A fixed-price line item for an applied-AI research phase looks, to a finance leader, like a single number to approve or reject. It is neither single nor a number you should read on its own terms. This whitepaper reads the commercial anatomy of a multi-year subsurface computer-vision programme built for a major operator in Oman and reframes it as the questions a CFO should put to any applied-AI R&D vendor. Six levers set the price. First, effort-obligation contracting, because the deliverable is a metric nobody can warrant at signing, so paying for a guaranteed result is paying for a payout the vendor cannot honestly price. Second, milestone-tranche payments gated to phase boundaries, which convert an open-ended bet into a sequence of small, cancellable ones: on this programme, five tranches of 84,047.60, 66,648.40, 64,584.00, 78,717.60 and 51,584.00 US dollars summing to 345,581.60, where the buyer's real exposure at any gate is only what is sunk. Third, a contingency for requirement change, a bounded 7% line (8,135.40 US dollars on a worked phase) that absorbs scope drift without reopening the price. Fourth, owned-versus-rented compute, the choice that decides whether the largest recurring cost is fixed or floating, which through the 2022 energy shock meant the difference between a rental market climbing from 12-15K to past 20K a month and a single bounded contingency of about 51,000 US dollars for twelve months. Fifth, academic-rate arbitrage, the single largest lever, where routing delivery through a university partner halved the day-rate card (a principal at 1600 against 3200, a senior researcher at 760 against 1727) and drove a 35% bulk discount worth 62,580 US dollars on a 178,800 role stack. Sixth, per-head training economics, which price capability transfer into the mandays so the operator ends up owning what it paid to build. We assemble one worked phase from these levers to 169,355.40 US dollars and argue that a disciplined price is engineered from the six, while a lump-sum quote hides which of them the buyer is, or is not, getting. The mechanics of each lever were treated at length in earlier chapters of this programme; here they are the columns of a single buyer scorecard.

Tarry Singh
Founder & CEO

Tannistha Maiti
Senior AI Researcher

APPLIED-AI-PROCUREMENT / CFO-SCORECARD / ENGAGEMENT-ECONOMICS / EFFORT-OBLIGATION / MILESTONE-TRANCHES / PHASE-GATES

CONTENTS

01 The lump sum is a category error

02 Lever two, made concrete: exposure is only what is sunk

03 Lever three: a contingency that is a line, not a surprise

04 Lever four: whether the biggest recurring cost is fixed or floating

05 Lever five: the arbitrage that is the largest single lever

06 Lever six: paying for capability you keep

07 Assembling one worked phase from the six levers

08 The same money, read in two currencies and two counterparties

09 The questions, as a buyer would actually ask them

10 Where this leaves the finance leader

11 Limitations

12 References

A finance leader who is handed a single fixed-price line for an applied-AI research engagement has been handed a summary of a decision, not the decision. The number is real. It can be defensible or hopeful, disciplined or lazy, and from the outside the two look identical. What separates them is not the digit at the bottom but the structure above it: how the work is contracted, how the money is phased, what absorbs the surprises, where the compute sits, which rate card the people are billed on, and whether the price buys the operator its own future capability or rents it forever. Read the line alone and you cannot tell. Read the six levers behind it and you can.

We write this from a specific engagement rather than a template. Over roughly three years we built and ran a subsurface computer-vision programme for a major operator in Oman: detection models for fractures and bedding planes in a fractured carbonate reservoir, a vug-quantification track, and the compute-heavy training a research-grade programme consumes over years. The technical work has its own write-ups. This paper is about the other envelope, the commercial one, and it is written for the person who has to approve the spend and later answer for it. The programme's own workbooks, cost memos and signed agreements are the source for every number here. The client's identity and its wells are withheld; nothing else is.

The claim is simple and it is the spine of everything below. A CFO should price an applied-AI R&D engagement on six levers, not on a single lump sum. The six are effort-obligation contracting, milestone-tranche payments gated to phase boundaries, a contingency for requirement change, owned-versus-rented compute, academic-rate arbitrage, and per-head training economics. Five of them are questions you put to the vendor before you sign. All six are things a lump-sum quote can quietly get wrong while still footing to a plausible total.

The lump sum is a category error

Start with why the instinct to buy a fixed outcome for a fixed price fails on this kind of work specifically. In a conventional procurement, a fixed price is a discipline: it forces the vendor to absorb its own estimation risk, and it gives the buyer a clean number. That logic holds when the deliverable is knowable at signing. An applied-AI research deliverable is not. The thing being bought is a metric, a detection accuracy or a throughput, whose value nobody can commit to before the labelled data arrives and the experiments run.

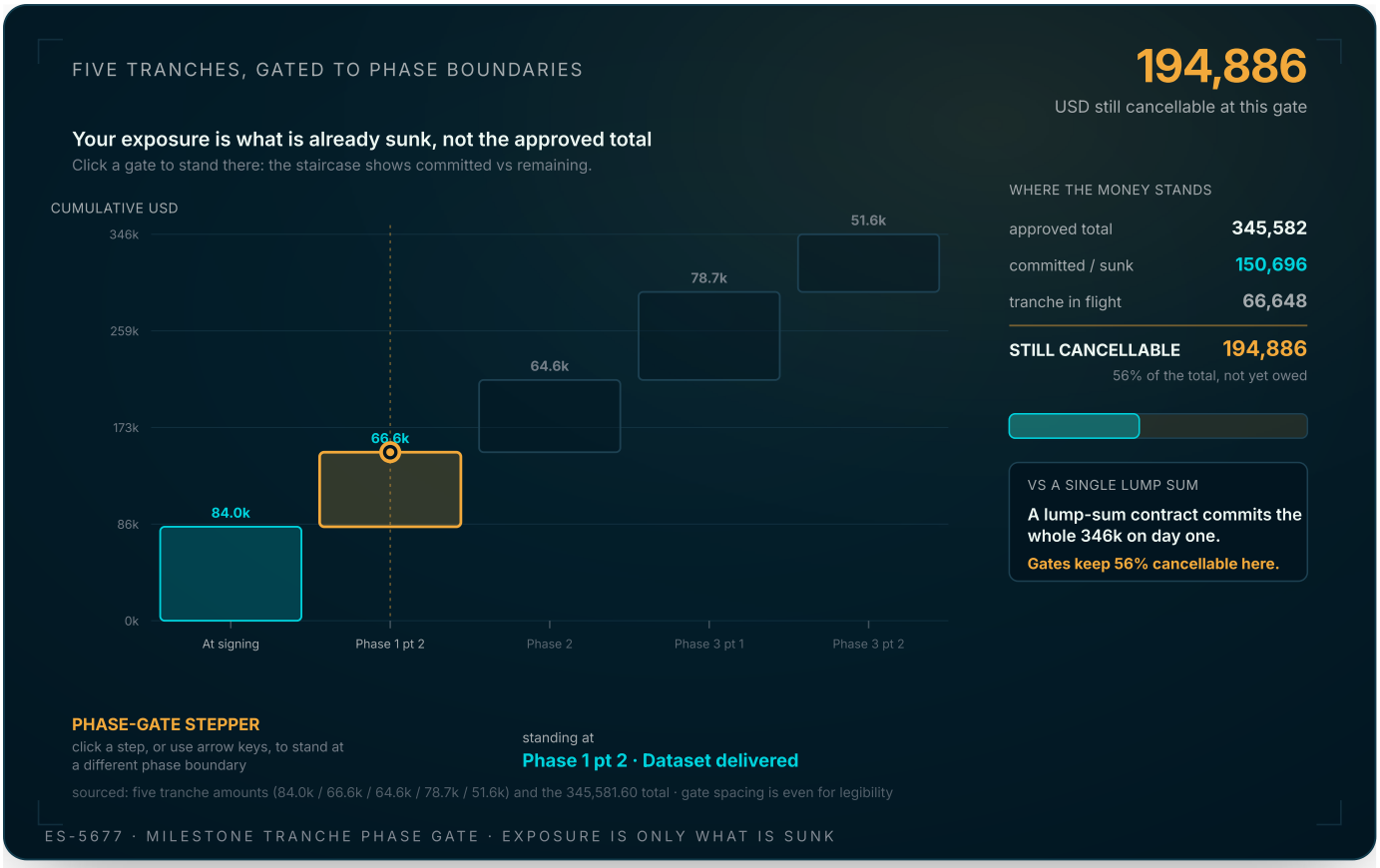
On this programme the point was not hypothetical. Fracture-recall accuracy at a five-centimetre depth tolerance climbed from roughly 10% at three labelled wells to about 40% at eight to roughly 75% at 16. That curve was unknowable when the contract was signed. A result guarantee written at kickoff would have committed the provider to a bar the science only cleared near the very end, and would have paid perverse incentives the whole way there: it rewards safe, unfalsifiable claims over honest experiments, because an honest experiment can fail and a vague one cannot be caught failing. The correct contractual object is therefore an effort obligation, explicitly "not responsible for achieving a specific result," which prices the work rather than the outcome. That is lever one, and the full clause-by-clause reading of how such a contract is papered, effort obligation, phase-capped liability, the data-timeliness clause, and the foreground-IP split, lives in our earlier chapters on contracting an AI R&D engagement and is a pointer here, not a re-derivation.

There is a subtler failure too, and finance leaders feel it after signing rather than before. A lump sum collapses the internal structure of the price into a single figure, and a figure cannot be audited. When the programme is a year in and someone on the board asks why the number was what it was, a lump sum has no answer beyond "that is what they quoted." A price built from named levers has an answer for every dollar: this much was people, at this rate card, discounted this much for the whole-phase commitment; this much was a bounded contingency; this much was compute, carved out and owned rather than rented. The difference is not cosmetic. It is the difference between a spend you can defend under scrutiny and one you can only apologise for.

The lump sum's second structural failure is that it hides the phasing. A single number paid on a single schedule gives the buyer no cancellation optionality, which on genuinely uncertain research is the most valuable thing the buyer can hold. That is lever two, and it deserves the first instrument.

Lever two, made concrete: exposure is only what is sunk

The signed year-1 service agreement on this programme carried a total fee of 345,581.60 US dollars, and it was not payable as one sum. It was five tranches, each released at a phase boundary: 84,047.60 at signing, then 66,648.40, 64,584.00, 78,717.60, and a final 51,584.00, invoiced on a five-month cadence. Each gate was a real decision point, with provider phase-exit ramps after Phase 1 and Phase 2. The structure is not administrative housekeeping. It is the risk-management instrument.



Paying an applied-AI research programme as milestone-tranches gated to phase boundaries turns an open-ended bet into a sequence of small, cancellable ones. This staircase plots the five sourced tranches of the signed year-1 agreement cumulatively, from 84,047.60 US dollars at signing to a final 51,584.00 at handover, summing to 345,581.60. Click a gate to stand there and read the split: the committed and sunk portion in teal, and the remaining tranches that a lump-sum contract would already have committed but a gated one leaves cancellable. The orange marker is the only element that argues: the cancellable remainder at the chosen gate, which at every point before the last is a large fraction of the total the CFO does not yet owe. All five tranche amounts and the total are sourced from the signed agreement; the gates are drawn at even horizontal spacing for legibility, while the real invoicing cadence was five months.

The staircase above is the CFO's view of that structure. Stand at any gate and the total resolves into three parts: what is already committed and sunk, the tranche currently in flight, and the remainder that is still cancellable. A lump-sum contract commits the whole 345,581.60 on day one. The gated contract, at every point before the last gate, leaves a large fraction of the total un-owed and walk-away-able. That cancellable remainder is the entire argument for phasing, and it is why the milestone schedule, not the rate card, is the thing a treasurer should read first.

Two properties of the tranche design are worth naming because a buyer can specify them. The first is that the gates fall on phase boundaries rather than on the calendar. A tranche released simply because three months passed rewards elapsed time; a tranche released because a phase deliverable landed rewards progress, and it gives the buyer a genuine

decision at each gate rather than a rubber stamp. The second is the exit ramp. On this programme the provider could leave cleanly after Phase 1 or Phase 2 on short notice, and the buyer could terminate on sixty days' notice either way. Exit ramps sound like they protect only the vendor, but they protect the buyer more: a vendor that is contractually free to walk after a completed phase has no incentive to pad a struggling programme along to keep billing, because it is not trapped. The mechanics of how these tranches map onto the phases, and onto the two-party funding structure denominated in Omani rial that sat behind them, are treated in full in our anatomy of the twenty-month engagement's money and milestones. Here the five tranches are simply the columns that make lever two legible, and the lesson for the scorecard is narrow: a buyer should be able to point at any moment in the programme and say exactly how much of the approved total is still theirs to cancel.

Lever three: a contingency that is a line, not a surprise

Research changes shape as it runs. Requirements drift, a well arrives in an unusable format, a promising approach dies and a second has to be stood up. A disciplined price does not pretend this will not happen; it prices a bounded amount for it, on the face of the contract, so that the drift is absorbed without anyone reopening the number.

On the worked Phase-3 build, that line was a 7% contingency: 8,135.40 US dollars carried on top of the discounted role stack. It is a small number by design. A contingency large enough to swallow any conceivable change is just a slush fund that erodes the discipline of the fixed price; one too small is theatre. Seven percent on a research phase is the vendor saying, on paper, "this much surprise is priced in, and beyond it we talk." A related but larger contingency appeared on the commercial engine of the earlier phases at 10%, which tells the buyer something useful: the contingency percentage is itself a negotiable dial, and a vendor that can articulate why it chose 7 rather than 10, or 10 rather than 20, is a vendor that has actually thought about the risk it is pricing rather than padding a round number.

When Phase-2 compute on the same programme ran materially over its budget, roughly 2,600 GPU-hours against a 1,200-hour plan, the overrun did not detonate the contract. It was carried through the contract's own change mechanism as an additional request rather than an emergency renegotiation, precisely because the structure anticipated that research consumes compute unpredictably. This is the part buyers most often get wrong. They treat a contingency as a sign of vendor weakness, a hedge against the vendor's own bad estimate, and negotiate it toward zero. It is the opposite. A named contingency plus a

written change mechanism is what keeps a genuine surprise from becoming a contract crisis, because the paper already says what happens. The contingency line and the change mechanism are two sides of the same discipline: the price is fixed, and the way it is allowed to move is written down in advance, so that movement is a procedure rather than a fight.

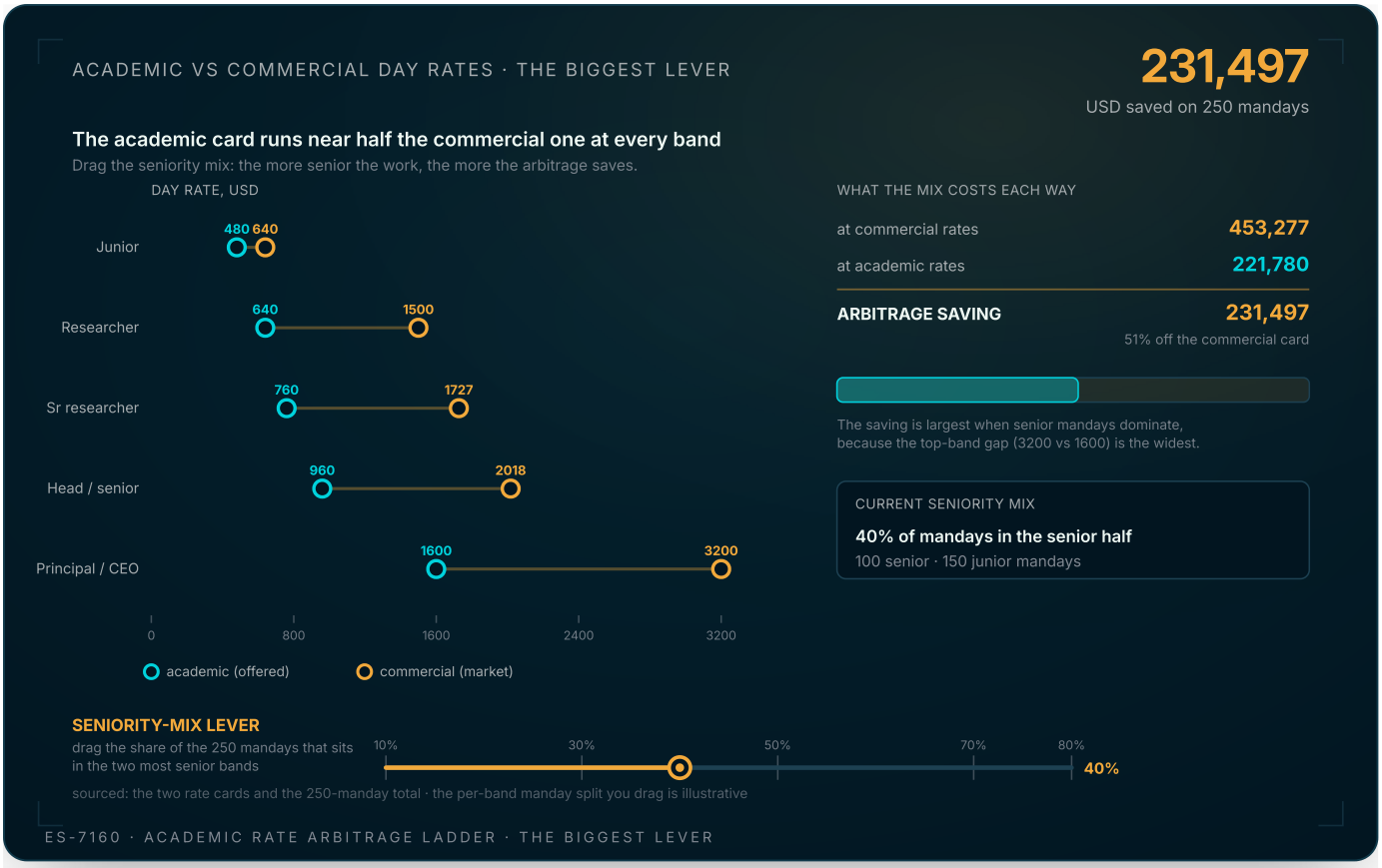
Lever four: whether the biggest recurring cost is fixed or floating

Compute is usually the largest recurring cost in a multi-year AI programme, and the decision to own it or rent it is made first, on day one, when the programme's risk is least understood. Framed as a finance question it looks like capex versus opex, and renting almost always wins the first meeting because it defers the capital and preserves the option to walk. Framed as a risk question it looks different, because renting pins your largest recurring cost to a market you do not control.

We made the owning choice in 2020, through a multi-year hardware partnership that put the cards on our own books and a local data-center deal that gave full cost visibility. Nobody forecast a war. But when the 2022 energy shock arrived, the consequences were measurable and asymmetric. The market for DGX and high-performance-compute hosting moved from about 12,000 to 15,000 US dollars a month before the shock to past 20,000 after it; energy prices rose roughly 400% and electricity bills about 394%. A programme renting its compute would have watched its single largest line climb without a ceiling for the duration. The programme we ran carried a different number entirely: a single bounded, transparent contingency of OMR 20K, roughly 51,000 US dollars, across a twelve-month window, and that was the whole of the shock's effect on our infrastructure cost. The full board-grade reconstruction of that decision, the three-way public-cloud, private-cloud and on-premise cost model behind it, is a pointer to our hardware-moat chapter. For the CFO scorecard the lever reduces to one question: is the largest recurring cost of this programme fixed the day I sign, or is it floating on a market I cannot bound?

Lever five: the arbitrage that is the largest single lever

Of the six, the one that moves the price most is the least visible on the invoice: which rate card the mandays are billed at. On this programme, delivery routed through a university partnership, which unlocked academic day rates that run roughly half the open-market commercial ones across every seniority band.



The single largest lever in an applied-AI research price is which rate card the mandays are billed at. This paired ladder plots the commercial (open-market) day rate against the academic (offered) rate for five seniority bands, from a junior at 640 versus 480 US dollars a day up to a principal at 3200 versus 1600. The gap is roughly half at every band and widest at the top. Drag the seniority-mix lever to shift the 250-manday allocation toward the senior bands and watch the arbitrage saving grow, because the top-band gap is where the discount bites hardest. The orange running figure is the only element that argues: the cash the academic card saves at the chosen mix. The two rate cards and the 250-manday total are sourced from the engagement's commercial workbook and Phase-3 effort sheet; the per-band manday distribution the reader drags is an illustrative allocation, not the exact staffing split.

The paired ladder above is the arbitrage stated plainly. A junior researcher costs 640 US dollars a day commercially and 480 academically; a full researcher 1500 against 640; a senior researcher 1727 against 760; a head-level researcher 2018 against 960; a principal 3200 against 1600. The gap is roughly half at every band and it is widest at the top, which is why the saving grows as the work skews senior. This is not a discount a vendor grants out of goodwill. It is a structural feature of routing an R&D engagement through an academic partner, and it is available to the buyer only if the engagement is set up that way from the start. A CFO who does not ask which rate card is in force is leaving the single largest lever unexamined. The arbitrage also carries a real obligation the buyer should want: academic routing typically comes with a local training and capability mandate, which on this programme meant Omanization was a first-class deliverable, not a footnote, and it connects directly to lever six.

Lever six: paying for capability you keep

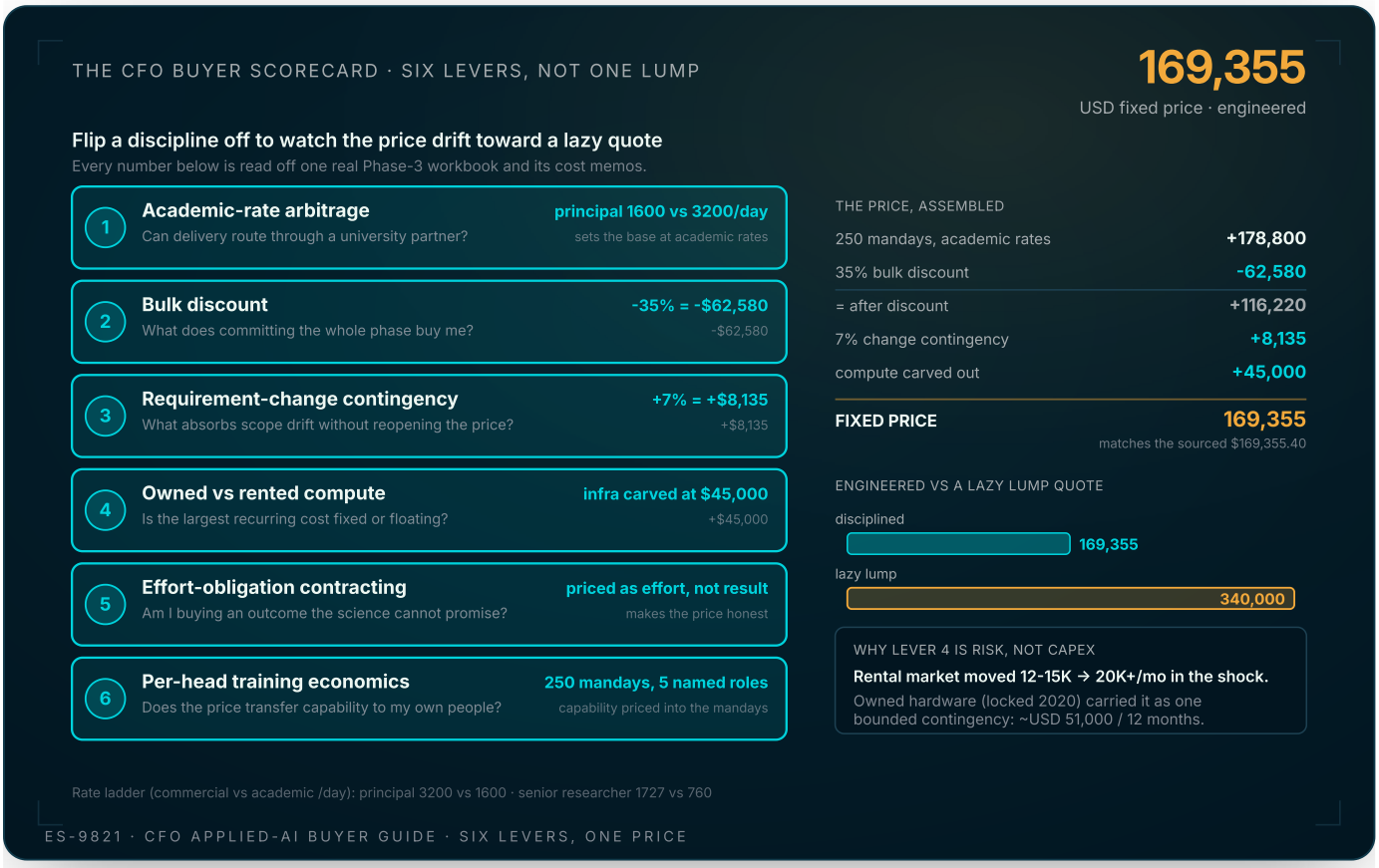
The last lever asks whether the price buys the operator its own future capability or merely rents an outcome. On a well-structured applied-AI engagement, the mandays are not just labour; they are the mechanism by which the vendor's people train the client's people. This programme's year-two agreement was built around exactly that: six live technology-transfer sessions, three online and three onsite, hands-on work with the client's own transformer model, and a two-part MLOps track, so that the operator ended the engagement able to run its own AI rather than dependent on the vendor to run it. That year-two agreement carried a total of 313,000 US dollars over twelve months, payable monthly or quarterly, and a meaningful share of it was capability transfer rather than model-building. Per-head training economics price that transfer into the day rates rather than bolting it on as a separate consulting line.

The distinction the CFO should hold is between two ways to spend the same money. One buys an outcome the vendor delivers and keeps the expertise to reproduce. The other buys the outcome and the operator's ability to reproduce it without the vendor next time. The second is more expensive per unit of immediate output and far cheaper over the life of the capability, because it ends the dependency. The programme's academic routing reinforced this, since a university partnership carries a training and localisation mandate as a matter of course, which on this engagement made Omanization a first-class deliverable rather than a courtesy. And the capability the operator built was theirs to keep: a 51/49 foreground-IP split handed the operator the majority of what was created, a structure treated in full in our chapter on the research-services agreement. The relevant question for the CFO is therefore not only what the mandays produce, but what the organisation can still do the day the vendor's people leave.

That is the whole scorecard. Effort obligation makes the price honest. Milestone tranches make it cancellable. The contingency makes it durable against change. Owned compute makes it stable against macro shock. Academic arbitrage makes it cheap. Per-head training makes it a purchase of capability rather than a rental of output. Now assemble them.

Assembling one worked phase from the six levers

Take the Phase-3 build and reconstruct its fixed price the way the workbook does, lever by lever. Five roles were budgeted across 250 mandays at academic day rates, summing to 178,800 US dollars. The 35% bulk discount removed 62,580, landing the stack at 116,220. A 7% contingency added 8,135.40. Phase-3 infrastructure was carved out and priced on its own at 45,000. The disciplined fixed price is 169,355.40 US dollars.



A finance leader's scorecard for pricing an applied-AI R&D engagement. Six levers, each anchored to a real sourced number, assemble one worked Phase-3 fixed price of 169,355.40 US dollars: a 250-manday role stack summed at academic day rates (178,800), a 35 percent bulk discount (minus 62,580), a 7 percent requirement-change contingency (plus 8,135.40), and a 45,000 US dollar compute package carved out and priced on its own. Toggle any lever off and the running total drifts away from the engineered figure toward the lazy lump a quote-first vendor would carry. Levers 5 and 6, effort-obligation contracting and per-head training economics, do not move the number - they change what kind of number it is, priced as effort rather than a guaranteed result and with capability transfer built into the mandays. The orange running total is the only element that argues. The role-stack sum, the discount, the contingency, the carved-out infrastructure, the academic-vs-commercial rate ladder (principal 3200 versus 1600, senior researcher 1727 versus 760), and the owned-vs-rented compute shock (rental market 12-15K to past 20K per month, absorbed as a single bounded contingency of about 51,000 US dollars over twelve months) are all sourced from the engagement archive; the lazy-lump comparison figure is the same role stack costed at commercial rates with no discipline applied, shown as a contrast and not a quote.

The scorecard above is the argument in one object. Every lever is a card anchored to a sourced number, and the running total on the right reassembles the price as you engage or disengage each discipline. Turn the arbitrage lever off and the base jumps to a commercial rate card. Turn the discount off and 62,580 US dollars comes back. Turn the compute carve-out off and the infrastructure vanishes from a line the buyer should have been shown. With every lever engaged, the total lands exactly on the sourced 169,355.40. Flip any one off and it drifts toward the lazy lump a quote-first vendor would carry, a number that foots to something plausible while hiding which disciplines the buyer is actually getting. The two structural levers, effort obligation and per-head training, do not move the digit at all; they change what kind of number it is, which is the point the whole paper is making. A fixed price is not a quantity to accept or reject. It is an artefact you can take apart, and the six levers are the seams.

The arithmetic of the discount deserves one line of explicit statement, because it is where a buyer most often mistakes a headline for a saving. The discounted price is the stack net of its bulk reduction:

THE DISCOUNTED ROLE STACK

Formula LaTeX

Copy LaTeX

$$178{,}800 \times (1 - 0.35) = 116{,}220 \text{ USD}$$

and the contingency is a bounded percentage of that discounted base, not of the pre-discount sticker:

THE BOUNDED CHANGE CONTINGENCY

Formula LaTeX

Copy LaTeX

$$116{,}220 \times 0.07 = 8{,}135.40 \text{ USD}$$

The order matters. A vendor who applies the contingency to the pre-discount number, or who folds infrastructure silently into the manday rate instead of carving it out, produces a total that looks like the same kind of thing but is not built the same way. Reading the price as an assembled artefact is the only way to catch that.

The same money, read in two currencies and two counterparties

There is one more view a finance leader should insist on, because it exposes whether the commercial structure is real or presentational. The dollar figures above are how the work was priced. The programme was also funded, and the funding view was denominated in Omani rial and split across two counterparties. Behind the roughly 165,000 OMR budget sat a two-party split: the operator-facing delivery firm carried 95,000 OMR, itself split across a Phases 1-2 runway and the Phase-3 build, and the university partner carried 70,000 OMR on the same phased basis. The academic rate card and the training mandate are not floating claims; they are the direct consequence of a real counterparty in the structure whose economics differ from a commercial vendor's.

Why should a CFO care about the funding view when the pricing view already foots? Because the two views cross-check each other. A price that is quoted in dollars at academic rates but funded entirely through a commercial entity with no academic counterparty is a price claiming a discount it has no structural right to. The rial split is the evidence that the academic-rate lever was genuine, that a university was actually in the delivery chain and actually bore part of the budget, and that the arbitrage the buyer was charged reflected a real cost structure rather than a marketing line. When you ask which rate card the people are billed on, the honest follow-up is: and who, structurally, is billing at that card? If the answer is a partner whose economics justify the rate, the discount is durable. If the answer is a commercial vendor simply asserting a lower number, it is a figure that can evaporate at renewal. The two-party, two-currency reading is how a finance leader tells the two apart.

The questions, as a buyer would actually ask them

The scorecard translates into a short list a CFO can carry into any applied-AI R&D vendor conversation. Each question maps to a lever, and each has a right kind of answer.

- Is this priced as an effort or as a guaranteed result? On genuine research, "effort" is the honest answer; a vendor promising a specific accuracy at signing is either naive or pricing in a payout you will fund.
- How is payment phased, and where are the gates? You want tranches tied to phase boundaries with exit ramps, so your exposure at any moment is what is sunk, not the whole

total.

- What absorbs a requirement change without reopening the price? You want a bounded contingency line on the face of the contract, in the single digits of percent, plus a written change mechanism for anything beyond it.
- Is the largest recurring cost, compute, fixed or floating? You want to know whether an energy or hosting shock lands on you as an open-ended bill or as a bounded, pre-agreed contingency.
- Which rate card are the people billed on? An academic or partnership rate card can halve the day rates; if it is available, the engagement should be structured to capture it.
- Does the price leave my organisation with owned capability? You want per-head training and a majority IP position priced in, so the money buys a capability you keep rather than a dependency you renew.

A vendor who answers all six cleanly is quoting a price you can defend. A vendor who cannot, or who insists the single lump sum is the only sensible unit of purchase, is asking you to approve a summary you are not allowed to inspect.

Where this leaves the finance leader

The temptation with applied AI is to treat the commercial structure as plumbing and the model as the substance. On this programme the opposite was closer to true. The model was uncertain until late; the accuracy curve was genuinely unknowable at signing. What made the programme fundable, survivable and ultimately ownable was the commercial structure: effort obligation instead of an unpriceable warranty, phased tranches instead of a single irreversible commitment, a bounded contingency instead of an open reopening clause, owned compute instead of a floating rental, an academic rate card instead of the open-market one, and per-head training instead of a rented outcome. The 169,355.40 US dollar phase and the 345,581.60 US dollar year are not interesting because of their size. They are interesting because each is an assembled artefact whose seams a buyer can inspect, and inspecting the seams is the whole job.

Price the six levers, not the lump. That is the field guide.

Limitations

This is a synthesis of one programme's commercial record, not a market survey. Every number is sourced from the engagement's own workbooks, cost memos and signed agreements: the 250-manday Phase-3 stack at 178,800 US dollars, the 35% discount removing 62,580, the 7% contingency of 8,135.40, the 45,000 infrastructure carve-out and the 169,355.40 total; the five year-1 tranches summing to 345,581.60; the commercial-versus-academic day-rate ladder; and the owned-versus-rented compute figures around the 2022 shock. Those are specific to a subsurface computer-vision engagement routed through a university partnership in Oman, and the exact percentages, rate cards and tranche sizes will differ in other sectors, currencies and partnership structures. The lazy-lump comparison in the scorecard is the same role stack costed at commercial rates with no discipline applied; it is a contrast to make the levers visible, not a competing quote anyone issued. The seniority-mix distribution the arbitrage ladder lets a reader drag, and the even horizontal spacing of the tranche gates, are presentation choices flagged in each instrument, not data claims. What generalises is the method, reading a fixed price as an assembled artefact and asking the six lever-questions, rather than the specific dollars.

References

Determined AI, 2021. Cloud versus on-premise total cost of ownership for deep-learning infrastructure. The public TCO framing referenced in the engagement's own three-way cost comparison behind the owned-versus-rented compute lever.
https://www.determined.ai/blog

Intercontinental Exchange (ICE), 2022. European natural-gas and electricity forward benchmarks through the 2022 energy shock. The market source cited in the programme's war-time contingency memo for the roughly 400% energy and 394% electricity moves.
https://www.ice.com

EARTHSCAN

Buying Applied-AI Research: A CFO's Field Guide to Pricing, Phasing, and De-Risking an R&D Engagement

Published September 2026. Generated 2026-09-08.
https://www.earthscan.io/whitepapers/buying-applied-ai-research-a-cfos-field-guide-to-pricing-phasing-de-risking
Copyright EarthScan. All rights reserved.