, Google Cloud and Amazon Web Services, in the four quarters ending in December 2031. Anthropic is my leading candidate. Its ability to sell intelligence to other firms is already measurable. The additional claim I want to investigate is that it can take the customer relationship and the price that goes with responsibility for an outcome. A buyer's second signature, the one that renews the service after the first year's mistakes have arrived, would persuade me more than its signature on the pilot.

There are five plausible places to find the money. A smaller seat order can release part of an application budget. A consultant’s next engagement can go to the firm that already performed the first piece of work. A bank can keep an installed system and change who gets paid to operate around it. The provider of a data platform may collect more for answering questions even as its customer spends less preparing each answer. A cloud operator can lose the prime contract and keep supplying the machines. Adding up the size of those industries would count some of the same customers and many of the same dollars several times.

There is also a customer outside the procurement department. The person who waited months for a reply might value an answer tomorrow even if the bank saves nothing. An employee who can deal with the exceptional file before a deadline may justify a new service with a larger annual budget. Those purchases belong in the growth story. Evidence of displacement requires a second document showing what the buyer stopped paying for. On a trading screen an increase in revenue looks the same either way. Inside the bank, one purchase has made a rival poorer and the other has brought an unanswered question within reach.

Before following that purchase into the bank, we should count what actually reached the new public company. The first mistake an investor could make with this story fits on the first page of its prospectus. Some of the proceeds belonged to people selling shares they already owned.

CHAPTER 01

What the float bought

Trace the cash from the public sale and private placement. Then change the share count underneath the price that made the headlines.

A thousand Anthropic shares cost $83,000 at Wednesday's offering price. At Thursday's first Nasdaq trade they were worth $111,500. Someone who bought at that price held a position worth $14,330 less by the close. Neither move changed what Anthropic received for its newly issued shares. On November 20 its closing announcement put the combined transaction at approximately $119.5 billion. The wording matters. Some of that money belonged to people selling stock they already owned. [01] [02] [03]

Three different claims on the business acquired a price that week. New shareholders paid for an interest in its future. Suppliers of chips and computing capacity paid to own a piece of a customer. Existing owners found a public market for part of their holdings. The transaction gave Anthropic the means to compete for work now sold by other companies. It also gave some of those companies a financial interest in its success. To understand who might ultimately collect the enterprise computing bill, start with who wrote the first cheques.

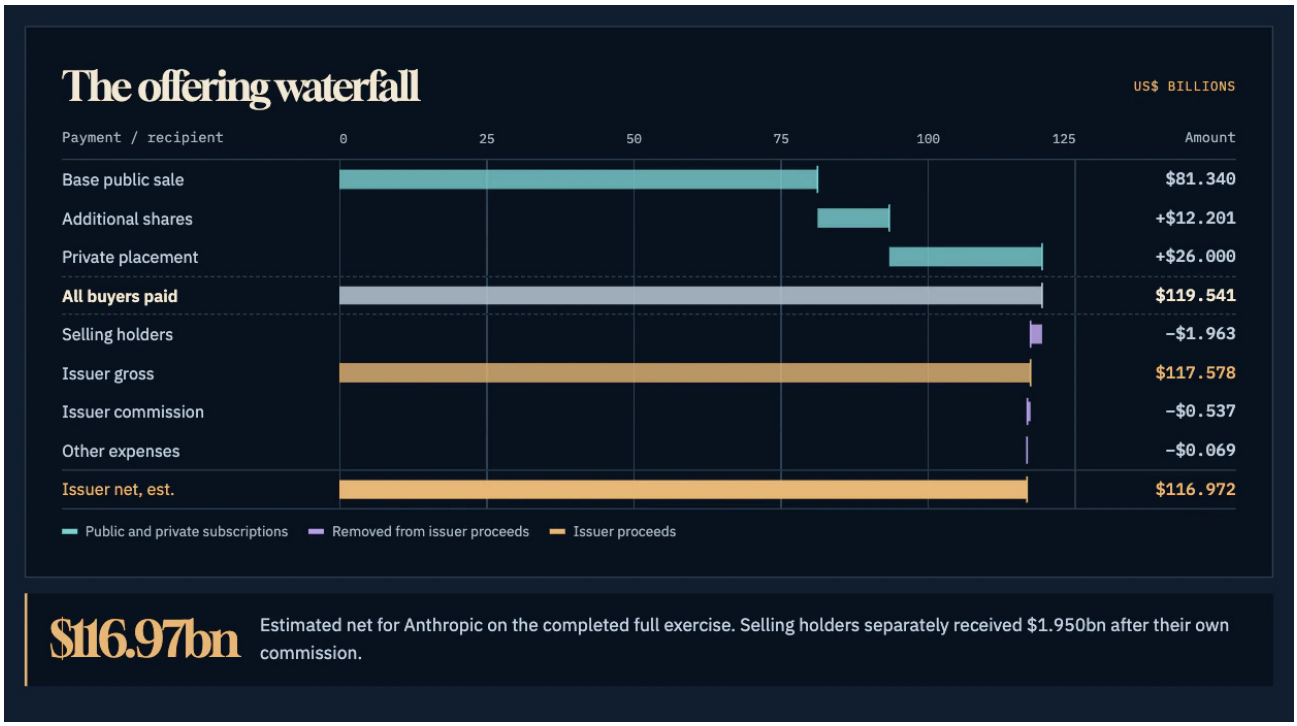
A hundred and nineteen billion dollars changes hands

The public offering priced on November 18 at $83 a share. The front page of the prospectus lists 980 million shares. Multiply and the familiar $81.34 billion appears. Turn to the division between sellers and the new capital shrinks. Anthropic sold 956,351,354 shares, for $79.377 billion before expenses. Existing holders sold the remaining 23,648,646 for $1.963 billion before their commissions. A secondary share changed hands at exactly the same price as a new share. It funded its previous owner. Issuing the other shares put cash at the disposal of the company and spread the claim on its future earnings across more owners. [01]

The underwriters held an option on 147 million additional new shares. When the press reported the first session, exercise was still described as a possibility. The closing notice the following day records exercise in full. That added $12.201 billion of primary capital and lifted the public sale to 1.127 billion shares worth $93.541 billion. A separate placement sold 313,253,007 newly issued Class N shares at the same $83. That brought in almost exactly $26 billion. Add public and private sales and the press release reconciles to $119.541 billion gross across all sellers. [01] [02] [03]

EXHIBIT 02 / OFFERING PROCEEDS

Where the offering went



An offering waterfall traces $119.541bn paid across the public sale and placement to estimated net proceeds of $116.972bn for Anthropic.

Buyer payments include secondary stock. Anthropic pays the commission on its base shares and an estimated $69m of other offering costs. Its underwriting commission is unchanged under full exercise. Historical September cash belongs to Figure 04.
Prospectus, offer and underwriting tables . Closing release

Open exhibit 02 in the interactive edition

Follow the money one step further. Removing the secondary block leaves $117.578 billion gross for Anthropic. Its share of underwriting discounts was $536.7 million. Its estimated share of other offering expenses was another $69 million. The underwriting table contains an unusual detail. The issuer's total commission is identical in the columns for no exercise and full exercise. Neither the additional shares nor the concurrent placement carried an underwriting commission. Deduct the disclosed charges and the company's estimated net comes to $116.97 billion, consistent with the prospectus figure of about $117.0 billion. The sellers paid their own commission. Charging that to Anthropic would move an expense between two different parties. [01]

The register of private buyers reads like a purchasing department's address book. Google and NVIDIA subscribed for roughly $10 billion apiece. AMD put up $2.5 billion. Broadcom, Micron and Microsoft took approximately $1 billion each and TSMC Partners about half a billion. These were nonvoting Class N shares carrying economic rights equivalent to Class A. Seven industrial investors thus put up nearly twenty-two cents of each gross dollar in the combined transaction. The documents distinguish the securities from the ordinary public sale, even though the price was identical. [01] [02]

Consider the investment from the other side of the table. A supplier can earn money selling infrastructure to a lab whose competition it also fears. An equity stake gives it an additional claim on the value of whatever the lab builds above that infrastructure. The same supplier can still press for attractive terms when the compute contract comes up. Nothing in the placement allocations

tells us which customers have signed a workflow contract, the price of a chip purchase, or how much profit either party will retain. Each of those questions calls for a different document.

What traded at two trillion

At the first Nasdaq trade on November 19, $111.50 bought one Class A share. By the close the price was $97.17. Reuters reported a valuation around $2 trillion on a fully diluted basis and up to approximately $1.5 trillion on outstanding shares. One can reproduce the scale of both claims from the prospectus without inventing a price between those observations. The choice of share count accounts for a larger difference in the opening valuation than the entire amount of fresh capital the company raised. [01] [03]

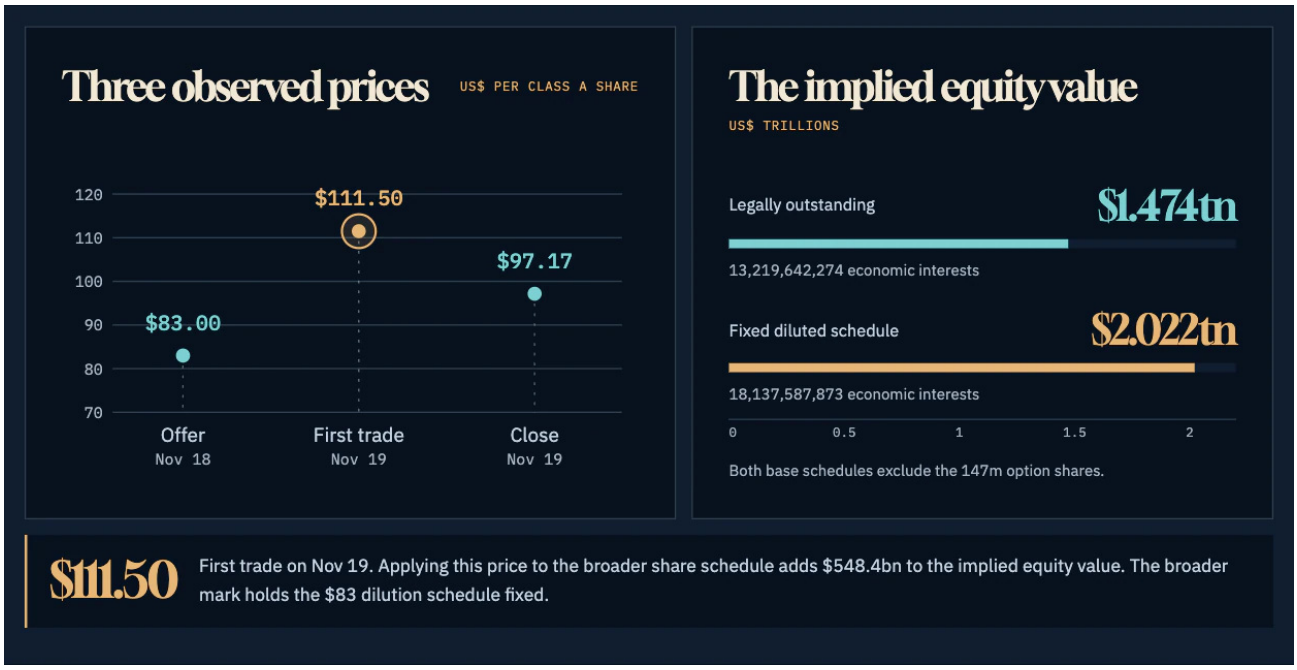
The base offering terms give 7,583,103,996 Class A shares and 5,636,538,278 Class N shares legally outstanding after the sale and placement. Give each economic share the traded price and the resulting 13.220 billion shares were worth about $1.097 trillion at the offer, $1.474 trillion at the open and $1.285 trillion at the close. A single Class T share and a single Class F share carried governance rights without economic rights. The company’s accounts also exclude from shares outstanding the Class N shares held by its wholly owned public benefit subsidiary while it holds them. The legal and accounting columns answer different questions. [01]

Suppose someone instead asks how many economic shares the offering terms imply after full option exercise. Adding 147 million takes the legal count to 13.367 billion. At $83 those extra shares contribute the additional cash raised because they were issued at $83. At Thursday’s closing price their marked value would be about $14.28 billion, against the same $12.201 billion paid to the company. Dating the share schedule and dating the trade independently makes that difference visible. [01] [02] [03]

A separate prospectus table offers an economic share count of approximately 18.138 billion after specified additional adjustments. Among them are shares underlying rollover notes, shares reserved under the donation matching plan, restricted stock units and options. The table credits the company for shares it could repurchase with the proceeds of option exercises at $83. Mark that fixed schedule at the three prices and the corresponding values are approximately $1.505 trillion, $2.022 trillion and $1.762 trillion. Both schedules use the base offering. Neither includes the 147 million extra shares confirmed at closing. Recalculating the option adjustment at a new market price would itself change the diluted count. [01]

EXHIBIT 03 / TRADING AND SHARE COUNTS

The first day on two share counts



Three discrete points mark $83.00 at offer, $111.50 at first trade and $97.17 at close. The first trade implies $1.474tn on legal shares and $2.022tn on the fixed diluted schedule.

The three price observations are discrete. Reuters separately reported around $2tn on a diluted basis and up to approximately $1.5tn on outstanding shares. No intraday maximum is inferred. Prices on the traded class are applied equally to nonvoting economic shares. Prospectus, pages 20 and 26 . Reuters first-session report

Open exhibit 03 in the interactive edition

These are marks on all economic interests at a price observed on the traded class. The calculation assumes the same price for a nonvoting restricted share and a publicly traded share. It gives no special discount to a large block whose holder might need months to sell. The buyer of a thousand shares could transact close to an observed price. No order book in the record establishes that every owner could exit together at that price. And when Thursday's buyers paid each other a premium over $83, the company did not acquire a second instalment of primary capital.

What the cash comparison can bear

The Financial Times graphic that prompted this investigation placed Anthropic beside seven familiar technology companies and separated an optional final financing block. The underwriters have since exercised that option. Rebuilding the comparison from financial statements requires one more decision. Corporate cash is often reported alongside securities that are easy to sell. A consistent comparison takes cash and cash equivalents from each balance sheet and leaves separately reported investments in a different column. The choice has a measurable effect even inside a single company's filing. 02 04

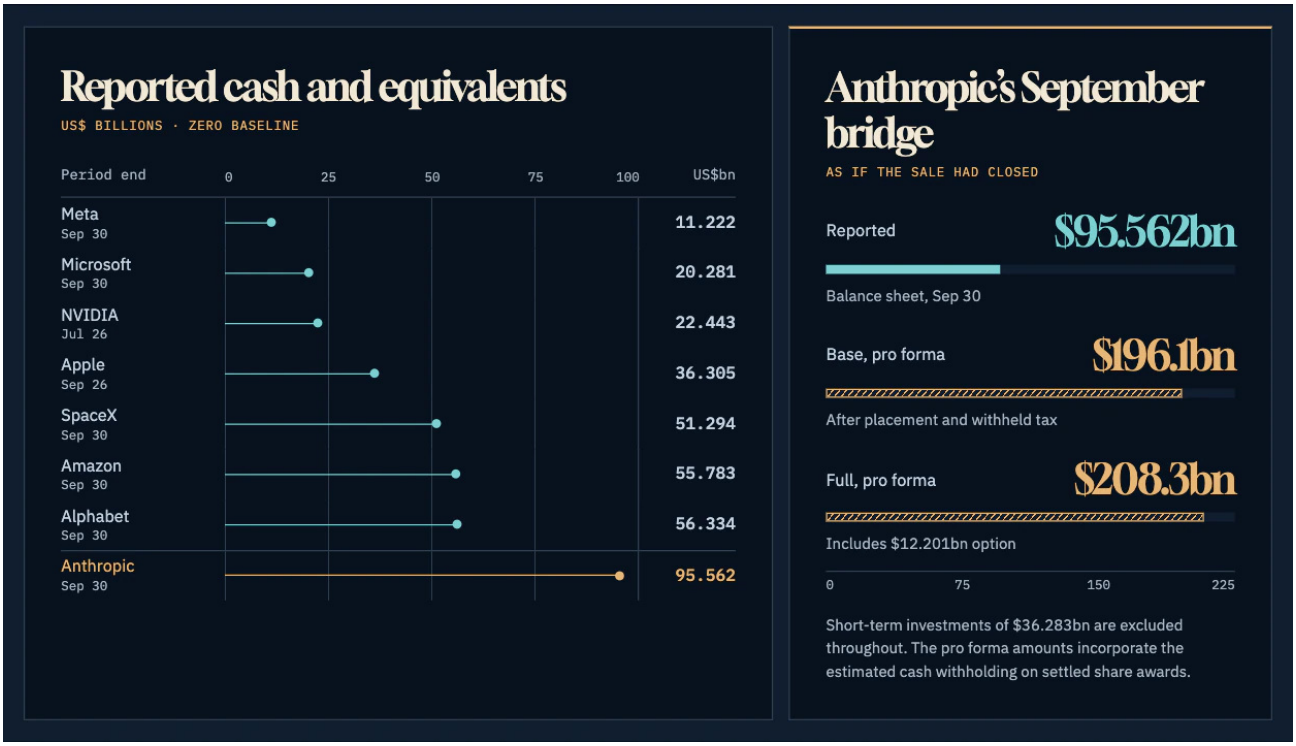
Take SpaceX. Its September balance sheet reports $51.294 billion of cash and equivalents and $38.561 billion of marketable securities. Its cash flow statement ends with $53.386 billion because that statement also includes restricted amounts. An analyst could lift three plausible liquidity figures from the same report and move the company substantially on the chart. We use $51.294 billion. The same rule gives $11.222 billion at Meta, $20.281 billion at Microsoft, $55.783 billion at

Amazon and $56.334 billion at Alphabet on September 30. Apple closed its financial year four days earlier with $36.305 billion. [05 06 95 96 99 100]

NVIDIA poses a question about time. The latest quarter it ended by September 30 closed on July 26, with $22.443 billion in cash and equivalents. Its next quarter ended on October 25. By November 17, before Anthropic's float, the company had reported $29.089 billion. Either comparison has an intelligible use. The earlier observation respects a common end date. The later observation captures the information available to an investor at the offering close. The chart lets the reader change that rule and see exactly which dot moves. [97 98]

EXHIBIT 04 / DATED CASH BALANCES

Cash on the same basis



Eight balance-sheet cash figures share a zero baseline. A separate hatched panel takes Anthropic from $95.562bn actually reported to about $208.3bn under full-exercise assumptions.

Reported ends in 2026 differ by fiscal calendar. No marketable securities or separately classified restricted cash are included. All seven comparison peers in the supplied FT graphic could be reconciled to their original accounts. Graphic supplied by the author . Anthropic balance sheet and capitalization . NVIDIA October update

Open exhibit 04 in the interactive edition

Anthropic itself had $95.562 billion of cash and equivalents on September 30, already more than any of the named peers on this particular measure. Its separately reported short-term investments were $36.283 billion. The prospectus presents the combined $131.845 billion and constructs a balance sheet as though the base offering and placement had occurred on that date. Among its adjustments is cash to meet roughly $4.3 billion of tax withholding associated with settlement of employees' restricted stock units. The filing prices the awards at $83 and assumes a 50.90 per cent withholding rate to calculate the cash the company remits. A shareholder trying to reconstruct cash by simply adding the new capital would overlook that payment. [01]

After those adjustments the filed base figure for cash, equivalents and short-term investments is $232.351 billion. Removing the same investments gives $196.068 billion of cash and equivalents. Full exercise adds $12.201 billion with no further underwriting discount, taking the reconstructed

cash figure to about $208.3 billion. The resulting combined figure rounds to the $244.6 billion specified in the prospectus. November's closing release confirms how many shares were sold. It supplies no closing-day cash balance. A treasurer would need the intervening receipts, disbursements and investment movements to turn either September calculation into one. [01] [02]

Buying time to compete

The chart also puts an awkward limit on the claim that a newcomer can now outspend a cloud incumbent. Microsoft's cash line belongs to the entire group. Its treasury can hold a smaller operating balance while placing funds in short-term investments. A profitable incumbent can use next quarter's customer receipts to finance construction, even if a competitor has more cash today. Measuring their relative buying power would require maturities, incoming cash, available borrowing and outstanding commitments on both sides. A rank on one balance sheet line settles none of those quantities. [96]

What can that sort of balance buy? The prospectus gives management discretion to finance compute, infrastructure, operating expenses and hiring. It allows acquisitions, while identifying no agreement for a material acquisition at the time. It also allows the company to hold proceeds in interest-bearing instruments pending use. A dollar that stays in a short-term security can come back with interest. A dollar committed to capacity with a minimum purchase has a different future. A shareholder owns both decisions, including the cost of equipment or capacity ordered for demand that arrives late. [01]

For an enterprise buyer, a large balance matters when a supplier asks to take responsibility for a process that cannot be switched off. Someone has to pay engineers to connect old systems, carry the expense of parallel operation until the buyer accepts the change, and stand behind failures after launch. A lab with substantial capital has the option of funding that transition before the first profitable renewal. A cloud provider or consulting firm can offer its own financing and point to years of running the work. The decisive evidence will be a contract showing which party carried the transition cost and which party earned the renewal.

An underwriting syndicate can clear a stock offering in days. A bank replacing a system that records customer claims must wait to learn which decisions survive scrutiny and which invoices disappear from its accounts. The transaction leaves those two clocks running at different speeds. My wager on AI computing begins at the point where the first clock stops. Follow the proceeds into the capacity contracts, then follow the capacity into customer accounts that keep paying.

CHAPTER 02

The income statement underneath Claude

Seven quarters of sales, the price of convertible financing and years of reserved computing capacity explain what a shareholder has actually bought.

In September 2026 Akamai signed two project plans to supply computing capacity to Anthropic. On its schedule for the final quarter of the year sat approximately $1.7 billion of capital expenditure and no revenue from the two plans. The supplier expected to spend money on components, including memory, before the first day of service. Each seven-year term would begin when its service began. The customer had made a commitment worth approximately $11.6 billion, subject to delivery, availability and termination provisions. Here was a sale big enough to change a public company's investment plans without putting a dollar into that quarter's revenue forecast. [107] [108]

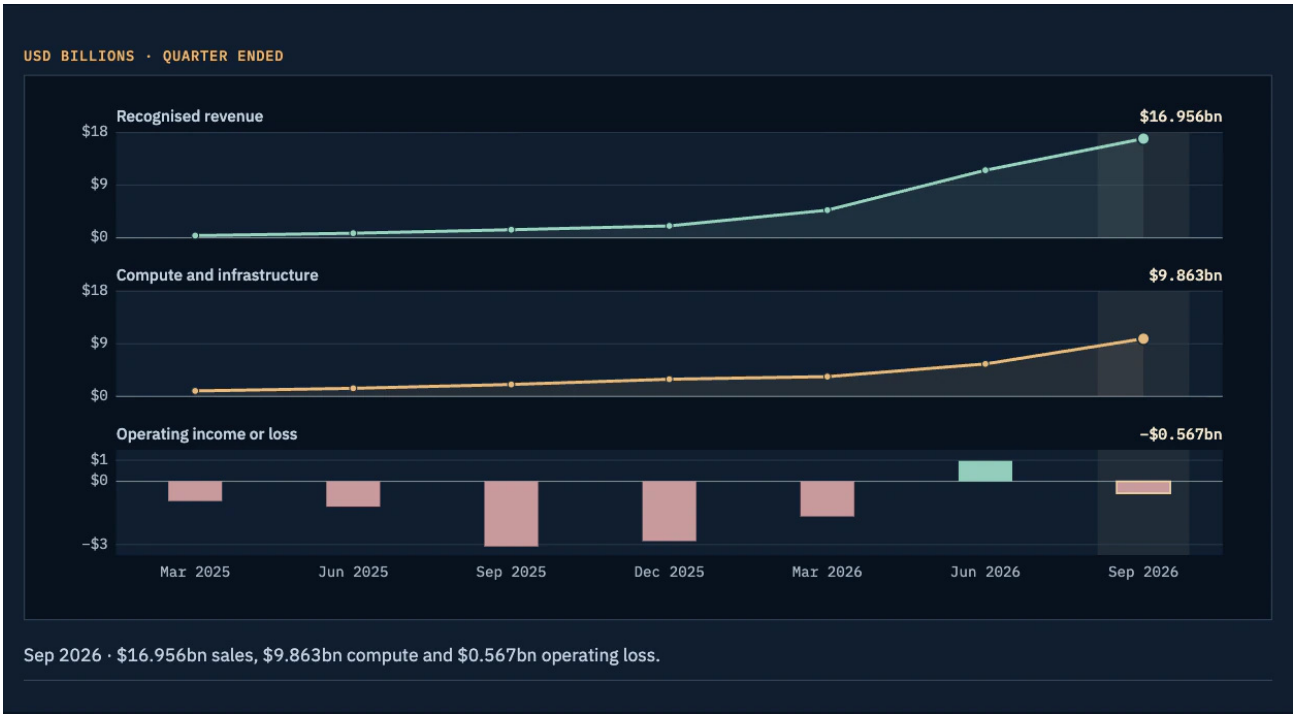
On November 20 Anthropic announced that it had closed its public offering. It received approximately $91.6 billion of gross proceeds from its own public shares and approximately $26 billion from its concurrent private placement. None of those dollars appear in the September income statement. They belong to a later financing transaction. Yet much of the claim on those dollars came into existence earlier, in contracts like Akamai's. Anyone trying to value the company has to keep three dates in view. When did somebody sign for capacity? When did the provider make it available? When did a customer use Claude and generate revenue? [01] [02]

The profitable quarter and the one after it

The final prospectus gives us seven consecutive quarters with which to begin. Revenue rose from $376 million in March 2025 to $16.956 billion in September 2026. In the first quarter the company spent $944 million on compute and infrastructure. A year and a half later it spent $9.863 billion. Their relationship changed more than either headline suggests. At the beginning, every revenue dollar faced about $2.51 of compute expense. In June 2026 the figure had fallen to 48 cents. Three months later it was back at 58 cents. [01]

EXHIBIT 05 / QUARTERLY STATEMENTS

Seven quarters of selling intelligence



Three aligned charts track seven quarters of revenue, compute costs and operating results. June 2026 shows a profit, followed by a $567m operating loss in September.

The revenue and computing panels share a scale. Operating results use a separate scale around zero to distinguish June's profit from the six loss quarters. Percentage mode divides each observation by its own quarter's sales. Compute includes training and serving. Final prospectus, pp. 165–167

Open exhibit 05 in the interactive edition

Revenue rose $5.401 billion between those last two quarters. Compute alone absorbed $4.305 billion of the increase. Personnel expense fell $146 million while sales, marketing and partnerships rose $226 million. Other operating and administrative costs increased $2.557 billion. Add those movements and June's $974 million operating profit becomes September's $567 million operating loss. The company identified $5 billion of September compute under short-term arrangements, significantly more expensive than comparable long-term capacity, alongside some $2 billion of noncash donation-matching charges. The full $5 billion is what that capacity cost. Only the difference against the alternative would be a premium. 01

June deserves as careful a reading as September. A reservation bought in anticipation of growth puts a fixed expense under the business. Until the reserved machines fill up, each additional paying user can improve the apparent efficiency of the whole estate. Once they are full, the next user may require a short-term rental at a much higher rate. Which quarter represents the economics of the next million users depends on which machines are ready when they arrive. Anthropic says it typically procures compute twelve to eighteen months ahead of availability. A June model of a September business can be wrong even if its forecast of customer demand is right. 01

Price moved too. Management estimated that new model versions and contractual discounts reduced the sequential revenue increase by about $600 million in the September quarter. The business was selling more work at a changing price. In its nine-month figures, consumption-based revenue supplied about $26.3 billion, or 79 percent of the total. An annual subscription and an

application that calls a model on demand put different obligations on the supplier. Neither tells us how much of the installed fleet was running an experiment on the day a user paid. The filing reports research, product development and serving costs together in the compute line. Subtracting it from sales cannot yield a gross margin on inference. [01]

The customers behind the curve changed as well. Anthropic counted approximately 400 customers generating at least $100,000 of revenue over the previous twelve months in March 2025. By September 2026 it counted more than 11,900. A buyer enters that group through a year of actual purchases. A supplier trying to decide how many machines to order needs to know whether a buyer that crossed the threshold in September will renew in March. Annualising the September quarter gives $67.824 billion. That calculation leaves the renewal question open. [01]

A claim that grows as the company succeeds

For the nine months to September the operating loss was $1.260 billion. Interest income added $2.359 billion, enough by itself to take that subtotal above zero. Then came a $94.324 billion fair-value charge, $10.397 billion of other net expense and a $417 million tax provision. The resulting net loss was $104.039 billion. An analyst who starts with the loss and divides it by nine months can manufacture a spectacular monthly burn rate. The cash-flow statement shows positive operating cash flow. Understanding how the business could report both requires looking at the people who supplied it with capital. [01]

EXHIBIT 06 / EARNINGS PER SHARE

What happened to earnings per share



A loss waterfall isolates the $94.324bn fair-value charge. Separate earnings-per-share cards give a $50.10 historical loss and an $8.78 pro forma loss on different share counts.

Bars move along a common dollar axis from the operating result to the GAAP net loss. Each card divides its own numerator by its own weighted-average share count. Both show basic and diluted EPS. The pro forma column assumes specified conversions and the listing-related award condition occurred earlier. The two reporting periods have different lengths. Final prospectus, pp. 28–29 and Note 4

Open exhibit 06 in the interactive edition

A convertible note can give its holder a claim on shares under agreed conditions. As new financings signal a higher company valuation, the estimated value of that claim can increase before a share has changed hands. Anthropic elected to remeasure its notes at fair value. Its modelling used assumptions about future financings, valuation and the likelihood of conversion. Note 4 attributes the entire nine-month $94.324 billion change in this category to convertible notes. The caption also names derivative liabilities and simple agreements for future equity, commonly called SAFEs. All $122 million of SAFEs issued in 2023 had converted to preferred shares by the end of 2024. None was outstanding during these nine months. [01]

The charge is a poor measure of electricity consumed and a good reason to examine who owns the company. A noteholder's claim does not lose its economic significance when settled in shares. If other things are equal, distributing the same future profits among more owners reduces what each common share can claim. Nor can an analyst remove the whole revaluation from the statutory loss and call the result the prospectus's pro forma earnings. The filing gives a narrower calculation with a precise answer. It changes both the loss attributable to the period and the number of shares across which that loss is spread. [01]

Start with the historical loss of $104.039 billion and a weighted average of 2,076,810,122 common shares. The result is a loss of $50.10 a share, basic and diluted. The pro forma calculation removes $15.541 billion of valuation expense on specified converting notes. It also records $7.487 billion of additional compensation for restricted stock units, applying the listing-related performance condition as if it had been met on January 1, 2025. The service condition determines which awards qualify. The revised loss is $95.985 billion. A company that employs people in return for stock can accumulate a large compensation charge before the event that releases it into the accounts. [01]

Now change the divisor. Conversion of preferred stock adds 8,561,242,048 weighted shares. Conversion of notes adds 292,768,653 and settlement of restricted units adds 3,674,719. The total becomes 10,934,495,542. Dividing the revised loss by those shares gives $8.78 of loss per share. The calculation treats specified conversions as having occurred earlier and weights the shares for the period. Using the offer's closing share count would mix two measurement dates. The striking improvement from $50.10 to $8.78 tells us a great deal about capital structure. The customers served, prices charged and machines rented in those nine months are identical in both columns. [01]

The share count has another trap. The filing describes 16,775,835 net shares issued under the assumed restricted-unit settlement for the nine-month calculation. Their weighted contribution to its denominator is only 3,674,719. A share that enters the calculation late carries less weight than a share present throughout. An investor can ask how much of the company she would own after every conversion and issue at closing. That question calls for a point-in-time capitalisation table. Earnings per share asks how much period income belongs to the shares attributed to that period. [01]

Management offers a third view in its adjusted net income of $5.591 billion. Among its exclusions are the $94.324 billion revaluation, $10.277 billion on certain other marked financial instruments, $3.255 billion of stock compensation and $2.086 billion of share-settled charitable contributions, with a tax adjustment. An operating manager can use this view to see through financing volatility. An owner still needs to know how employees will be paid and what future issuance will do to her interest. As of September, the company identified $24.1 billion of unrecognised compensation on units with an unmet service condition outside the pro forma EPS calculation. [01]

When the money moves

Operating activities provided $6.569 billion in the nine months. The cash-flow statement adds the noncash revaluations back to net loss and then records what happened between the sale and collection. The movement in receivables absorbed $7.076 billion. Accounts payable, accruals and other liabilities contributed $4.377 billion, and deferred revenue added $1.702 billion. A client paying in advance can fund some of the service it will consume next quarter. A growing unpaid supplier balance can finance today's activity until its invoice falls due. Neither timing benefit is a general licence to spend six and a half billion every nine months. [01]

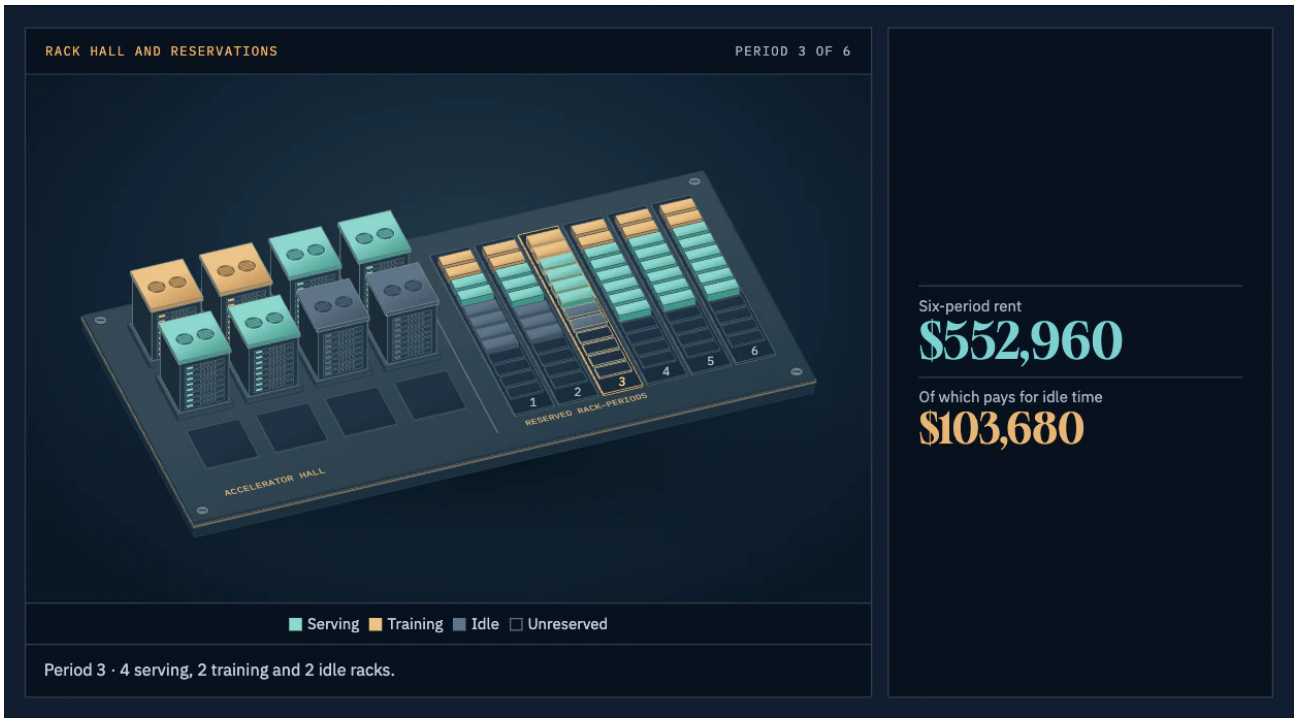
Investing activities used $36.546 billion, including $40.318 billion of short-term investment purchases and $6.900 billion of sales and maturities. Treasury management explains much of the outflow. Purchases of property and equipment were $925 million and prepayments for leased assets and equipment were $1.955 billion. Those are interesting numbers for a business preparing to put more infrastructure under its own direction. There were also $110.735 billion of financing inflows, net, before the public offering. A pile of investable cash can grow while the enterprise is taking on claims against cash it has yet to earn. [01]

The purchase order beneath the rack

At September 30 Anthropic's contractual maturity table totalled $319.993 billion. It included $313.819 billion of purchase commitments with more than a year remaining, $4.224 billion of operating leases and $1.950 billion of finance leases. The table assigned $5.231 billion to the rest of 2026 and $29.165 billion to calendar 2027. Readers sometimes take those figures for the company's entire capacity bill. The table had a narrower purpose. A lease signed for a building that had yet to open was outside it. So were certain short agreements that matter a great deal when the next long-term delivery arrives late. [01]

EXHIBIT 07 / FILED OBLIGATIONS AND A CAPACITY MODEL

The capacity contract



Eight reserved racks are coloured by use. In billing period three, four serve requests, two train models and two sit idle. Idle time costs $103,680 across the six-period example.

Eight identical accelerators per rack, 720 hours per period, no carryover, no on-demand alternative. The six periods' serving requests start at 2, 3, 4, 6, 7 and 8 rack-periods. Adjusted requests round to whole racks. The model charges every reserved accelerator-hour, including idle time. Training is occupied capacity with no assumed current sales. Prices, workloads and hardware here are illustrative.

Open exhibit 07 in the interactive edition

The wider account adds approximately $17.4 billion of short-term commitments at September, $451.2 billion of signed leases yet to commence, $105.7 billion of subsequent longer commitments and $6.6 billion of subsequent short commitments. Together with the table these give the prospectus's approximately $900.9 billion. The company expected about $70.7 billion over the twelve months following September 30. An estimated $186.2 billion within the overall figure depended on commencement conditions or was subject to termination rights. Another estimated $150 billion to $175 billion of payments lay outside the total under arrangements without fixed dollar commitments. Actual capacity, delivery terms and financing determine those sums. [01]

Consider three versions of the promise. Microsoft's original announcement in November 2025 specified $30 billion of Azure compute. The agreement as amended through November 2026 gave a minimum of $47.2 billion and up to $49.2 billion between November 2026 and June 2033, regardless of usage and subject to stated remedies. Google hosting carried a minimum of approximately $121.3 billion from April 2026 through July 2033. Amazon's minimum was $110 billion from May 2026 through April 2036. Their shortfall provisions can require a payment when actual spending falls below the commitment. Adding Microsoft's original announcement to its amended agreement would sell the reader the same capacity twice. [01 105]

Amazon and Anthropic had described their arrangement in April as securing up to five gigawatts. Gigawatts measure power. They do not measure paid accelerator-hours, successful customer requests or the number of kilowatt-hours ultimately delivered. There are contingencies on the

other side, too. Lambda had missed a facility lease deadline on an August agreement representing $34.9 billion of the $41.3 billion across its agreements. At the prospectus date either party could terminate that August contract, and neither had given notice. A supplier's ability to deliver is as consequential to this business as a customer's willingness to pay. [01] [106]

The exhibit gives our illustrative lab eight racks for six equal billing periods. It can run training on a quiet rack or hold it for a paying customer's request. It pays for the reserved time in both cases. An hour left idle in the first period cannot catch a request that arrives four periods later. The supplier has made a parallel calculation. Akamai expected to invest about $5.5 billion across the build for its two plans and to recognise roughly $150 million to $300 million of revenue in 2027. Those were management's forecasts in September. A delay can shift the return for both sides of the same contract. [107] [108]

At the displayed rental price of $2 per accelerator-hour, eight accelerators in each rack and 720 hours per period produce a six-period bill of $552,960. Nine idle rack-periods cost $103,680. Even so, three rack-periods of customer demand go unserved at the end. Paying for spare machines and turning away business can occur on the same contract.

My proposition that an AI computing company can take a larger share of enterprise spending depends on a demanding sequence. The lab has to assemble capacity before demand is certain, sell useful work at a price that may fall, and collect enough of the proceeds to pay the suppliers that made delivery possible. The public offering bought Anthropic time in which to do that. Its next operating results will tell us how much of that time the customers are prepared to buy back.

CHAPTER 03

When the computer sells the work

A token, a user licence and an accepted file buy different promises. Read the invoices, then follow a proposed bank posting to the person who can release it.

South Downs National Park Authority bought 175 one-year Copilot licences. It named the supplier, Digital Devices Ltd, and gave the amount to the penny. The contract, signed on 13 October 2025, was worth £56,555.10 including VAT. That works out to £323.17 per licence for the year. Open the notice and try to find the number of planning applications processed, hours returned to a ranger or letters sent to residents. The purchaser could have received excellent value on all three counts. None of them determined how many licences it bought. [113]

A financial statement can put a different sale beside that contract under the heading of AI revenue. One business buys enough computation to read a file twice. Another buys a year of access for the person reading it. A third agrees to pay for a specified piece of work once the file has passed inspection. An enterprise could acquire all three on the same morning. Its ability to move money out of a consulting contract, a software renewal or an operations budget depends on which promise it has bought.

A receipt for an attempt

In October 2025 Anthropic introduced Claude Haiku 4.5 at one dollar per million input tokens and five dollars per million output tokens. Take those dated prices and give a machine a six-thousand-token packet to read. Allow a thousand output tokens to prepare a recommendation. A completed call costs 1.1 cents. Another completed call of the same size costs another 1.1 cents. The price rewards economical use of context and shorter answers. It gives the person who writes the purchase order no claim that a beneficiary name has been matched correctly. [111]

Suppose one hundred files arrive in a month. Twenty extra runs push the charge for one hundred and twenty completed calls to $1.32. This narrow calculation leaves out document retrieval, storage, integration and the employee who acts on the answer. A supplier can perform exactly the operation purchased even when the employee rejects every recommendation. A well-run team might still prefer this arrangement. It can change its own evidentiary standard, keep the savings from better document preparation and switch the underlying model if its test set exposes a weakness.

For the model company an improvement in quality can have an awkward first effect. Preventing those twenty repeats removes twenty-two cents of usage from the specimen. It may win the next hundred files, or a much larger customer, by doing so. The order for those files has to arrive. Anthropic's prospectus describes revenue as primarily consumption-based and cautions that much of it lacks long-term customer commitments, while its own infrastructure commitments usually run for several years. A better answer and a higher quarterly bill can therefore move on different schedules. [01]

A receipt for access

Microsoft lists its US enterprise Copilot add-on at $30 per user per month on an annual subscription paid yearly. The customer needs a qualifying Microsoft 365 plan. Four seats represent a $1,440 commitment for the year. Allocating a twelfth to the month produces $120. A quiet February changes the cost per useful file without changing that commitment. Comparing this US offer with the South Downs award would require the local terms, tax treatment and reseller agreement. Dividing a sterling contract by an American list price cannot supply them. [112] [113]

The British notice names a small or medium-sized enterprise as the supplier and gives 12 October 2026 as the end date. Digital Devices could receive the purchase order while Microsoft supplied the software. If the authority bought another year through a different reseller, a change in the name on the order could look like a competitive loss to one intermediary while leaving demand for the product intact. Nothing in the notice establishes whether a renewal took place. [113]

An integrated assistant may earn that subscription by meeting the employee where the underlying documents already live. Imagine a credit officer who spends a morning reconciling figures in a spreadsheet and writes the committee paper after lunch. The same seat can help with both. The buyer keeps the discretion to send a draft back as often as its policy requires. For included work the supplier bears the additional inference cost of a retry. The buyer still pays for each inspection its officer performs. Metered extensions or separate agents can change that allocation.

There is already competition inside this middle category. Salesforce publishes options including user access and credits for actions. Anthropic has announced finance agents that work with Microsoft tools. An officer might open the same spreadsheet through either supplier. A demonstration that a product can complete a spreadsheet tells us little about which contract the bank would retire or which supplier would own the relationship at renewal. Look for the invoice address, the right to change the workflow and the person whom the buyer calls when the formula is wrong. [13] [16]

EXHIBIT 08 / THREE SALES CONTRACTS

Three ways to sell a machine decision



Three illustrated invoices price the same 100 submitted files by tokens, seats and retained case fees. The respective charges are $1.32, $120 and $640 under different contracts.

The input and output charges use Anthropic's October 2025 Haiku 4.5 launch tariff. Each completed run uses 6,000 input and 1,000 output tokens. The seat allocation uses Microsoft's US enterprise add-on price, paid annually and requiring a separate eligible subscription. The $8 case price and volumes are illustrative. The three offers cover different work. Token tariff . Seat terms

Open exhibit 08 in the interactive edition

A receipt with a return window

Gorgias gives retailers a concrete example of a third sale. Its billing guide describes a shopper asking about an order and an automated answer supplying a tracking link. The guide counts an automated interaction when the customer needs no human involvement within seventy-two hours. Under the current arrangement a fully automated answer can consume both a helpdesk ticket and an automated interaction. A handoff can still consume a helpdesk ticket. Some older accounts retain terms under which a ticket takes one charge or the other. Two merchants comparing automation rates ought to compare their invoices first. [114]

A shopper who has received the right parcel and a shopper who gave up can both leave the queue silent. The supplier can count its commercial unit consistently while an operator needs additional evidence about service quality. A retailer has reason to link the ticket to later contacts, returns and payment disputes. The relevant horizon may be the delivery date. A Thursday answer about a parcel due the following Wednesday will have passed a seventy-two-hour test before anyone can tell whether the parcel arrived. That is a reason to observe what the tariff leaves outside its window.

Give our hypothetical bank a stricter bargain. One hundred files arrive, eighty-four meet the initial contractual test and four later receive qualifying credits. At eight dollars per accepted file the service raises $672 of charges and gives back $32. The net invoice is $640. The service provider pays for its own repeated attempts. The bank retains the cost of its authorised checker. Sixteen files have yet to earn a fee at all. Dividing $640 by the hundred incoming files yields $6.40. Dividing it by

the eighty fees still standing yields eight dollars. Both numbers are useful to different budget holders.

The signature belongs to the bank

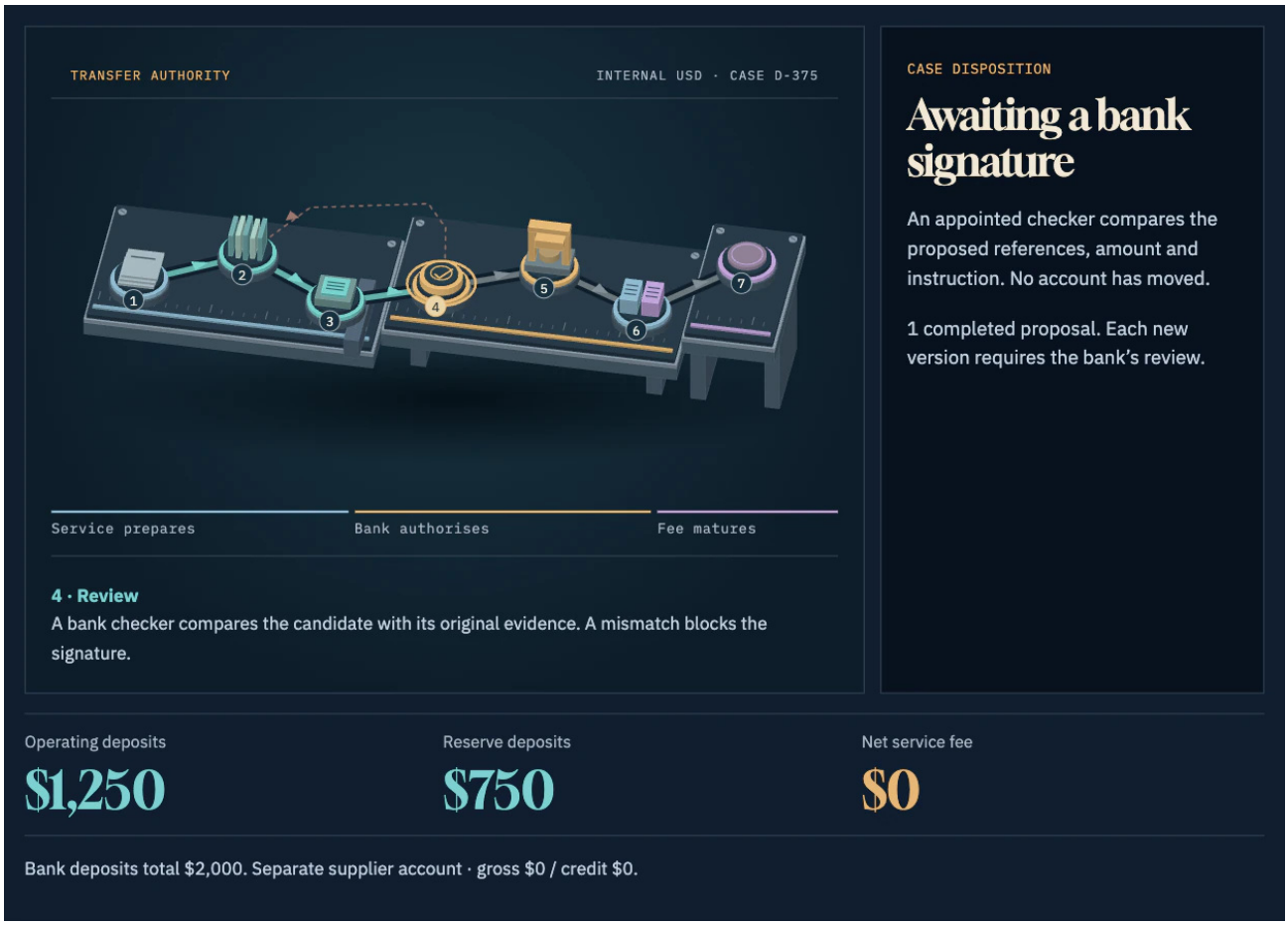
Consider an instruction to move $375 from a customer's operating account to the same customer's reserve account at one bank. The specimen starts with $1,250 in the first and $750 in the second. A service retrieves the signed instruction, looks up the two account references and proposes an internal transfer. A discrepancy in the reserve reference sends the packet back for evidence. The service has consumed resources and produced a proposal. At this point the customer can still spend the original balances and the supplier has earned nothing under our terms.

An appointed bank checker compares the proposal with the instruction and records approval against that version. Only the bank's posting service can then release the instruction to the core. Modern Treasury's payments documentation illustrates the distinction. Its configurable approval system ordinarily requires fresh review after an object changes. For manually created or updated objects it prevents the latest editor from reviewing that object. Its ledger documentation describes a separate control. Once posted, the entries are immutable. Undoing their effect requires another transaction. These are product rules a bank can examine beside its own policy. [115] [116]

On release the bank debits its operating-deposit liability and credits its reserve-deposit liability by $375 apiece. The customer's balances become $875 and $1,125. The total remains $2,000. A matched journal reference and two business days without a qualifying return earn the service its eight dollars. Suppose an authorised investigation on day ten establishes a booking error. The example allows an internal correction, with funds available, and a full service credit through day thirty. A second approval releases a compensating entry. The first journal remains visible. Getting the principal back into place and crediting the service fee answer different claims.

EXHIBIT 09 / AUTHORISATION AND BOOKING

A case earns its signature



A seven-station transfer model pauses at the bank checker. The $375 instruction awaits a signature, both deposit balances still total $2,000, and no service fee has matured.

Scenario D-375 follows a single transfer on the bank's own books. The fork returns a mismatched file to evidence. The brass gate represents a signature bound to one version. The third platform holds a separate commercial claim. Both accounts belong to the same customer. A correction assumes authorisation and available funds. The case price and return periods are hypothetical. Approval controls . Posted and reversing journals

Open exhibit 09 in the interactive edition

Where another supplier can lose

A bid for this work reaches beyond the technology budget because somebody already pays to investigate instructions. Suppose the bank could retire $560 of purchased processing for eighty matched files, with equivalent quality and turnaround, while retaining the same manual service for the other twenty. The new service also requires $80 of incremental bank checking. The largest fee the bank can pay while matching the old cash cost is $480, or six dollars per retained case. A technically successful offer at eight dollars loses by $160. Nothing about that result diminishes the model's ability to read an instruction.

The bank could rationally pay nonetheless if it previously incurred verifiable late charges that faster handling removes. Avoiding three dollars per retained case would release a further $240 and leave an $80 advantage. If those charges belong to different files, or if ordinary rule changes already avoid them, strike the corresponding benefit. A bank that keeps both the old processor and the new service through a parallel run adds expenditure during the overlap. Calling that

investment in better service is defensible. Calling all of it revenue taken from the processor would leave the old invoice unexplained.

An operations manager can test the bid before choosing a supplier. Give each bidder the same closed files with the final disposition withheld. After scoring, restore the files that would have fallen outside each bidder's contract and count the staff needed to finish them. A provider can improve its reported acceptance rate by admitting only straightforward instructions. The bank still owns the difficult ones. A price per completed file means more alongside the proportion of the incoming queue that the supplier will take.

I use the term AI computing company for a supplier willing to operate enough of the system to sell the buyer a defined piece of work. A laboratory can earn that position, purchase it through a services business or rent its intelligence to someone else who earns it. From the customer's side the decisive documents will be unglamorous. One establishes what the bank can stop paying. Another establishes who bears the cost of reopening a file. The shortest of the three may be the list of people allowed to sign. Its length places a practical limit on how much work anyone can sell.

CHAPTER 04

How large is the bill today?

One bank's technology outlook, six worldwide spending lines and four sellers' invoices show where a market estimate can count the same dollar twice.

JPMorganChase put a number on its technology plans for 2026 that would make a respectable market all by itself. Management expected technology expense of roughly $19.8 billion, up about $1.9 billion from the year before. In the same presentation it identified roughly $600 million of efficiencies, about a quarter of them related to AI. For a seller looking for money to displace, there is a temptation to circle the smaller number. For a seller trying to understand what the customer will actually buy, the more interesting fact is the direction of the larger one. The bank expected to save money on particular activities and to increase its technology expense. [118] [119]

Look at who owns that estimate. The consumer and community bank accounted for $8.2 billion of the outlook. The commercial and investment bank accounted for $8.8 billion. Asset and wealth management accounted for $2.2 billion and corporate functions for $600 million. Within the projected increase, the firm identified higher prices for labour, hardware and other inputs, more demand for cloud infrastructure and software, and investment in new capabilities. Those are rounded management figures. They describe an organisation buying and making things for different customers. A model developer that wins a mandate in one division has earned a place in one budget negotiation. An analyst who treats the whole bank as a single subscription will miss the next three. [118]

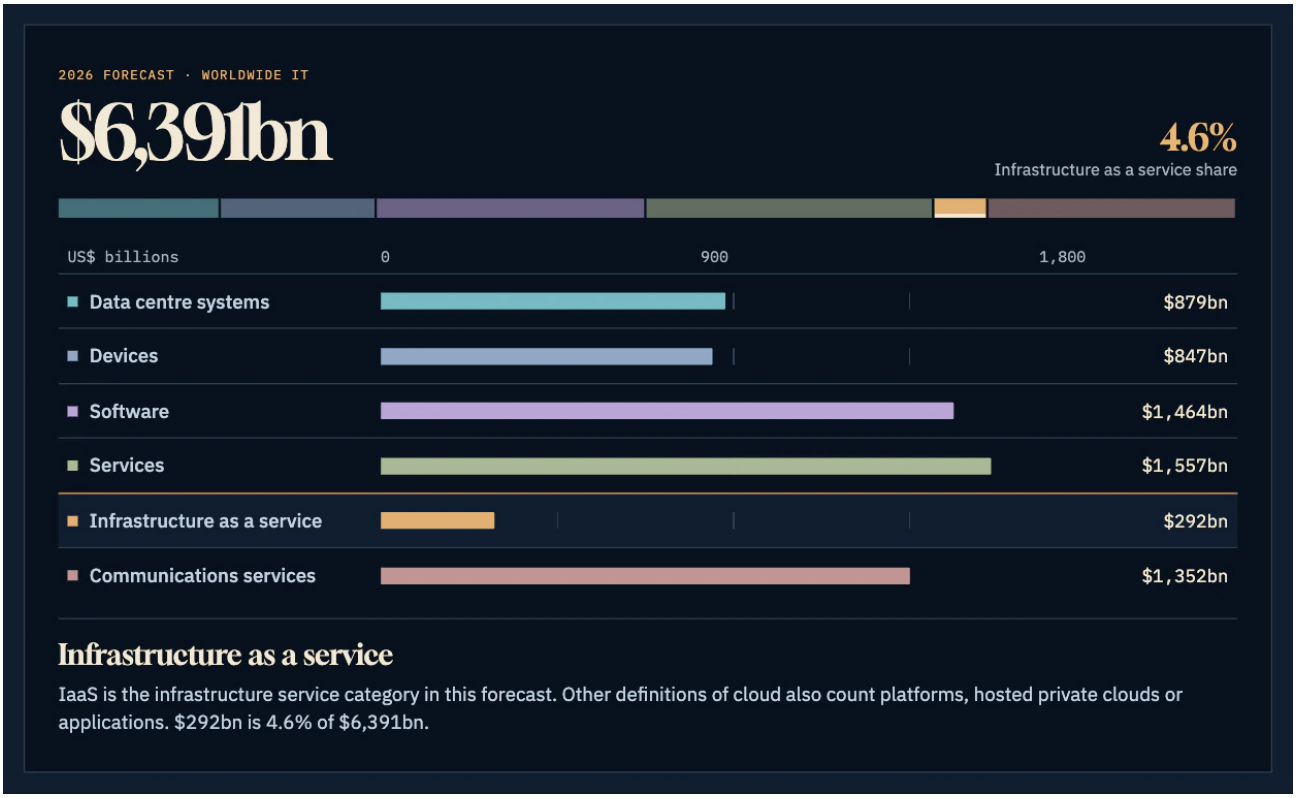
The roughly $150 million one can infer by multiplying the two management estimates is particularly easy to misuse. It is an approximate AI-related component of identified efficiencies in a technology expense outlook. The slide gives no corresponding invoice from Anthropic, no population of redundant employees and no division of the benefit between better purchasing, faster engineering and changes in service. Management told investors that efficiencies enabled additional investment. A bank can use cheaper infrastructure to add a facility for its clients. The firm selling that facility could be the most important commercial beneficiary without having made the infrastructure cheaper itself. [118] [119]

Six spending lines and a tempting division

In October 2026 Gartner put worldwide IT spending for the calendar year at $6.391 trillion. The forecast for 2027 was $7.352 trillion. Both estimates came from the same release, used the same six categories and were expressed in US dollars. That makes them suitable for a comparison of what this particular forecaster expected to change between two years. The first figure was published while part of the year still lay ahead. When someone describes it as the cash enterprises had already spent, ask where the final two months of invoices came from. [08]

EXHIBIT 10 / WORLDWIDE EXPENDITURE FORECASTS

Which IT expenditure counts?



Six bars divide Gartner’s $6,391bn worldwide IT forecast for 2026. Infrastructure as a service accounts for $292bn, or 4.6%.

Forecast published 21 October 2026. The segmented strip divides the selected total. Each row below it starts at zero. In the 2027 view a white mark gives the 2026 level. The increase is the difference between two forecasts, in billions of current US dollars. Category explanations are the author’s reading of the scope. Gartner, Table 1

Open exhibit 10 in the interactive edition

Infrastructure as a service accounts for $292 billion in the 2026 table, about 4.6 per cent of the total. It is a remarkably small fraction for an industry that dominates so much discussion of corporate computing. Yet assigning the other 95.4 per cent to technology outside the cloud would produce a remarkable mistake. A customer renewing a hosted payroll application has bought software. A consultant moving a bank's records into a cloud environment can issue a services invoice. A data centre supplying cloud capacity buys equipment. A product, a delivery method and the buyer of a machine answer different questions about the same system. 08

The composition of growth is more revealing than a single share. Gartner's two columns imply $961 billion of additional worldwide spending. Data centre systems account for $487 billion of that difference, just over half. The IaaS line rises by $75 billion, software by $215 billion and services by $76 billion. These are differences between rounded forecasts. An equal dollar of extra equipment and extra software revenue has a different life thereafter. Equipment can be put to work for years before the purchaser has recovered its cost. An application must keep enough customers to support its next renewal. An investor trying to identify the winner of a computing cycle needs to know where the equipment goes and who gets to set the price of using it. 08

Five markets seen from different tills

Take the cloud first. Synergy Research estimated $167.7 billion of cloud infrastructure service revenue in the third quarter of 2026. Its definition includes infrastructure services, platform services and hosted private cloud. Four quarters at the September pace would give $670.8 billion. The same researcher reported $558 billion for the twelve months actually covered by its estimate. The difference is $112.8 billion before anyone has changed a contract, simply because the industry was growing and the two calculations answer different questions. Putting the annualised Synergy figure beside Gartner's $292 billion as rival estimates of the same business would add a second error involving the products each researcher counted. [25] [08]

Enterprise applications offer a case where two numbers can legitimately be combined. Futurum's Table 2 forecasts $323.3 billion of 2026 sales delivered as SaaS and $91.8 billion delivered on premises across the applications it studies. Within that particular table the two delivery categories amount to approximately $415 billion. A customer replacing a licensed application with the same vendor's hosted version moves between the categories. To establish that an AI company captured revenue, we would need to identify a change of supplier or a new paid AI component and trace it through the agreement. A shift of delivery method on its own cannot tell us whose shareholders got the money. [26]

There is less reason to add the data market. Gartner's 2025 forecast for database management systems put 2026 at $161 billion. A bank can run a database on its own equipment, buy it as a managed cloud service or consume it inside an application. Its application vendor may be the party paying for the database. There are also analytical services outside the definition of a database management system. The estimate gives a sense of the scale of one contested activity. A survey of bank purchases would still need to identify whether the bank actually sees the database invoice or whether the application company decides what engine to buy. [28]

Consulting presents a further problem of time. Source Global Research's March outlook put the market around $315 billion. In August it expected the worldwide market to pass $350 billion in 2026. Without a published reconciliation it would be unwise to assign the difference to AI demand, new geographies or any particular service. Nor can the entire consulting category simply be added to a worldwide IT services estimate. Some engagements help a client reorganise a business, others implement software, and an outsourcing contract can contain both work and software rights. To price a bid against a consultant, start with the statement of work that expires this year. Its boundaries are at least negotiable by the people in the room. [27] [50]

The customer who never issued an invoice

Technology bought from an outside supplier leaves a price and a payee. A tool written by the bank's engineers for use in its own branches can enter the economy through another door. The US Bureau of Economic Analysis counts qualifying software developed for a business's own use as investment. Where there is no arm's-length sale, its method estimates production costs, including the labour that goes into developing the software. It distinguishes new or substantially enhanced software from ordinary running activity. An archive of vendor invoices and a national estimate of software investment can therefore both be carefully assembled and cover different work. [121]

This matters to the proposed AI computing company because the first cheque it receives can come from an employer that has never purchased the task before. Think of a group that builds the internal tool for assigning payment disputes to specialists. Employees write the application and keep its rules up to date. The bank then buys a service to prepare the evidence for those specialists. Its external technology bill rises immediately. Its internal developers may go on maintaining the same assignment tool. If a new job has been bought, the bank can be better served, and the lab can earn recurring sales, without a dollar coming off an incumbent software invoice.

A different bank might already employ an outside firm to assemble those files. There the same product could replace a renewal. The commercial evidence in the two cases would look almost identical to the lab's finance team. Both customers could have the same accepted volume, the same price and the same implementation cost. Even a perfect statement of the lab's recognised revenue would leave the origin of the two payments undetermined. The buyer has to supply that part of the story. A staffing plan, an expiring services contract or an order authorising a previously unfunded job will say more about the source of growth than a list of impressive customer names.

Even a purchase with a named supplier can reach the financial statements on a different timetable from the cash. The IFRS Interpretations Committee considered a customer paying to configure a cloud application. In the circumstances it examined, the customer received access to software controlled by the supplier. The committee asked first whether the additional work created an identifiable asset the customer controlled. If it did not, the timing of expense depended on who supplied the configuration and whether it was a distinct service. A separate service could be expensed when performed. Work inseparable from access could be expensed as the supplier provided access over the contract term. [122]

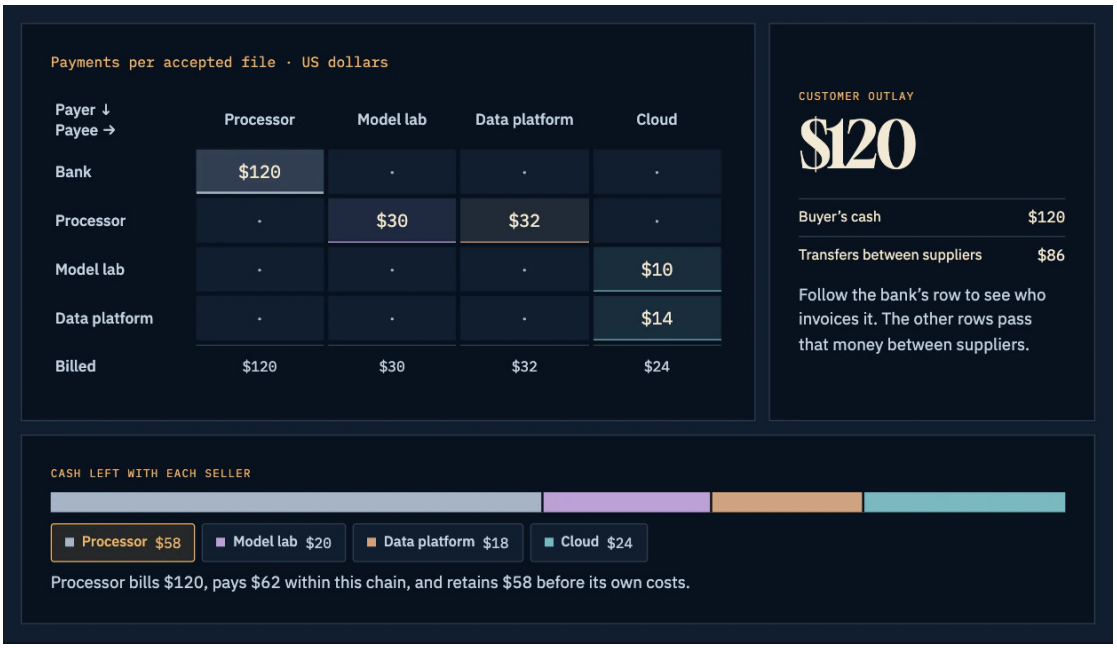
For a reporter examining a migration, that distinction changes the documents worth asking for. A bank may pay for installation in December, begin receiving service in January and show part of the expense well after the original project team has dispersed. Comparing two quarters of expense could then suggest a price reduction that the cash ledger never experienced. The order, the payment date, the expiry clause and the period of service each belong in the file.

One purchase, several sellers

The national accountants have a way of thinking about this problem. At industry level the Bureau of Economic Analysis distinguishes gross output from the contribution left after purchased intermediate inputs are deducted. A material can appear first in the sale of a component and again inside the price of the finished article. Both producers made actual sales. The intermediate input has to be dealt with before adding up what industries contributed to domestic product. A bank buying one service has a simpler question. How much left its own account? The answer can be smaller than the sum of the perfectly genuine sales made down the chain. [120]

EXHIBIT 11 / SUPPLIER INVOICES AND CUSTOMER CASH

How a dollar gets counted twice



A payment matrix traces the bank's $120 invoice and $86 of subsequent supplier transfers. Closing holdings across the four suppliers sum to the original $120.

Illustrative US-dollar contracts for the same accepted file. The bank can assume some or all of the $24 cloud bill. Other suppliers reduce their gross prices by the cloud charges transferred. Each supplier keeps its original balance after listed transfers. Its wages, taxes, capital costs and profit are outside this comparison. A dot means no payment between those parties. BEA on intermediate purchases

Open exhibit 11 in the interactive edition

Give a processor $120 for a completed file. It buys a $30 model service and a $32 data service. The lab spends $10 of its receipt on cloud capacity and the data supplier spends $14 on the same cloud. Add what the four sellers billed and the total is $206. Deduct the $86 of purchases that moved money among them. The balances are $58 for the processor, $20 for the lab, $18 for data and $24 for infrastructure. They add to the customer's $120. Each seller still has wages, equipment and other costs outside the illustration. Its balance after the listed payments gives us no entitlement to call it operating profit.

Now have the bank procure all $24 of cloud capacity directly and have the other parties reduce their quotes by exactly the charges they no longer incur. The processor bills $96. The lab and data suppliers bill $20 and $18. The cloud still bills $24. The customer spends the same $120, every supplier retains the same balance and the sum of supplier invoices falls to $158. Put a revenue chart beside the two contracts and the industry appears to have lost $48 of business. Put the transaction records beside the chart and you can identify the change as a new billing address for the same work.

A sensible starting point for a market forecast is therefore a small set of records that can actually be matched. Pick a buyer and a period. Record the party issuing each invoice and whether the charge sits inside a larger purchase. Keep the contracted service level, the accepted volume and the old renewal quote. Then ask what the prospective supplier can collect after the buyer has exercised its own options for a better price. There is ample money in the world for the prediction in this essay to succeed. The test begins with the customer whose next payment can be followed twice, first into the seller's account and then out to the companies that make the sale possible. In the cloud those companies can include the seller's most determined competitor.

CHAPTER 05

The cloud sends two invoices

One bank pays for a resolved case. Follow the money through the marketplace and down to the machines that did the work.

Put a bank's disputed payment on the table. A customer paid once. The account shows two entries. The person who takes the call can see both, but only the payments team can tell whether the second is a temporary authorisation, a duplicate posting or a transaction the customer has forgotten. In the meantime the customer has money missing and the bank has a file that somebody must own. Imagine offering the bank a price for returning that file with a supported explanation, the necessary correction approved and a record that will stand up when the customer calls back. The procurement question is which department used to pay for all of that.

I would start there if I were trying to understand Anthropic's prospects after its flotation. Start with one job that already has a cost, even though nobody has ever seen that cost on a single invoice. Look at who pays for the screen in front of the employee, the database behind it, the time spent looking for the second entry and the engineer who maintains the connection to the ledger. Then ask which of those expenses the buyer can actually stop paying. A machine capable of following the entire dispute through the bank acquires a claim on a much larger budget than a machine that rents out a few seconds of processing. Collecting on that claim is a commercial achievement in its own right.

The cloud provider has an unusual seat at this table. It can quote the bank for a finished service. It can collect a payment on a model developer's behalf. It can also sell the developer the capacity on which the service runs. Those roles give it distinct interests in the same decision. The bank might move its contract to an AI company on Monday and send more business to its old cloud supplier on Tuesday, without anyone in procurement changing their mind. The second purchase takes place upstream, where the bank's accounts payable team seldom sees it.

What arrives on the bill

Microsoft gives us a particularly clear view of the distinction. Under its Claude Consumption Unit arrangement in Foundry, model use is converted into billing units after applicable discounts and appears on the customer's Azure invoice. The model detail is available elsewhere. Eligible consumption counts towards the customer's Microsoft Azure Consumption Commitment. A finance officer studying the invoice can see Microsoft, while the application owner choosing the model and negotiating its price may be dealing with Anthropic. Two people in the same company can give different, perfectly sensible answers when asked who sold them the intelligence. [35]

Now put yourself in that finance officer's position near the end of the contract year. A team proposes to move its model calls to a cheaper direct agreement. The team has timed its own workflow, checked the answers and secured a price below the one it sees in the cloud console. Finance wants one more piece of information. How much of the existing commitment will expire unused if those calls move? Spending the money with a new supplier could leave the bank paying the old one as well. If another department has work ready to consume the released commitment,

the decision changes again. A comparison of prices per token cannot settle a question about two departments' capacity to use the same prepaid or committed pound.

A second complication appears when the bank crosses a border. Google documents regional and multi-region options for Claude that let a buyer choose where processing takes place. A bank can value the ability to keep processing within an approved geography while distributing work across eligible locations. To evaluate a competing offer, it needs to know whether that offer covers the same locations when a site fails. A cheaper supplier can win easily on an ordinary afternoon and prove expensive on the one afternoon that determines the contract. The analyst cutting costs and the officer signing the continuity plan are purchasing different attributes of the same call. [36]

Amazon can bring a different argument to the meeting. Its Bedrock catalogue gives a buyer access to competing model families within one managed service. A bank that expects next year's best model to come from a different lab has a reason to value this arrangement. It can keep parts of its existing operations and change what runs inside them. The lab, in turn, has a reason to earn the loyalty of the bank's payments staff directly. A product that remembers how that bank records a reversal and which disputes its supervisors reopen is harder to replace than a generic model behind a common interface. The contest begins with a choice about where that knowledge will live. [05]

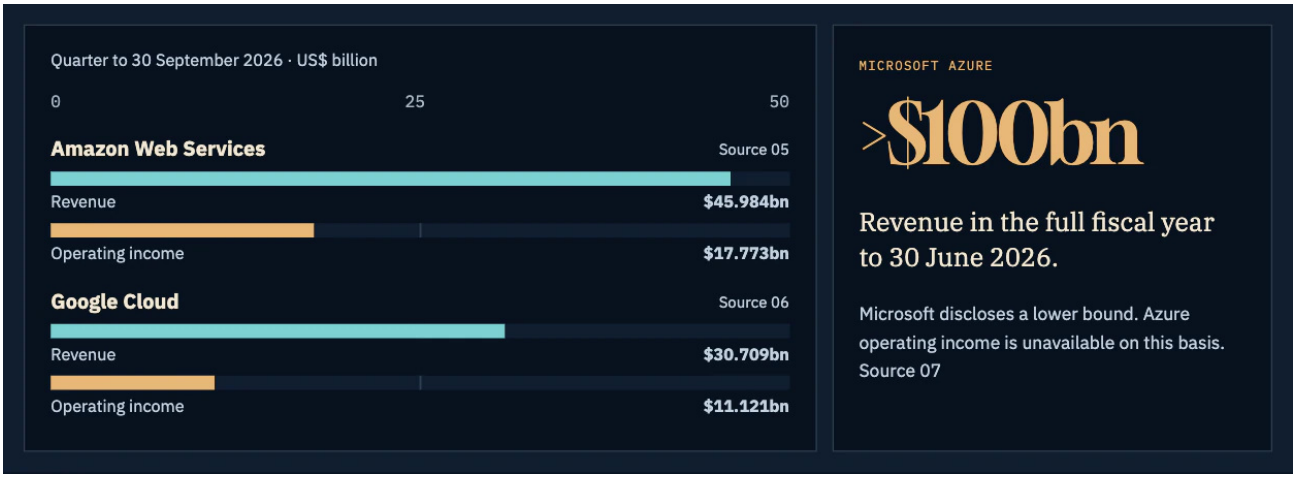
Three different rulers

Amazon reported AWS sales of $45.984 billion for the quarter to September 2026 and operating income of $17.773 billion. Take the revenue over four actual quarters and AWS recorded $161.382 billion. Take the September quarter alone and multiply by four and the pace was almost $184 billion. Both figures have appeared in serious conversations about the size of the cloud. The difference between them exceeds $22 billion, simply because one looks back over the previous year and the other assumes that the September pace persists. An essay predicting a change in leadership can move the finish line a considerable distance by switching rulers mid-race. [05]

Alphabet disclosed $30.709 billion in Google Cloud revenue and $11.121 billion in segment operating income for that same quarter. Its account of the period includes sales of TPU systems as a contributor to growth. Microsoft says Azure revenue exceeded $100 billion in the fiscal year ended in June. The word "exceeded" matters. So does the word "year". Plotting a bar at exactly $100 billion beside a September-quarter number would quietly introduce an error in both magnitude and time. Microsoft Cloud encompasses other businesses as well. A credible test for a new market leader must retain all three companies' reporting boundaries, even where that leaves an empty cell on the page. [06] [07]

EXHIBIT 12 / COMPANY DISCLOSURES

Three clouds, three disclosures



Quarterly sales and operating-income bars compare AWS and Google Cloud. A separate panel records Azure's annual revenue lower bound of $100bn.

Both paired bars use a zero-to-$50bn scale. Google Cloud includes more than infrastructure services. The separate Azure panel preserves the annual reporting period and the open upper bound.

Open exhibit 12 in the interactive edition

There is another reason for keeping profit on the page. A cloud operator has to fill a site through the useful life of the equipment, and a model supplier can buy capacity on a schedule that begins long before its end customer signs up. One can report rising sales while granting progressively deeper discounts to keep machines busy. The other can announce a large enterprise contract whose early years are spent integrating systems for a fee that barely covers the engineers. The bank keeps its budget in cash. Both suppliers report revenue. Their shareholders ultimately need a return on what each had to commit to earn it. Those are three separate vantage points on the same growth.

The comparison I propose for 2031 starts with four consecutive quarters of recognised sales for each business and then repeats the exercise for operating profit wherever the disclosures allow it. If a lab overtakes one cloud segment while reselling a great deal of that segment's infrastructure, it will have passed an important test of commercial reach. We will need the second comparison to establish how much economic power travelled with the customer contract. A distributor can become very large by passing expensive inputs through its accounts. The customer's willingness to pay a premium for what the distributor adds is a more demanding achievement.

The channel that shrank and grew

The most revealing figure in Anthropic's prospectus is tucked into the financial statement notes. In the first nine months of 2025, roughly 48 per cent of its revenue came from sales through cloud marketplaces. A year later the corresponding share was about 44 per cent. Put those percentages beside nine-month revenues of $2.545 billion and $33.242 billion. The implied marketplace amounts rise from about $1.22 billion to $14.63 billion. The percentages are rounded and the derived amounts are estimates. These are sales of Anthropic's service through a channel. The amount retained by the hosts of those marketplaces requires a different disclosure. [01]

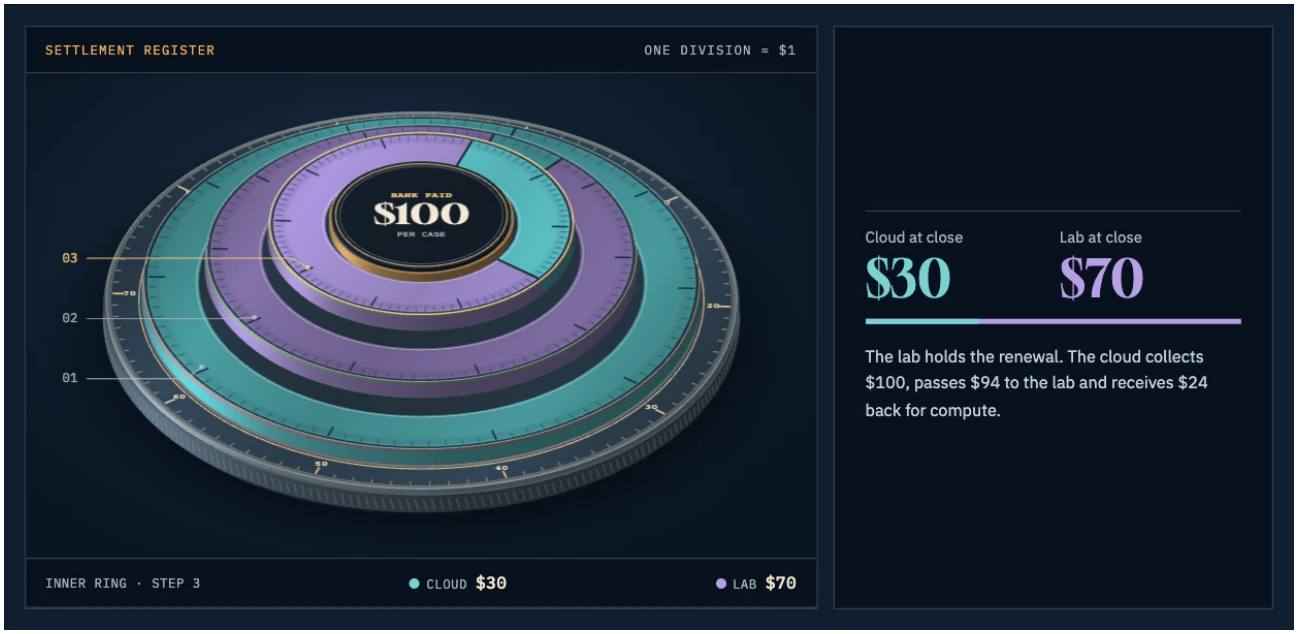
Consider what an executive could conclude from either half of that comparison. A sales director who sees only the percentage decline might instruct the team to arrest a deterioration in the channel. A partner manager who sees only the dollar increase might argue that the marketplace's

position is now impregnable. Both would be reaching beyond their information. A customer may have increased its purchases through its existing channel, bought a second service directly, or moved existing work between the two. The public numbers combine all three behaviours. Only matched customer histories could tell us how often one supplier took a contract away from another.

The filing describes the cloud providers as collection agents on these marketplace sales. That observation settles a question about whose revenue the filing is measuring. It leaves open the terms of each infrastructure contract and the identity of the party that will control renewal. The cloud that collected the bank's payment need not even be the cloud that supplied all the computing behind it. For clarity the instrument below initially puts both roles in one company and prices a case at a round hundred dollars. The bank pays once. Open the instrument and one payment separates into the fee for distribution, the rent for computing and the balance available to deliver the work.

EXHIBIT 13 / ILLUSTRATIVE SCENARIO

The marketplace beneath the model



Three concentric settlement rings trace one $100 bank payment. After the marketplace fee and wholesale computing charge the cloud retains $30 and the lab $70 before other costs.

Read the rings from the rim inward. Each turn represents the same $100. The teal fraction belongs to the cloud and the violet fraction to the lab after that transfer. Closing balances precede other costs. The same company handles distribution and infrastructure in this scenario. All fees are illustrative. Anthropic's reported 44% marketplace share measures the distribution of its sales. Prospectus, Note 2

Open exhibit 13 in the interactive edition

At the default settings the marketplace keeps six dollars for the channel and receives twenty-four for infrastructure. Seventy remains with the lab before its other costs. Calling that seventy profit would erase the people who installed the workflow, the support contract and the cost of a difficult case that has to be run again. Change to a direct sale and hold the underlying machine usage steady. The cloud still receives twenty-four. The bank still pays a hundred. Six dollars changes hands because the route to the customer has changed. A lab may spend more than six dollars acquiring and supporting that customer itself. Direct distribution buys more control even on occasions when it earns less money.

Now give the cloud the primary contract. In this hypothetical bid it promises the result to the bank and buys the lab's service for sixty-five dollars. The lab in turn buys the same twenty-four dollars of computing from the cloud. Follow the transfers in sequence and the cloud ends with fifty-nine dollars, the lab with forty-one. The cloud has a larger balance and a different obligation. If the bank refuses to accept the case, it knows which counterparty it can pursue. The fee attached to taking that call is one reason the enterprise software industry has remained more durable than generations of technically superior substitutes might suggest.

What a bank has already done

There is a real bank behind a nearby, more modest proposition. Trust Bank in Singapore built a customer service application on Amazon Bedrock. The architecture described by AWS uses a containerised agent, retrieved customer and product context, and a database of material against which to ground an answer. AWS reports that more than 52 per cent of customer enquiries reached resolution without human intervention, and that some service agents were retrained for AI operations. These are results reported in the supplier's case study. They concern customer enquiries, a wider and less specialised category than the disputed payments used in our thought experiment. [37]

The interesting piece of evidence is the list of things that had to work together. Somebody had to maintain the bank's approved information, decide when an answer counted as satisfactory and feed troublesome conversations back into the evaluation process. A resolved enquiry created demand for several technical services as well as for the model call. An incumbent with a working deployment has a chance to sell the next improvement to a buyer with whom it already shares an operating history. A new supplier asking to replace it would have to demonstrate that it could carry over the hard-won lessons embedded in that history.

Imagine interviewing the person who funded the rollout. I would ask for the number of customers who contacted the bank again about the same issue within seven days. I would ask to see overtime, outsourced support bills and the size of the backlog before and after the change. At a growing bank, keeping headcount steady can represent a considerable economic gain. Quantifying it means constructing a credible account of how many people would otherwise have been hired. I would also want a cloud invoice for the same cohort of enquiries. The reported resolution rate alone cannot tell a reader whether the resulting gain went to the customer, the bank, a model developer or Amazon.

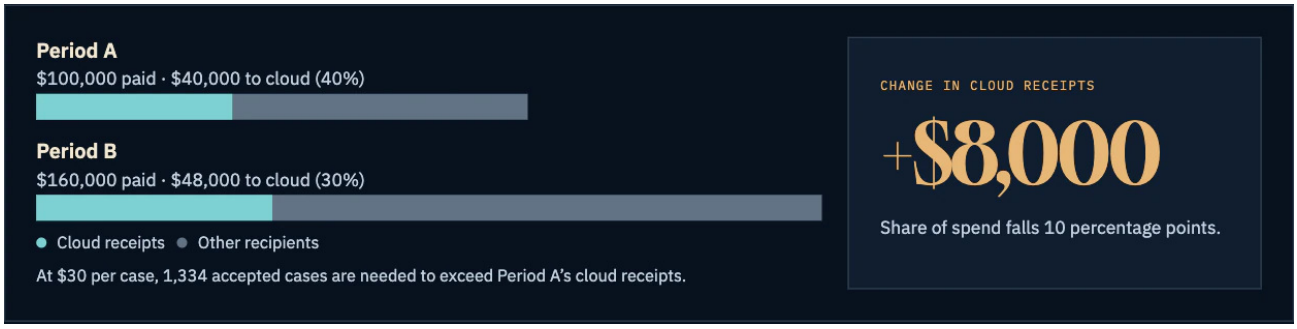
One can learn a great deal from the customer who declines to switch. Suppose the payments team has a year of decisions linked to the ledger, and supervisors can reproduce the reasoning behind any adjustment in an inspection. The incoming bidder can read the same records and write a better explanation. Can it preserve the links when a case moves across systems? Will a supervisor still be able to recover the evidence after a model is retired? Somebody must pay to qualify the replacement against old cases that customers can still contest. As that stock of prior work grows, portability of the decision record becomes a substantive term of the next tender. A bank that secured the right to export that record in usable form has already bought itself a more competitive renewal.

How much cheaper must the work become?

A cloud operator can receive fewer dollars per case and more dollars in total. Assume a thousand accepted cases at a hundred dollars each. Forty dollars from every case reaches the cloud, giving it forty thousand dollars over the period. In a second period its receipts fall to thirty dollars per case. If the bank buys sixteen hundred completed cases, the cloud now receives forty-eight thousand dollars. Its share of the buyer's expenditures falls ten percentage points and its receipts rise by a fifth. The volume increase is an assumption. Changing it is the quickest way to find out how much of a growth story rests on new demand.

EXHIBIT 14 / ILLUSTRATIVE SCENARIO

A cloud can grow as its share falls



Two total-spend bars grow from $100,000 to $160,000. The cloud's share falls from 40% to 30% as its cash receipts increase by $8,000.

Both bars use the same dollar scale. Each bar spans total buyer spend for its period. The coloured portion is the amount received by the cloud. Volume and unit receipts vary independently. The model holds the customer price and acceptance standard constant.

Open exhibit 14 in the interactive edition

Where would those extra six hundred cases come from? A bank may have a backlog of low-value discrepancies that it has historically written off because investigation costs more than recovery. Cheaper investigation can make some of them worth opening. In that event the appropriate baseline for an additional case is the loss the bank previously accepted, together with the cost of any customer redress it avoided or deferred. There might be no consultant invoice to displace. There might instead be a new service whose expenditure is justified by previously forgone recoveries. A claim that the lab won money from consulting would require finding a consultant the bank had actually intended to hire.

The opposite can happen just as easily. The number of payments may be essentially fixed over the life of a contract. A better upstream validation rule may eliminate many exceptions altogether. An AI supplier paid per resolved case has a commercial reason to expand its scope as the original queue shrinks. A cloud supplier paid by usage has a similar exposure. The bank, by contrast, might earn its largest return by fixing a reference-data error that prevents a thousand disputes from being opened in the first place. A procurement system that rewards every closed file and assigns no value to prevention can make the supplier's dashboard look excellent while keeping the bank unnecessarily busy.

This is where I would press a team promising to remove waste. Compare its proposal with the best affordable change the bank could make without buying a new AI service. Try the validation rule, remove a duplicate data feed or renegotiate an idle reservation. Keep the required quality, capacity for busy periods and time to correct an error constant. Count the staff hours released separately from expenditure the finance office can remove. A team that saves an hour every morning has

achieved something real. A budget that pays the same staff to clear an overdue queue has also achieved something real. It takes an additional piece of evidence to say that an outside company has captured an hour’s wages as revenue.

The terms of the next bid

My prediction depends on the tender that comes after the pilot. In the first year the lab may help the bank design a service at an introductory price, while the cloud provider supplies discounted capacity and a consultancy puts its best people on the project. At renewal the buyer will know the actual case mix, the probability of reopening and the cost of an unusually difficult month. It can ask bidders to quote for the same accepted work and name the party responsible for repairs. The firm that can make a credible price under those conditions has a plausible route into the bank’s operating budget.

If the bank invites only its cloud account team to that negotiation, the cloud has preserved a valuable advantage. If the payments director calls the lab first and asks it to obtain competing infrastructure prices, something significant has moved even if the bank’s cloud bill rises in absolute terms. If the bank owns the evidence and can run a fair competition between both, it may keep most of the benefit itself. The question I would put to a board is which of those three positions its contracts are creating today. By the time the next invoice arrives, the answer may already have been written into the right to take the work elsewhere.

CHAPTER 06

Two routes into the client

The labs bought delivery teams and financed consulting partners. A bank's second purchase order reveals which of them actually won the work.

The name to look for is on the second purchase order. On the first, the bank buys an experiment. Somebody has to interview the staff, connect the case system and teach a model what an admissible piece of evidence looks like. Several suppliers can share the credit for getting that far. On the second order the bank buys another year of service. One company gets the call when a file has gone missing. One executive can put a price on taking that call for another twelve months. If you want to understand why model developers are buying consulting firms, start with the distance between those two signatures.

Imagine receiving the two documents with the names of the businesses covered. The first gives a project manager sixteen weeks and a small team. The second specifies an answer by the close of the next banking day. Halfway down its terms is an obligation to reconstruct the answer if a customer challenges it six months later. You could learn more about who stands to make money from the second page than from an afternoon spent watching the software. Whoever wrote that clause knows something about the customer that the developer of an exceptionally capable model might never discover at its own office.

Buying the people who know where to start

On 4 May 2026 Anthropic announced a services company with Blackstone, Hellman & Friedman and Goldman Sachs. Its intended customers included mid-sized businesses with operations important enough to justify custom engineering and too few people available to do the engineering themselves. Anthropic described its own applied engineers working alongside the firm's staff. Its existing consulting partners would continue to work with large enterprises. Here were two ways into the same economy, one through other people's practices and another through a business the lab helped establish. [10]

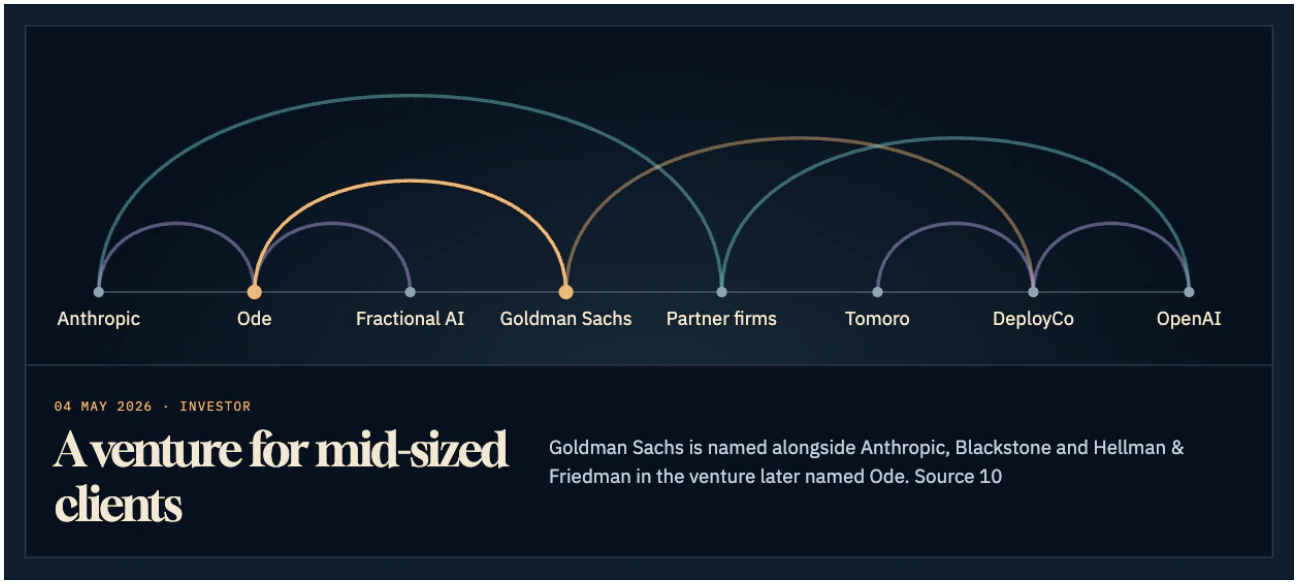
By July there was a name and an identifiable team. Ode with Anthropic said that the purchase of Fractional AI in May supplied its operating core. Fractional's co-founders became Ode's chief executive and chief technology officer. The difference between a launch presentation and an acquisition is quite tangible to a buyer. A team arrives with relationships, examples of things that broke and people accustomed to fixing them together. The purchase can save years of learning who is reliable under pressure. It buys no advance consent from the next customer to use the same design. [69]

OpenAI announced its own deployment company on 11 May. It described majority ownership and control, investment of more than $4 billion at launch and an agreement to acquire Tomoro. It expected Tomoro to bring about 150 people experienced in deployment. On 8 July the deployment company itself described the Tomoro acquisition as completed in an announcement about another proposed purchase, Northslope. The July document put the two deals at different stages. The distinction matters to anyone trying to estimate how many people are actually available to send to a client. [11] [70]

Consider the investors. TPG led the partnership for the OpenAI business. Its participants included consulting and integration firms that might otherwise have sold the entire engagement themselves. Goldman Sachs appears in the accounts of both ventures. An investor can own an interest in a supplier, introduce it to a company, compete with it for a separate project and buy its software for internal use. Each transaction has a different payer. In particular, an asset manager's connections to portfolio companies give a new business an audience with buyers. Those companies still have procurement officers who can choose another offer. [10] [11]

EXHIBIT 15 / DATED RELATIONSHIPS

Two labs take a seat at the client



Arcs connect two labs, investors, operators and delivery partners across eight dated announcements. The selected May 4 record identifies the investors in the venture later called Ode.

Selected 2026 disclosures. Each arc represents a documented relationship. All use the same stroke width. “Partner firms” groups the participating firms in each respective programme. Dates identify the announcement of the selected fact. Acquisition dates are identified separately in the record.

Open exhibit 15 in the interactive edition

The March and June announcements add a puzzle. Anthropic committed an initial $100 million to its partner network in March, including training and technical support. OpenAI announced a $150 million investment in its own partner network in June. A consultant trained by the lab can charge the client for the resulting work. Financing the training can therefore increase someone else's services revenue. The lab makes money if the work brings it model usage or creates an opening for a larger direct contract. How often that happens matters more than the size of either commitment. The announcements give the spending intention. A buyer's order would give the sale. [68] [12]

The practical choice facing both companies is how often to send their own people. The first engineer who learns how a bank releases a disputed payment may identify a product improvement useful to a hundred banks. The hundredth engineer dispatched to explain a field unique to one client's archive may spend a week producing something nobody else can use. A deployment company has to distinguish those jobs before it quotes a price. The workforce that opens a market also places a limit on its economics if every subsequent customer needs precisely as much attention as the first.

What one bank is willing to say

There is a buyer's account against which to test the story. In March 2025 NatWest announced a collaboration with OpenAI. The bank already had digital assistants and put the new relationship in the context of its existing work. Its own 2025 annual report subsequently associated AI with about nineteen minutes less handling time per complaint, equivalent to more than 22,000 hours over the year. It also identified AWS, OpenAI and Google among its technology partners. A customer can name its partners and measure a work improvement while keeping the commercial allocation private. [71] [72]

The distinction has consequences. A shorter complaint can mean that someone waiting for help speaks to a colleague that afternoon. It might allow the bank to forgo a temporary hire. It could help an experienced employee handle the cases that otherwise reached a supervisor. An external investor cannot tell which of these happened by assigning a wage to the reported hours. Nor does the bank's figure give the share attributable to any one partner. A supplier forecasting its fees off those hours would still need to identify who owns the corresponding budget and how much of it the bank could release. [72]

Think about the employee who knows which complaints must return to the same caseworker. A consultant interviewing her can discover the rule in an hour. Turning it into a durable business involves several other people. Someone must give a definition that works during staff absences. Someone must preserve the reasoning when the customer calls again. A buyer must decide whether a file counts as complete on its first closure or after the period in which it may reopen. These decisions determine both the denominator of a productivity measure and the numerator of a supplier's invoice. Some of the most valuable work in a deployment consists of getting those two people into the same room.

Three ways to write the second order

Give our illustrative bank a $240,000 implementation contract with an integrator. The integrator pays the lab $40,000 out of that fee. The bank's accounts payable department has spent $240,000. Supplier receipts include the lab's $40,000, taken from the money the integrator collected. Adding the two as enterprise expenditure would make the bank pay an imaginary extra invoice. At the end of implementation the bank owns an acceptance pack, a tested connection and the right to put the operating contract out to tender. For this example all three bids must serve the same 12,000 accepted files over the next year, including the same reopening provision.

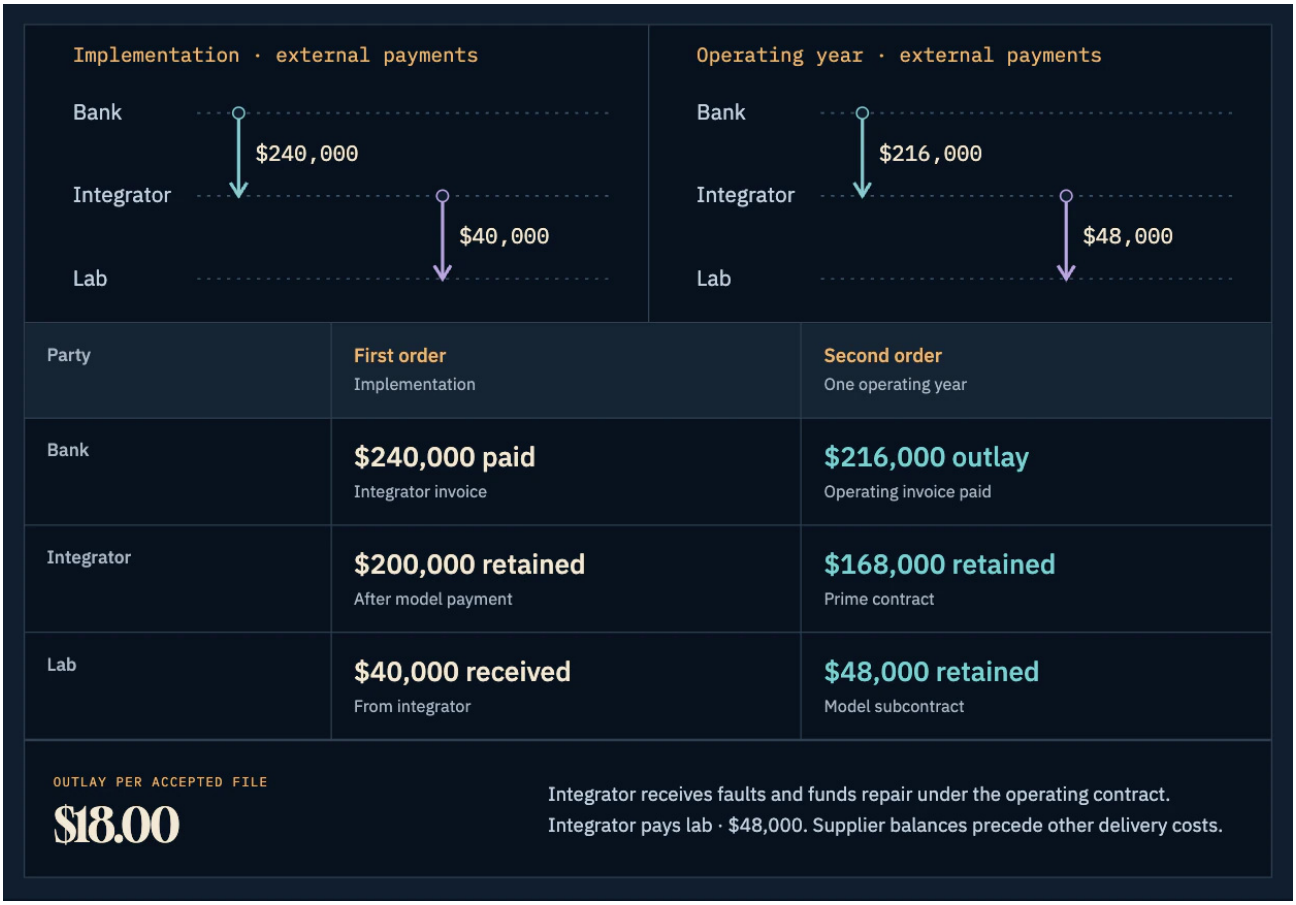
On the first bid the integrator keeps the prime contract. Its annual price is $216,000. It spends $48,000 of that on the lab's service and keeps $168,000 before its other delivery expenses. An investor in the lab sees recurring business and has good reason to be pleased. The more aggressive claim about displacing the consultancy has met its first difficulty. The integrator still sets the price, manages the bank relationship and can change model supplier subject to the agreed tests. It has put a new input inside a familiar service. The revenue associated with providing that input can grow quickly.

On the second bid the lab signs for $192,000. It hires the integrator for $60,000 of incident cover and systems work. The bank saves $24,000 against the first bid and the lab now controls a much larger payment. It also inherits the obligation to pay for investigations that end with its own error. The $132,000 left after the subcontract has to cover inference, support, staff and capital. Anyone

describing that sum as profit owes the reader quite a few more invoices. The integrator has lost the prime role and acquired a smaller, useful piece of work with an unusually well informed customer.

EXHIBIT 16 / OPERATING CONTRACTS

Who keeps the second engagement?



Party	First order Implementation	Second order One operating year
Bank	$240,000 paid Integrator invoice	$216,000 outlay Operating invoice paid
Integrator	$200,000 retained After model payment	$168,000 retained Prime contract
Lab	$40,000 received From integrator	$48,000 retained Model subcontract

OUTLAY PER ACCEPTED FILE
$18.00

Integrator receives faults and funds repair under the operating contract.
Integrator pays lab · $48,000. Supplier balances precede other delivery costs.

Two contract diagrams follow implementation and annual operation for 12,000 accepted files. The integrator keeps the prime contract and pays the lab $48,000 on the second order.

Illustrative contracts in US dollars. Both phases have their own period. The buyer's payment is allocated once. “Retained” gives cash after the displayed subcontract payment and before other expenses. All routes assume the same acceptance, incident cover and reopening obligation.

Open exhibit 16 in the interactive edition

The bank can also split the order. It buys $78,000 of platform service and $72,000 of integration and response cover directly. On these terms it must spend another $32,000 in cash on its own duty rota and release management. Total annual outlay falls to $182,000. The apparent $66,000 bargain on the two outside invoices shrinks to $34,000 against the first bid. The buyer gets control over which provider it replaces. It also gets the telephone call from each provider suggesting that the other is responsible. The extra internal money buys somebody with the ability and authority to settle that disagreement.

There is an instructive negotiation hiding in the last two figures. The lab-led bid costs only $10,000 more than the bank-run route. A bank short of experienced supervisors might gladly pay the difference. If its internal duty cover ultimately costs $43,000 instead of $32,000, the bank-run option goes from $10,000 cheaper to $1,000 dearer. The ability to deliver the team's promised rota is worth real money. A smaller bank could have exactly the same technical possibilities and choose the opposite supplier because its qualified people already cover a different service at night.

Ask each bidder to price one more clause. A batch arrives late and the prime must give the bank a $12,000 credit on its next invoice. On the lab-led route that would reduce the $132,000 left after the subcontract to $120,000, before the cost of repairing the work. The bank might also incur a separate expense contacting customers. The credit settles precisely one part of the loss. An experienced integrator can offer a higher credit because it already has staff on call for other clients. A smaller firm can match it by negotiating reciprocal cover or buying insurance with an appropriate limit. Now the buyer has a revealing comparison. Both providers ran the same model during the demonstration. One of them is willing to put another twelve thousand dollars behind the result.

How a partner defends the account

The incumbent's best defence is hidden in the acceptance pack. It knows which fields changed their meaning after an acquisition, who can authorise a refund on a public holiday and which monthly report will detect a quiet error. It can demonstrate those facts in the tender using anonymised cases the buyer has approved. If it uses a lab's model to reduce preparation time under a fixed annual fee, some of the saving initially lands with the incumbent. A procurement officer can reclaim part at the next rebid. Between those two dates, a successful AI deployment can improve the business it was supposed to threaten.

A specialist has a different advantage. It can insist on a narrow contract in which it knows the cost of nearly every exception. It can decline the heterogeneous cases that make a global systems contract expensive. That lets it put an attractive price on the common files. It also leaves the incumbent with a more difficult population. To evaluate the offers honestly the bank has to reprice the remaining service. A ten-dollar discount on an easy case can be consumed by the higher unit cost of keeping an expert available for the five hard ones left on the desk. Holding the old average price constant would credit the specialist with a saving its customer will never see.

The bank can improve its own bargaining position by keeping the examples that justify payment. Suppose the first supplier retains the only executable acceptance tests. A second bidder must charge for reconstructing them and the bank must pay its staff to adjudicate old arguments again. Suppose instead that the bank can hand every bidder a versioned set of test cases, accepted results and known exceptions under agreed access terms. An outsider can now price the obligation, and the incumbent can show where its judgement is worth a premium. The asset that makes switching possible may have cost less than a week of the original engagement.

The price of a reference customer

A client that agrees to appear in a launch announcement can be worth more to the supplier than the first year's operating fee. A deployment firm may therefore pay for integration, assign unusually senior engineers or absorb early losses to win a recognisable name. For that customer the bargain can be excellent. For the next buyer it is an unreliable price list. I would look at the first extension after the launch team has moved on. Which concession expires? Does the fee rise with the number of files submitted or the number accepted? Can the bank change the model if a rival performs better on its own tests?

For a lab considering an acquisition, the same questions run in the other direction. Buying a team gives access to customers whose existing agreements may have been written for that team as an independent adviser. Some clients may value its freedom to recommend a rival platform. Others may pay more for a direct line to the lab's engineers. Assessing the purchase requires following

renewals on both sides. The number of people acquired is publicly legible. The number of contracts that become easier or harder to renew after the acquisition is the number a shareholder actually needs.

My prediction is that the first decisive transfers will be visible on quite ordinary documents. An implementation partner disappears from the top of an operating order and reappears as a subcontractor. A lab's fee carries a service credit for a missed deadline. A bank renews an integrated service while keeping a second model qualified for emergencies. The strongest evidence against my prediction would be equally ordinary. The consultancies keep signing the extensions, buyers retain their tests and model providers compete for a replaceable input. The announcements establish that the labs have bought a way into the room. The next contract will tell us what they can sell there.

CHAPTER 07

The consulting partner's arithmetic

A returned milestone payment, a constrained review desk and three competing bids put a price on the promise to stand behind the work.

A$97,587.11 makes an unusual line in a history of artificial intelligence. It was the last instalment on an Australian government contract. Deloitte had been engaged to examine the system behind employment-services compliance decisions. Questions about references in its final report reached the department. On 8 September 2025 a senior official wrote to the firm asking for the money back. The public contract record reports that the instalment was repaid. This is one place where a reader can follow an argument about the quality of AI-assisted work all the way to a change in the client's bill. [46] [47] [49]

The department signed for A$439,142 including GST. Its briefing lists 25 interviews, more than 150 documents and 370 business rules examined. It records A$341,554.89 of expenditure after the repayment. There was a system to understand, a large body of evidence to evaluate and an official entitled to decide whether the final report met the terms of the engagement. Compress the episode into a story about a machine writing a document and nearly every important participant disappears.

The correspondence is unusually specific about the work. The department had agreed to the use of particular tools to help analyse code. According to its letter, Deloitte subsequently described using generative AI on two publishing tasks involving a legal case and citations, and acknowledged that the tools likely contributed to the errors. The department tied its demand to the professional standard required for the final deliverable. The refund tells us how much one buyer recovered. The costs of checking the references again, handling correspondence and correcting the published record extend beyond that number. The documents supply no complete price for those costs. [46] [49]

Start there when a model company announces its intention to disrupt consulting. Its commercial opportunity includes a signature that Deloitte had supplied to the Australian government. Will the new supplier put its own name on the report? Will it repay a fee when the evidence fails? The answers determine whether it can bid for the engagement or must sell a much smaller service to the firm that accepts those terms.

The size of the cheque

Source Global Research put worldwide consulting at about $315 billion in its March 2026 forecast. By August the same research firm expected the year to finish above $350 billion. Both articles put growth at roughly six per cent. They give no reconciliation of the market levels. I use the August statement as a dated estimate of scale. Reading the gap between the two estimates as $35 billion of new contracts would require a common base and evidence of what changed between publications. [27] [50]

The client has a more immediate problem of definition. A bank may hire an adviser to diagnose a queue, an integrator to rebuild the case system and a service company to operate it at night. One

master agreement might cover all three. An industry estimate of advisory work and a second estimate of outsourced operations can each be defensible. Treating the entire master agreement as up for grabs in both markets puts the same purchase in two places. Before asking for a percentage of a consulting market, I would ask the head of operations which budget paid the team and which deliverable renews.

EXHIBIT 17 / COMPANY DISCLOSURES

One consultant invoice, two businesses



Paired panels show $9.28bn of Accenture consulting revenue and $9.40bn of managed-services revenue alongside separate consulting and managed-services bookings.

Worldwide Accenture results, fiscal year ending 31 August 2026. Each pair shares a zero-based scale. The chart uses the amounts rounded to billions in the release. Bookings include obligations extending into later periods. Earnings release, pp. 2 and 4

Open exhibit 17 in the interactive edition

For the quarter ended 31 August 2026 Accenture reported $9.28 billion of consulting revenue and $9.40 billion of managed-services revenue. Together they account for the quarter's $18.68 billion. Holding that quarter constant for a year gives $37.12 billion of consulting revenue. The four quarters the company actually delivered produced $36.88 billion. There are good uses for an annualised quarter, especially when growth is fast. Here the full year is available beside it, and the reader can see precisely what the approximation costs. [09]

Consulting bookings over the same quarter came to $9.40 billion. By coincidence that is also the reported managed-services revenue, rounded to the nearest $10 million. One number concerns selling work, the other concerns earning fees. An agreement for several years of service can put business on the books on the day of signature, while the people responsible for it will be working across several sets of accounts. Accenture explains that it uses several fee arrangements and that recognition depends on performance obligations. A buyer's payment date, a project's recognised revenue and the cash the firm ultimately collects can fall in three different periods. [09] [44]

Consider another service provider. Globant's 2025 filing assigns $1.639 billion of revenue to time and materials, about $686 million to fixed-price contracts and roughly $130 million to licences, resales and other work. Time and materials generated 66.7 per cent of its $2.455 billion total. Globant is itself an international business. I use it because a reader can inspect its division by contract type. A heading such as digital engineering gives a buyer little guidance about who takes the first financial benefit when a task becomes quicker. Two contracts for the same business can give opposite answers. [45]

A timesheet with two owners

Imagine an engagement at our bank. An analyst finds the original payment instruction, retrieves an amendment, attaches the dates to the right entries and passes the file for review. The customer asked a question about one debit. The manager wants the answer to survive a complaint several weeks later. When the analyst has to retrieve the same amendment a second time because its identifier was missing from the first handover, there is a plausible saving. When a second person checks whether the debit was authorised, there is a control. A proposal to eliminate both steps under the heading of duplicate work deserves two separate prices.

Ask the team to mark the hours on actual files. Suppose it can remove 400 hours of repeat evidence assembly from a time-and-materials engagement billed at $150 per hour. The immediate exposure on that line is $60,000. Give the AI supplier $20,000 to operate the replacement service, then allow $12,000 for the additional senior validation the new route requires. The client retains $28,000 if the old hours genuinely disappear, quality holds and the bank can cancel them under the order. The three amounts belong to three different parties. Only the $20,000 is a sale by the AI supplier.

Give procurement another afternoon. The existing partner might repair the handover with a compulsory identifier, fetch the amendment once and cut the same hours by half without introducing a model. That would save $30,000 with no new purchase in this example. A proposal previously credited with saving the bank $28,000 now leaves it $2,000 worse off than ordinary repair. A bidder can still win by improving the price, handling harder cases or delivering something else the bank values. It cannot use the avoidable cost of a bad handover as an entitlement to a new fee.

The empirical case for faster preparation is quite specific. Fabrizio Dell’Acqua and colleagues tested AI assistance with 758 consultants in a preregistered experiment with Boston Consulting Group. On a set of eighteen suitable tasks, consultants given GPT-4 completed 12.2 per cent more tasks and worked 25.1 per cent faster on average. A different assignment required a correct recommendation from financial data and interviews. The unassisted group chose correctly 84.5 per cent of the time. The two assisted groups reached 70.6 and 60.0 per cent. A firm making the staff faster on the first set still needs a way to recognise the second kind of assignment when it arrives disguised as ordinary work. [48]

The experiment used particular tasks, tools and evaluators. It gives us a reason to test two columns on the bank's own return. In one, enter time spent on work the client accepts. In the other, enter what happened when somebody challenged the answer. The second column needs a closing date. If a complaint takes twenty days to arrive, a pilot that closes its books after five days can award the supplier a large improvement merely for sending the next team the consequences.

The narrow desk

EXHIBIT 18 / STAFFING MODEL

The staffing pyramid after automation



A tiered staffing model places five analysts below two reviewers and one partner. Review capacity limits the weekly cohort to 120 accepted files and a $29,520 invoice.

Each tier carries one desk per person. The coloured inlay fills the proportion of that desk’s reserved hours assigned to accepted files. A gold edge marks a tier at its file capacity. The table gives the hours and charges. All staffing, rates and acceptance terms are illustrative.

Open exhibit 18 in the interactive edition

Give the bank 180 comparable files in a weekly cohort. Five analysts have 150 hours between them. Two reviewers reserve 48 hours for this client. A partner has eighteen hours for sign-off. For the purposes of the example, acceptance becomes final after the same thirty-day reopening window in both routes. We are looking back at the outcome of a mature cohort. The team releases only as many files as all three desks can finish, so a file held for the next cohort accrues no billed preparation this week.

At an hour of preparation per file, the analysts have room for 150. Each reviewer needs 24 minutes per file, so the middle desk can get through 120. The partner takes six minutes and can handle all 180. The engagement delivers 120 accepted files and holds sixty. Give the analysts a tool that cuts preparation time in half. They could prepare 300 files in their allocated hours. They still send 120 through to acceptance. A buyer who simply multiplies the improvement at the first desk by the size of the queue has promised sixty customers service that nobody is available to approve.

Now issue the invoices. At hourly rates of $120 for preparation, $225 for review and $360 for sign-off, the original 120 files cost $29,520. With faster preparation, billed analyst hours fall from 120 to sixty. The $7,200 difference in the buyer's invoice is exactly the hour reduction multiplied by the agreed rate. Fixed salaries can continue across the week. A partner who measures success by utilisation now needs another paying engagement on which those freed hours can be worked. Moving them to internal training might be a sensible investment in the business. It produces no replacement invoice this week.

Under a different contract the bank pays $15,000 for each milestone of sixty accepted files. With two milestones delivered, the same customer owes $30,000 under either preparation method. The firm can benefit if its cost of delivery falls, although the size of that benefit depends on staffing, tool

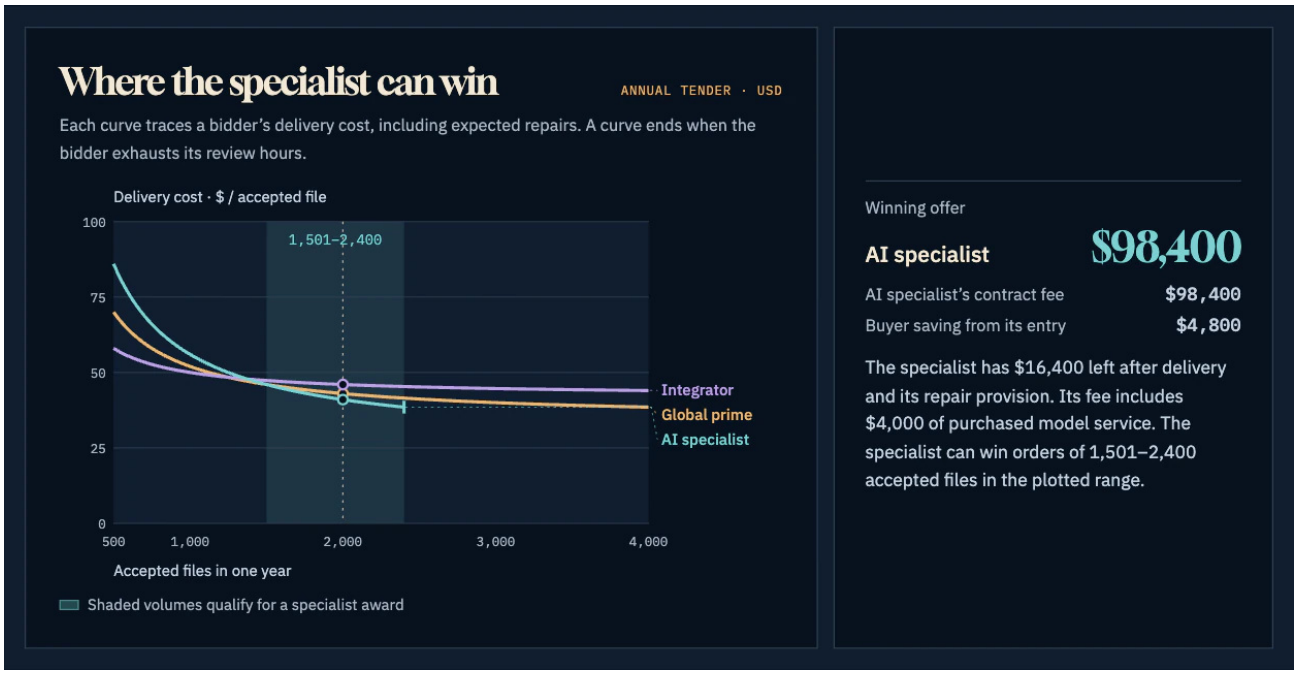
charges and what the freed people do. At other volumes the timing of payment matters. A completed file in a partially filled milestone can carry into the next invoice. The chart shows work done and milestones earned separately so that moving a payment across a billing date cannot manufacture a productivity gain.

Add reviewer time and the bottleneck moves. At 72 review hours the middle desk can release 180 files. With sixty minutes of preparation, the first desk can still produce only 150. With thirty minutes, all 180 can clear. That difference is material to a bank facing a backlog. Before it buys the faster tool on this evidence it should price the extra 24 hours of professional review and identify where they will come from. The second engagement may be won by the firm able to supply a reviewer at the right moment, even where the bank can buy the preparation software from several vendors.

The specialist enters the tender

EXHIBIT 19 / TENDER MODEL

The small specialist’s counterbid



Three delivery-cost curves end at each bidder’s review capacity. The specialist can win between 1,501 and 2,400 annual files and bids $98,400 for 2,000.

Hypothetical tender for the same evidence, acceptance tests and 30-day repair warranty. Shading marks accepted volumes at which the AI specialist can both deliver and strictly underbid its competitors. The global prime has priority on equal prices, followed by the integrator. The bank awards the whole order. Buyer saving is measured against the cheapest qualifying bid without the specialist.

Open exhibit 19 in the interactive edition

Move from staffing one weekly cohort to pricing a separate annual contract. This time three firms must quote for the whole order. The bank specifies the same evidence, service window and thirty-day repair obligation. A small deployment firm proposes to run the work itself. A regional integrator and a global prime can deliver it through larger teams. Give all three the same twenty per cent markup on their expected delivery costs, including provision for repairs. That choice makes the differences in cost visible. In a real negotiation each bidder is free to price its advantage wherever it thinks the buyer's alternative permits.

The specialist has the most expensive entry ticket in this example. It must spend $30,000 to qualify, connect and support its first bank account. Its underlying delivery cost is $16 per file, which includes $2 of purchased model service. It allows $200 per reopened file for investigation and repair. A regional integrator already serving the bank carries only $8,000 of account cost, pays $38 per file and allows $80 per return. The global prime puts $18,000 against the account, $29 against each file and $100 against a return. Incumbents can spread some infrastructure and incident cover across customers. A small firm might have an excellent automated process and still have to buy an expensive hour from a specialist when it fails.

At a five per cent reopening rate those figures give variable costs of $26, $42 and $34 per accepted file. For an order of 2,000 the small firm therefore expects $82,000 of delivery cost. Its $98,400 bid leaves $16,400 of contribution before corporate overhead. The global prime asks $103,200 and the integrator $110,400. If the bank awards the lowest qualifying offer, it saves $4,800 against the next offer and the new prime receives the $98,400 contract fee. Of that, $4,000 pays for the purchased model service. If a reader's definition of AI company includes only the upstream model developer, the associated sale here is $4,000. Adding the two receipts as a measure of the bank's technology spending would count the model service twice.

Trace the curves from left to right. A small order favours the integrator's low entry cost. At 1,250 files the global prime ties it and wins under the stated tie rule. At 1,500 files the specialist just matches the global prime. The next file tips the price advantage towards the specialist. The advantage lasts until its team runs out of review time. At the assumed return rate each file requires an average fifteen minutes of specialist review and repair cover. Six hundred hours cover 2,400 accepted files. For an order of 2,401 the specialist must supply more capacity before it can submit a qualifying whole-order bid. The bank would have to amend the tender to award a fraction.

Increase the estimated reopening rate to ten per cent. The specialist now expects to spend $36 per accepted file before account costs, and its reviewer can cover 2,000. At that volume its price rises to $122,400. The prime's price is $115,200. An extra five returned files per hundred has changed the identity of the winning contractor. These are conditional prices built for inspection. The actual rate at which files reopen, the correlation of failures in an exceptional week and the ability of a reviewer to reach the underlying evidence would all belong in a serious bid.

Where the partnership can still win

The most uncomfortable prospect for the big consulting firm is a smaller rival that learns something expensive on its first account and can use the lesson on its second. If the regional payments scheme changes one field name, the specialist might update the mapping once and correct many client files. Its return-handling cost could fall. However, one shared mistake can also reopen files at every customer on the same morning. The economic value of a warranty depends on the team available that morning. A proposed reserve that covers the average repair bill can leave a small firm short of the people required on its worst day.

The incumbent owns useful assets in precisely that part of the distribution. Its people may already have permission to access historical records. A named partner may have the authority to halt a rollout. It can subcontract model access, offer the bank a lower price at renewal and keep the larger prime fee. A bank would need the supplier invoices to discover how much revenue reached the lab in that arrangement. The answer might be an attractive recurring software business. It would offer much weaker evidence for a prediction that labs are taking over the consulting relationship.

My test is the second purchase order. Record who signed the first engagement, the hours and expenses the bank paid, the share of files still accepted after thirty days and the refunds and repairs that came later. Then look for the same work in the next tender. A specialist that wins the renewal at a price covering its service and its warranty has taken a piece of the consulting market. A firm that sells its software to the incumbent has won a customer of another kind. And if the bank learns to do the work itself and buys only a modest amount of compute, it is the bank that has captured most of the opportunity. All three outcomes can begin with a demonstration in which an analyst works twice as fast.

CHAPTER 08

What becomes of a seat?

A seat disappears from the renewal. Follow the money that returns to the buyer, the fees the platform keeps and the price an AI supplier can defend.

The most expensive person on the bank's software bill may be somebody who left in February. Their manager returned the laptop. Payroll stopped the salary. A name survived on an order form, where it costs exactly as much as the name of the colleague who opens the application every morning. An AI company bidding for the bank's work would love to lay claim to that money. The first person I would invite to its sales presentation is the buyer who can take the name off for free.

Consider a different employee. She opens the case system to find the disputed transaction, checks a second window for the merchant's response and types a summary into a third. Her account shows activity every working day. A software inventory will give her seat a clean bill of health. Yet an agent that can assemble the evidence, preserve the links to the original entries and submit the file for her approval might remove the need for her to hold the full licence. The seat that looks indispensable in the usage log is the more interesting commercial prospect. Winning it requires someone to change how work gets done. The buyer can settle the departed employee's account with an email.

A portfolio, a market and a seller

What does a business actually spend? Among organisations in Zylo's 2026 SaaS Management Index, median annual software-as-a-service spending was $20.6 million. The mean was $55.7 million. The corresponding portfolio counts were 240 applications at the median and 305 on average. Those are customers visible to a software management provider. A provincial architecture practice with twelve employees has little reason to expect either bill. Even inside this dataset, the mean would give a procurement director a substantially larger comparison budget than the median. Neither tells her what a given application costs without the contract. [34]

For a worldwide measure I use Futurum's forecast for the SaaS delivery of enterprise applications across its twelve product segments. The April 2025 forecast put calendar 2026 at $323.3 billion. Its publication date was July 2025. This gives us a denominator with a date and a boundary. Payroll for the employees using those applications is outside it. So are the consulting invoices for redesigning their jobs. Those expenses may fund the same AI purchase. A seller that gets paid from three budgets can report one sale while an analyst who claims the entire sale against each industry has manufactured two extra markets. [26]

Look from the other side of the invoice. Salesforce reported $10.8 billion of subscription and support revenue in the quarter ended July 31, 2026, including $440 million from Informatica. Four quarters at that pace would yield about $43.2 billion. This annualisation holds the July quarter constant. Salesforce separately reports recurring-revenue metrics for particular products, and neither measure is a count of removable human seats. A company can lose a thousand logins at one bank, add a data subscription at another and collect a larger support payment from a third. Looking at aggregate growth, an investor would have no way to reconstruct any of those three renewals. [29]

What a buyer has already taken back

Flexera published an unusually revealing account of buying software for its own employees. It signed a three-year agreement worth $46 million, covering a mixture of SaaS, cloud and installed software. Its report assigns $1 million to hard savings on SaaS and $1.5 million to hard savings on the other licences. It records a further $4.5 million under cost avoidance. More than 300 duplicate productivity seats came out after an acquisition. Flexera sells the management tools its team used. The figures are its account of the negotiation, and it has not released the two invoices that would allow a reader to reproduce the calculation. [38]

What the $7 million included

SaaS hard savings $1.0m · Other licence savings $1.5m · Cost avoidance $4.5m
Flexera’s own three-year account of a mixed $46 million renewal. Source 38.

The distinction matters even if you accept every dollar. One set of terms reduces what the buyer has to pay. Another prevents a price increase the seller wanted. Both make the finance director better off than signing the opening offer. Only a fraction of the headline belongs to SaaS, and none of the reported savings establishes a fee paid to a model developer. Procurement has bargaining power of its own. Treating the entire improvement as a prospective AI sale gives the challenger credit for a negotiation it never attended.

Zylo puts a value on the discrepancy it sees across customer estates. Its buyer guide reports that 54 per cent of purchased licences are used and estimates about $19.8 million of annual unused spend for a typical large organisation. The percentage counts licences. The dollar figure values entitlements. Applying the unused fraction to every dollar of the global forecast would assume the same price for every seat, comparable activity across every vendor and a right to cancel every idle entitlement. The evidence establishes none of those conditions. It does tell a buyer where to start looking. [43]

An audit of our hypothetical bank would begin with an awkward conversation about the definition of use. A recovery officer might sign in only when a payment system fails. A supervisor who approves quarter-end adjustments may look dormant in a February extract. A test account may be indispensable to the team that releases changes. Someone needs to establish what happens when each identity goes away. The quickest route to an impressive saving on a spreadsheet is to remove the people the business needs on its worst day. The first live incident is an expensive time to discover the error.

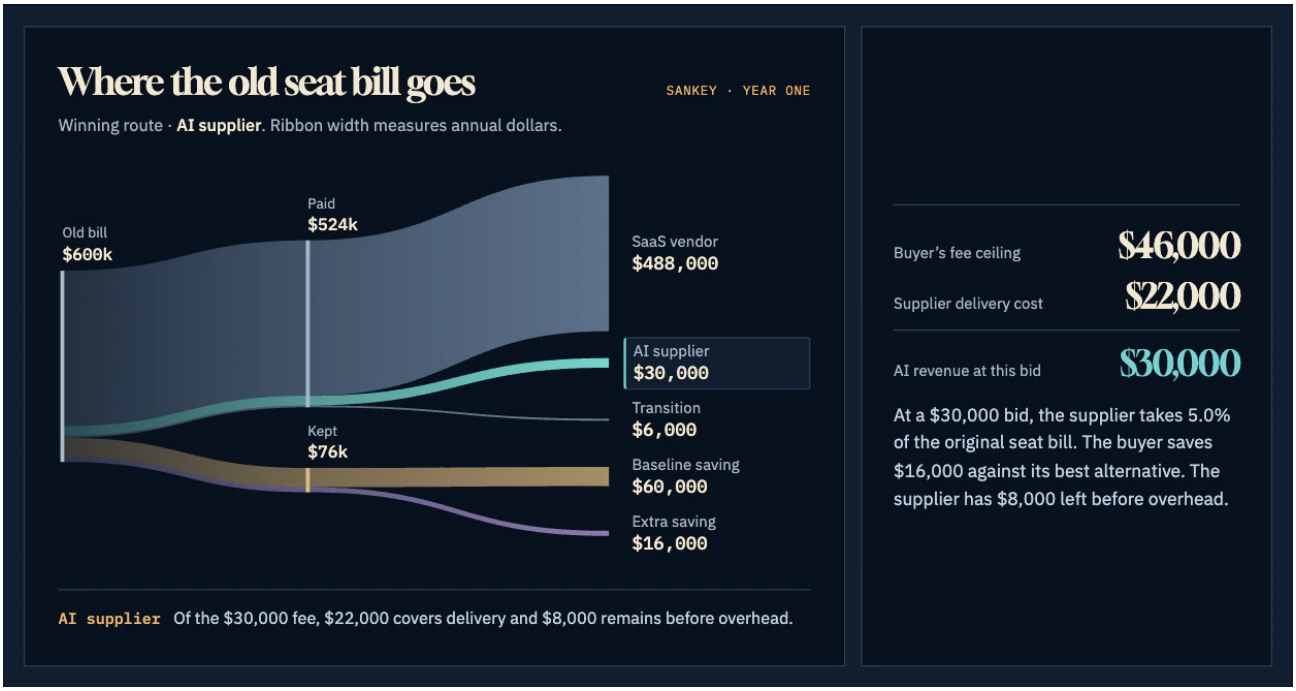
The same care applies to time. Suppose a revised workflow saves three minutes on each of 20,000 accepted cases. That is 1,000 staff hours over the year. Ask the branch manager where those hours go. If the team closes a backlog during regular shifts, customers receive faster service. If an agency agreement falls by a thousand billed hours, cash comes off a different invoice. If nobody changes the roster or the queue, follow the cases to find out whether the saving was measured end to end. In the subscription comparison below I give the bidder zero credit for payroll. A second budget can be introduced when its owner can identify the expenditure it will release.

The price of the hundred and first seat

Give the bank a thousand seats at a negotiated $600 per year. Its annual invoice is $600,000. An internal review finds 160 seats with very little recent activity. Sixty have a continuing purpose, including seasonal reviews and emergency cover. The other hundred can go in the ordinary renewal. Ask for two quotes on the same afternoon. One simply trims the estate. The other changes how the bank prepares the file, so another 120 employees can do their jobs without full access. Assume the access rights for the substitute workflow have been agreed in writing. Otherwise the buyer has a demonstration and a licensing question.

EXHIBIT 20 / RENEWAL MODEL

The missing seat and the new action



A Sankey traces the prior $600,000 seat bill. The AI supplier receives $30,000, transition costs $6,000 and the buyer keeps $76,000, including $16,000 beyond ordinary cleanup.

Illustrative bank, first year. The Sankey reconciles the old $600,000 bill to first-year outlay and money the buyer keeps. The benchmark saving is available through the best competing renewal. The frontier holds discounts and action charges fixed while varying both contract terms. Each award requires a lower total cost than the best alternative and a positive supplier contribution.
Action price basis

Open exhibit 20 in the interactive edition

At an 800-seat minimum, ordinary cleanup takes the bank to 900 seats and $540,000. The revised workflow needs 780 full seats, yet the supplier will bill 800. The seat order therefore comes to $480,000. Of the $120,000 apparently removed from the original renewal, exactly half was available to the bank without the lab. The last twenty technically redundant seats still have a buyer. This is why the size of a price concession can turn on a purchasing clause that nobody on the AI project team has read.

Now let the application vendor charge for some of the work the agent performs inside it. Salesforce lists 100,000 Flex Credits at $500 and a standard Agentforce action at 20 credits. The resulting list price is ten cents per standard action. Our contract assumes five metered actions for each of 20,000 accepted cases, with 20,000 actions included in the annual agreement. That leaves 80,000 to buy, or $8,000. An actual bidder would need a trace showing which operations invoke those

meters. Ordinary API requests have their own terms. Counting every database read as a paid Agentforce action would be as careless as assuming all execution is free. [16]

The bank pays $6,000 in the first year to qualify the new route and migrate the files. Before the AI supplier collects a cent, the buyer has committed $494,000. Against the $540,000 cleanup alternative, the most the bank could pay for the AI service and still break even is $46,000. The lab quotes $1.50 for each accepted case, or $30,000 for the year. Its customer spends $524,000 and retains a $16,000 advantage over ordinary cleanup. The lab earns five cents of each dollar on the original seat bill if it wins. The customer's total saving against the old invoice is $76,000. Those two gains have different owners.

When the deal can clear

There is a seller's reservation price too. Assume that inference, exception review and support cost the lab seventy cents per accepted case, and that it must spend another $8,000 a year to keep this customer running. At the stated volume, delivery costs $22,000. A $30,000 contract leaves $8,000 to cover corporate overhead and reward the capital at risk. At $22,000 the buyer would like the deal more, and the supplier would have nothing left for either purpose. This gives both sides of the bargaining range. On these assumptions, a price above $22,000 and below $46,000 gives the parties something to divide.

Invite the incumbent to the table. It puts an all-inclusive $520,000 offer in front of the bank for the same accepted work. The most the lab can now charge at buyer indifference falls to $26,000, because the other $494,000 of its route is already spoken for. Its original $30,000 bid loses. A revised price of $1.20 per accepted case would produce $24,000 of revenue, $2,000 of contribution before overhead and $2,000 in savings for the customer against the incumbent bid. A single difficult implementation could consume the supplier's entire cushion. A business that can persuade investors it has access to a large pool of annual subscription spending may still walk away from this particular bank.

Try the 900-seat floor in the exhibit. At the original seat discount, the bank would owe the same $540,000 for seats under both designs. Actions and transition then make the external route more expensive before anyone charges for intelligence. Or keep the 800-seat minimum and remove the volume discount on the smaller AI-assisted order. At $750 each, the remaining seats alone cost $600,000. This is the commercial force of a bundle. Engineers can show a hundred employees finishing their work without the full application. The vendor can acknowledge the demonstration and present a revised price schedule that absorbs the benefit.

An AI supplier has a defensible answer when it can change more than the count. It might keep the expensive application for authorised writes and give the bank a cheaper independent work surface for preparing evidence. It might demonstrate that a light-access licence really covers the supervisors who approve its work. It could finance the transition against a longer commitment. Each proposal moves a specific term of the comparison. Financing pushes a cost into a later year. A cheaper licence requires written rights. A longer commitment asks the customer to surrender some of the competition it would otherwise enjoy at renewal. Those concessions have prices of their own.

The incumbent can sell the result

On September 10, 2026, Salesforce completed its acquisition of Fin, formerly Intercom. A week later, Fin's billing guide described a $49 monthly plan including 50 resolutions on a customer's helpdesk and a charge of $0.99 for additional outcomes. Read the definition before comparing that number with the bank's $1.50. Fin can count a resolution when a customer confirms success or stops asking for help after an answer. A configured procedure that successfully hands the matter to a human can also count as an outcome. Our bank pays on an accepted case file. Silence, a handoff and a reviewed disposition are three different things a supplier can competently deliver. The procurement document must specify which one the buyer is purchasing. [39] [40]

Salesforce's acquisition changes the arithmetic of market share. A company can retire seats, move support work onto an outcome-priced service and still send the new revenue to the same corporate group. For the product owner, that is a considerable change in how software works. For the shareholder predicting a transfer out of Salesforce, the identity of the payee settles the question. There is a related possibility for Anthropic. Its August partnership announcement with Salesforce describes both Salesforce skills inside Claude and Claude reasoning inside Salesforce. The document gives a prospective customer two places to begin work. It gives an analyst no price split from which to infer who earns more on a completed case. [41]

EXHIBIT 21 / CONTRACT SCENARIOS

Who can raise the price?

Capability	Bank	AI supplier	SaaS platform
Identity Who may act?	Sets authority	Uses a delegated identity	Enforces permissions
Orchestration Who sequences the work?	Sets acceptance rules	Runs the case and quotes the next one	Exposes approved actions
Execution Who can commit a change?	Approves exceptions	Calls the authorised action	Validates and writes
Case record Who can reproduce the decision?	Requires an export	Keeps the work trace	Retains the case history

The lab can reprice the completed case. The platform can reprice the actions the lab needs. A bank that cannot export the history has two renewals to negotiate.

A matrix assigns identity, orchestration, execution and recordkeeping across the bank, AI supplier and SaaS platform. The supplier runs the case while the platform validates and writes the change.

Each coloured cell marks the lead party for one decision in a hypothetical agreement. One party can control the workflow while another controls the authorised write. Product documentation supports the capabilities. The allocation of bargaining power is the author's inference.

Open exhibit 21 in the interactive edition

For the bank, the most valuable part of an application may sit several steps away from the interface its employees see. An external system calling Agentforce needs an authorised client and an identity permitted to act. Someone still owns the rules for changing an account. A financial workflow can also invoke services outside the CRM. Snowflake's lending-agent guide distinguishes charges for AI from warehouse usage. Databricks documents permissions over the data and tools an agent can

use and traces that can be stored for later examination. A lab bidding a fixed price has to find out which of those bills comes to it, which goes to the customer, and who answers when a case crosses the boundary between them. [42] [17] [18]

I would ask the three bidders to return with last month's hardest reopened case. Give each the original evidence, the policy version in force that day and the supervisor's recorded objection. Ask for a corrected file and an itemised price for producing it a second time. A platform that owns the history may be able to do that cheaply using its existing support staff. A lab working across several applications may recover evidence the incumbent never captured. The bank that can export the whole file can put both accounts of their advantage to a test. A product demonstration built from a clean new case leaves the costliest part of the competition unresolved.

What I would count at the next renewal

There is good evidence that assistance can change throughput. Brynjolfsson, Li and Raymond studied 5,172 customer-support agents and found an average increase of about 15 per cent in issues resolved per hour with AI assistance. It is an important result for someone responsible for a queue. It also shows why employee count makes such a poor shortcut for pricing this market. A service team can use the gain to shorten waits or train new staff, still renew every seat, and willingly pay for the assistant on top. That would create a new revenue stream from better service. Calling it cannibalised SaaS revenue would assign it to the wrong budget. [20]

My prediction for this market is narrower than a wholesale abandonment of enterprise software. A lab can take the renewal when the job crosses applications, the buyer can remove a contractual expense, and the lab will stand behind a result the incumbent finds costly to match. At $30,000 apiece, a hundred wins like the illustrative one would bring in $3 million over a year at the stated volume. Finding a hundred banks with the same removable seats, service level, acceptance rate and willingness to switch is the expensive part of that multiplication. A national market-share figure requires evidence on the frequency of those conditions and the price of the bids that actually win.

To count one win I would want the expiring order, a credible cleanup quote, a competing offer where available, the replacement invoices and the acceptance log for the same case mix. I would follow the contract into its second year. The migration charge should disappear, while a discount granted for the pilot might expire. If the original suite collects more in actions than it lost on seats, it has successfully changed the unit on its invoice. If the lab earns a recurring case fee and the bank keeps the saving after paying for exceptions, the predicted transfer has a customer, an amount and a date. Those are the entries worth adding up.

CHAPTER 09

The database collects its toll

A bank can buy a cheaper answer and still run up its warehouse bill. Follow one approved question all the way to the meter.

The bank has issued two receipts for the same payment. One says Friday. The other says Monday. A customer wants to know which date counted towards her bill. A model can give a perfectly sensible account of weekends, settlement windows and posting delays. Someone at the bank still has to establish which rule applied to this payment, what the ledger contained on Friday evening and whether the adjustment made on Monday changed the answer. The customer has already supplied the bank with everything it needs. The bank has put those things in different places.

Suppose an outside company offers to deliver the answer and its supporting entries to a case officer for ninety cents. Its sales director might reasonably see a route into an established data platform's business. I would ask her to bring a second person to the meeting, the engineer who can tell us how many times the warehouse started while her agent was doing the work. The platform can lose the customer's first click and still get paid for almost everything that happens afterwards. In a contract priced by consumption, an enthusiastic competitor can be an excellent customer.

Three numbers a buyer could recognise

In its 2025 update, Gartner forecast $161 billion in worldwide database management system spending for 2026. That category covers a great deal of work unrelated to our question, including databases that record sales as they happen. Some of the work our agent needs, such as maintaining a business glossary or correcting an old payment rule, can sit in a services contract outside that market. I use the forecast to locate an established budget. Applying a prospective percentage loss to it would require evidence about the jobs buyers actually move. [28]

Snowflake's filing offers a more precise view of one company. It earned $1.491861 billion in product revenue in the three months to July 31, 2026. Its professional services and other revenue brought the quarter's total to $1.546793 billion. Holding product sales at the July quarter's pace for four quarters gives $5.97 billion. An investor wanting the trailing four actual quarters can add the four rounded quarterly figures in the same filing and obtain about $5.21 billion. Growth accounts for the difference. Choosing between those numbers changes the apparent size of the business by roughly three quarters of a billion dollars without changing one customer order. [31] [51]

Databricks said on August 13 that it had crossed a $7 billion revenue run rate. In that announcement it also reported more than a thousand customers consuming its products at annual rates above $1 million, including more than a hundred above $10 million. A reader can take from this something more concrete than a valuation. There are individual accounts in which changing how a fraction of the work gets done can matter to both parties. Databricks supplies thresholds at a particular time. Snowflake supplies recognised sales over a fiscal period. Precise relative market shares would require comparable numerators and a common market definition. [32]

The Snowflake filing also puts a name to the party beneath the platform. Third-party cloud infrastructure, including resources used for inference, represented about 75 per cent of its $434.4 million cost of product revenue that quarter. The denominator is the cost of delivering the product. The share of the buyer's dollar going upstream will differ. When a platform customer and a cloud operator both announce a new AI contract, it is worth checking whether there was only one buyer at the start of the chain. The same inference run can appear in the platform's revenue and the infrastructure supplier's sales. [51]

A saving the engineers already found

A good outside bid should have to beat what a competent data team can do for itself. Zendesk engineers Neilson Zanotti and Anatoly Mikhaylov published an instructive example. They copied three of their largest data models, cloned the production workflows and ran them in parallel against different warehouse configurations. Each configuration processed the same input. The engineers reported a reduction of over 80 per cent in credit costs from sizing the warehouses appropriately and described bringing additional work onto Snowflake with the capacity they had freed. The comparison concerns particular workloads and configurations. The engineers gave no company-wide before-and-after cash invoice. [52]

There are two buyers an AI sales team could meet at that company. The first owns a set of costly jobs and can document what it took to reduce their consumption. The second owns an unmet demand for analysis and can use the headroom immediately. The first buyer's evidence establishes a saving on the work tested. The second buyer can cause total consumption to rise in the same month. It would be quite a trick to sell an AI service against the former's supposed waste while leaving the latter's order out of the budget.

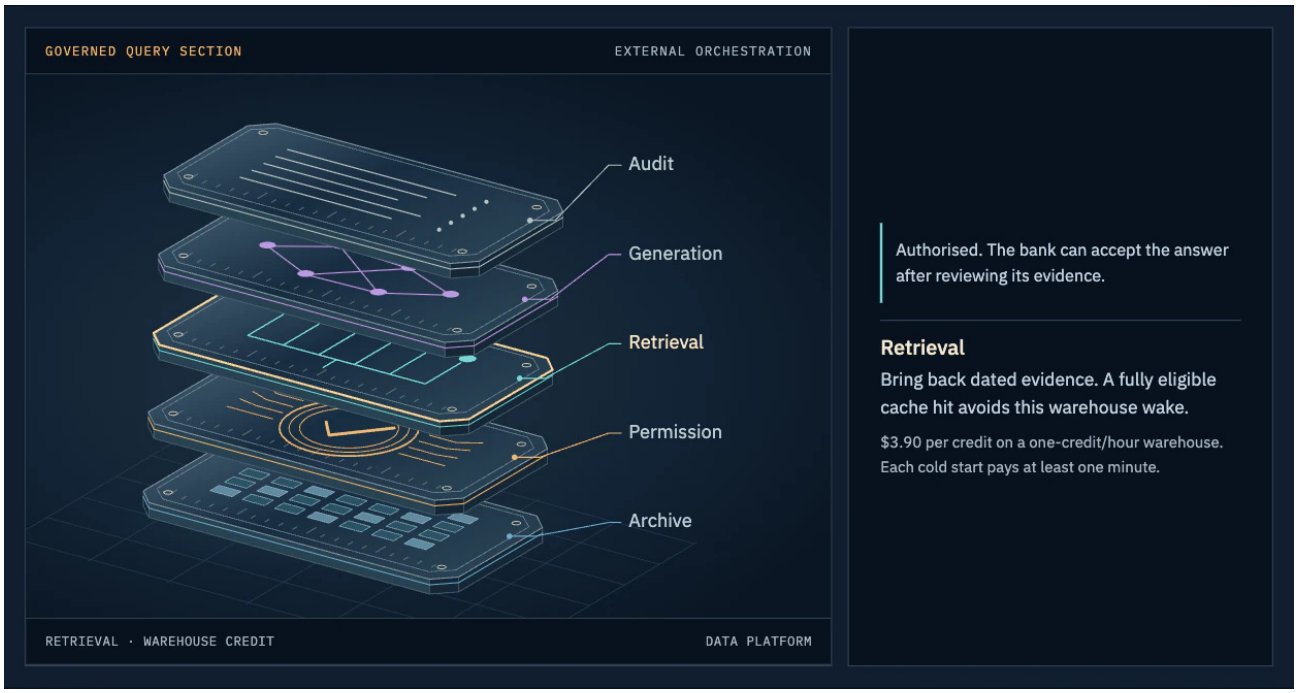
Back at the bank, a plausible source of waste is a daily job that reconstructs the payment calendar independently for each department. Another is the analyst who reruns the same extract because a case officer cannot see when the first one was made. Remove the redundant job and publish a dated extract. Those improvements deserve a price of their own in the tender. Some repeated reads will survive. The Monday officer may need both Friday's official position and the later corrected record. Combining them into one current view can shorten the query log while making it impossible to establish what the bank told its customer.

What the officer is allowed to know

Follow the question through five layers. The bank retains a versioned copy of the ledger. A permission check determines which entries the officer may inspect. Retrieval brings back the payment, the posting history and the dated policy. A model composes an explanation. The bank keeps a record of the sources used, the decision and the reviewer. In one arrangement the platform supplies the whole service. In another the outside company organises the search and prepares the answer, while the bank continues paying the platform that holds the records. Select either arrangement in the cutaway and inspect what each party supplies.

EXHIBIT 22 / GOVERNED QUERY MODEL

The governed query



Five layers show archive, permission, retrieval, generation and audit. The highlighted retrieval layer charges warehouse credits unless the bank can reuse eligible evidence.

A single request, read from the archive upwards. The native offer bundles all five stages. The external bid leaves the warehouse, access control and evidence trail with the platform. The route is an illustration of capabilities described by both vendors. Prices in the model are stipulated bank terms. Snowflake evaluation . Databricks authorisation

Open exhibit 22 in the interactive edition

The permission check comes before reuse. Yesterday's senior officer had access to two accounts. Today's officer has access to one. Both can type exactly the same question. Databricks documents an option for an application to act on behalf of its user. The user's existing Unity Catalog permissions apply within the scopes granted to the application. The company separately describes how an agent can call tools and retain traces for inspection. These mechanisms give the buyer something to test with the actual roles in its bank. An application running solely under its own service identity can have a different entitlement from the employee at the keyboard. [18] [53]

Snowflake's rules for reusing a stored query result illustrate the same point from another direction. Privileges, unchanged underlying data, query text and configuration can all affect eligibility. A reusable database result also differs from a previously generated letter to the customer. The figures here count a cache hit when all the evidence for an attempt can be reused under the current permissions and freshness rules. The agent still prepares a response and the reviewer still decides what to accept. Revoking access stops a retrieval. An audit entry can record that the system refused the request. There is no payable accepted answer on that file. [54]

A supplier engineer might object that a good agent can find another way to answer. That is precisely what the bank has to test. Snowflake published a synthetic lending evaluation in which eight of forty correct baseline answers took an unexpected tool path. In a separate run, evaluating forty questions at on-demand prices came to about twelve dollars, including both judges and warehouse work. The example uses an Enterprise account on AWS in Dublin. It tells us how the authors priced a particular evaluation and why they examined the route to an answer. No customer reached a bank decision for thirty cents on the strength of that experiment. [17]

A bid for nine thousand answers

Give our illustrative bank ten thousand authorised questions a month. The buyer accepts nine thousand after a thirty-day reopening window. The other thousand go to its existing exception desk under either contract. The bids stipulate the same mix of cases, the same escalation time and the same evidence and acceptance standard. One buyer payment covers the service even if an accepted case needed two attempts. Both suppliers have already qualified at this renewal. First-time installation, unchanged bank salaries and the cost of the common exception desk sit outside the comparison. A buyer changing any of those terms should reopen the prices too.

The current platform bill for this work is $12,600 a month. The platform offers a new all-inclusive price of $10,350, equivalent to $1.15 for each accepted case at the stated volume. The outside company quotes ninety cents for the answer. A director comparing the two quotes might conclude that the newcomer saves a quarter on every case. That quarter has to pay for a considerable amount of work. Under the split contract, the bank still pays the platform $900 a month to retain the records and another $300 for authorisation and the audit trail. These are assumed commercial terms. The two vendors in the reporting above have offered this hypothetical bank no such tariff.

Now turn to the warehouse. Assume one question in ten needs one fresh attempt. That produces eleven thousand retrieval attempts across the month. Half can reuse all the evidence they need. Each of the remaining 5,500 wakes a dedicated warehouse, makes six consecutive twenty-second queries, then suspends it. Under the illustrative one-credit-an-hour configuration at $3.90 per credit, two minutes cost thirteen cents. The month’s bill for the warehouse comes to $715. The rate uses the Dublin platform-credit example in Snowflake’s guide. The figures use its documented minimum of sixty seconds per provisioning and subsequent billing by the second. The duration, isolation of workloads and immediate suspension are our assumptions. [17] [55]

EXHIBIT 23 / COST SENSITIVITY

How often will the warehouse wake?



A heatmap prices accepted answers by cache reuse and queries per uncached attempt. At the selected terms the external route costs $10,015 in total, or $1.1128 per accepted answer.

Horizontal movement changes the queries per uncached attempt. Vertical movement changes the share of attempts with fully reusable evidence. Cells show proposed cost; the award uses the unrounded amount. Retries change warehouse activity while accepted output stays fixed. Each row's first three cells share the same price because they fit within the sixty-second minimum.

Billing rule . Regional price example

Open exhibit 23 in the interactive edition

Add the invoices. The bank pays $8,100 to the lab and $1,915 to the data platform, a total of $10,015. The bank keeps $2,585 compared with its current agreement. It could already have kept $2,250 by accepting the native offer. Winning the argument for this particular lab therefore puts an additional $335 in the buyer's hands. Of the $8,435 in platform receipts at stake relative to the native renewal, $8,100 changes payee and $335 stays with the customer. A headline about the bank saving $2,585 would give the challenger credit for a concession the incumbent had already put on the table.

The outside supplier needs another calculation before it signs. Suppose its own inference, support and review cost sixty cents for each accepted case, including its spending on attempts that failed, and it carries $900 of fixed monthly delivery cost. The contract yields $1,800 before corporate overhead. The most the lab could charge at buyer indifference in the default workload is $8,435 a month, just under ninety-four cents per accepted case. A competitive fee has to leave something below that ceiling for the bank. A lab with a more expensive remediation desk might find a customer willing to buy at ninety cents and still be wise to decline.

How a cheaper model can enlarge the bill

The first three columns of the heatmap have identical prices in each row. A cold warehouse running one twenty-second query is charged the sixty-second minimum. Three consecutive queries still fit within that first minute. A fourth runs the clock to eighty seconds. Batching enough useful work into an already paid minute can improve the result without another cent of warehouse expenditure. Sending each of those queries to its own cold cluster would give entirely different economics. So would running them on a warm shared warehouse that has paid for the minute anyway. The exhibit exposes the assumption because it can decide the winner.

Choose twelve queries per uncached attempt and leave the cache hit rate at one half. Warehouse spend rises from $715 to $1,430. The platform receives $2,630 on the proposed external route. The buyer would spend $10,730 in total and can accept the platform's $10,350 renewal instead. A lab hoping to get in by cutting the price of inference now has to take at least $380 out of its quoted fee to reach a tie. In this model the bank keeps the incumbent at a tie, so the lab has to concede a little more to win. The heatmap shows the cost of the proposed route even when the buyer would decline it. A losing proposal earns zero.

There is another way for an AI company to increase platform revenue. A regional office that once put questions on a list for Friday can now submit them throughout the day. Its staff discover errors before statements go out. A successful pilot spreads to five other offices. Even if the warehouse cost per accepted case falls, the number of accepted cases can rise faster. That is valuable new demand. A financial account that silently treats every new question as a query transferred from an existing workload will exaggerate the old vendor's loss and understate what its infrastructure helped create.

The incumbent's best bid

A data platform has a credible route into the customer's work surface. On November 11, Databricks announced a version of Genie intended to carry out long-running work and bring decisions back to business users. Its materials describe access across sources well beyond the company's own tables. A buyer can use that announcement to invite an additional bidder. The commercial test is whether the platform can complete the same mixed queue, subject to the same release rules, for less than the outside team. An outside lab answering questions across several competing platforms could still have an advantage with a bank that wants a portable workflow. The price of moving its evidence and its acceptance tests becomes part of that advantage. [57]

I would request one month of linked records before accepting either account of the opportunity. For each case, show the submission, authorisation, attempts, query identifiers, evidence version and reopening outcome. Join those to the charges the buyer actually incurred. Snowflake's per-query attribution view excludes warehouse idle time. Its credit totals cannot by themselves price a warehouse that spent much of a shift waiting for work. Reconcile the warehouse meter to the account invoice, then check the commitment. Credits left unused under a minimum purchase can become capacity for a later job. They turn into a near-term cash saving only if the contract lets the buyer pay less or avoid another purchase. [56]

The hardest question at the second renewal is who still knows how to prove an answer. A bank that owns the definitions of posting date and value date, the permission map and a set of difficult cases with agreed outcomes can invite fresh bids for the reasoning work. A bank that let its supplier keep the only working versions of those things can still switch model providers on paper. Its staff will spend the changeover reconstructing the institution's memory. The prize for the AI company is the right to organise recurring, valuable work. The data company can keep collecting every time that work goes back to the ledger. Which one earns more turns on a fact the buyer can now inspect, how much it paid for an answer it could use.

CHAPTER 10

A bank that keeps its mainframe

Follow a proposed correction through the people who can release it. The line an AI service can win sits beside the core licence.

Give the payments desk a receipt for $240 and a customer who cannot see the money. The bank has received the funds. The reference that would ordinarily tell it which account to credit did not survive a change of format. Its books still balance. The sum sits in an account for money received and awaiting allocation. To the customer, the balance that matters is $240 short. To the bank, there are now two jobs. Someone must establish who owns the money. Someone with authority must change what the bank records that it owes each party.

This is an invented case. It has a surprisingly concrete procurement consequence. A supplier can put a fine new screen in front of the payments officer, find the missing reference and prepare both entries. The bank can buy all of that and pay exactly the same annual fee for the software that keeps its books. An investor looking for evidence of the sale in the incumbent's maintenance revenue would be watching the wrong invoice. Start with the purchase the supplier made possible. Which expense would the bank have incurred to resolve the same receipt without it?

The estate behind a single entry

Installed software is unusually easy to overvalue as a prize because the word covers so much. A bank may operate the code, rent the machines from a service provider, license the database from a third company and retain the engineers who wrote its original deposit application. For the next renewal, those are four different negotiations. An operations team can stop paying an agency to work the evening queue. It cannot thereby terminate a perpetual licence that was bought years ago, cancel this year's support agreement or reclaim the salaries of engineers who have already moved to another queue.

There is a defensible, narrower denominator. Futurum's April 2025 forecast, published that July, put calendar 2026 on-premises enterprise applications at $91.8 billion across twelve product segments. Cloud-delivered applications appear separately in the same forecast. An internal banking program written by the bank has no software vendor sale simply by virtue of continuing to run. The figure supplies a scale for a commercial applications business. It leaves unanswered how much the owners of those systems pay to operate them, answer queries about them or keep the next release compatible with them. [26]

SAP provides the view from a supplier's books. In the September quarter of 2026 it recorded €147 million in software licence revenue and €2.434 billion in software support revenue. Together that is €2.581 billion, or €10.324 billion at four identical quarters. Support accounted for roughly ninety-four cents of every euro in the quarterly sum. The annualisation is an arithmetic comparison. It makes no allowance for seasonal buying. SAP sells applications to many industries, and its accounts do not identify the installed base at our specimen bank. The example tells a bidder why support deserves its own line in a commercial analysis. [30]

The term of that line can outlast several automation pilots. SAP offers mainstream maintenance through the end of 2027 for specified Business Suite 7 core applications and eligible enhancement packages. Its optional extended period through 2030 carries a premium of two percentage points on the maintenance basis. A buyer who puts a new service alongside an eligible installation still has to decide how to support that installation. The working interface that wins the AI pilot may even make the older system easier to retain for another year. In that event the challenger has created value for a supplier it once expected to displace. [64]

Where a customer has already done the engineering

In Italy, Gruppo BCC Iccrea connected services outside the bank to banking applications running on its IBM Z infrastructure. IBM's account describes an important detail of the direction of travel. Programs on the mainframe invoked external services through z/OS Connect. The team also sent operational records into the bank's existing Elasticsearch monitoring environment. A person investigating a problem could follow work across the boundary without relying on separate teams to reconstruct each end by hand. The bank retained the core and changed the cost of reaching across it. [60]

IBM reports 65 per cent less integration effort in a comparison with alternative technologies. It also describes forty person-days saved annually in incident management and half a full-time equivalent saved in monitoring. These observations refer to different activities. Adding them would require schedules showing whether the same people and hours appeared in both estimates. The account includes a named bank representative and identifies the products the bank used. It gives us no invoice for an AI supplier and no payroll comparison. I take it as evidence that the incumbent can deliver part of the efficiency a challenger might promise in its pitch. [60]

This is a serious competing bid. In our bank, an engineer might add the missing reference to the queue with a deterministic lookup. Once the relevant fields have been mapped, ordinary software can assemble the common receipts cheaply. A vendor relying on language interpretation has to find enough difficult but resolvable cases to cover what it charges. Its best trial would include messages that arrived under an earlier format, ambiguous remitter names and receipts for two customers with the same trading name. A trial drawn entirely from the easy queue would establish very little about why the bank needs this particular supplier.

What the machine may read

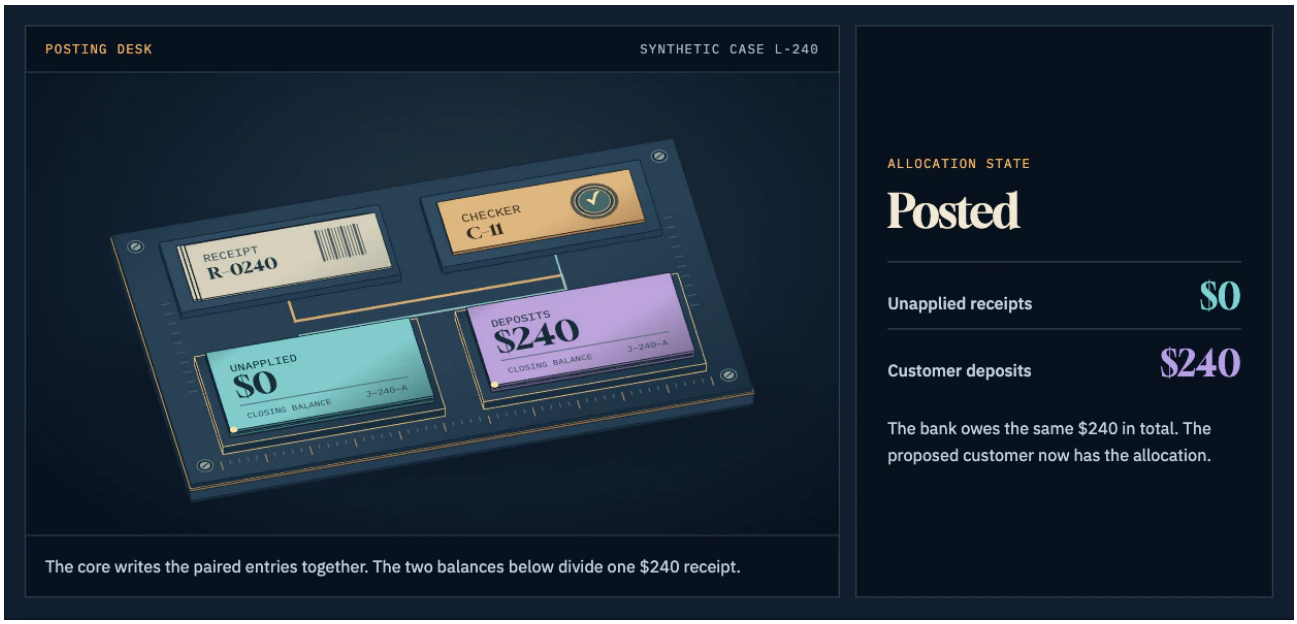
Put the arrival record, the bank statement and the customer's remittance advice on one desk. The captured amount is $240 on all three. Two documents carry the payment reference. One identifies the intended account. A payments officer can now ask a useful question of an assistant. Which two pieces of evidence link this receipt to this beneficiary, and where in the original documents can I inspect them? A useful answer returns the records together with the proposed match and any inconsistency. A match that relies only on the amount remains in the queue. Other customers can send $240 on the same day.

Payment networks have developed identifiers for this reason. Swift describes a 36-character Unique End-to-end Transaction Reference that accompanies payment instructions on its network and can be carried through the chain to help track and reconcile the payment. The reference travels with the payment. A bank still decides which of its own records corresponds to that payment and whether a proposed internal allocation is authorised. In the exhibit the shortened receipt number is a local specimen identifier. There is no connection to a payment network. [62]

The assistant has a copy of the relevant evidence and permission to propose the entry. A bank employee acts as maker by checking the identity, amount, currency and account, then attaching the proposal to the case. Another authorised employee acts as checker. She can return it if a document is missing. The existing posting service accepts her instruction and commits both sides to the bank's journal together. The machine that prepared the evidence has no ledger-writing role in this design. IBM documents that z/OS Connect can expose existing functions through APIs and allow applications on the platform to call out. The identity that may call a particular function, and the circumstances in which it may do so, remain decisions for the institution. [19]

EXHIBIT 24 / DOUBLE-ENTRY MODEL

The bank’s double entry



A posting desk and two balances show one authorised $240 receipt. The bank has credited deposits and cleared the unapplied liability.

The two front registers show the split of one received payment. The rear tray holds the evidence and the brass desk represents the bank's independent approval. Each journal adds equal debits and credits. Repeating a receipt adds no entries. A reversal requires fresh approval. The diagram models an internal held allocation. Interface context

Open exhibit 24 in the interactive edition

Watch the accounts from the bank's side. Money held for its customers is a liability. Debiting the unapplied receipts liability reduces it by $240. Crediting customer deposits increases that liability by $240. The bank still owes the same amount in total. It has changed its record of whom it owes. Neither entry by itself would describe the transaction, and the bank does not release a half-posted journal for someone to complete later. In the exhibit the customer column begins at zero for this receipt alone. It makes no claim about the customer's other deposits or available balance.

When the same receipt arrives twice

Imagine that the posting succeeds and the acknowledgement arrives late. The caseworker's screen still looks open. A second copy of the same receipt then appears. If the desk treated each arrival as new work, the customer would get two credits for one incoming payment. A well-designed core records a stable posting reference and returns the existing result when that same authorised instruction is presented again. Both ledger balances stay where the first posting left them. The history can record another message received without claiming another dollar received. This is a feature of our specimen, which the buyer should require a prospective supplier to demonstrate against its own systems.

Look at the supplier's invoice too. A service priced on accepted corrections can earn one fee here. If it bills the second message or a read-back as another resolution, every interruption becomes a sales opportunity. A buyer could prevent that by defining the billable event as one accepted outcome against a unique case, subject to a stated reopening window. It would also need a rule for genuinely separate receipts made by the same customer for the same amount. A counter so crude that it suppresses the second legitimate payment would introduce another class of complaint. This distinction is worth settling before a volume discount obscures the unit being purchased.

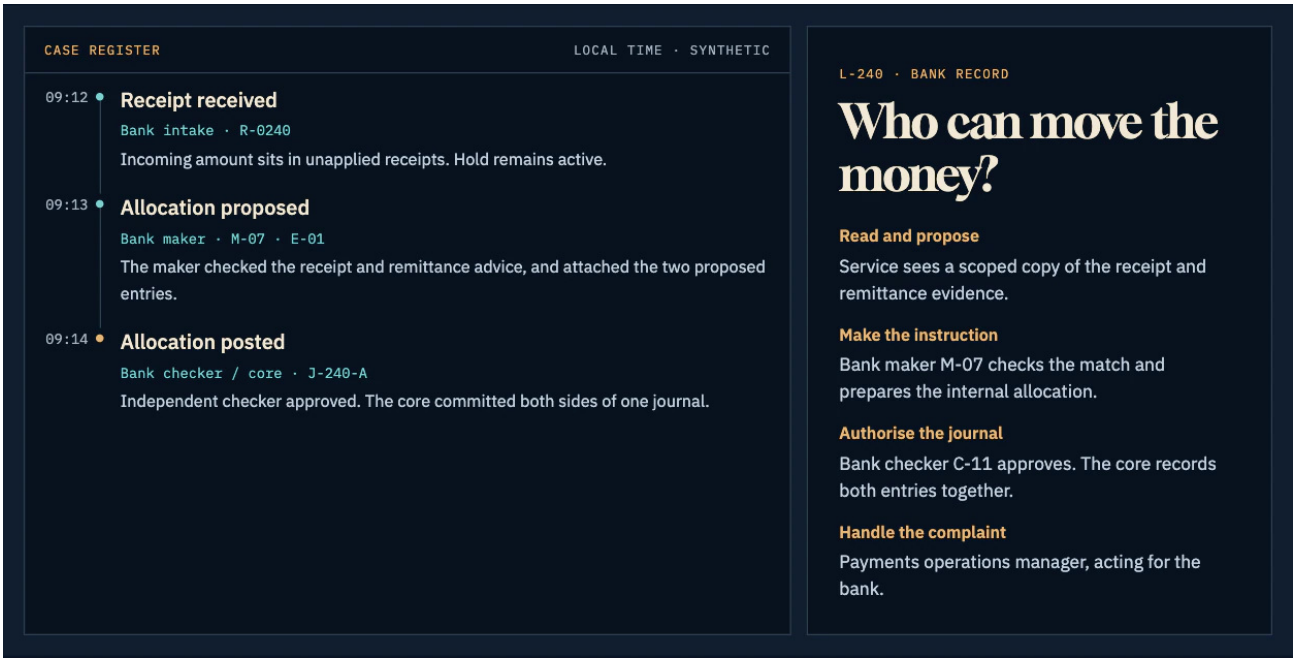
Now let the desk discover that the remittance advice belongs to a different beneficiary. In our specimen the funds remain under an internal hold, and the bank can correct this internal allocation under its agreed procedures. The maker records the new evidence. The checker independently approves a linked reversing journal. The second pair debits customer deposits and credits unapplied receipts. The original pair remains legible with its own time and authorisation. The customer service team can explain which entry appeared first, why the money returned to suspense, and who owns the open investigation. Money already sent onwards or made available under other terms would require a different procedure.

A record that survives the supplier

I would ask the person negotiating the contract to imagine a complaint six months later. A customer has a statement showing the credit and the subsequent debit. The model version has changed. The employees who ran the pilot have moved to other work. The customer asks why their money was moved. A record held only in the supplier's chat history would put the bank in the position of asking permission to read its own decision. A case file that carries durable receipt, evidence, maker, checker and journal references lets an authorised reviewer give a specific answer, even if the service no longer runs.

EXHIBIT 25 / AUDIT AND CASH

What a bank must see before it pays



A timed bank register follows receipt, proposal and posting over three minutes. The facing record names the maker, independent checker and complaint owner.

The left-hand register follows the selected specimen in time. Each decision retains its evidence and record reference. A repeat changes the event history. An approved reversal adds a second journal. Identifiers, times and account values are synthetic. Payment reference context . Responsibility under EU DORA

Open exhibit 25 in the interactive edition

The reviewer should also be able to reconstruct the decision that was available at the time. A document corrected on Tuesday can make Monday's approval look indefensible, unless the evidence attached to Monday's file has its own version. A trail that preserves the record used on each occasion protects the customer who needs a correction and the employee who acted in accordance with the information then available. It gives a procurement manager something much more valuable than the vendor's general promise of explainability. She can hand a colleague the file and see whether the colleague reaches the same journal and the same two amounts.

There is a legal reason for many European banks to care who can do this after the supplier leaves. Article 28(1)(a) of the EU's Digital Operational Resilience Act keeps covered financial entities responsible for their obligations under the regulation and applicable financial services law when they contract for ICT services. It does not prescribe our particular two-person desk. In this example the checker authorises a posting under the bank's policy, the payments operations manager owns unresolved cases and the bank remains the customer's counterparty. Commercial terms can make a supplier reimburse certain failures. The customer still has a bank to call. [63]

A thousand hours and a small cheque

Put a price on the desk. Assume it handles 12,000 comparable reconciliations each year. Suppose that after ordinary improvements to the existing queue, a proposed service still saves a net five minutes of bank effort on each accepted case, including review, returns and support. At constant volume and service quality that releases 60,000 minutes, or 1,000 hours. A schedule reconstructed from the roster shows that a quarter of those hours would have incurred overtime at an incremental cash cost of $50 an hour. The other 750 fall within salaries the bank has committed to pay. On these assumptions the expense it can remove is $12,500 per year.

The lab asks $9,000 annually for the workflow, equivalent to seventy-five cents per accepted case at this volume. The bank expects to pay another $4,000 once to qualify access and test the integration, with no further increase in access charges. Its core licence, maintenance and ordinary staff payroll continue on both routes. In the first year the bank spends $13,000 to avoid $12,500. In the second it spends $9,000 to avoid $12,500. At the end of two years the lab has billed $18,000 and, ignoring discounting, tax and any change in volumes, the bank is $3,000 ahead in cash. These are proposed terms and an illustrative award, conditional on the bank agreeing to fund the first-year shortfall.

Change the purchase order by a single word. Make the $4,000 access and test charge annual. The bank now loses $500 in each year. If it awards only contracts that save cash within two years, the lab at this price gets no sale. If payments management values shorter customer waits, it could still fund the service and write that objective next to the cost. A bank that promises shorter waits should measure the time from receipt to corrected customer balance, including the time a case spends waiting for a checker. Measuring only the seconds to prepare a draft would make it easy to deliver the wrong improvement.

There is a limit to the lab's bargaining power even if the trial goes well. Over two years the roster releases $25,000. A one-time qualification expense leaves $21,000 for a supplier at buyer indifference. That is $10,500 a year, or 87.5 cents a case. If qualification recurs, the ceiling drops to $8,500 a year. Within the first ceiling the lab still has to pay for model calls, its support desk and any rework it has promised to absorb. Until the supplier supplies those costs, the $18,000 on its two invoices measures revenue. The bank's success on the overtime budget tells us nothing definitive about the return to the lab's shareholders.

The upside is just as instructive. If one more quarter of the hours replaces overtime on the same terms, the removable expense doubles to $25,000 per year. The AI fee stays at $9,000 unless the supplier can renegotiate. The bank keeps most of the improvement. If released hours instead go into investigating older cases, depositors may obtain long-awaited corrections without any reduction in this year's payroll. Both outcomes could justify the work. Only one puts substantially more cash into the next negotiation. Ask who owns the roster before accepting a business case priced at a thousand hours times an average salary.

The offer at the next renewal

A technically sound replacement can still impose large costs on customers. In April 2018 TSB transferred its data to a new platform. The regulator found that customers then had difficulty obtaining service across channels. Its final notice records 225,492 complaints between 22 April 2018 and 7 April 2019 and £32,705,762 paid through the redress programme. The episode concerns a particular migration and its governance. It supplies neither a failure rate for mainframe replacements nor a price for introducing an assistant. It does explain why the manager responsible for deposits might reasonably insist on a reversible trial before committing the entire book. [61]

I would start the commercial comparison at a branch that handles both common receipts and disputed ones. Give the existing team time to improve its matching rules. Have the prospective supplier prepare evidence in parallel on a defined sample. Release the incumbent queue or overtime booking only after the bank observes the same accurate closures, stable reopening rates and acceptable waits through a demanding settlement period. Include the time needed to handle a service outage and hand the queue back to staff. Then ask the incumbent to price its own version of the improvement. Its knowledge of the bank's existing records may buy it a shorter implementation, even if its model costs more to run.

The rates in this chapter give the challenger a credible route into an old estate. They also give us an unusually clear way of recording a loss. A signed workflow subscription alongside continuing maintenance establishes a sale into the bank's operating budget. A reduction in the incumbent's maintenance invoice for the same scope would establish a different achievement. A faster demonstration with no change to the order book establishes neither. At this bank the earliest sign of an industrial shift may be a modest overtime line that disappears from the payments manager's budget. To learn who captured it, turn the page of the accounts until you find the new payee.

CHAPTER 11

Five industries rewrite a purchase order

Give five suppliers the same buyer. The winning bids show how much of a million-dollar budget an AI company can collect, and how much the bank keeps.

Give me the bank's accounts payable file and I can tell you quite a lot about the business of artificial intelligence. Give me five market forecasts and I can tell you how large five analysts think their subjects are. The first file contains something precious. Each payment has left the bank exactly once. That property disappears as soon as we start adding the sales of companies that buy from one another. A dollar that pays a model company can turn up again in a cloud provider's results. A dollar that pays a consultancy can appear in both. All three firms have done real work. The bank had one dollar.

This chapter puts five claimants around the same table. Our buyer is an illustrative bank with one million US dollars of annual spending under review. It handles the same volume and mix of work after renewal. Every bidder faces the same requirements for service, evidence, reopening and recovery. The cash comes out of a single year's budget. The amounts were constructed to make the contracts independently awardable, a condition that often takes considerable work to establish in a real business. They describe a fresh procurement exercise. The separate worked examples in the preceding chapters have different scopes and cannot be stitched into a customer.

Across the table sit a cloud account director, a software sales representative, a consulting partner, a supplier supporting an installed system and a data platform vendor. Their existing claims are $280,000, $260,000, $220,000, $150,000 and $90,000 respectively. A sixth firm wants to provide AI services using some of the same suppliers. The interesting number is the fee it can collect once every person at this table has been given a chance to respond. The easy intellectual mistake would be to send the other five home before asking for their prices.

Clearing the table

Start with the biggest line. Synergy Research estimated worldwide cloud infrastructure service sales of $167.7 billion in the September 2026 quarter and $558 billion over the preceding twelve months including that quarter. At the quarterly pace the industry would have been running at about $671 billion a year. The difference between that extrapolation and the trailing total is large enough to change an investor's sense of the hurdle a challenger must clear. Synergy already includes platform activity by AI labs where it meets its definition. A supplier can enter this market statistic before it takes a procurement decision away from anyone. [25]

Amazon reported $45.984 billion of AWS segment sales in that quarter. Some of a model company's success can become AWS revenue through purchased computing power even when a bank deals exclusively with the lab. For the bank's own $280,000 line in our example I include only money it pays directly for its runtime and recovery capacity. The lab may also rent machines and the bank may pay a separate data provider that does the same. Both of those upstream cloud bills

belong inside their respective suppliers' costs. Taking them out of the buyer's cloud column makes the chart smaller and the displacement claim considerably more precise. [05]

Futurum's enterprise application forecast offers a rare chance to compare two categories on the same basis. Its 2025 study put the SaaS delivery of those applications at $323.3 billion in 2026 and on-premises delivery at $91.8 billion. A database engine can sit underneath either application. Forecasting the engine market independently tells us something about the supplier's opportunity, yet putting that total beside the application totals on a single pie chart would give the reader a false sense of where the slices meet. Our bank avoids the problem by giving each contract line a single purchasing owner. [26]

Consulting presents a different problem. Even a single market researcher can change its account of the year. Source Global Research's March 2026 estimate put worldwide consulting around $315 billion. In August the firm expected the market to pass $350 billion by year-end. Both were forecasts. The two publications give a reader too little information to reconcile that difference to contracts, currencies or a change in coverage. Choosing the later estimate gives us a dated point of reference. Treating the gap as $35 billion of newly signed business would turn a change in the researcher's account into money in somebody's till. [27] [50]

The vendor disclosures come in different currencies and periods. Accenture reported $9.28 billion in quarterly consulting revenue, with managed services reported separately. SAP's software licence and support line was €2.581 billion for its September quarter. Salesforce's subscription and support line reached $10.8 billion in its July quarter, including Informatica. They imply annual paces of roughly $37.12 billion, €10.324 billion and $43.2 billion if the respective quarters repeat. A firm moving an installed customer onto a subscription could move sales out of one category and into another while keeping the same customer. Getting those distinctions right is part of finding the business at risk. [09] [30] [29]

The data market resists a neat boundary. Gartner's December 2025 forecast put 2026 database management system spending at $161 billion. Some modern data platforms sell services beyond that definition, and a substantial part of the database business already lives in the cloud. Snowflake reported just under $1.492 billion of product revenue in the three months to July. Four times that figure is about $5.97 billion. In August Databricks said it had crossed a $7 billion revenue run rate. These are different disclosures. The procurement question survives the imperfect market comparison. When a new agent answers a hundred questions, whose meters turn? [28] [51] [32]

The buyer gets the first bid

I would bring the bank's head of procurement to this meeting an hour before the AI seller. Give her the authority to negotiate the five current services without changing the way a single case gets handled. She obtains prices of $268,000 for cloud, $246,000 for subscriptions, $200,000 for consulting, $150,000 for the installed estate and $88,000 for the data platform. The business can therefore keep $48,000 on its own. That saving sets the price of entry for any bidder who wants credit for changing how work gets done. A board that compares every new proposal with an unchallenged million will buy a very expensive opinion of its own innovation.

An ordinary renewal deserves as much scrutiny as an AI pilot. Suppose a cloud engineer turns off a service over a quiet fortnight and the bill falls. Was it idle or was it holding the replica that allows the bank to survive a regional failure? A software buyer sees an inactive login. Did its owner leave the firm, or is she one of two people authorised to release an emergency correction? A consulting

team spends a day documenting an old exception. Was it reconstructing lost work or creating the only evidence that protects an individual customer from being charged twice? A good procurement office puts names against these questions before the owner of a budget takes credit for the saving.

There is real evidence that ordinary intervention matters. Flexera's account of renegotiating its own mixed software agreement identified $1 million of hard SaaS savings and separated that result from both installed-licence savings and cost avoidance. Engineers at Zendesk describe setting up parallel tests on three large data models to examine actual warehouse consumption. Neither buyer needed an AI vendor to discover that a commercial or technical configuration could be improved. Their methods differ, and neither case supplies a representative percentage of waste. They show why the baseline for a proposition about displacement deserves a competent buyer throughout the exercise. [38] [52]

EXHIBIT 26 / MATCHED BUYER WALLET

Five bills before and after renewal



Five variable-width columns allocate a $1m prior budget among incumbents, AI fees, transition and cash kept by the bank. The winning mix pays AI suppliers $79,000.

Illustrative matched wallet in US dollars. Tile area represents dollars from the original million. Quoted extra expenditure is printed separately. “Winning mix” minimises the same-service bill across all five lines under the selected discount terms. Ordinary renewals include procurement's own savings. Industry-wide estimates have overlapping boundaries.

Open exhibit 26 in the interactive edition

The five columns describe a million dollars that the bank really could have paid under our assumptions. The width of the cloud column is twenty-eight cents in each original dollar. Consulting takes twenty-two. Height shows what happens inside each contract. A large column made almost entirely of retained supplier charges is a larger market and a poorer immediate

opportunity than a smaller column with a generous, contestable service fee. Select consulting and the distinction becomes visible. A $220,000 line sustains a $54,000 AI bid. The bigger cloud line sustains only $10,000. The share of spending that can move depends on what the bank is prepared to stop buying.

Five decisions at renewal

Take cloud first. A lab offers scheduling improvements and accepts responsibility for the timing of a defined batch of bank work. The bank expects its direct infrastructure charge to fall to $250,000. Add the lab's $10,000 fee and $4,000 to qualify the new schedule. The buyer pays $264,000, only $4,000 below the price its own team obtained. This contract is worth bidding for if the supplier can serve it at that fee. The former cloud supplier may receive part of the $10,000 back as an upstream payment. A buyer interested in controlling its bill needs the $264,000. An investor interested in the cloud provider's growth needs both the direct loss and that second receipt.

Seats come next. The bank can purchase its existing setup for $246,000 once obvious excess rights have gone. The alternative keeps $220,000 of software and metered actions, buys a $15,000 agent service and incurs $5,000 to change access and train the staff. At $240,000, the buyer is another $6,000 ahead. Its scarce asset may be the contractual right to use a different interface. If the software agreement requires a licensed identity for every person who benefits from the work, the bank needs written confirmation of the proposed terms. In the example the parties have agreed them. A photograph of an agent producing the right answer would be a curious substitute for that letter.

Consulting accounts for the largest transfer. The incumbent offers $200,000 for the same annual programme after taking duplicate evidence collection out of its ordinary bid. Using the specialist leaves $120,000 with the existing consultancy, pays the new supplier $54,000 and costs $12,000 to qualify the work, for a total of $186,000. Of the $34,000 saved against the original contract, $14,000 requires the specialist's offer. The fee the entrant earns is considerably larger than the saving it creates for the buyer. It can be paid partly for replacing valuable work that someone else used to perform. Counting only waste would miss that business. Describing the whole fee as newly created economic value would exaggerate it.

The installed estate gives the salesperson a harder afternoon. Most of the $150,000 is hard to move because the bank still needs the system and the people who can restore it. On the proposed route those retained services cost $140,000, the lab asks $6,000 and the first-year qualification costs $5,000. Somebody is going to have to explain to the chief financial officer why spending $151,000 counts as a saving. A later year may justify the transition if the qualification does not recur and both suppliers hold their prices. The one-year buyer in this exercise keeps the present arrangement. A multiyear buyer could quite rationally make a different decision.

The data provider has an even better counteroffer. Its current $90,000 line comes down to $88,000 after ordinary changes. Paying $84,000 for the retained platform, $7,000 for an outside answer service and $2,000 for transition would bring the cost to $93,000. There are attractive reasons to buy better answers, especially if they allow more employees to investigate more problems. In our tender everyone promised the same accepted work. A broader service deserves a separate purchase request from someone prepared to pay for the extra output. It cannot make an expensive bid for this service cheaper. The agent wins three lines and loses two.

The agreed payments now add to $928,000. Existing suppliers receive $828,000. The AI supplier invoices $79,000 and the bank pays another $21,000 for transition. The bank keeps $72,000 against the old bill, of which $24,000 is incremental to the best ordinary renewals. One useful check is to award all five AI proposals in the exhibit. The sales organisation can now point to $92,000 of fees it proposed. The buyer would spend $934,000, or $6,000 more than when it rejected the last two offers. An addressable pipeline can grow through bids whose acceptance would make the customer worse off. A sensible revenue forecast has to model rejection.

A clause that can move two columns

One assumption in that drawing deserves a lawyer's pencil. The five lines can be awarded independently. Change a single clause. Suppose the software distributor withdraws a $9,000 discount on the remaining subscriptions if the bank gives the cloud scheduling work to the lab. That condition makes the three-winning-bid combination cost $937,000. The bank has a better response. It can accept the ordinary cloud renewal at $268,000, keep its software discount and award the other two winning bids as before. It now spends $932,000. Its AI supplier books $69,000. A discount written on one invoice has taken $10,000 of prospective revenue off another firm's order.

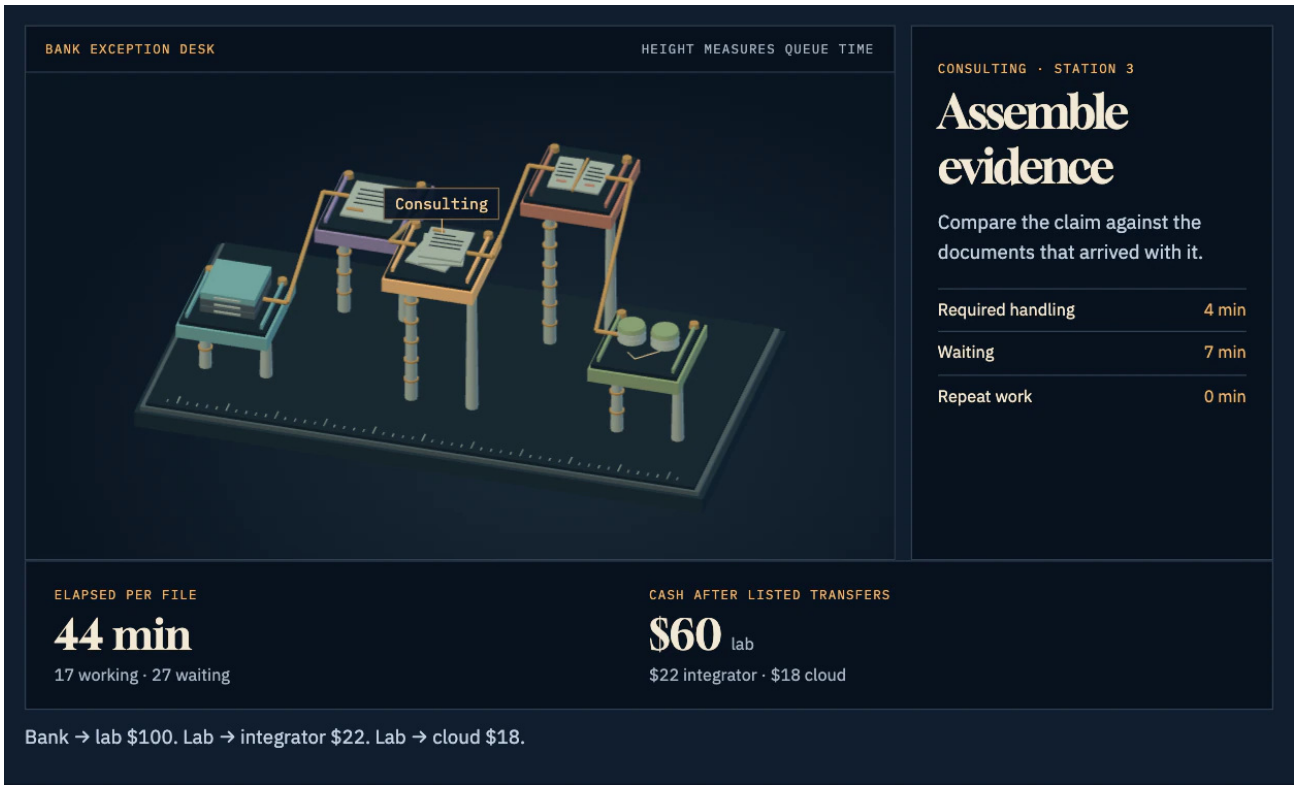
The point at which the choice changes is unusually easy to miss in a large negotiation. The lab's cloud route was $4,000 cheaper before the condition appeared. A lost discount above $4,000 consumes that advantage. The lab could cut its fee to try to get back in. Its ability to do so depends on a cost of service invisible to the bank's spreadsheet. A sales director looking only at her cloud opportunity might sign a concession that loses money. Her opposite number at the distributor may happily give up some software margin to retain a much larger cloud commitment. There is a reason procurement departments ask for the complete set of private offers on the same day.

One file, five clocks

Time gives us another way to see what the six businesses are selling. A customer has challenged a transaction. The bank must retrieve the original evidence, open a case, assemble the material, authorise the correction and verify the resulting balance. Five systems or service providers participate. Some tasks take minutes with a person or a computer working on the file. Others take minutes because the file is sitting in somebody's queue. If you buy a faster engine for a desk that spends most of the morning waiting for authority to proceed, the customer may have difficulty noticing the purchase.

EXHIBIT 27 / CASE AND SETTLEMENT MODEL

Five contracts for one banking operation



Five stations elevate the longer queues. The selected banking path spends 17 minutes working and 27 waiting. A separate $100 service ticket pays the lab, integrator and cloud.

Synthetic serial path, minutes per accepted case. Each brass collar on a station's support marks one minute of waiting. When repeat work is included, each amber bead on the front of a tray marks another minute of handling. Each station preserves its required checks and waiting time. The separate $100 ticket illustrates settlement under two prime contracts. Upstream payments are funded from that ticket. This per-file exercise uses a separate scale from the annual wallet in Figure 26.

Open exhibit 27 in the interactive edition

In the illustrative serial trace there are seventeen minutes of required handling, twenty-seven of waiting and ten of repeated work. The customer experiences fifty-four minutes. Removing all ten repeated minutes gets the result to the customer in forty-four. The single largest opportunity left is waiting. A revised shift pattern or a clearer delegation may release more of that time than another model upgrade. A buyer could pay for a different service level, or retain the twenty-seven minutes as an acceptable cost of keeping two authorised people involved. A shorter clock on a dashboard earns its way into the economic argument when someone knows which minute changed and what else depended on it.

The prime contract changes the cash trace without moving a minute on the clock. In a separate hundred-dollar example the lab collects the payment, sends $22 to an integrator and $18 to the cloud. It keeps $60 before other costs. Give the prime to the integrator and it pays $32 to the lab, which in turn pays $12 for cloud. The integrator now retains $68, the lab $20 and the cloud $12. In both arrangements the bank spends $100. None of the three balances includes every expense of delivering the service. A change in who signs has moved more cash between suppliers than the removal of a single queue, because in this example nobody renegotiated the service price when the queue moved.

That is a practical reason executives should care about the contract even when the technology works superbly. The party that can demonstrate how long a file sat and who had the authority to

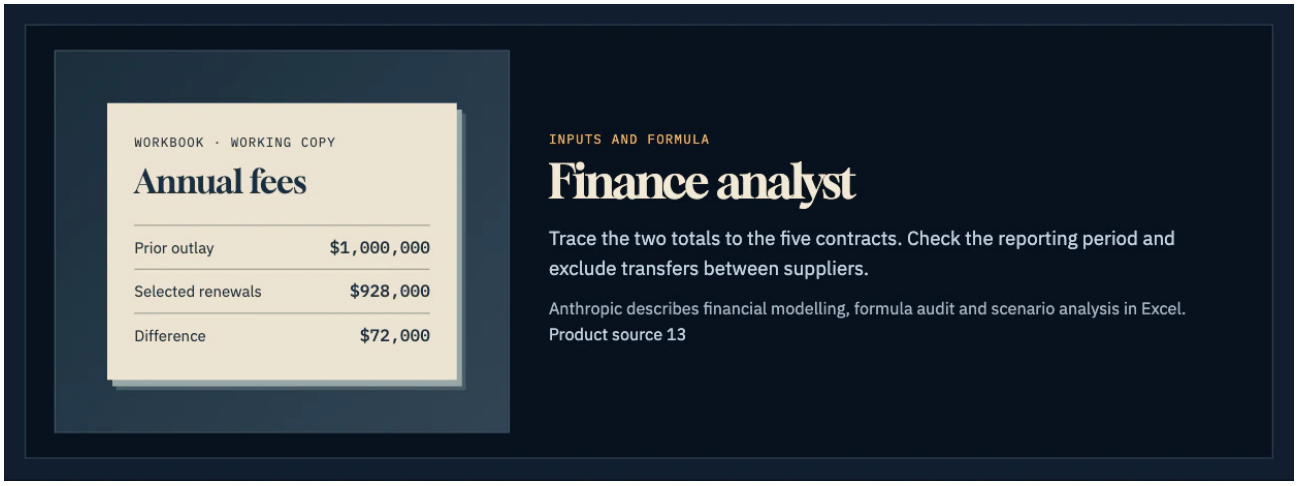
release it can defend its price against the next bidder. The bank can preserve that evidence itself. It can ask for time stamps in a form a replacement supplier can read and a documented method for calculating reopened cases. If it hands the record exclusively to its prime, it may be buying better service today at the cost of a weaker tender next year.

Where the work surfaces

It would be easy for a reader to imagine that all these decisions travel through a grand new software platform. Consider a more familiar ending to the exercise. An analyst puts the five awards in a workbook. Her manager uses a paragraph from it in a spending memo. Counsel corrects a term that lets a supplier count a reopened file as finished. The committee sees a one-page illustration of the saving. The number passes through four documents and changes its meaning in every one. The workbook can show $72,000 below last year while the committee needs to know that only $24,000 is attributable to the new route. Somewhere between the two documents a human being has to notice the difference.

EXHIBIT 28 / WORKING DOCUMENTS

The desktop handoff



A workbook specimen reconciles a $1m prior bill with $928,000 of selected renewals. Its review note asks the finance analyst to verify the five underlying contracts.

The four specimens are an editorial reconstruction using the illustrative wallet in Figure 26. The cited product materials document individual capabilities. The handoffs, approval rules and bank documents are illustrative.

Open exhibit 28 in the interactive edition

There is commercial substance to this apparently prosaic sequence. Anthropic has described financial workflows inside Microsoft's Excel, Word, PowerPoint and Outlook. Its legal plugin repository includes a vendor agreement review against a firm's playbook. Claude Design can take document inputs and export a presentation or PDF for review. Each capability provides a place from which to contest a piece of paid work while the existing desktop applications keep running. The depicted route through this particular bank is mine. The vendors document the individual tools. An uninterrupted, approved journey across all four desks would require a customer willing to demonstrate it. [13] [14] [15]

The incumbent has several ways to benefit. A data platform can charge for the queries behind the workbook, preserve access controls and offer a native agent that costs less than a separate service. A software vendor can keep a smaller seat order while selling a larger volume of paid activity. A consultant can sell independent testing to a bank that quite reasonably declines to let an AI

provider mark its own work. A mainframe operator can expose an approved transaction through a supported interface and sell more capacity when more good cases reach it. Each response puts a price on something the bank still needs. Disruption can increase one supplier's unit volumes at precisely the moment another supplier wins the customer relationship. [17] [19] [57]

There is also a way for the buyer to lose. Picture a contract in which nearly every performance credit depends on whether a file was eligible for automation. The vendor can make its resolution rate look excellent by removing difficult cases from the eligible population. Staff now receive a smaller, harder queue while an old per-case price falls apart. The budget holder discovers the error only after the department pays for overtime or commissions a new support agreement. A fair comparison keeps the entire original cohort under observation and assigns the cost of each exception to whoever actually paid it. The expensive case deserves the same place in the ledger as the easy sale.

I expect AI computing companies to win operating budgets that once seemed inaccessible to a software supplier. This example puts a size and a condition on the claim. A single bidder can take $79,000 from five categories of spending at one bank even though two of its offers fail. Its advantage survives a competent purchasing department by $24,000. Its own profitability depends on costs the bank never sees. To take that prediction out of the model and into the market I would want the winning orders, the losing offers and a year's worth of case outcomes from the same buyer. At that point we could give the finance director a number to sign, the customer a service to judge and the investor a business worth pricing.

CHAPTER 12

The wage bill beyond the CIO

A shorter shift, a cancelled hire and an afternoon spent on older cases leave different marks on the accounts. Follow the hours to payday.

A board can approve a software subscription in an afternoon. It can take a year to discover whose Saturday paid for it. NatWest Group reported that AI cut the handling time for complaints by about nineteen minutes per case, equivalent to more than 22,000 hours during 2025. A bank put its name to that figure in its annual report. An investor trying to price the opportunity for an AI supplier still needs the next page of the calculation. How many hours came off paid shifts? How many went into resolving other complaints? Who received the money the change made available? [72]

Those are different questions for the bank and its employees. A customer whose disputed payment is settled on Thursday instead of the following Tuesday has gained something tangible. An employee who depended on overtime to cover the rent might lose income even while keeping the same job title and base pay. A junior employee released from gathering documents might finally get to hear the difficult calls. Another could spend the year checking generated summaries and never hear what a customer sounds like when the summary is wrong. Our accounts have a line for the software fee. They have little to say about either career.

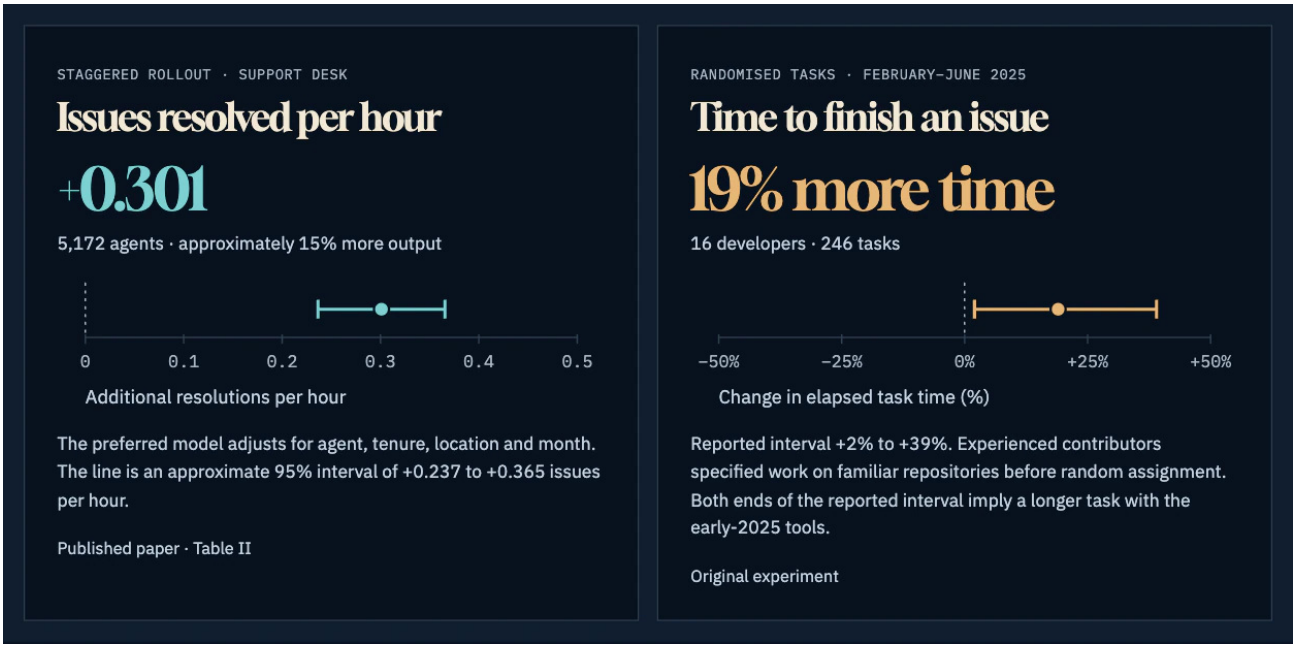
What the clock measured

The cleanest productivity figure in this debate comes from a place where there was already a good reason to count the work. A software company wanted to know whether customers' problems had been resolved. It could observe how many conversations an agent handled, whether the issue was settled and how much time the agent spent at the desk. A worker could conduct several chats at once. A shorter conversation and a larger number of successful resolutions in an hour therefore measured different achievements.

Erik Brynjolfsson, Danielle Li and Lindsey Raymond studied the staggered arrival of a conversational assistant across 5,172 support agents. The main rollout occurred in autumn 2020 and winter 2021. The system drew on past conversations, giving extra weight to chats involving strong performers. Agents could use or discard a suggestion. In the authors' preferred specification the tool increased output by 0.301 resolved issues per hour, roughly fifteen per cent of the pretreatment mean. Newer and previously lower-performing agents benefited most. Among highly skilled agents the small improvement in speed came with some deterioration in quality. [20]

EXHIBIT 29 / STUDY EVIDENCE

Two studies that cannot be averaged



Separate interval plots show 0.301 more issues resolved per support hour in one study and 19% longer issue duration for developers in an early 2025 experiment.

Points show study estimates, whiskers their uncertainty. Left, additional successful resolutions per support hour. Right, percentage change in issue duration when developers had access to AI. A positive value on the right means the work took longer. Each estimate belongs to the people, tasks and tools in its study. Support study . Developer experiment . Subsequent design review

Open exhibit 29 in the interactive edition

That last detail would matter to the people negotiating a bank contract. The senior employee whose calls taught the assistant how to respond might have little use for its script. A supervisor who rewards adherence could make her worse at precisely the difficult conversations through which her experience earns its return. The support study also documents what happened during outages, when suggestions temporarily stopped. The patterns in chat duration suggest some learning persisted. Outages were rare, and the measure available at that level was duration. For a bank interested in training, the relevant follow-up would include whether a colleague could reach the right decision when the assistant went dark.

The firm’s own agents received hourly pay and performance bonuses. Managers told the researchers that performance bonuses could account for twenty to forty per cent of typical take-home pay. The researchers lacked individual compensation records. A higher resolution rate could move an agent across a bonus threshold. Redrawing that threshold could return the gain to the employer. Two workers could gain equally on the operational measure and receive different amounts at payday. The study identifies where to look for that division. Resolving it requires the bonus schedules and the payments. [20]

Look at the other clock. METR recruited sixteen experienced open-source developers to work on their own projects, where they had about five years of experience on average. They specified issues before each was randomly assigned to a condition permitting or excluding AI tools. Across 246 tasks completed with the early 2025 toolset, access increased completion time by nineteen per cent. Before starting, developers expected a twenty-four per cent reduction. Afterwards, they believed they had been about twenty per cent faster. A buyer who interviewed the people doing the work would have come away with the opposite sign to the estimate from the experiment. [21]

By the follow-up, the experiment itself had become harder to run. Ten original participants returned and 47 new developers joined. The raw time estimates pointed towards faster completion, with intervals spanning either side of zero. Some developers declined to enter tasks they especially wanted to do with AI. Others were reluctant to take part under terms that could deny them a tool they used daily. The hourly study payment had also fallen from $150 to $50. Concurrent agents complicated timekeeping because a person waiting for one job could already be working on another. The researchers called for a change in design. [22]

Put the two clocks beside each other and the tempting national forecast dissolves. One counts accepted customer outcomes per hour after access arrived at different times. The other estimates time required for a prespecified programming task under random assignment. Even the arithmetic runs in opposite directions. At constant demand and quality, a fifteen per cent rise in output per hour would allow 100 hours of identical work to be done in about 87 hours. It would take a further decision to cut thirteen paid hours from the rota. Nothing in either study establishes who could sign for that reduction at our bank.

The entries on the payslip

Anders Humlum and Emilie Vestergaard could follow workers across the boundary that frustrates an enterprise pilot. Working with Statistics Denmark, they linked surveys of chatbot adoption to administrative records of earnings, hours and jobs. Their late 2024 survey included about 25,000 people across 7,000 workplaces in eleven exposed occupations. In the March 2026 revision of their working paper, differences between adopters and comparison workers before and after ChatGPT's arrival yielded estimates near zero for earnings and recorded hours. The authors could rule out average effects larger than two per cent in their observation window, which ended in December 2024. [73]

The same investigation found employees writing material, checking machine output and helping integrate the tools into existing processes. Some of these duties even reached people who never used the systems. Most chatbot users reported assigning time saved to other tasks. The source for that last finding was a survey. The source for earnings was an administrative record. Taken together they explain how a firm can change the contents of a job quite substantially before it changes what it pays for the job. A study with an early closing date also leaves room for a later round of hiring decisions. [73]

An American payroll series raises a sharper question about entry. In their August 2026 working paper, Brynjolfsson, Bharat Chandar and Ruyu Chen compared young workers in occupations with different degrees of AI exposure using ADP data through June 2026. Employment among 22-to-25-year-olds in exposed work stood nineteen per cent below the level implied by keeping pace with their less-exposed peers. Reduced recruitment accounted for most of the divergence. An employer could contribute to that gap year after year without announcing a single redundancy. A graduating class would experience the change one unanswered application at a time. [74]

The authors describe their result as a pattern requiring explanation. The estimates weaken with controls for education, some differences predate widespread generative AI use and the divergence is more pronounced in their payroll sample than in national surveys. Occupational exposure also leaves open whether a specific employer adopted any tool. A separate New York Fed analysis of US vacancies found relative weakness in postings for exposed occupations already visible before ChatGPT, with little evidence of a new divergence between junior and senior postings. A vacancy, a

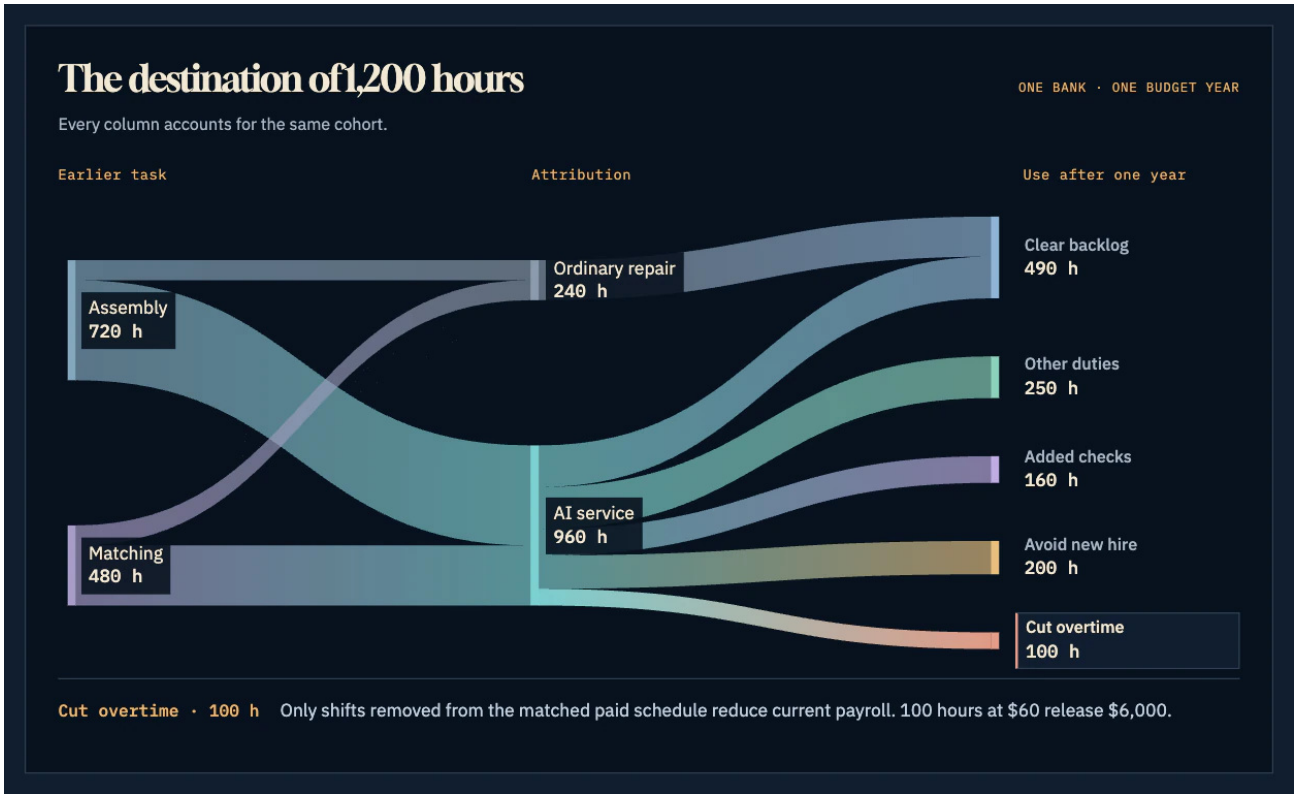
hire and an employee on payroll are three successive events. The studies observe different ones. [74] [75]

Three minutes in the bank

There is a test at the boundary between departments. A call centre can remove a minute from its record by sending a doubtful payment directly to investigations. If the investigator spends four minutes finding material omitted from the referral, the apparent improvement at the first desk has increased total labour. Match referrals to downstream records using the same case identifier, including files that cross a month end. Draw the comparison around the customer's completed case. Departmental averages will otherwise reward the team that moves work across the reporting boundary.

EXHIBIT 30 / ANNUAL HOURS MODEL

Where recovered hours go



An alluvial diagram follows 1,200 released hours through ordinary repair or AI and into five uses. The 100 cancelled overtime hours release $6,000 in current cash.

An illustrative annual cohort traced from prior tasks through the reason each hour was released to its eventual use. Ribbon thickness measures hours. The right-hand routes describe an assumed allocation within each group. The fee and labour rates are scenario inputs. The two buyer comparisons use different budgets and each counts the service fee once.

Open exhibit 30 in the interactive edition

Give the bank a cohort small enough to audit. For a full budget year assume 24,000 payment files of comparable difficulty. The old process takes an extra three minutes per file to assemble evidence and match records. On an accepted file the evidence must still resolve to the original entries, with the same controls and the same thirty days in which a correction can reopen. Count work against a mature cohort, including returns. The three minutes yield 1,200 annual hours. They measure two steps in these files. The bank has many other reasons to employ the people who perform them.

Divide the old work into 720 hours spent assembling the evidence and 480 spent matching it. Put a competent manager in charge of the old system before offering the bank anything new. Assume a common case identifier can save 120 hours in each step. That releases 240 hours for the existing backlog with the incumbent process. The proposed service can take over the remaining work, freeing a further 960 hours before checking. Budget 160 of them for examining exceptions and maintaining the comparison files. The bank now has 800 hours whose destination can reasonably be attributed to its AI purchase.

Put the shifts that affect the cash budget first. Suppose one hundred hours of overtime would otherwise have been paid this year and the new rota cancels them. At an assumed marginal $60 per hour, the bank's wage bill falls by $6,000. The same transaction can take $6,000 of gross pay from people who previously worked those hours, if the whole amount was wages. Where the employer's rate includes payroll contributions, the workers' own reduction will be smaller. An annual salary divided by a standard number of hours gives an accounting rate. It rarely tells you what the next cancelled shift actually saves.

The next 200 hours belong to a temporary assignment the bank had budgeted at $40 an hour. Existing employees can cover it in their ordinary shifts. The bank can withdraw an approved requisition worth $8,000 provided the work and the budget really existed. No employee on last year's payroll has had $8,000 taken from their pay for that reason. Somebody who might have been hired loses a possible piece of work. The evidence a reporter would want here is a dated requisition for the right skills and season, followed by its withdrawal. A manager's recollection of intending to hire carries less weight.

Assign half the remaining 500 hours to other duties and half to the waiting queue. The queue now receives 490 hours in all, 240 from the simple repair and another 250 from AI. Customers might value those hours more than the entire contract price. To turn them into cash the bank would need a specific consequence, such as a service credit it would otherwise have paid. Value them at zero in the cash calculation while recording how long the oldest complaint has been open. A service team deserves a measure of what it delivers even when the customer never receives a separate bill for patience.

The supplier wants $7,000 for the year, including setup and support. Compared with the current paid schedule after ordinary repair, the bank spends $1,000 more once the fee is included. Against a funded hiring plan it spends $7,000 less. Lose the requisition and the second case for buying disappears. Increase checking time until it consumes all 960 incremental hours and the bank owes the fee while its ordinary backlog improvement survives. This is a plausible route from a wages budget to AI revenue. The transfer depends on an employment decision that a technology director acting alone may have no authority to make.

The person learning to sign

Removing preparation also removes a place where a junior employee used to encounter errors at relatively low cost. Some mistakes belong in training. A new payment officer can learn the difference between a genuine reversal and a second debit by searching for the initiating instruction, comparing the dates and explaining the discrepancy to a colleague. If every clean case is prepared automatically and every hard case goes straight to a senior colleague, where does the next senior colleague acquire that experience? The buyer drawing up the service specification gets to decide whether any of the recovered hours answer that question.

Anthropic ran an experiment with 52 mostly junior software engineers learning an unfamiliar Python library. Participants could use it to complete two features, then took a quiz on what they had just encountered. Those assigned AI assistance scored fifty per cent on average, compared with sixty-seven per cent for participants who coded without it. The difference in completion time was statistically insignificant. An immediate quiz on a single library has a short reach. It does, however, separate today's submission from the participant's ability to diagnose the next failure. A bank could design the same distinction into an assessment with anonymised historical files. [76]

Training has an expense on the other side of the desk. Natalia Emanuel, Emma Harrington and Amanda Pallais examined engineers at one large firm around office closures and later returns to the office. Sitting near teammates increased coding feedback by 18.3 per cent. Less experienced colleagues gained most, while senior engineers wrote less code when near teammates. This was a study of proximity. Its relevance to an AI rollout is the accounting problem it exposes. A manager measuring the senior employee only by personal output could classify time spent developing somebody else's judgment as a loss. [77]

Who gets the recovered afternoon

An organisation can also fill a newly available hour several times in expectation. Berkeley researcher Xingqi Maggie Ye described eight months of observation and interviews at a US technology company with about 200 employees. Workers used generative tools to attempt a broader range of tasks, and work could extend into more of the day. The study traces a mechanism within one company. It offers no population estimate of unpaid hours. For our bank it suggests a practical check. Ask the person who signed the file when they reviewed it, especially if the system records the draft as complete at the instant the assistant produced it. [78]

Consider the worker on the closing shift. A queue can fall during the day while complicated returns accumulate as branches shut. Averages over all employees will look better even if one team loses control of its evenings. Compare hours by shift, seniority and case severity. Check reopening and customer complaints after the same interval. Record cancelled overtime against payslips and withdrawn hiring against approved budgets. Ask workers taking on the new checking whether the preparation guide gives them enough authority to send a doubtful file back. Somebody whose only option is to approve can generate an astonishing acceptance rate.

At renewal I would put the person who owns the rota next to the person who owns the software contract. Give them the gross minutes, the matched queue, the payment records and a list of responsibilities that changed hands. Ask which part of the supplier's next invoice can be financed from a smaller cash bill, which part buys a better service and which part buys the ability to train people for tomorrow. The employee who gains an afternoon, the graduate who never receives an offer and the duty manager who takes the call on Saturday each know something the procurement system cannot tell us. The economics of AI computing depends on whether anyone asks them.

CHAPTER 13

Who keeps the saving?

Medicare recovery accounts and a payments company's annual report pose a question for the AI buyer. Once appeals, suppliers and internal work are paid, who keeps the money?

An auditor found an overpayment. A contractor collected money. An appeal gave some of it back. Each event happened to the same public buyer, and the first two would make excellent material for a sales presentation. The United States Centers for Medicare & Medicaid Services published all three in an account that a prospective buyer of AI services might profitably take into a negotiation. For fiscal 2022 its fee-for-service recovery auditors collected $464.74 million in overpayments. The amount returned to the Medicare trust funds was $220.74 million. What happened to the other $244 million is a small education in the ownership of a result. [79]

The program returned $41.94 million to providers that had been underpaid. Decisions overturned at the first level of appeal accounted for $115.70 million. The recovery auditors earned $72.94 million in contingency fees and the agency spent $13.42 million on administration. Subtract those four amounts from collections and the trust funds receive precisely the published balance. The contractor's revenue, the money properly due to providers and the public benefit all have names in this account. Applying the recovery figure to a technology supplier's sales forecast would give that supplier credit for money belonging to the Medicare trust funds, medical providers and program administrators. [79]

There is even a fourth number a seller could have advertised. The auditors identified approximately $583.06 million in overpayments that fiscal year. Dividing collections by that figure gives about 80 percent. CMS reports an 83 percent recovery rate. Match cases across fiscal years. Cash from earlier determinations can reach the accounts this year. CMS describes negotiated fees and requires the contractor to give its contingency fee back when an improper-payment determination is overturned at any level of appeal. The rule reaches into cash already paid. Think about the difference that makes to a bidder deciding whether to take on a doubtful case in the last week of the quarter. [80]

The same report describes a market that contractors declined. Recovery vendors found Medicare Part C audits commercially unattractive because payment arrangements differed and appeals could continue without a time limit. CMS used a different contracting arrangement to do the work. A laboratory selling accepted decisions faces the same question a bidder could ask here. How long after delivery can a challenge reach its balance sheet? Which result has the customer agreed to buy? Who pays to establish that it happened? For the prediction I am making about AI companies, the answers are worth more than a league table of model scores. [80]

What a buyer can honestly claim

Klarna gives us a company with an actual AI deployment and enough public accounting to complicate an easy story. Management said the assistant powered by OpenAI handled 80 percent of customer service chats in 2025 and estimated approximately $59 million of cost savings that year. Customers could still reach a human. In the same filing, customer service and operations expense rose from $203 million to $207 million. A reader can believe both disclosures. The business processed 25 percent more transactions and its volume grew 32 percent. Providing a larger service cheaply can increase the amount a company actually spends. [81]

The dates deserve a second look. The annual filing says repeat inquiries fell 25 percent between December 2023 and January 2024. It dates the launch to February 2024. Klarna’s May 2024 results announcement placed the launch in January. Neither document gives the deployment timetable by country and channel for the customers in that comparison or a repeat-inquiry rate for 2025. The expense line also covers work outside chat. A transaction can proceed without a conversation. To price the supplier’s share of the reported gain, an investor would need the contract fee, the accepted volume and the cost of serving those customers. Klarna’s estimate supplies none of those terms. [81] [88]

Consider what would persuade a bank's controller. She would want the number of cases received, the distribution of difficulty and the outcomes after customers had time to contest them. She would ask payroll which shifts were cancelled and procurement which contract stopped billing. Then she would ask the operating manager what happened to the freed hours. Perhaps an experienced clerk used them to investigate complaints that previously expired unanswered. That can be an excellent return on a purchase. On the day the clerk receives the same salary as before, the payroll saving is zero. The unanswered complaint and the salary both deserve a place in the decision.

Follow a case to its last owner

Take a fresh hypothetical purchase. A bank outsources evidence-file preparation for $100 per accepted file. The incumbent price includes the work needed to answer a challenge. Three bidders promise the same evidence, deadline, correction rights and access to a human. Each quotes $80 for a file finally accepted. A thousand files qualify each year for three years. The bank keeps its people and its ordinary overhead. After its own checks it regains twelve minutes of a clerk's time per file. At constant volume the bank removes $20,000 of external spending and gains two hundred staff hours each year. Its accounts show the first amount. Its service managers can put the hours to work.

EXHIBIT 31 / SETTLEMENT MODEL

The settlement of a hundred dollars



Three stacked account strips subdivide the same $100, first at the buyer, then at the prime and finally by use. The bank keeps $20 and the lab earns $38 of contribution.

Illustrative three-year service at constant annual volume and quality. Each transparent layer describes the same money at a different stage. The middle layer divides the prime's receipt after its immediate payments. The lowest layer separates direct costs and supplier contributions. An owner funding a loss is shown as a separate source. Amounts are rounded for display.

Open exhibit 31 in the interactive edition

Begin with the laboratory as principal supplier. Every file gets one automated run. One in four needs a second attempt. The laboratory buys computing capacity at $12 an attempt and reimburses a specialist for verifying each attempt at $8. Independently of that retry classification, one in ten files needs a manual finish costing $50. A human completing a retried case earns that fee for different work. The bank pays once for the accepted outcome. In a thousand cases there are 1,250 cloud charges, 1,250 verification charges, 100 manual finishes and 1,000 customer invoices if the files are billed separately. Aggregate billing changes the paperwork and leaves those quantities intact.

The specialist advances $18,000 to qualify the service, recoverable in equal annual instalments across the three-year term in our calculation. At a thousand files a year this absorbs $6 per accepted file. The specialist also earns $6 per file for coordinating delivery after recovery of its specified costs. Its average invoice to the lab is consequently $27. Checking takes $10, manual completion $5, qualification $6 and coordination accounts for the remaining $6. The lab pays $15 for computing power. Against its $80 sale it therefore keeps $38 toward model development, general overhead, financing, taxes and whatever profit survives those expenses. Calling the full $80 a margin would erase two suppliers.

The date on the cheque matters to the specialist. In the first year it collects $27,000, pays $10,000 for checking and $5,000 for manual work, and advances the entire $18,000. Its cash balance from this assignment is negative $6,000 before other expenses. In each of the next two years the same annual receipts and ordinary costs leave $12,000. Across the full term the specialist has earned $18,000 for coordination. If the buyer cancels after twelve months, the specialist may have financed a deployment whose remaining instalments never arrive. The party that advances the

money has a perfectly good reason to ask who pays the remaining qualification balance at termination.

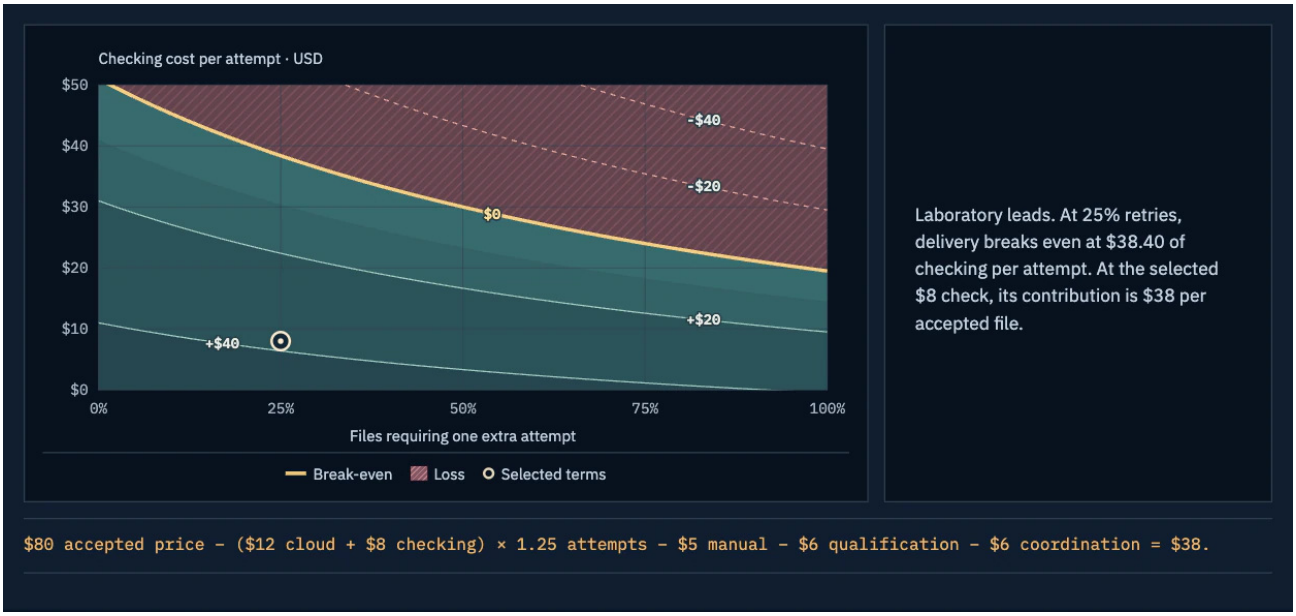
Working backward from the bank's original hundred gives a further check. Twenty stays with the buyer. Fifteen reaches the infrastructure company. Ten funds checking, five covers manual completion and six recovers the initial outlay. The specialist retains six and the lab retains thirty-eight. Those amounts sum to one hundred. An economist could follow the infrastructure company's own electricity, equipment and wages another level down. At this boundary the cloud purchase is a single outside cost. Adding its employees' wages to the hundred after including the cloud bill would count some of the same money twice. Each layer of the illustration repartitions the preceding one.

Bad runs have a creditor

An output contract changes the consequences of a difficult Tuesday. Under our terms the bank pays the same $80 for an accepted difficult case as for an ordinary one. The cloud supplier charges for both attempts and the checker must inspect both. The delivery specialist recovers those costs. The lab has effectively written the bank a small insurance policy on the difficulty of the incoming work. It can do this profitably when easy cases fund hard ones. To price the promise it needs the distribution of retries and the cost of finding out that a seemingly convincing first answer was wrong.

EXHIBIT 32 / COST SENSITIVITY

The price of a failed run



A contour plot shows how extra attempts and checking costs change the prime's contribution. With 25% retries, checking breaks even at $38.40 per attempt.

Each line holds the owner's contribution constant as checking cost and retry frequency change. Above the gold boundary the signed price fails to cover delivery. The specialist incurs the qualification, verification and manual costs. When subcontracted it passes those amounts on and receives $6 for coordination. Price, volume and manual fallback follow the settlement plate. The model assumes every file ultimately meets the same acceptance terms.

Open exhibit 32 in the interactive edition

At the original retry frequency the lab breaks even on delivery when checking reaches $38.40 an attempt. The arithmetic starts with the $80 sale, reserves $5 for fallback, $6 for qualification and

$6 for the specialist's coordination. That leaves $63 to fund 1.25 attempts. Each could cost $50.40 altogether before the lab's residual disappears. The machine already takes $12 of that. Now move to a portfolio where sixty percent of files need another attempt and each check costs $30. Attempts consume $67.20 per accepted file. Add the other $17 and the lab loses $4.20 on each delivery. The bank still receives its $20 discount. The signed price determines who finances that result.

Difficulty often arrives in batches. A single employer might submit a year's worth of irregular records just before a deadline. That produces a queue of precisely the files likely to need a second look. An annual average can price the labour while concealing a week in which no experienced checker is free. The bank's recourse might be to reopen the old service at an emergency rate. A credible bid states who staffs the peak, which party pays for an overflow desk and how a case moves there without losing the evidence already collected. Those are economic terms even when procurement puts them in a service appendix.

Cheap verification can disguise a different cost. Suppose the bidder proposes to sample fewer files. Its contribution rises immediately in the model if the buyer holds accepted quality constant. Whether that assumption survives a trial depends on errors among the cases nobody chose to examine. Have an independent reviewer draw a random sample from the apparently clean files as well as inspecting the obvious exceptions. Follow reopened cases to an agreed maturity date. Separate confirmed errors from changes in policy. Charge the review to the party that commissioned it. A price comparison across bids then has a common basis for calling a file complete. A fast first answer is worth exactly what that evidence says it is worth.

A workable contract gives each accepted file a reference that follows a subsequent complaint. It gives the buyer a period to challenge the file and specifies the remedy if the challenge succeeds. A replacement report might restore the promised service. It may also arrive after the bank has paid compensation to its customer. The contract must say who bears that second loss. Otherwise the cheap bidder can exclude a serious expense that another bidder has included in its fee. An early trial can test this clause without waiting for a catastrophe. Give every bidder the same previously overturned file and ask its commercial director to price the correction against the proposed terms.

The company whose name is on the contract

Now put the same work through three different owners. Keep every volume, price to the bank and quality requirement fixed. Where another party leads, the lab offers its runs wholesale at $24 an attempt and pays its own $12 cloud charge. At 1.25 attempts it receives $30 and retains $15 per accepted file. When the specialist leads, it collects $80 from the bank, pays $30 to the lab and covers $21 of checking, manual work and qualification. It keeps $29. Compare its $6 when it worked for the lab. The extra compensation pays for risks that arrive with its signature.

EXHIBIT 33 / THREE OPERATING CONTRACTS

Three owners of the workflow



Paired marks compare lab revenue and contribution across three possible prime contractors. The lab takes $80 and retains $38 when prime, compared with $30 and $15 as a wholesale supplier.

Square marks show what the laboratory earns for its specified service. Circles deduct its own delivery costs. The prime controls and supplies the completed file under the assumed contract. A specialist or incumbent prime buys attempts wholesale from the lab and bears the cost of additional runs. Contributions exclude shared overhead, financing, tax and model development. Principal and agent guidance . A distributor's reported figures

Open exhibit 33 in the interactive edition

Put the incumbent software company in charge. It receives $80, pays $30 for the lab's wholesale service and $27 for the specialist's work. It retains $23. Across the three routes the customer saves the same $20 and the three bidders together retain the same $44 before unallocated costs. Yet the laboratory reports $80 of revenue as prime and $30 as a wholesale supplier under the assumed contracts. Its contribution changes from $38 to $15. A price cut by the lab on wholesale attempts could make the incumbent's route more attractive while leaving the bank's bill untouched until the next renewal. A competitive response can transfer money between sellers before a customer sees a cent.

An accountant would first inspect the promise that each seller controls. The IFRS Interpretations Committee examined a software reseller that advised a customer and arranged delivery of standard licences. Its May 2022 decision asks what good or service the customer was promised and whether the reseller controlled it before transfer. Responsibility for fulfilment, inventory risk and discretion over price can help resolve that question. Their weight depends on the facts. Our prime assumes responsibility for the integrated file, contracts the inputs and delivers the accepted work. A broker who collects $80, owes $74 to the actual provider and earns a $6 arranging fee presents a different transaction. Putting the whole collection into that broker's sales would radically distort its apparent scale. [82]

Consider a listed distributor with real numbers. Climb Global Solutions reported 2025 gross billings of about $2.105 billion and net sales of $652.5 million. Gross billings, a company-defined operating measure, includes the gross value of certain transactions for which revenue is recorded net. Billings grew 18 percent and net sales grew 40 percent. The same company, during the same period, produced two defensible growth rates for different purposes. Neither difference measures an AI margin. A prediction that an AI computing company can outgrow a cloud company has to specify which reported measure crosses which line. The face value of every invoice passing through a distributor would set a wonderfully easy hurdle to clear. [83]

Read the second order

Double the bank's volume to two thousand files in each year of the term. The $6,000 annual qualification allocation now costs $3 per file. Under the lab-led price and cost reimbursement above, the specialist's invoice falls from $27 to $24 per case. The lab's contribution rises from $38 to $41 and its annual revenue reaches $160,000. If the bank would have bought all two thousand files at the old price, it saves $40,000. If it would only ever have bought the first thousand, the extra thousand purchase work that previously went undone. The old $100,000 budget has grown to $160,000. Better service may justify every dollar of that expansion.

Volume can fall too. At five hundred files a year qualification absorbs $12 a case and the lab's contribution drops to $32. A seller anticipating that exposure may offer the bank a lower unit price in exchange for a minimum annual purchase. The buyer accepting the minimum then acquires some of the demand risk that previously sat with the laboratory. Look for a line in the purchase order showing the payment due if the bank tenders fewer files than expected. In a quiet year, dividing the cash actually paid by the files actually accepted will reveal an effective price quite different from the prominent figure on the opening page.

At renewal I would ask for two lists. The first identifies the old contracts, shifts or charges that actually disappeared and the buyer who can confirm their removal. The second identifies work the customer can now afford, its recipients and what happened after delivery. A vendor with durable pricing power ought to be able to produce examples from both lists. Once a customer knows how much of the saving it can keep and how many additional cases that saving buys, the negotiation changes. The company that can quote the next accepted case, pay for a bad run and stand behind its answer has earned a credible claim on the next order.

CHAPTER 14

A company larger than the cloud

Start with four quarters of filed sales. Give AWS five more years to grow. What would Anthropic have to sell to catch it, and what would those sales have to earn?

Put a date on the wager. For the four quarters ending in December 2031, I expect an AI computing company to report more recognised revenue than at least one of the two cloud businesses whose accounts let readers check the result. Anthropic is my leading candidate. Google Cloud is the nearer target today. Overtaking Amazon Web Services would settle the stronger claim. A company that reaches either milestone by taking responsibility for recurring enterprise work would change what a technology supplier can sell. A company that reaches it by reselling ever larger amounts of someone else's compute would have accomplished something financially and industrially different.

A price on a trading screen can settle neither question. An increase in the share price changes the value ascribed to future earnings. An issue of new shares puts money in the company's account. Someone still has to agree to buy the service, the company has to deliver what it promised, and the resulting sales have to enter its accounts. For the calculation below I start the clock with the four quarters ended September 30, 2026. An exact five-year interval ends in September 2031. The December reporting period gives the prediction its final calendar-year check. I would use the reporting period, even when the accounts appear in the following February.

The starting distance

Anthropic reported $2.040 billion of revenue in the December 2025 quarter. March, June and September brought in $4.731 billion, $11.555 billion and $16.956 billion. Add the four entries and the starting line is $35.282 billion. Multiplying the last quarter by four would put the company at $67.824 billion. That shortcut would put almost twice the revenue Anthropic actually recognised into the twelve months under comparison. The distinction is particularly consequential for the fastest-growing company in a comparison. The businesses that have already achieved scale receive much less of a head start from the same shortcut. [01]

Amazon provides the check on our arithmetic. The quarterly AWS sales in its supplement add to the trailing figure the company itself prints, $161.382 billion. AWS operating income across those quarters was $61.020 billion. For Google Cloud, adding four published segment tables gives $93.169 billion in revenue and $31.846 billion in operating income. Alphabet's nine-month segment results plus the preceding December quarter produce the same two totals. Both incumbents had profitable businesses with which to finance a response. Their shareholders already own a claim on some of the growth in AI computing, however the contest for the prime contract turns out. [05 06 84 85 86 87]

Anthropic recorded a $4.096 billion operating loss over the same year. The June quarter contributed positive operating income of $974 million, between losses in March and September. Amazon's figure belongs to a segment inside a group. Alphabet allocates certain expenses to its

operating segments and leaves some shared and research costs at the parent. Anthropic's figure belongs to the whole company and includes expenses such as stock compensation. A comparison of reported operating results tells an investor what the accounts say. Estimating what either cloud would earn as a separate listed company would require its own set of accounts. [01 05 87]

Microsoft reported Azure revenue above $100 billion for the fiscal year ended June 30, 2026. That disclosure is relevant to the scale of the industry. It supplies neither an exact amount nor an Azure operating result for the twelve months to September. Microsoft Cloud includes other businesses. Substituting its revenue for Azure's would turn improved disclosure from one company into an arbitrary penalty on another. If Microsoft publishes a comparable series, there will be a third opponent to track. Until then the two named tests can be decided on their own terms. [07]

The opponent gets five years too

A stationary AWS would require Anthropic to compound revenue by about 35.5 per cent a year just to draw level after five years. A stationary Google Cloud requires about 21.4 per cent. Freeze the challenger's present scale in one part of a discussion, declare it capable of transforming every customer's business in another, and an important connection goes missing. Suppliers of storage, model hosting and distribution can grow precisely because the challenger succeeds. Both sides of the race deserve a growth assumption.

EXHIBIT 34 / REVENUE AND EARNINGS HURDLES

The hurdle to exceed AWS



Growth paths extend from September 2026 to 2031. At 15% annual AWS revenue growth, Anthropic needs 55.87% a year to draw level, passing $85.72bn on the parity path in 2028.

Quarter ends run from December 2025 through September 2026. Reference paths assume 0% and 30% annual incumbent growth. Margins belong to the operating-profit comparison. All future observations are scenarios. [Anthropic prospectus] . [AWS supplement] . [Alphabet segment filing]

Open exhibit 34 in the interactive edition

Give AWS 15 per cent annual growth as an exercise. Its trailing sales reach $324.597 billion in September 2031. Equality then requires Anthropic to grow by 55.868 per cent a year from the disclosed starting line, with a fraction more needed to take the lead. Apply the same incumbent growth assumption to Google Cloud and the target becomes $187.396 billion, requiring 39.650 per cent annual growth from Anthropic. The reader can replace either rate. These numbers describe the size of a bet somebody has chosen. They contain no estimate of the probability of winning it.

One way to see the burden is to work backwards from the order book. An annual increase of fifty or sixty per cent on a modest base can come from a handful of large customers. The same percentage near the end of the exercise requires tens of billions of dollars of additional sales in a year, on top of retaining the existing buyers. A bank that expands its deployment from one business unit to six can contribute to that increase once. In the following year, the supplier must renew all six and find the next set of operating budgets.

Consider a lab that compounds at 60 per cent. Its sales reach $369.959 billion. At a 20 per cent operating margin, that business earns $73.992 billion in the terminal year. Give both cloud segments their starting operating margins and the assumed 15 per cent sales growth. AWS generates about $122.733 billion in operating income and Google Cloud about $64.054 billion. The lab wins both revenue comparisons and one profit comparison. Relative to AWS at those sales, the lab would need to earn more than 33.17 cents of operating income per dollar of revenue to win the second contest.

There is no sensible positive compound growth rate for earnings that starts with a loss and ends with a profit. Instead, take the incumbent’s terminal operating income and divide it by the operating margin the lab could earn then. At a 20 per cent margin, matching AWS’s $122.733 billion requires around $613.665 billion in lab revenue. Its sales would have to grow by about 77.0 per cent a year to get there. This arithmetic gives due weight to the compute bill, the engineers who maintain the service and the promises made to buyers whose cases come back.

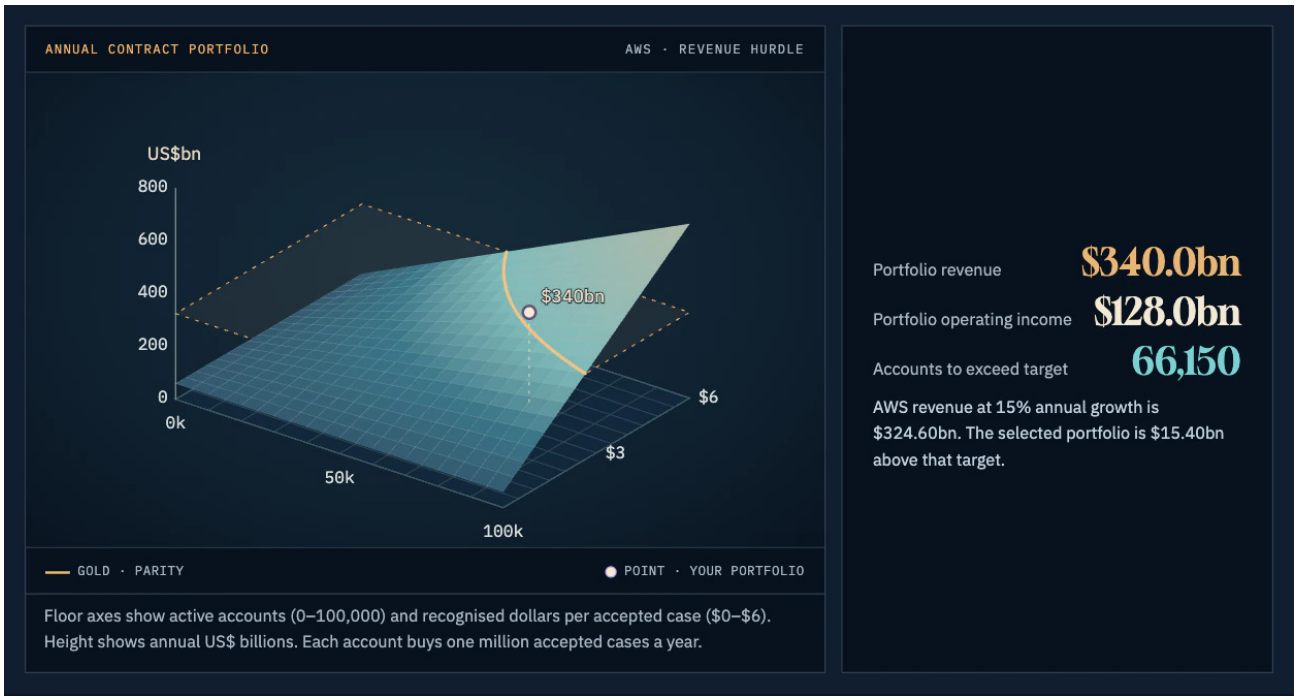
What an accepted case would have to earn

A procurement director can do more with a price and a volume than with a compound growth rate. Construct a separate possible business for 2031. Assign $60 billion of annual revenue to subscriptions, model access and work outside the outcome contracts in this exercise. Assume those lines leave $18 billion after direct delivery costs. Put $30 billion of annual operating expense beneath the combined businesses for research, selling, administration and other fixed expense, including relevant equity compensation. These round numbers are terms of the illustration. Every dollar of expense belongs to one category and the other lines exclude the cases we are about to price.

Give the company 70,000 accounts, each buying a million accepted cases a year at four dollars of recognised supplier revenue per case. These contracts add $280 billion to the $60 billion earned elsewhere. If delivering an accepted case costs two dollars, the case business contributes $140 billion before fixed expenses. Add the $18 billion from the other lines and deduct the $30 billion operating expense. The result is $128 billion of operating income on $340 billion in revenue. That clears both AWS hurdles in the selected scenario, with substantially less room to spare on profit.

EXHIBIT 35 / UNIT ECONOMICS MODEL

The growth surface



A gold contour crosses a contract-revenue surface at the projected AWS target. The selected hypothetical portfolio earns $340bn at 70,000 accounts and $4 per accepted case.

This is a portfolio assembled from assumptions. Its customer count is hypothetical. Revenue equals accepted volume times the recognised fee plus $60bn of other, mutually exclusive sales. Operating income starts from $18bn of other direct contribution, adds case revenue less direct costs and deducts $30bn of fixed operating expenses. The two figures share the selected incumbent's 2031 target. The portfolio replaces the lab's growth extrapolation with an independent calculation.

Open exhibit 35 in the interactive edition

Look closely at the unit of adoption. A million accepted cases per customer per year is about 2,740 a day, counted through weekends and holidays. That could describe a large payment operation with a tightly defined task. A legal team commissioning a handful of consequential opinions each month has a different price, volume, review cost and responsibility for error. A million inexpensive data lookups and a million signed opinions are equivalent in the model only if they command the same recognised fee and incur the same full delivery cost. The distribution across contracts is a question for reporting.

Move the delivery cost to three dollars per accepted case. Sales still come to $340 billion. Operating income falls to $58 billion. A second model call on a difficult file, a review by somebody with the appropriate authority, an allowance for remedying late errors or an expensive minimum compute purchase can all take part of that extra dollar. The supplier holding the case contract can give the bank a price cap only by carrying the costs that breach it. If a vendor reports margin on accepted cases while charging failed attempts to another business unit, follow the costs back into the consolidated accounts.

Fee compression has a different signature. A customer that obtains a reduction from four dollars to three cuts recognised revenue at the original volume by $70 billion and also reduces operating income by $70 billion before any compensating savings. The service may become popular enough at the lower price to win new customers. To recover the original $280 billion of case sales at three dollars, the company would need over 93 billion accepted cases a year. Recovering the original case

contribution at two dollars delivery cost takes 140 billion. Growing into a price reduction is possible. The required sales and delivery capacity have different magnitudes.

Who controls the dollar that enters the accounts

A four-dollar fee needs a definition. Anthropic's prospectus says it recognises the intelligence service gross in specified cloud marketplace arrangements. It establishes the price, contracts directly with the customer and takes responsibility for the service. Under those circumstances, money it spends on infrastructure enters the cost calculation. In an arrangement where a company only arranges somebody else's service, the amount it earns can instead be the commission. The model asks for revenue the seller can recognise per accepted case. It places the cost of providing its own promised service on a separate line. Changing the agreement can change the first number even when the bank writes the same cheque. [01]

Suppose a bank buys ten dollars of service and one dollar belongs to an agent for arranging the sale. Counting ten dollars of lab revenue in a commission business would inflate its scale tenfold. Conversely, deducting wholesale compute from a principal's recognised sales and also charging it as a delivery expense would make the seller pay twice on paper. In comparing public companies I would retain each reported figure and identify related purchases alongside it. In reporting total customer expenditure I would follow the original ten dollars once. Whether the AI company became bigger and whether the buyer spent more are separate empirical questions.

Acquisitions pose a related test. Imagine the lab buys a deployment firm billing the bank ten million dollars a year, six million of it for work that has never used the lab's models. After the purchase closes, acquired external sales can lift the lab's consolidated revenue. Payments between the two units disappear on consolidation. I would show the reported total, the acquisition's disclosed contribution and the date control changed hands. A takeover can put the outcome contract in the lab's name immediately. Evidence that the combined company can deliver it more efficiently comes later, when the bank receives a comparable offer and decides whether to renew.

What I will look for in 2028

At the September 2028 review the smooth path to sales parity with AWS under the 15 per cent assumption reaches $85.718 billion of annual lab revenue. The corresponding Google Cloud path reaches $68.808 billion. A slower opening can be recovered with faster subsequent growth. Record the actual trailing sales for each side, the remaining three years and the new rate required. A lab can fall behind the initial reference and still improve its prospects if it has accumulated customers with demonstrated expansions, lower delivery costs and purchase commitments that convert into recognised revenue on a credible timetable.

For example, suppose the 2028 accounts show the lab at seventy billion dollars and AWS at two hundred billion. Allow AWS another three years at fifteen per cent. Its target is now $304.175 billion. Starting from seventy billion, the lab needs about 63.2 per cent annual growth to reach it. This update asks how much more the lab has to sell from that date onward. The original calculation survives in the historical record, so readers can see which assumptions changed. A board investing against this route can also translate the new target back into accepted volume, prices and support obligations.

I would also follow a bank through its first renewal after twelve months of service. Ask for the accepted volumes and the fee actually realised after rebates. Ask who paid the warehouse bill, how many files reopened and whether the supplier can recover its integration cost at the renewal price.

A lab can license a large model to the winner of a contract and collect a useful wholesale fee. That is a less extensive claim on the customer than becoming responsible for the outcome. The supplier’s ability to retain that responsibility after a competing bid should carry weight in any revision to the forecast.

The deciding entry in 2031 will be a dated pair of trailing revenue figures. Report the operating result beside them and name the group that holds the prime contracts behind the growth, wherever the evidence permits. If neither revenue comparison clears, I will record the miss. If one clears and the larger seller still has poorer operating economics, the revenue prediction will have succeeded and the advantage claimed for its business model will remain in dispute. The bank that renews an accepted outcome at a price both parties can live with supplies a more durable reason for expecting the next billion.

CHAPTER 15

The case for the incumbent

A cloud commitment, an engineering team with its own machines and a provider willing to pay for repairs can each change who wins the work.

On 16 October 2026 Britain's competition authority described an extraordinary piece of competitive progress. Two of the world's richest companies had made it easier for their customers to leave. Amazon and Microsoft had updated the terms on which UK customers could transfer data out when switching providers. The window for a qualifying move had been extended from ninety days to a hundred and eighty. The authority had examined internal company data and spoken to customers. Those customers continued to describe material obstacles to switching or operating across providers. Moving the bytes was one part of moving the business. [90]

A six-month departure provision belongs in a discussion of frontier intelligence for a simple reason. A bank can evaluate a new model over a weekend and still need most of those six months to transfer the work. The people who keep the existing operation running can spend that interval improving it. They know where the files are, where the supervisors sit and what the bank has already agreed to buy. Any prediction in which AI computing companies take revenue from the cloud has to let the cloud companies use those advantages. An incumbent that politely holds still until its challenger finishes the installation is a peculiar invention of forecasters.

There is a practical lesson in what the British authority chose to investigate after prices changed. Amazon and Google had launched a direct connection between their systems, and Amazon had introduced a free tier. At the October reporting date, a direct link between Amazon and Microsoft remained in progress. The difference matters to a bidder who has thirty days to deliver a working service. A promised connection can be a sensible choice for the long-term design and an expensive one for the first release. Write two dates on the estimate, the day the bank pays the engineer and the day the connection is available for the bank's region. Until the dates meet, someone must fund a temporary route. [90]

The cheque already written

Microsoft's own billing documents give an unusually clear view of the defence. Claude usage through Foundry can appear on an Azure invoice in Claude Consumption Units. Under the documented arrangement the meter counts towards the customer's Microsoft Azure Consumption Commitment. Discounts enter the conversion from model tokens into the billing unit. A company can therefore decide it wants Anthropic's intelligence and find that the easiest place to buy it is an agreement the purchasing department signed long ago. Both companies can make a sale. Only the buyer knows whether the purchase consumed an otherwise stranded commitment or displaced some other eligible work. [35]

That last distinction can change the result of a tender. A prepaid unit that the bank would otherwise use for a scheduled migration has a real opportunity cost. Charge it to the new project and the migration team will eventually have to purchase another unit. A unit expiring unused next month is a different proposition. An account director talking about a customer's commitment might know the exact contractual amount while having little idea which of those two situations applies. The person who can settle it is often a finance manager with the next six months of project forecasts and the authority to challenge them.

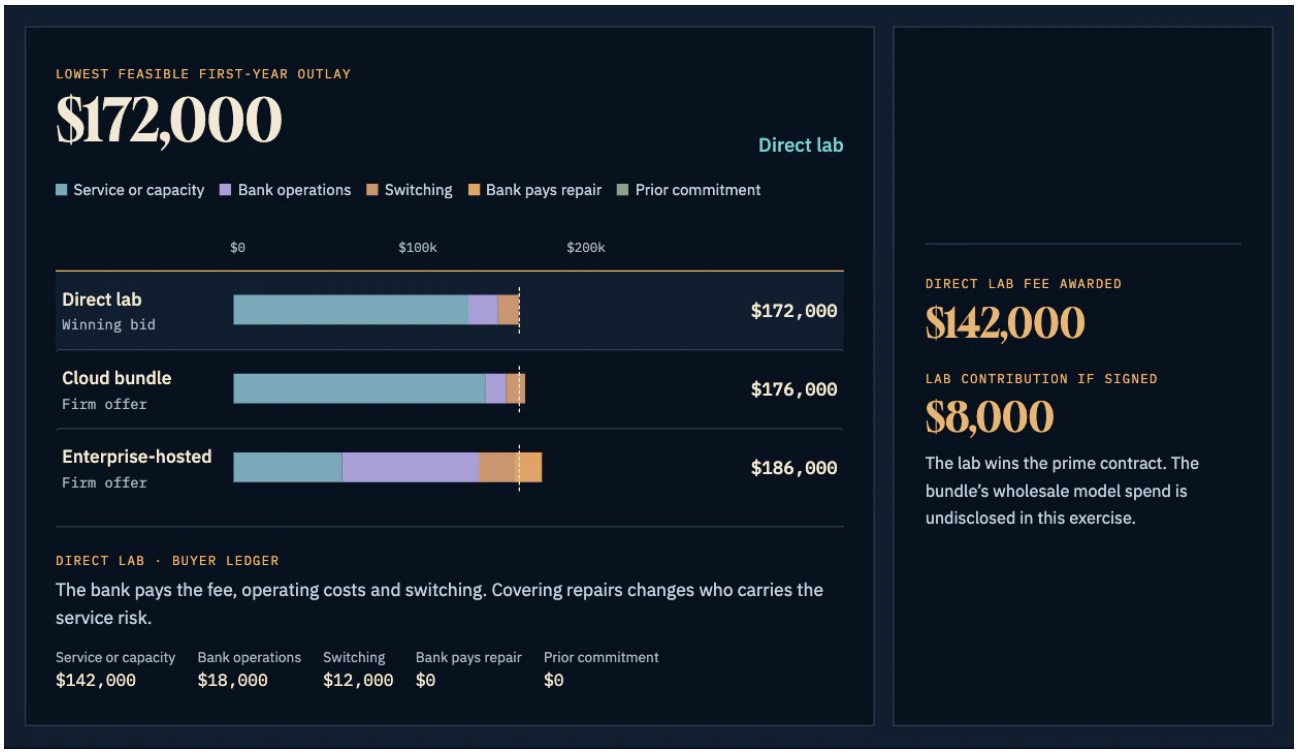
A platform has another response. It can make the choice of model a feature of a product the buyer already likes. In September GitHub described three settings for automatic model selection in Copilot. A user could favour efficiency, balance or intelligence. Charges depended on the model selected, with a ten per cent discount for paid subscribers on qualifying automatic usage at the time of the announcement. A developer selecting a preference delegates some purchasing discretion to the platform. The model developer can win the underlying call while losing the occasion to negotiate directly with the person whose work generated it. [91]

Put three bids on the same desk

Give the bank ten thousand accepted files for one operating year. All three bidders must deliver the same checked results and bring reopened files back to the same standard. A direct lab wants $142,000, to which the buyer adds $18,000 of retained operating work and $12,000 for switching. A cloud bundle costs $152,000, requires $12,000 of bank operations and costs $12,000 to switch. The bank can also host an open-weight model. It anticipates paying $65,000 for the machines and power, $83,000 for operations and $22,000 to set up the service. These are three invented bids. Their prices describe no particular company's offer.

EXHIBIT 36 / COMPARABLE TENDER MODEL

Which part can the incumbent bundle?



Three strips break competing bids into service, bank operations, switching and repair. The direct lab wins with $172,000 of total buyer outlay and a $142,000 fee.

Three illustrative US-dollar bids for one year. The vertical rule marks the lowest feasible outlay. Shading on a withdrawn quote preserves the cost comparison. The buyer requires equal quality and accepts the cheapest firm bid. The lab requires positive contribution. The cloud quote includes repairs and stays firm across this exercise.

Open exhibit 36 in the interactive edition

At a five per cent reopening rate, five hundred files require repair. Allow $32 per repair. The lab includes that service in its direct quote. So does the cloud bidder. On the self-hosted route the bank pays the $16,000 itself. It is looking at annual cash outlays of $172,000, $176,000 and $186,000 respectively. The lab wins this contest by $4,000. One per cent of the files being much harder than

expected could mean very little to the buyer's contracted cash outlay and a great deal to the supplier that promised to pay for repairs.

Look inside the winning offer. Assume ordinary delivery costs the lab $118,000, including the work needed to support the bank throughout the year. Its $142,000 fee leaves $24,000 before reopening costs and $8,000 afterwards. At a reopening rate of seven and a half per cent the entire $24,000 is gone. A supplier that requires a positive contribution would withdraw that offer at that point. It can quote a higher fee next year. It can also ask the bank to bear the repair bill. At five per cent the latter change takes the direct route to $188,000. The cloud's $176,000 offer now wins.

An old commitment can reverse the award without a single bad file. Assume $15,000 of unavoidable cloud expenditure will expire unused and the bundle's invoice qualifies against it. On the direct and self-hosted routes the bank still pays that $15,000. Their total outlays become $187,000 and $201,000. The cloud route stays at $176,000 because the same $15,000 settles part of its invoice. The bank has paid it once in all three cases. The balance due on the bundle's $152,000 service charge is $137,000 after applying the commitment. Describing the commitment as a discount on next year's bill would mislead both the buyer and the investor.

The company that bought its own machines

Nutanix has described a decision of precisely this kind about its own engineering work. The infrastructure vendor says it invested $20 million in a self-operated system for open-weight models after the economics of its external coding tools changed. It describes a design that directs routine work to its own equipment while sending the more demanding jobs to external frontier services. The company's stated allocation is eighty per cent routine and twenty per cent external. This is a supplier describing itself as a customer of the equipment it sells. Its account establishes a purchasing decision and a design intention. An audited saving per accepted engineering change would require different evidence. [92]

Buying the machines changes the shape of the problem. On a quiet Sunday the bank under our self-hosted proposal will still pay for the installation and the people who can restore it. On a busy Monday an extra request within the available capacity can be cheap. A fleet large enough for an ordinary week may be poorly matched to the month-end peak. Outsourcing the tail can make the internal installation more attractive by reducing the capacity it needs to reserve. In that case an outside model supplier can help a customer reduce its outside purchases while earning healthy revenue from the smaller, harder set it retains.

The buyer also has to count the jobs it sends twice. If an internal model attempts a file, fails the validation and passes it to an external service that starts again, both providers have consumed resources. Routing ninety per cent of initial requests internally could be entirely consistent with paying an external supplier for the most expensive half of the work. The invoice and the routing log measure different things. I would ask the owner for costs by accepted file, the distribution of those costs on deadline days and the number of cases that needed both routes. A press release describing a target allocation supplies none of these measurements.

There is a financial test for owning the equipment too. Spread an illustrative $20 million capital outlay over four years of service with no residual value. Recovering that investment alone requires $5 million a year before electricity, financing, maintenance or engineering. If the equipment supports ten million accepted jobs annually, capital recovery amounts to fifty cents each. At two million jobs it is two dollars and fifty cents. This calculation uses the size of the investment Nutanix reported to illustrate the sensitivity to utilisation. Its own service life, use of the machines and avoided external payments would be needed to calculate the company’s actual payback. Resale proceeds and a longer working life would lower the annual hurdle. [92]

Who can reserve the scarce hour?

Power introduces a constraint neither a benchmark nor an elegant contract can remove. Lawrence Berkeley National Laboratory estimated that US data centres used 176 terawatt-hours in 2023, about 4.4 per cent of national electricity consumption. Its 2024 report gave a wide range of possible consumption by 2028, roughly 325 to 580 terawatt-hours. Those numbers cover data centres broadly. They measure national energy over time. They cannot tell a bank whether a particular operator has spare capacity in the required region on the evening it must close the books. That local question is often the one for which a buyer will pay a premium. [93]

A cloud operator that can demonstrate capacity at the contracted time can charge for a promise a model developer would otherwise have to buy from it. A lab with a long-term reservation can offer the bank comparable assurance but must carry the reservation through quiet periods. A bank willing to delay some work can get a better deal from either party. The commercial advantage of scale comes partly from mixing customers whose peaks fall at different times. To price that advantage a supplier has to know whether the next customer's busiest afternoon coincides with everyone else's. Ten more banks with the same reporting deadline are a less attractive addition than a simple volume model suggests.

What would make me change my mind

The strongest incumbent response may be to sell the same result the challenger has promised, through a route already qualified. Salesforce's acquisition of Fin brought an outcome-priced business inside an established application company. Databricks describes its newer Genie as able to take on work across enterprise sources. The fact that a lab can build a better agent gives it a bargaining position. It gives the incumbent a reason to buy, partner, integrate or improve its own. A bank accepting any of these offers can record a genuine gain in productivity while renewing with the business that supplied the old interface. [40] [57]

There is a test that involves considerably less speculation than comparing model announcements. Read the tender after the pilot credit has expired. Establish whether the bank could have used the rest of its cloud commitment elsewhere. Identify the owner of the reopening obligation. Compare the actual repair cost against the price at which the winning supplier promised to carry it. Then give the runner-up the same incident record and ask for its next price. A lab that wins the first year, pays for its mistakes and wins the extension has defended something more valuable than a successful demonstration.

I would narrow the original prediction if established firms repeatedly kept those extensions while labs competed for replaceable inputs. The incumbent could lose the privilege of charging for every human seat and gain the privilege of choosing the model on every accepted job. A supplier of intelligence might become a much larger business in absolute terms under that arrangement. Its shareholders would still own a different business from the one that had captured the customer's operating budget. The decisive document would be readily recognisable to anyone who has bought enterprise software. Above the signature, next to the price, it says exactly who the bank expects to come back tomorrow.

CHAPTER 16

The terms of the next renewal

The prospectus tells us what Anthropic earned. A bank's signed renewal can tell us whose contract it replaced. The prediction needs both records.

A buyer can save herself a year of argument by writing down one sentence before the pilot begins. It should say precisely what would have happened to the old invoice if the pilot had never taken place. The ordinary renewal quote is worth preserving even when everyone is keen to accept the new offer. Six months later the sales team will remember the opening price, the procurement team will remember the discount it negotiated and the operations team will remember the temporary people it needed to keep both routes running. All three recollections could be honest. Only one comparison will answer the question the board thought it was asking.

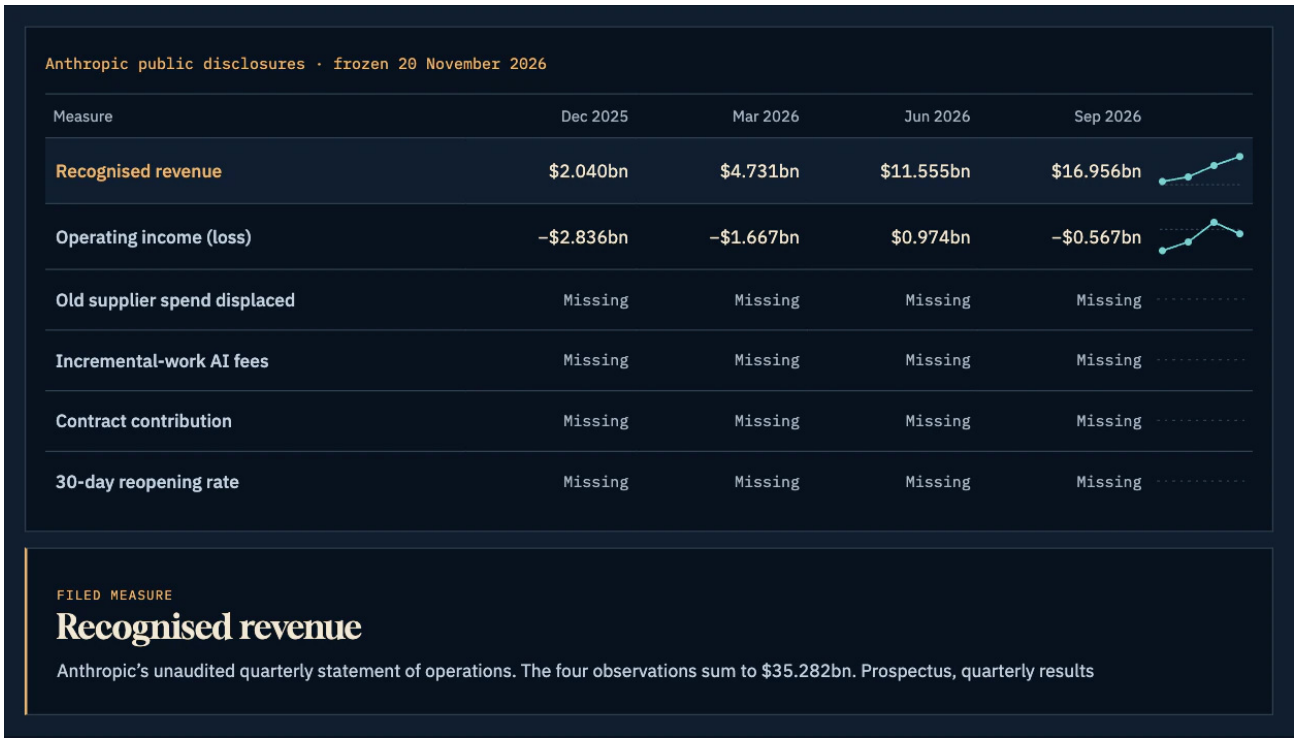
American banking supervisors had a version of this problem in mind when they issued their interagency guidance on third-party relationships. Their discussion of service measures asks banks to consider what an incentive will encourage. Paying for processing volume or speed without sufficient regard for accuracy or customer consequences is one example. Their discussion of termination covers remedies and the ability to change providers. The guidance dates from 2023. A bank can apply the underlying purchasing questions to a model provider, a traditional processor or the integration company standing between them. [94]

What the quarterly report can tell us

Anthropic's prospectus gives a precise answer to one important question. Across the four quarters ending in September 2026, its recognised revenue went from $2.040 billion in the first quarter of that sequence to $16.956 billion in the last. The four observations total $35.282 billion. The corresponding operating results include a profitable June quarter and a loss in September. This is enough evidence to establish rapid growth, to construct a trailing comparison and to ask why a larger business returned to an operating loss. It leaves open how much of the increase came from an enterprise cutting a previous supplier's invoice. [01]

EXHIBIT 37 / FILED RESULTS AND A BUYER SPECIMEN

The renewal scoreboard



A four-quarter ledger plots disclosed sales and operating results. It explicitly marks displaced spend, incremental AI fees, contract contribution and 30-day reopenings as missing.

Public record frozen on 20 November 2026. Financial observations use US dollars and are shown in billions. Each trace has its own vertical scale, with gaps for missing observations. "Missing" identifies an unavailable measure. The specimen is a separate, hypothetical buyer in its first four operating quarters. Public financial data

Open exhibit 37 in the interactive edition

Open the public ledger in the exhibit. A missing value is an important observation about the evidence available to a reader. Filling the blank for customer displacement with zero would claim knowledge that no customer cut a contract. Filling it with the company's revenue growth would claim that every additional dollar came out of an old supplier's pocket. Newly funded work, consumer spending and heavier use by existing customers can all increase revenue. The field becomes measurable for a particular buyer when somebody produces the relevant old and new orders and can explain what work each covered.

The second ledger is a specimen of the records that would answer the question. Its quarters are four periods in an invented bank contract. In the first, the lab receives $3,000 to clear a new backlog and the existing supplier still gets $40,000 for ordinary service. The measured reduction in the old bill is zero. In the second, the bank pays a lab $6,000 to try the existing work while retaining the original supplier at full price and paying for the changeover. A trial can produce valuable information and two simultaneous invoices. Attributing the trial fee to displaced spend would record a commercial victory before the buyer had withdrawn a dollar.

In the third period the old supplier accepts $30,000 for its reduced scope. The $8,000 paid to the lab for replacement work can be traced to the same budget. The lab also receives $2,000 for genuinely additional work and the bank incurs $1,000 of transition cost. The continuing service therefore costs $39,000 against the old $40,000 benchmark. Total cash spent reaches $41,000. Both the purchasing director who says the original service got cheaper and the controller who says

spending went up have read the accounts correctly. Their disagreement disappears when the additional work gets its own column.

The fourth quarter of the specimen supplies another test. The incumbent now receives $28,000, the lab gets $9,000 for existing work and $3,000 for new work. The transition charge has ended. The bank spends exactly $40,000. A line chart showing total expenditure would record a flat quarter against the benchmark and conceal a $12,000 fall in incumbent sales. On the replacement service alone, the buyer keeps $3,000. The same company can win a growing share of a fixed budget without causing the total to grow. An analyst who looks only at the aggregate could miss the event this essay sets out to predict.

Quality needs a denominator beside the money. In the third specimen quarter eight of four hundred accepted files reopened within thirty days. In the fourth, six of four hundred reopened. This is a difference of two cases. Before attaching a performance payment to the improvement, inspect those cases and the maturity of the follow-up. A file closed on the last afternoon of the quarter has had less opportunity to come back than one closed on the first. Freeze the cohort, wait for the agreed window and preserve its original membership. A supplier whose difficult files migrate silently into the next reporting period can collect a bonus for clearing a queue it has merely moved. The contract should credit the correction to the original reporting cohort.

Put four people on the order

The first signature belongs to the budget holder. She needs a dated alternative, an enumerated workload and a price for bringing every rejected file back into scope. Customer volumes vary, so preserve the unit rates as well as the total invoice. If the old supplier used a monthly minimum, record the minimum in its own field. The bank may be able to eliminate thousands of operations without taking one dollar off that month's bill. It may still earn the saving at the anniversary. Those are two different dates, and an honest forecast can show both.

The second signature belongs to the operating manager. Give him a week of difficult files and a way to trace the people who dealt with them. Have him write down what counts as a successful closure and how long a customer has to challenge it. The sample needs the files rejected by the automation as well as those it completed. A supplier can improve its published success rate simply by narrowing the population on which it is willing to be measured. The original queue gives the manager a way to spot that change and price the work left behind.

The third signature belongs to the supplier that accepts responsibility for repair. A public agency in Australia recovered the final instalment on a consulting report after problems with its references. The public record supplies the amount. A service credit in the bank's own contract needs an equivalent trigger, an owner who can invoke it and an accounting entry if it is paid. One contract may give the buyer a refund while leaving the expensive work of contacting customers with bank staff. Trace both amounts before deciding which bidder has the better warranty. [46] [47]

The fourth signature belongs inside the bank. Someone needs to know what became of the people who used to do the work. NatWest could report equivalent hours saved in complaints handling because it had a measure of time per case. A labour budget asks a further question. Did the team stop paying overtime, change its hiring plan, clear older complaints or take on another obligation? A worker who no longer assembles an evidence pack may spend the same hours reviewing disputed outcomes. The bank can have improved its service and preserved every salary. An investor seeking a transfer from the wage bill has more reporting to do. [72]

Two dates for the prediction

In 2028 I would ask for a cohort of operating contracts that have passed their first renewal. Include terminated pilots, unchanged renewals and customers that added an AI service for a previously unfunded backlog. Match the surviving contracts to a full year of reopening costs. Report how many firms supplied information and which types refused. A sample made entirely of suppliers' reference customers can establish that a mechanism exists. A representative estimate of how often it occurs needs a population from which the unsuccessful cases can also be drawn.

At the end of calendar 2031, compare four quarters of recognised AI company revenue with the corresponding four quarters at each named cloud business. Keep the comparison with AWS separate from the comparison with Google Cloud. Record the operating profit where the respective accounting boundaries allow it and explain the remaining differences. Microsoft’s disclosed Azure annual threshold cannot fill a missing quarterly series. Market capitalisation supplies another view of investors' expectations, but a price on the stock exchange cannot settle a prediction about four quarters of customer purchases. [01] [05] [06] [07]

These two dates answer different questions. If the AI business passes the revenue hurdle selling entirely new services, the size prediction succeeds and the displacement claim still needs its bank records. If buyers document substantial transfers from consulting or software yet the AI business stays smaller than the named cloud segment, the mechanism works and the size prediction misses. If the data needed to make a comparison remain private, record the missing evidence alongside the original test. My reason for writing down the terms now is to let a reader make those judgements later without needing my permission.

CODA

Whose signature counts

A signed decision leaves one choice with the customer. Two receipts for the opening bank file reveal who can earn a fee, who can refuse it and what would change the prediction.

On the last page of the decision about the food order, beneath a finding that Metro Bank should pay £100, there is one more piece of business. Christopher Convery, the ombudsman, asks Mr O to say whether he accepts or rejects the decision before 22 July 2025. Timely acceptance would bind both parties. We have the investigator's view, the bank's offer and the ombudsman's reasons. The published account ends before we learn what the customer chose. It would take only a keystroke to mark this file paid. It would take evidence from outside the file to justify the entry. [117] [123]

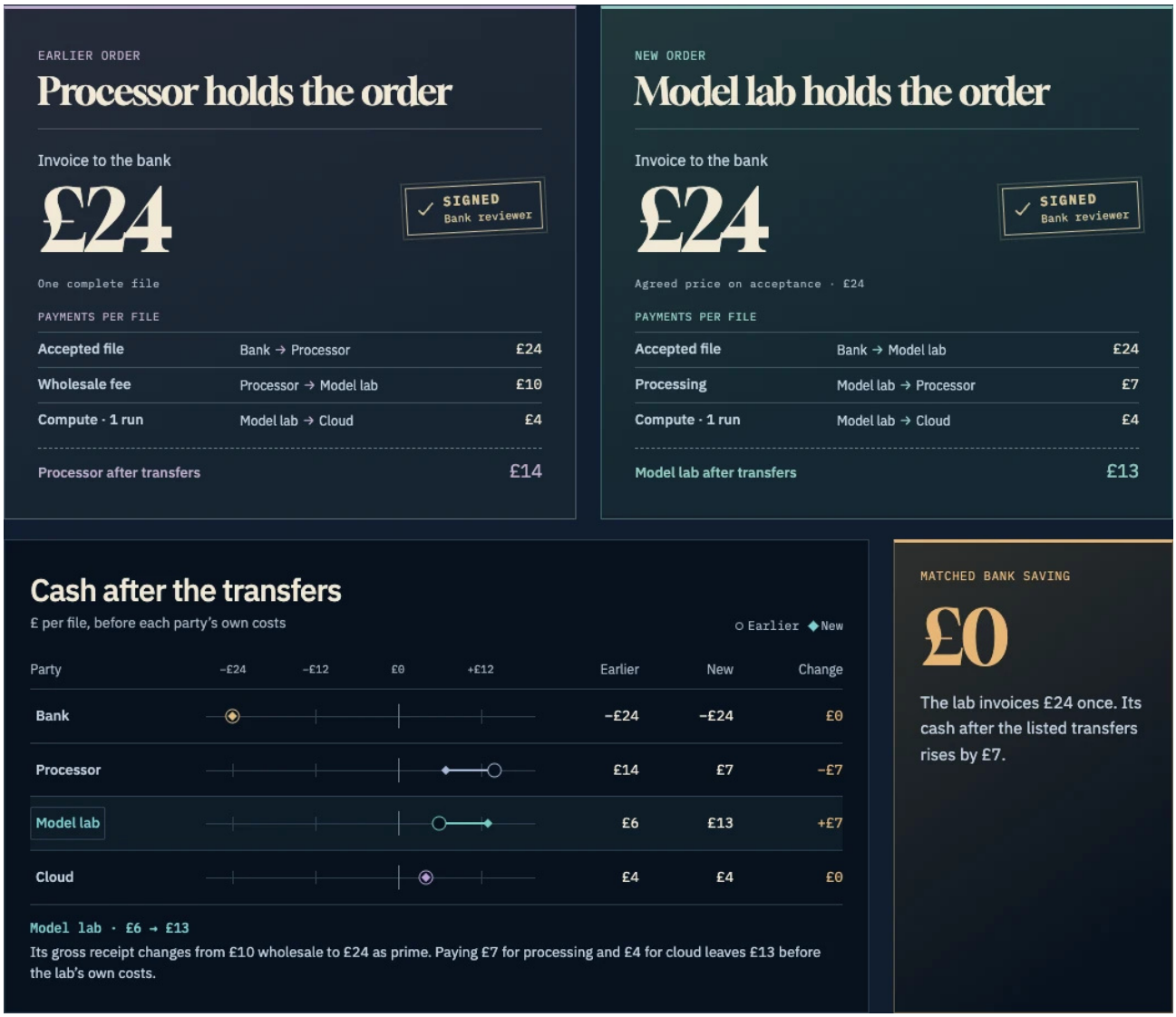
A detail on the preceding page explains why the bank had asked for more information. In these circumstances it had one opportunity to raise a chargeback on each transaction. There was a reason to get the receipts straight. There was also a telephone call Metro acknowledged it could have made, and at least one complaint it had failed to open when asked. A machine that finds a missing receipt could improve the bank's submission. A machine that finds the customer has already tried to complain could put him through to someone. Those are different services, even if the same model supplies both. [117]

An investor can put a seductive price on that call. In the financial year ending March 2026, the Financial Ombudsman Service kept its standard case fee for firms outside its group arrangement at £650. The first three cases were free. Seven in ten businesses whose customers referred complaints paid no case fees that year. The tariff helps explain why a bank might pay to resolve trouble early. It tells us little about the cash that a particular call would release. The decision about Mr O contains no invoice to Metro for adjudication. A supplier claiming £650 saved on every early closure would have to produce one comparable bill after another. [124] [117]

Consider the two receipts in the final exhibit. They belong to the invented £24 service order from the prologue. In the first, the processor owns the contract and buys £10 of work from the lab. The lab buys £4 of cloud and keeps £6 before its own costs. In the second, the lab wins the bank's order, pays the processor £7 and buys the same £4 of cloud. It keeps £13. The lab's receipts rise by £14, its balance after those purchases rises by £7, and the bank saves nothing. Ownership of a customer can change hands without a discount appearing anywhere on the customer's statement.

EXHIBIT 38 / TWO ILLUSTRATIVE SETTLEMENT RECEIPTS

Who signed, who earned, who stayed



Two signed receipts compare the same £24 service. The model lab keeps £6 under the processor and £13 as prime, the cloud keeps £4 in both, and the bank saves £0.

Illustrative sterling payments for one service file. Both contracts leave the bank responsible to the complainant. The receipts exclude any refund or compensation owed to the customer. The separate Metro Bank dispute concerned a March 2024 order and a £100 compensation offer. The published decision records no AI vendor fee. Ombudsman decision

Open exhibit 38 in the interactive edition

Now let the examiner withhold her signature. The bank owes no fee on the unfinished file. The first run has already cost the lab £4. If another run produces an acceptable file, the bank still pays once, the cloud supplier gets £8 and the lab keeps £9 at the original price. A procurement officer who counts a held file as £24 saved can improve her figures simply by leaving customers waiting. Give her an independent sample of the waiting files and the original dates on which people asked for help. An overdue answer will then carry its age into the next report.

The examiner has an economic interest in the contract too. If she can see which receipt the supplier used, she can spend her time on a disputed fact. If all she receives is a recommendation and a countdown to approval, her signature is worth less to the bank precisely when it appears more often. I would give her team authority to send back a defective file without losing credit for the time spent inspecting it. I would also ask who takes her calls when she is away. A process that

depends on one experienced employee has a cost that survives the cancellation of a software licence.

In 2028 I want to revisit buyers after a full renewal and ask the losing suppliers for their version of events. An incumbent may have surrendered work it priced too cheaply, then earned more selling infrastructure to the winner. A consulting partner may have kept the engagement by cutting the price and giving the bank ownership of the tools. If buyers repeatedly retain the savings while labs sell interchangeable inputs and incumbents keep setting the terms, I will withdraw the claim that control of these workflows is passing to the labs. A powerful model can be a very good purchase for the company I expected it to displace.

Climb Global Solutions, a technology distributor, gives investors both figures. For 2025 it reported $2.105 billion in gross billings and $652.5 million in net sales. At a lab bidding to take over a workflow, the accountant would ask which services the lab controls before delivery and which it merely arranges for another supplier to provide. [83] [82]

My separate wager comes due with the four quarters ending in December 2031. I expect an AI computing company, most likely Anthropic, to report more recognised revenue over those quarters than at least one of Google Cloud and AWS, measured separately. Passing AWS would clear the higher of the two starting points. Today’s filings establish those starting points. Future accounts will have to settle the comparison on consistent terms. Where the figures are available I will publish operating earnings alongside sales. If comparable accounts put every contender at or below both cloud businesses, I will record the miss. If the requisite disclosures disappear, the result will remain open. [01] [05] [06]

Mr O had no reason to care which company owned the machines that might one day read a case like his. He had reason to care whether somebody saw the evidence he said he had already sent. The company that can answer him, show its work and afford to answer again has something to sell on the scale of an industry. Whether it earns that position will be visible in a succession of ordinary documents. The supplier sends an invoice. The bank decides whether to renew. The last page bears the ombudsman’s name. The last choice belonged to the man who placed the order. [117]

THE REPORTING RECORD

Sources and reading notes

01 · sec.gov

Anthropic final prospectus, 18 November 2026

Use of proceeds, summary statements, quarterly results, financial statement Note 2 Other Concentrations and commitments. Nine-month statements unaudited. Keep GAAP and pro forma denominators.

02 · anthropic.com

Anthropic closing announcement, 20 November

Final public shares including exercise and secondary shares, concurrent placement. Aggregate transaction gross includes selling holders.

03 · investing.com

Reuters reporting on the first session, syndicated

Open, close and the distinction between a fully diluted valuation near $2tn and a value up to $1.5tn on outstanding shares. These are reported estimates.

04 · ft.com

Financial Times, supplied graphic

User supplied the image. Its option was conditional. Full linked text unavailable. Research lead only. Reconstruct data, dates and methodology independently.

05 · ir.aboutamazon.com

Amazon third-quarter 2026 earnings

AWS segment revenue and operating profit. Quarter to September. Exclude equity gains in Anthropic from AWS operating performance.

06 · sec.gov

Alphabet third-quarter 2026 earnings, SEC exhibit

Segment table, Google Cloud revenue and profit. Product scope includes more than infrastructure and results include TPU system sales.

07 · microsoft.com

Microsoft FY2026 annual report

Azure revenue exceeded $100bn for FY to June. Microsoft Cloud has broader scope. Keep threshold separate from exact quarterly series.

08 · gartner.com

Gartner October 2026 worldwide IT spending forecast

Table 1. Forecast vintage, $bn globally. Contains separately reported IaaS and services. Services is broader than consulting. Total includes communications and devices.

09 · investor.accenture.com

Accenture FY26 Q4 earnings

Consulting and managed services by type of work. New bookings can span years. One firm's revenue cannot establish total industry size.

10 · anthropic.com

Anthropic, enterprise AI services venture

Announced partners and implementation strategy. A press release cannot establish customer renewals or venture profitability.

11 · openai.com

OpenAI Deployment Company launch

Majority control, partners and agreement to acquire Tomoro. The May agreement precedes the July acquisition confirmation in S70.

12 · openai.com

OpenAI Partner Network

Announced $150m ecosystem investment. Planned spending, distinct from earned fees.

13 · anthropic.com

Anthropic finance agents

Named tasks and Microsoft 365 integration. Track availability per product/platform. No independent productivity estimate.

14 · github.com

Claude for Legal

Plugin inventory and supported workflows. Reproduce version before testing and record human legal sign-off.

15 · anthropic.com

Claude Design by Anthropic Labs

Design artefacts and handoff. Availability is evidence of entry, independent adoption needs additional reporting.

16 · salesforce.com

Salesforce Agentforce pricing

Pricing units, credits and seat alternatives. Public list price is different from negotiated unit price.

17 · snowflake.com

Snowflake Cortex evaluation guide

Distinct AI and warehouse meters, geography-specific example and usage history. Do not generalise effective pricing.

18 · docs.databricks.com

Databricks agent platform documentation

Unity Catalog governance, runtime and tools. Product evidence. Unlisted Databricks financials require attributed company disclosures.

19 · ibm.com

IBM z/OS Connect

API access to existing mainframe applications. Example architecture requires client authorisation and testing.

20 · academic.oup.com

Brynjolfsson, Li and Raymond, QJE 2025

Published version, staggered introduction, 5,172 support agents, +15% issues resolved per hour. Use this version consistently.

21 · metr.org

METR, experienced developers randomised study

16 developers, 246 real tasks in familiar repositories, early-2025 tools, 19% longer elapsed completion with AI allowed.

22 · metr.org

METR, subsequent study design and selection

Follow-up explains changing participation and task selection. Avoid extrapolating across model vintages and samples.

25 · srgresearch.com

Synergy Research, 29 October 2026

Worldwide Q3 enterprise spend on IaaS, PaaS and hosted private cloud. Report text gives $167.7bn, or $670.8bn at that quarterly pace, and $558bn trailing twelve months. Different scope from Gartner IaaS.

26 · info.futurumgroup.com

Futurum, enterprise applications market, H1 2025 edition

Table 2, April 2025 forecast vintage. 2026 forecasts $323.3bn SaaS and $91.8bn on-premises enterprise applications. Excludes many infrastructure and middleware categories.

27 · sourceglobalresearch.com

Source Global Research, 2026 service line forecasts

About $315bn worldwide consulting market by this firm's definition. Forecast, not audited turnover. Test sensitivity to inclusion of systems integration.

28 · gartner.com

Gartner, worldwide DBMS forecast, 2025 update

$161bn forecast for 2026 DBMS. Partial proxy for data platforms. Includes cloud databases, excludes some engineering, orchestration and governance.

29 · investor.salesforce.com

Salesforce FY27 Q2 earnings

Quarter to July 2026. $10.8bn subscription and support includes $440m Informatica. Separate acquisition contribution from organic change.

30 · sap.com

SAP Q3 2026 consolidated results

Table A.1, EUR millions. €2.581bn software licences and support. Vendor revenue category is a proxy for exposure to installed software.

31 · investors.snowflake.com

Snowflake fiscal 2027 quarterly results

FY27 Q2 to July 2026, $1.49bn product revenue. Consumption revenue differs from contracted ARR.

32 · databricks.com

Databricks Q2 announcement, 13 August 2026

Company-reported run rate above $7bn. Method and audited annual sales are separate questions.

34 · zylo.com

Zylo 2026 SaaS Management Index

Instrumented customer estate. For a renewal, distinguish entitlements, assignment, recent activity, workflow necessity and cancellability.

35 · learn.microsoft.com

Microsoft Foundry, Claude consumption unit billing

Billing unit, marketplace private offers and application to a Microsoft Azure Consumption Commitment. Other account arrangements can differ.

36 · cloud.google.com

Google Cloud, multi-region endpoints for Claude

Regional processing and eligible multi-region routing on Vertex AI. Scope depends on model, region and mode.

37 · aws.amazon.com

Trust Bank on Amazon Bedrock, AWS customer study

Supplier-published Singapore bank case reporting over 52% enquiry resolution without human intervention. No matched invoice or independent causal estimate.

38 · flexera.com

Flexera's own three-year software renewal

First-party account of a $46m mixed software and cloud agreement. $1m hard SaaS savings, $1.5m other hard savings and $4.5m cost avoidance over three years. No matched invoices disclosed.

39 · fin.ai

Fin outcome billing rules

17 September 2026. Distinguishes confirmed or assumed resolution and configured procedure handoff. Base plan, included resolutions and incremental unit price matter.

40 · salesforce.com

Salesforce completes its acquisition of Fin

Closing announcement dated 10 September 2026. Buying an outcome-based service can still pay the incumbent group.

41 · investor.salesforce.com

Salesforce and Anthropic announce Claudeforce

26 August 2026 product and partnership announcement. Salesforce in Claude and Claude inside Salesforce illustrate alternative locations for orchestration. Commercial splits are undisclosed.

42 · developer.salesforce.com

Salesforce Agent API setup

External client app and authorised identity for calling Agentforce. Specific access depends on configuration.

43 · zylo.com

Zylo's buyer guide to subscription management

Zylo's interpretation of the 2026 index reports 54% of licences used and estimates $19.8m yearly unused spend for a typical large organisation. Telemetry and estimated exposure, not audited cancellable cash.

44 · sec.gov

Accenture FY2026 Form 10-K

Item 7 on bookings and Note 2 on revenue recognition. Billing, cash collection and revenue recognition can occur in different periods.

45 · sec.gov

Globant 2025 Form 20-F

Page 48 and Note 5, page F-34. Contract-type revenues in thousands of US dollars. A technology-services comparator with a different product and geographic mix.

46 · dewr.gov.au

Australian Department of Employment and Workplace Relations, Targeted Compliance Framework assurance review briefing

Ministerial briefing SB25-000331 cleared 17 September 2025, printed p.108. Original fee and requested final-instalment repayment in Australian dollars including GST, plus scope and residual expenditure.

47 · api.tenders.gov.au

Australian Department of Finance, AusTender contract CN4118426

Contract record reports repayment of the last A$97,587.11 instalment against an original A$439,142 contract. No measurement of supplier repair hours or public cost of delay.

48 · doi.org

Dell’Acqua et al., Navigating the Jagged Technological Frontier, Organization Science 2026

Preregistered experiment with 758 consultants and GPT-4. Distinct tasks and grading methods. Figure 5 gives correct-answer rates on one diagnostic task. No buyer invoices or staffing decisions.

49 · dewr.gov.au

Australian Department of Employment and Workplace Relations, correspondence on the final assurance report

Department letter of 8 September 2025, document 3A, pp.15–16. Approved code analysis, reported uses in publication work and contractual basis for repayment.

50 · sourceglobalresearch.com

Source Global Research, August 2026 global consulting market forecast

20 August 2026. Forecast to pass US$350bn worldwide in 2026. March forecast S27 was about US$315bn. The articles do not reconcile the two estimates.

51 · sec.gov

Snowflake, Form 10-Q for the quarter to 31 July 2026

Note 3 gives $1,491.861m product revenue. The segment table reports third-party cloud infrastructure expenses at approximately 75% of cost of product revenue in the quarter. Those percentages use different denominators.

52 · medium.com

Zendesk engineers on parallel warehouse tests, 25 November 2025

The authors cloned three large data models and their production workflows. They report more than 80% lower credit costs on right-sized warehouses. Customer-authored on a Snowflake publication. The whole-company invoice and the cost of the newly added workloads are unavailable.

53 · docs.databricks.com

Databricks Apps, authorisation on AWS

User authorisation enforces the user's Unity Catalog permissions within the app's approved scopes. App authorisation uses its own service principal. Product capability depends on configuration.

54 · docs.snowflake.com

Snowflake, reusing persisted query results

Conditions include query text, data, role privileges and configuration. Eligibility does not guarantee result reuse. This differs from a cache of a previously generated answer.

55 · docs.snowflake.com

Snowflake, warehouse billing and sizing

One-minute minimum on provisioning, then per-second billing. One-credit-per-hour X-Small example. The exhibits assume an isolated, immediately suspended warehouse episode. Shared production workloads have different effective costs.

56 · docs.snowflake.com

Snowflake, query attribution history

Compute attribution per query excludes idle time and other service categories. Reconcile with warehouse metering and the customer contract before claiming cash savings.

57 · databricks.com

Databricks announces the second generation of Genie, 11 November 2026

The company describes proactive work and connections across enterprise data sources. Product announcement. Platform usage, customer cash savings and retention are separate measurements.

60 · ibm.com

Gruppo BCC Iccrea integration, IBM customer study

IBM describes an Italian banking group's existing core invoking external services and its monitoring integration. The supplier reports a 65% reduction in integration effort, 40 person-days annually for incident management and half an FTE for monitoring. No matched invoices, payroll or AI attribution.

61 · fca.org.uk

FCA final notice to TSB Bank plc, 2022

Paragraphs 6.8–6.12 document service disruption, 225,492 complaints and £32,705,762 redress following the 2018 migration. This is a named operational failure, with no general estimate of migration or AI risk.

62 · swift.com

Swift on the Unique End-to-end Transaction Reference

A 36-character payment reference travels along the chain. It aids tracing and reconciliation. Our specimen's local case and journal references are invented and use a different format.

63 · eur-lex.europa.eu

EU Digital Operational Resilience Act, Regulation 2022/2554

Article 28(1)(a) preserves the covered financial entity's responsibility for its obligations when it uses third-party ICT. The illustrative maker and checker are a local design choice.

64 · support.sap.com

SAP Business Suite 7 maintenance strategy

Mainstream maintenance through 2027 for specified core applications and enhancement packages. Optional extended maintenance in 2028–30 adds two percentage points to the maintenance basis. Different products and versions have different terms.

68 · anthropic.com

Anthropic launches the Claude Partner Network, 12 March 2026

Initial $100m commitment for partner training, support and market development. Programme investment is distinct from customer receipts.

69 · ode.com

Ode with Anthropic, operating launch on 15 July 2026

Standalone firm identifies Fractional AI, acquired in May, as its operating team and names its leadership. No fee or ownership percentage is inferred.

70 · deploy.co

DeployCo announces its agreement to acquire Northslope, 8 July 2026

The company describes its recent acquisition of Tomoro as completed by this date. Northslope was an agreement subject to closing in this announcement.

71 · natwestgroup.com

NatWest Group, collaboration with OpenAI announced in March 2025

Buyer-side announcement sets out collaboration in the context of existing assistants. Gives no division of renewal revenue among technology suppliers.

72 · investors.natwestgroup.com

NatWest Group, 2025 Annual Report and Accounts

Scaling our capabilities in technology. Group reports around nineteen minutes less complaints handling time and over 22,000 equivalent hours in 2025. No supplier-specific attribution or payroll cash release is given.

73 · nber.org

Humlum and Vestergaard, Still Waters, Rapid Currents

NBER working paper 33777, revised March 2026. Surveys linked to Danish administrative records through December 2024. Eleven exposed occupations. The revised abstract rules out average effects above 2% for earnings and recorded hours over the study window.

74 · digitaleconomy.stanford.edu

Brynjolfsson, Chandar and Chen, Canaries in the Coal Mine?

Stanford working paper, August 2026 revision using US ADP payroll observations through June 2026. The 19% is a relative employment gap for ages 22–25 compared with less-exposed peers. Authors describe selection, pretrends, education sensitivity and limits to causal inference.

75 · libertystreeteconomics.newyorkfed.org

Audoly, Guerin and Topa, US vacancies and occupational AI exposure

New York Fed researchers' 14 May 2026 analysis of Lightcast postings. Relative weakness of high-exposure openings predates ChatGPT. Advertised vacancies and realised payroll employment measure different stages of hiring.

76 · anthropic.com

Anthropic, randomised study of coding skill formation

29 January 2026. Fifty-two mostly junior Python engineers learning Trio. Immediate quiz means were 50% with assistance and 67% without. The timing difference was statistically insignificant. No long-term competence or earnings measure.

77 · doi.org

Emanuel, Harrington and Pallais, The Power of Proximity to Coworkers

Quarterly Journal of Economics, August 2026. Changes in proximity at one Fortune 500 employer, 2019–2024. Co-located teammates gave 18.3% more coding feedback. A study of learning from colleagues, outside the evidence on AI treatment effects.

78 · universityofcalifornia.edu

Xingqi Maggie Ye on the Berkeley field study of work intensity

26 February 2026 interview with the researcher. Eight months of ethnography at one US technology company with about 200 employees. Observations and interviews identify changes in pace, scope and boundaries. No causal estimate of paid hours or population prevalence.

79 · cms.gov

CMS, Medicare fee-for-service recovery audit program, FY2022 appendices

Appendix B reconciles overpayments collected, underpayments restored, first-level appeals, contractor fees and program administration to $220.74m returned to trust funds. A public operating account for a specific year. No AI attribution.

80 · cms.gov

CMS, Medicare and Medicaid integrity programs, FY2022 report to Congress

Section 4.2.1 describes negotiated recovery fees, their return after an appeal and independent validation. Identifying an overpayment differs from collecting it.

81 · sec.gov

Klarna Group, 2025 Form 20-F

Operating expenses and operating leverage discussions, especially pp. 132–134 and 184–185. Management estimates $59m of 2025 assistant savings. Customer service and operations expense rose $4m to $207m. The model supplier's fee is undisclosed.

82 · ifrs.org

IFRS Interpretations Committee, software reseller decision, May 2022

IFRS 15 B34–B38 require identification of specified goods or services and assessment of control before transfer. The indicators require judgment under each contract.

83 · climbglobalsolutions.com

Climb Global Solutions, full-year 2025 results

FY2025 net sales $652.5m and gross billings $2.105bn. The company's operating definition of gross billings includes values of certain transactions reported net in revenue. Company-wide figures span several products and contract types.

84 · sec.gov

Alphabet, fourth-quarter 2025 results, SEC exhibit

Quarter to 31 December 2025, segment results in USD millions. Google Cloud revenue 17,664 and operating income 5,313. The other three quarters bring the comparison to September 2026.

85 · sec.gov

Alphabet, first-quarter 2026 results, SEC exhibit

Quarter to 31 March 2026, Google Cloud revenue 20,028 and segment operating income 6,598, USD millions.

86 · sec.gov

Alphabet, second-quarter 2026 results, SEC exhibit

Quarter to 30 June 2026, Google Cloud revenue 24,768 and segment operating income 8,814, USD millions. Product scope includes sales of TPU systems beginning in this quarter.

87 · sec.gov

Alphabet, September 2026 Form 10-Q

Segment note reconciles the first nine months and explains how shared costs and Alphabet-level activities are treated. Google Cloud segment operating income has a different cost boundary from the lab’s consolidated operating result.

88 · investors.klarna.com

Klarna, first-quarter 2024 results announcement, 30 May 2024

Management dates the assistant launch to January 2024. The 2025 Form 20-F gives February and reports a December-to-January repeat-inquiry comparison. Neither account supplies the exposure dates of the comparison cohort.

90 · assets.publishing.service.gov.uk

UK Competition and Markets Authority, cloud actions update of 16 October 2026

Pages 7–8. UK switching egress terms, longer windows and interconnect availability. Findings are specific to geography, provider and date.

91 · github.blog

GitHub, cost and quality settings for automatic model selection, 14 September 2026

Three routing tiers and a 10% discount on eligible automatic usage for paid subscribers at announcement. A product setting gives no market share measure.

92 · nutanix.com

Nutanix describes its own private AI deployment

Nutanix reports $20m invested and an 80/20 routing strategy for its own engineering workloads. Self-reported implementation by an infrastructure seller. No audited return on capital or cost per accepted change is supplied.

93 · eta-publications.lbl.gov

Lawrence Berkeley National Laboratory, 2024 United States Data Center Energy Usage Report

Executive summary. Estimated US data centre electricity use in 2023 and 2028 scenarios. Annual energy at national scale cannot establish site-specific power availability or a particular provider’s cost.

94 · federalreserve.gov

US federal banking agencies, Interagency Guidance on Third-Party Relationships, June 2023

Contract negotiation sections on performance measures, subcontracting and termination. US supervisory guidance. The essay’s proposed tender and evidence schedule are illustrative.

95 · sec.gov

Meta Platforms Form 10-Q, September 2026

Balance sheet cash and cash equivalents of $11,222m at September 30, 2026. Excludes marketable securities.

96 · sec.gov

Microsoft Form 10-Q, September 2026

Balance sheet cash and cash equivalents of $20,281m at September 30, 2026. Short-term investments appear on a separate line.

97 · sec.gov

NVIDIA Form 10-Q, July 2026

Cash and cash equivalents of $22,443m at July 26, 2026. Latest fiscal period ending on or before the September 30 cutoff.

98 · sec.gov

NVIDIA fiscal third-quarter 2027 results, November 17, 2026

Cash and cash equivalents of $29,089m at October 25, 2026. Subsequent observation available by the November 20 offering close.

99 · sec.gov

Apple Form 10-K, September 2026

Cash and cash equivalents of $36,305m at fiscal year-end September 26, 2026, four days before the comparison cutoff.

100 · sec.gov

Space Exploration Technologies Form 10-Q, September 2026

Balance sheet cash and equivalents of $51,294m at September 30, 2026. Cash flow statement includes restricted amounts; marketable securities of $38,561m are separate.

105 · blogs.microsoft.com

Microsoft, NVIDIA and Anthropic announce strategic partnerships, 18 November 2025

Original $30bn Azure purchase statement. The later amendments appear in S01. Dollars and a stated gigawatt quantify different commitments.

106 · anthropic.com

Anthropic and Amazon, expanded compute collaboration, 20 April 2026

Original announcement of more than $100bn over ten years and up to five gigawatts. S01 specifies the $110bn minimum and payment window.

107 · sec.gov

Akamai, Form 8-K on Anthropic project plans, 24 September 2026

Item 1.01. Two September 18 project plans under a May 5 master agreement. Approximately $11.6bn aggregate subject to delivery, service and termination terms. Seven years from each start of service.

108 · ir.akamai.com

Akamai, expanded strategic relationship with Anthropic, 24 September 2026

The slide titled “Deal Timing & Cash Generation Phasing” models $5.5bn cumulative supplier capex, including roughly $1.7bn in Q4 2026 with no associated revenue that quarter. Management estimates at September 24.

111 · anthropic.com

Anthropic, Introducing Claude Haiku 4.5, 15 October 2025

Archived launch terms of US$1 per million input tokens and US$5 per million output tokens. Used as an explicit dated tariff in the specimen, before caching, tools or discounts.

112 · microsoft.com

Microsoft, Copilot plans and pricing for enterprise in the US

Microsoft 365 Copilot enterprise add-on, US$30 per user per month, paid yearly on an annual subscription. Qualifying Microsoft 365 plan required. Local prices and negotiated agreements differ.

113 · find-tender.service.gov.uk

South Downs National Park Authority, UK contract details notice 2025/S 000-081156

Published 9 December 2025. Digital Devices Ltd contract signed 13 October for 175 one-year Copilot licences. £56,555.10 includes VAT. Notice reports contract value, no active usage, realised savings or payment schedule.

114 · helpcenter.gorgias.com

Gorgias, How you are billed for using Gorgias

Billing guide distinguishes helpdesk tickets, automated interactions, human involvement within 72 hours and legacy terms. Applies to that commercial definition only.

115 · docs.moderntreasury.com

Modern Treasury, Approval Rules Overview

Manually created or updated objects, last-editor review restriction, renewed approval on an update and audit trail. Describes a configurable product. The chapter's bank policy is an explicit stipulation.

116 · docs.moderntreasury.com

Modern Treasury, Transaction Status and Balances

Contrasts pending, posted and archived states. Undoing posted entries requires a separately posted reversing transaction. The example is on the bank’s own books and stipulates authorisation and available funds.

117 · financial-ombudsman.org.uk

Financial Ombudsman Service, Metro Bank decision DRN-5452535

An online food order, two disputed debit card amounts and £100 compensation for service failings. The ombudsman found the request for more evidence reasonable. There is no AI deployment or vendor fee in the decision.

118 · jpmorganchase.com

JPMorganChase, 2026 company update, firm overview

Technology expense outlook slide. Approximately $19.8bn planned for 2026, with the split by division and rounded drivers of growth. Approximately one quarter of $0.6bn in identified efficiencies was AI-related. Forecast, before attribution to any vendor.

119 · jpmorganchase.com

JPMorganChase, company update transcript

Management discussion accompanying the technology outlook, including demand, unit costs, headcount and additional investment. Describes the bank’s plans.

120 · bea.gov

US Bureau of Economic Analysis, gross output and value added by industry

Methodological explanation of purchased intermediate inputs and value added. The essay’s closed supplier chain is an invoice consolidation with its own boundary.

121 · bea.gov

US Bureau of Economic Analysis, NIPA Handbook chapter 6, Private Fixed Investment

Section on software production, particularly pages 6-37 and 6-38. Qualifying own-account software is estimated from production costs. Applies to US national accounts.

122 · ifrs.org

IFRS Interpretations Committee, configuration or customisation costs in a cloud computing arrangement

IAS 38 agenda decision on a specified SaaS arrangement. Separate asset, nature of the service, supplier and delivery date affect recognition of customer expense.

123 · legislation.gov.uk

Financial Services and Markets Act 2000, section 228

Subsections 3 to 5 cover the signed reasons, the invitation to accept or reject and the binding effect of acceptance. The specific customer's response is outside the published case record.

124 · assets.publishing.service.gov.uk

Financial Ombudsman Service, Annual Report and Accounts 2025/26

Financial report, pages 38–39. Group arrangements and the standard £650 case fee for non-group firms, the annual free case allowance and the proportion of businesses paying no case fees. The rate does not establish the cost of a particular complaint.

EXHIBIT INDEX

The 38 exhibits

Exhibit 01	One bill through the enterprise	5
Exhibit 02	Where the offering went	8
Exhibit 03	The first day on two share counts	10
Exhibit 04	Cash on the same basis	11
Exhibit 05	Seven quarters of selling intelligence	14
Exhibit 06	What happened to earnings per share	15
Exhibit 07	The capacity contract	18
Exhibit 08	Three ways to sell a machine decision	22
Exhibit 09	A case earns its signature	24
Exhibit 10	Which IT expenditure counts?	27
Exhibit 11	How a dollar gets counted twice	30
Exhibit 12	Three clouds, three disclosures	33
Exhibit 13	The marketplace beneath the model	34
Exhibit 14	A cloud can grow as its share falls	36
Exhibit 15	Two labs take a seat at the client	39
Exhibit 16	Who keeps the second engagement?	41
Exhibit 17	One consultant invoice, two businesses	45
Exhibit 18	The staffing pyramid after automation	47
Exhibit 19	The small specialist’s counterbid	48

EXHIBIT INDEX / CONTINUED

Exhibit 20	The missing seat and the new action	53
Exhibit 21	Who can raise the price?	55
Exhibit 22	The governed query	59
Exhibit 23	How often will the warehouse wake?	61
Exhibit 24	The bank’s double entry	65
Exhibit 25	What a bank must see before it pays	67
Exhibit 26	Five bills before and after renewal	72
Exhibit 27	Five contracts for one banking operation	75
Exhibit 28	The desktop handoff	76
Exhibit 29	Two studies that cannot be averaged	79
Exhibit 30	Where recovered hours go	81
Exhibit 31	The settlement of a hundred dollars	86
Exhibit 32	The price of a failed run	87
Exhibit 33	Three owners of the workflow	89
Exhibit 34	The hurdle to exceed AWS	92
Exhibit 35	The growth surface	94
Exhibit 36	Which part can the incumbent bundle?	98
Exhibit 37	The renewal scoreboard	102
Exhibit 38	Who signed, who earned, who stayed	106

Copyright tarrysingh.com ©