

How to Paper an AI R&D Engagement

The standard commercial instinct for a paid engagement is to buy an outcome: define the deliverable, fix the price, and hold the vendor to a result. That instinct is exactly wrong for AI research and development, where the deliverable is a number nobody can commit to at signing. We reconstruct the contract we papered for a multi-year subsurface computer-vision programme with a major operator in Oman, and read every clause against the programme's own measured trajectory. Fracture-recall accuracy at a 5 cm depth tolerance climbed from roughly 10% at three labelled wells to 40% at eight to 75% at sixteen, a curve that was unknowable when the ink dried. 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 all the way there, rewarding safe, unfalsifiable claims over honest experiments. The architecture we used instead prices the work, not the outcome: an explicit effort obligation ('not responsible for achieving a specific result'), five tranches gated to phase boundaries, a data-timeliness clause that shifts delay cost onto whoever owes the data (which on this engagement was the client, whose labelled wells arrived five, then six, then eight of twenty-five across the year), provider-retained intellectual property with client indemnity on supplied data, a liability cap of USD 10,000 with all indirect damage excluded, invoicing every five months on 14-day terms, a two-year term, and provider phase-exit ramps after Phase 1 or Phase 2 on short notice. Three interactive instruments let a reader walk the argument: the map of the two contract modes against the science, the ledger of who absorbs a delay the provider did not cause, and the ramp that shows why bounded downside, not a performance promise, is what makes uncertain R&D signable at all.

Tarry Singh

Founder & CEO

Quamer Nasim

ML Research Engineer

AI-RD-CONTRACTS / EFFORT-OBLIGATION / RESULT-OBLIGATION / PHASE-GATED-PAYMENT / DATA-TIMELINESS-CLAUSE / LIABILITY-CAP

CONTENTS

01	The number nobody could promise
02	Five gates on a two-year clock
03	Why a guaranteed result pays for the wrong behaviour
04	Effort, stated in the plainest possible words
05	The data-timeliness clause, and who really caused the slip
06	The other clauses that hold the shape
07	The economics of the effort structure, stated once
08	What a reader should take from this into their own negotiation
09	Limitations
10	References
11	Get the full whitepaper

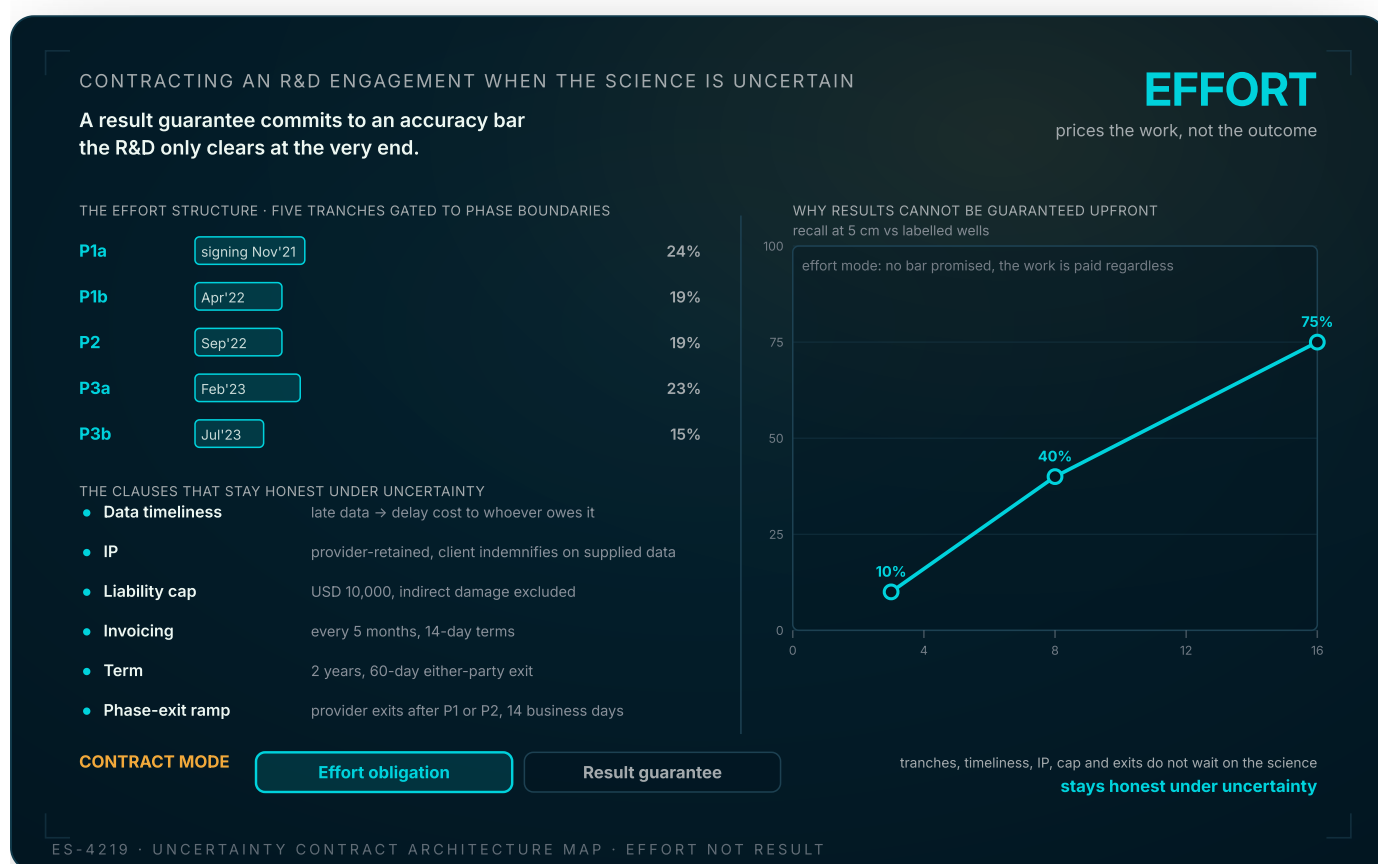
A client asked us, part-way through negotiating a multi-year subsurface-AI programme, to guarantee an accuracy number. It was a reasonable request in the way that most wrong requests are reasonable. They were paying real money; they wanted to know what they were buying; a percentage on a page is easier to defend to a board than a promise to try hard. The trouble is that the number they wanted guaranteed did not exist yet, and would not exist until the research had run its course. Guaranteeing it would have been guaranteeing the weather.

This whitepaper is about the contract we signed instead, and why its shape is not a lawyer's decoration but the load-bearing structure of the whole engagement. The work was a computer-vision programme to detect fractures, bedding planes, and vugs on borehole image logs from a fractured carbonate reservoir, run with a major operator in Oman and an academic partner. The technical story lives in other pieces. This one is about the commercial architecture: an effort obligation rather than a result obligation, five payment tranches gated to phase boundaries, a data-timeliness clause that moves delay cost onto whoever owes the data, provider-retained intellectual property with a client indemnity, a tight liability cap, and phase-exit ramps between the phases. The claim running through all of it is narrow and, we think, correct: result-guarantees on AI R&D create perverse incentives, and an effort obligation with phase-gated tranches, a data-timeliness clause, and phase-exit ramps is the only structure that stays honest when the science is genuinely uncertain.

The number nobody could promise

Start with the thing the client wanted guaranteed, because its behaviour over the programme is the entire case. The models detect sinusoids on the unrolled borehole image, and the headline quality metric is recall at a fixed depth tolerance: of the fractures a human interpreter picked, how many did the model also find, counting a hit only if it landed within a few centimetres of the true depth. At a 5 cm tolerance, fracture recall climbed from roughly 10% when the model had three labelled wells to train on, to 40% at eight wells, to 75% at sixteen wells. The same shape shows up in the controlled comparison: stepping the fracture model from nine to fourteen wells moved F1 at 5 cm from 40% to 75%, with dip-within-three-degrees rising from 70% to 91%.

That curve is the argument. At the moment of signing, with three wells in hand and a model producing 10%, no one on either side of the table could honestly say whether the programme would end at 75%, at 50%, or stall. It was a research question with a research answer, and the answer arrived in monthly increments as the data and the method matured. A result guarantee written at kickoff is a bet on the endpoint of that curve, placed before the first data point is in. Worse, the curve was not even monotone in the naive direction: adding a fifteenth well whose interpreter had labelled in a different style dropped F1 at 5 cm from 60% to 57%. More data hurt. If a contract had pinned the fee to hitting a fixed bar, that single well would have turned a genuine scientific finding into a financial default.



Two ways to paper the same R&D engagement, read against the science it actually produced. The recall curve on the right is measured: at a 5 cm depth tolerance, fracture recall climbed from 10% at three labelled wells to 40% at eight to 75% at sixteen. Toggle to Result guarantee and the orange bar marks the accuracy a client would want promised at kickoff, met only at sixteen wells and unknowable at signing, which is the perverse incentive the piece is about. Toggle to Effort obligation and the left column is what gets paid instead: five payment tranches gated to phase boundaries (Nov'21, Apr'22, Sep'22, Feb'23, Jul'23), a data-timeliness clause that shifts delay cost to whoever owes the data, provider-retained IP with client indemnity on supplied data, a USD 10,000 liability cap, invoicing every five months on 14-day terms, a two-year term with 60-day either-party exit, and provider phase-exit ramps after Phase 1 or Phase 2 on 14 business days notice. The five phase gates and every legal term are sourced from the draft service agreement; the recall figures are sourced from the progress-report bundle; the tranche percentages are illustrative anonymised shares, not amounts, and the whole map is a structure diagram, not a price.

The map above puts the two contract modes on the same axis as the science. Toggle to the result guarantee and the orange bar sits at the accuracy a client would want promised, met only at the last data point and unknowable at the first. Toggle to the effort obligation and the left column is what actually gets paid: the tranches, the clauses, the caps, the exits, none of which wait on the outcome the science cannot promise. The instrument is not decoration for the argument; it is the argument, drawn against the programme's own record.

Five gates on a two-year clock

Before the incentive theory, look at the schedule the contract actually imposed, because the shape of the payment matched the shape of the science. The fee was divided into five tranches, and each was tied to a phase boundary rather than a calendar date pulled from the air: a first tranche at signing in November 2021, a second in April 2022, a third in September 2022, a fourth in February 2023, and a fifth in July 2023. The programme itself was organised into three phases, Dataset Development, machine-learning operations on the image logs, and well-to-well correlation and integration, with the first and third phases each split into two funded parts. Invoices went out every five months on 14-day terms, the whole thing sat inside a two-year term, and the provider held a payment-suspension right if the client fell more than fourteen days behind.

None of that is arbitrary. Gating the fee to phase boundaries does something a lump-sum or a pure time-and-materials arrangement cannot: it makes each commitment short enough to price honestly and long enough to produce evidence. At the November 2021 signing, nobody was asked to commit to the July 2023 endpoint; they were asked to fund Dataset Development and to look again at the next gate. By April 2022 there was real data on how the first phase had gone, and the second tranche was a decision made with that evidence in hand rather than a promise made in its absence. The structure turns one large, unpriceable bet into five smaller, sequential ones, each taken with more information than the last. That is the correct response to an incomplete contract: not to pretend you can specify the whole future, but to build in the points where you will look again.

The payment-suspension right is the small clause that keeps the whole schedule enforceable. A phase-gated structure only works if the gates are real, and a gate is only real if falling behind on payment has a consequence. Fourteen days is short enough to matter and long enough to absorb an ordinary administrative delay, and pairing it with the

provider's separate right to suspend when data is late makes the two obligations symmetric: money late, work pauses; data late, work pauses; neither side can starve the other of what it owes while still demanding delivery.

Why a guaranteed result pays for the wrong behaviour

The obvious objection to an effort obligation is that it sounds like a licence to be lazy. If the provider is paid for trying rather than for succeeding, what stops them from trying badly? The answer is that the guaranteed-result alternative is worse on exactly this axis, and the reason is well understood in contract theory. When you attach a strong payment incentive to a metric that is hard to measure and partly outside the agent's control, you do not buy more of the thing you want. You buy more of whatever is easy to measure, and you buy distortion everywhere else.

On an R&D programme the distortions are specific and ugly. A provider on the hook for a fixed accuracy number has every reason to negotiate the measurement rather than improve the model: to widen the depth tolerance until the recall clears the bar, to quietly drop the hard wells from the test set, to define a fracture such that the easy ones count and the ambiguous ones do not. Every one of those moves raises the reported number and lowers the actual value of the deliverable, and every one of them is rational under a result guarantee. We saw the temptation in miniature and resisted it precisely because the contract did not reward it: a correctly-shaped sinusoid predicted at the wrong depth was scored as a false positive, the permissible human-error relaxations on depth, dip, and azimuth were to be agreed with the client's own expert interpreter rather than set unilaterally, and the honest metric was the one written down. That discipline is cheap to hold when nothing financial rides on gaming it. It is expensive to hold when the fee depends on the number.

The formal version of this is old. When a task has multiple dimensions and some are far more measurable than others, high-powered incentives on the measurable dimension pull effort away from the rest [2]. Fracture detection is exactly such a task: recall at one tolerance is measurable, but the value the geologist actually needs is spread across depth accuracy, dip and azimuth fidelity, behaviour in the crowded intervals that matter for flow, and honest uncertainty in the intervals where the image is ambiguous. Guarantee the one number and you deform all the others. Pay for competent, documented effort against an honestly-defined metric, and the incentive to cheat the measurement simply is not there.

There is a second, structural reason the guarantee misfires, and it comes straight from the economics of incomplete contracts. A contract over an uncertain future cannot enumerate every state of the world; it is necessarily incomplete, and the question becomes who holds the right to decide when the unwritten states arrive [1]. A result guarantee pretends the contract is complete, that the endpoint is known and only the effort to reach it is in doubt. Reality then forces renegotiation the first time the science surprises everyone, which on this programme was roughly monthly. An effort obligation with phase gates builds the renegotiation points into the structure from the start, so the surprises are handled at planned boundaries rather than as contract breaches.

The surprises were not rare. One month the team found that jumping to a much larger overlapped-and-augmented training set actually degraded the model, and the fix was to double the patch stride and generate less synthetic data, not more, as new wells arrived. Another month a change of interpretation model for depth localisation improved one metric while badly hurting dip and azimuth, so the earlier approach was retained. A guaranteed-result contract would have treated each of these as a schedule failure to be litigated; the effort obligation treated them as exactly what they were, which is what research findings look like from the inside. Every one of them was written down in a monthly signed report, and every one of them was progress precisely because it was an honest negative result rather than a hidden one. That is the behaviour the contract structure has to protect, because it is the behaviour that eventually produced the 75%.

Effort, stated in the plainest possible words

The contract said what it was. The operative sentence, in the draft service agreement, is that the provider is "not responsible for achieving a specific result," that the engagement is a research project. That is not weasel language; it is the honest description of the transaction, and putting it in plain words at the top of the agreement does more work than any downstream clause. It sets the client's expectation correctly before the first invoice, it frames every progress report as evidence of work rather than a scorecard against a promised bar, and it makes the whole rest of the structure coherent.

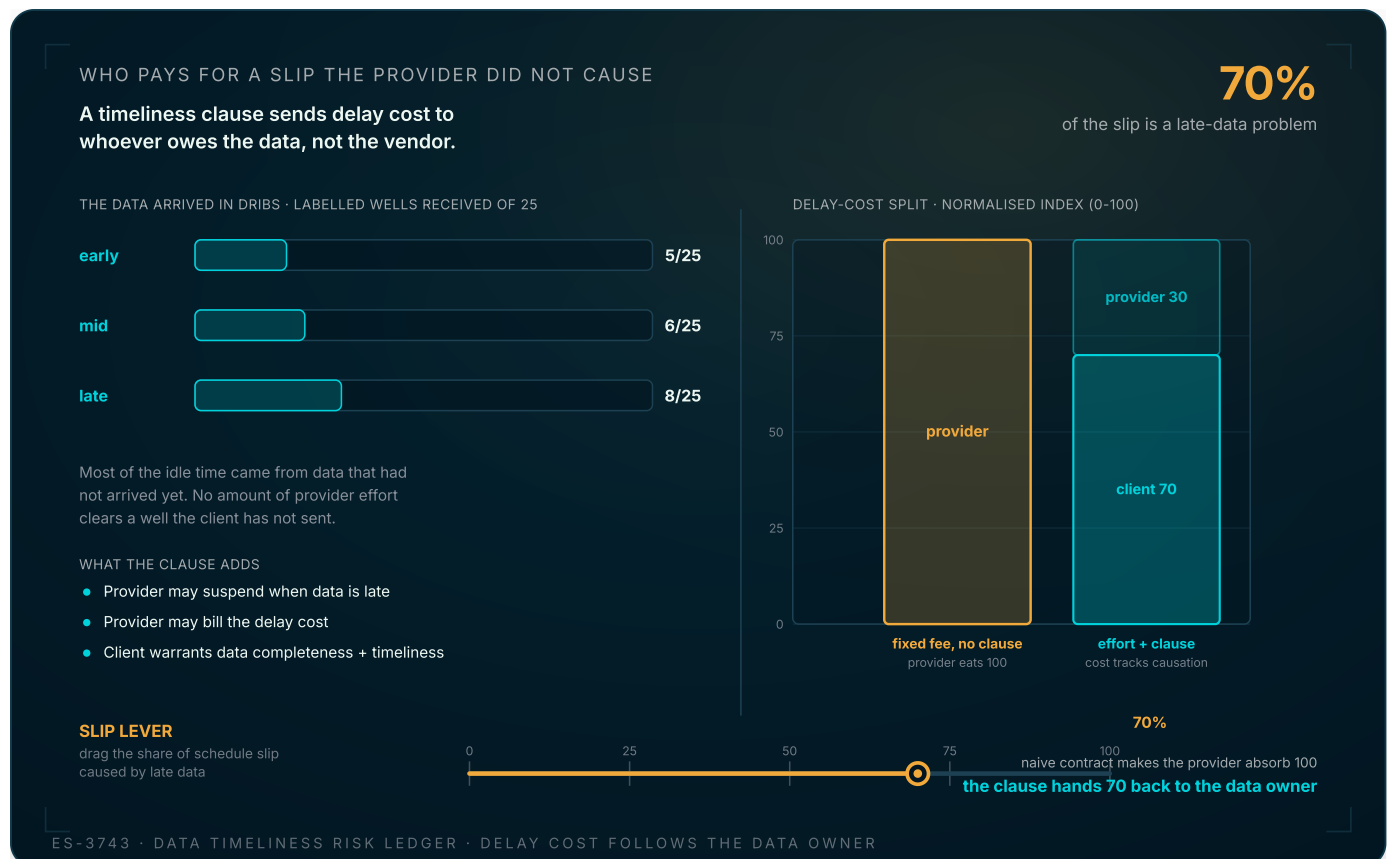
An effort obligation is not a lower standard than a result obligation. It is a different, and for research a more demanding, standard. The provider still owes competent, professional, documented work: the monthly signed progress reports, the reproducible training configurations, the versioned datasets with recorded quality-control exclusions, the

ablations that show each design choice earning its place. What the provider does not owe is an outcome the physics might refuse to deliver. The line between those two is exactly the line between a professional-services engagement and a bet, and R&D belongs firmly on the services side of it.

The data-timeliness clause, and who really caused the slip

The single clause that earned its place most visibly was the one governing data timeliness. It reads as boilerplate risk allocation: the customer warrants the correctness, completeness, and timeliness of the data it supplies, the provider may suspend services if data arrives late, and the provider may bill the resulting delay costs. On this engagement it was the opposite of boilerplate. It was prophetic.

The programme was starved of data for long stretches, and the shortage was on the client's side of the fence. Labelled wells arrived in dribs. The data-management dashboard recorded five wells of twenty-five received at one point, six of twenty-five at another, eight of twenty-five months later. The model's accuracy curve was gated not by the provider's effort but by how many wells had actually shown up, because the dominant lever on this problem was labelled wells, not model capacity. No amount of provider diligence clears a well the client has not sent, and the field evidence across machine-learning projects is that data availability, not modelling cleverness, is what sets the timeline [4].



Who pays for a delay the provider did not cause. On this engagement the labelled data arrived slowly, five then six then eight wells of twenty-five across the year, so most of the idle time was a supply problem no amount of provider effort could clear. Drag the lever to set the share of schedule slip caused by late data. The left bar is a naive fixed-fee result contract with no timeliness clause: the provider silently absorbs the entire delay cost (the orange bar that should not exist). The right bar is the effort obligation with a data-timeliness clause, which splits the same cost by causation, so the client carries the delay it caused by owing the data and the provider carries only its own. The data-receipt counts and the clause terms (suspend on late data, bill the delay cost, client warrants completeness and timeliness) are sourced from the engagement archive; the delay-cost figure is a normalised 0-100 index, not an amount, and the share-of-slip lever is a reader input.

The ledger above makes the allocation concrete. Drag the share of schedule slip caused by late data and watch who absorbs the cost under two clause designs. A naive fixed-fee result contract with no timeliness clause silently parks the entire delay cost on the provider, who eats the idle weeks and still owes the deliverable. The effort obligation's data-timeliness clause splits the same cost by causation, so the party that owed the late data carries the delay it caused. That is not the provider protecting itself at the client's expense; it is the contract tracking reality. The delay was real, it had a cause, and the cause was a data-supply problem the provider could not fix by working harder.

Two things make this clause fair rather than one-sided. First, it is symmetric in spirit: the provider owes competent work on the data it has, and the client owes the data on time; each is on the hook for its own side. Second, it converts an invisible, unpriced risk into a visible, priced one. Without the clause, the delay cost does not vanish; it just lands on whoever failed to write the clause. With it, both parties know at signing who pays for a slip and why, which is precisely the certainty a procurement lead should want even when the accuracy number cannot be promised.

The other clauses that hold the shape

Around the effort obligation and the timeliness clause sit four more terms, each of which looks like standard commercial hygiene and each of which is doing specific work in an R&D context.

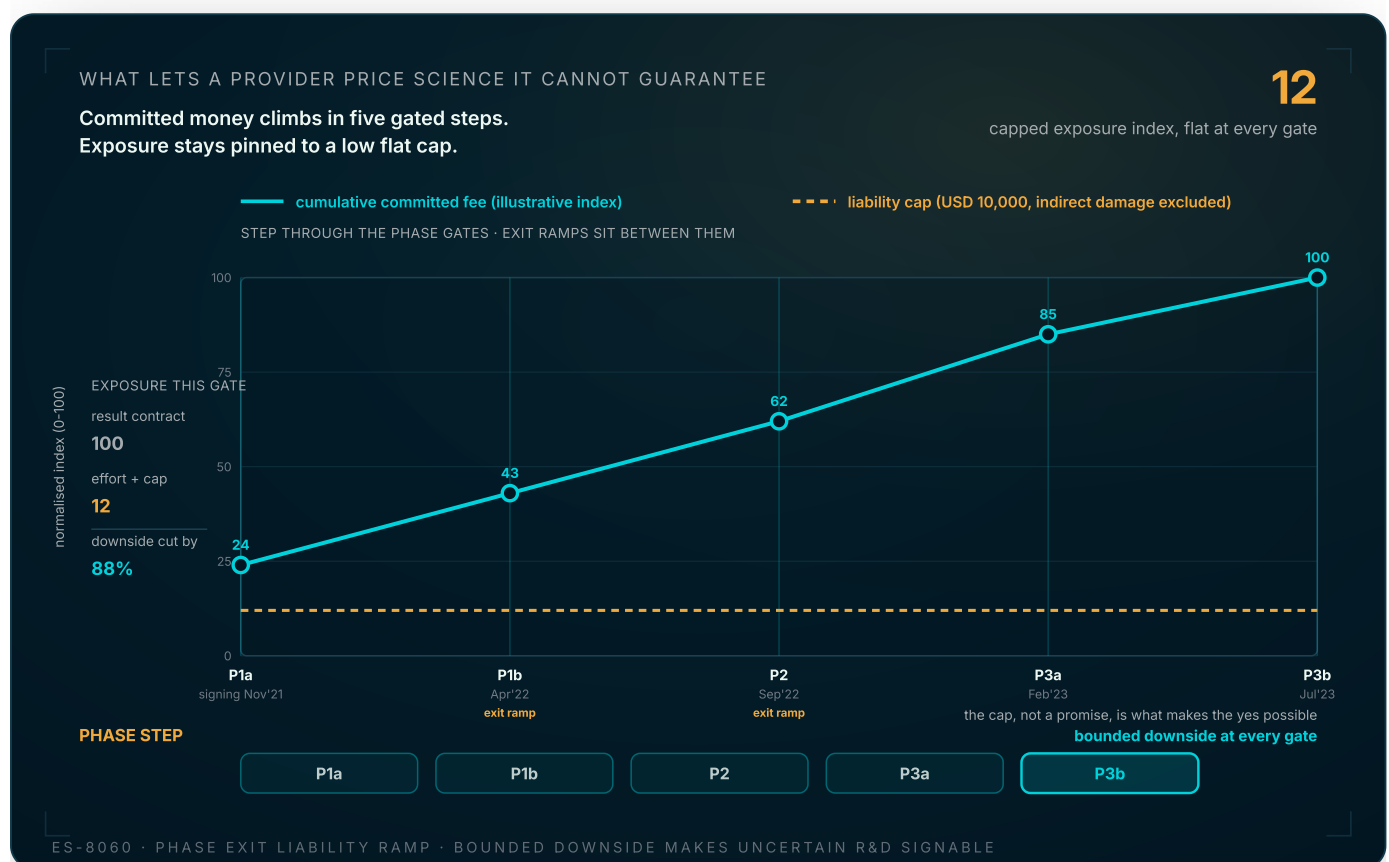
Provider-retained intellectual property, with a client indemnity on supplied data. The provider retains the intellectual property in everything it develops unless agreed otherwise in writing, and the client indemnifies the provider against third-party IP claims arising from the data the client supplied. This matters more for research than for delivery work, because the value of an R&D engagement is disproportionately in the reusable method rather than the single instance. A detector architecture, an augmentation recipe, a training pipeline: these are the assets that let the provider do the next engagement better, and a contract that quietly assigned them to the client would price the provider out of ever doing research at all. The indemnity is the mirror image, and it is fair for the same reason the timeliness clause is fair: the client warrants and stands behind the thing it controls, which is its own data.

A tight liability cap with indirect damage excluded. Liability is capped at USD 10,000 and all indirect and consequential damages are excluded. On a programme where a dip prediction eventually feeds a structural model that feeds a drilling decision, uncapped consequential liability would be uninsurable and, for a research provider, existential. A team will not honestly explore a method that might fail if the failure carries unbounded downstream liability; they will hedge, over-claim, and refuse the hard questions. The cap is what makes candour affordable.

Phase-gated tranches, invoicing, and term. The fee was split into five tranches, each tied to a phase boundary, with invoicing every five months on 14-day payment terms, a two-year term, and a payment-suspension right if the client fell more than fourteen days behind.

The gating is the mechanism that turns a long, uncertain programme into a sequence of shorter, bounded commitments. At each boundary both parties can look at the evidence and decide whether the next phase is worth funding, which is the incremental-commitment structure that an incomplete contract over uncertain science actually needs.

Phase-exit ramps. The provider could exit after Phase 1 or Phase 2 on fourteen business days notice, and either party could terminate the whole agreement on sixty days notice. These are not escape hatches for bad behaviour; they are the off-ramps that make the on-ramp signable.



Why bounded downside, not a performance promise, is what makes uncertain R&D signable. Step through the five phase gates (P1a at signing in Nov'21, then Apr'22, Sep'22, Feb'23, Jul'23) and two lines move: the cumulative committed fee climbs in gated steps (teal), while the provider's liability exposure stays pinned to a low flat cap (the orange dashed line, USD 10,000 with all indirect and consequential damage excluded). A single fixed-fee result contract would expose the provider to the whole committed value if the science fell short; the effort obligation caps that exposure and adds exit ramps, so the provider can leave after Phase 1 or Phase 2 on 14 business days notice and either party can terminate on 60 days. The phase gates, the exit ramps, the 60-day termination, and the USD 10,000 liability cap are sourced from the draft service agreement; the cumulative committed fee is an illustrative normalised 0-100 index, not an amount, and the cap is drawn as a flat exposure floor rather than a fraction of the fee.

The ramp above steps through the five phase gates and plots two things at once: the cumulative committed fee climbing in gated steps, and the provider's liability exposure staying pinned to a low flat cap. A single fixed-fee result contract would run the exposure line up with the fee, so that by the final phase the provider is betting the whole programme value on the science landing. The effort obligation flattens that exposure and adds the exit ramps, so the downside is bounded at every gate. The point the instrument argues is the one a general counsel should internalise: it is the bounded downside, not any promise about accuracy, that makes the provider able to say yes to research it cannot guarantee.

The economics of the effort structure, stated once

It helps to write the incentive comparison down rather than assert it. Model the provider's expected payoff under the two regimes. Under an effort obligation the provider is paid a fee for competent work, so the expected payoff is the fee less the cost of that work, independent of the realised accuracy. Under a result guarantee the provider is paid the fee only if the realised accuracy clears a bar, and pays a penalty otherwise, so the expected payoff weights the fee by the probability of clearing the bar and subtracts the penalty times the probability of missing it.

PROVIDER EXPECTED PAYOFF UNDER A RESULT GUARANTEE VERSUS AN EFFORT OBLIGATION

Formula

LaTeX

Copy LaTeX

$$\mathbb{E}[\pi_{\text{result}}] = p(a \geq \bar{a}) \cdot F - (1 - p(a \geq \bar{a})) \cdot L - c \quad \mathbb{E}[\pi_{\text{effort}}] = F$$

The difference between the two is driven entirely by the probability that the realised accuracy a clears the bar $\bar{a}$, a probability that on this programme swept from near zero at three wells to high only at sixteen. When that probability is both low and unknowable at signing, the result-guarantee payoff is dominated by the penalty term L , and the rational provider does one of two things: refuses the work, or accepts it and then manages the measurement rather than the model. Neither serves the client. The effort payoff $F - c$ has no such term, so the provider's rational move is simply to do the work well. The equation is not a flourish; it is the whole perverse-incentive argument compressed into one line, and it is why the structure is not merely fairer but produces better science.

What a reader should take from this into their own negotiation

The specifics here are ours, but the pattern generalises to any engagement where a client is buying research rather than delivery. Three tests separate a contract that will stay honest from one that will not.

First, does the agreement price the work or the outcome? If the fee is contingent on hitting a performance number that depends on data the vendor does not control and science nobody can predict, the contract is a bet dressed as a purchase order, and it will pay perverse incentives until it is renegotiated in acrimony. Price the work, name the effort obligation in plain words, and hold the vendor to competent, documented, reproducible practice against an honestly-defined metric.

Second, does the agreement put each risk on the party that controls it? Late data is the client's risk if the client owes the data; method risk is the vendor's; IP in the developed method belongs with whoever will carry it forward; downstream-decision liability should be capped because it is unbounded and uninsurable. A contract that allocates each risk to its controller is stable under surprise. One that dumps every risk on the vendor because the vendor happened to sign as vendor will break the first time the science does something interesting.

Third, does the agreement build its renegotiation points into its structure? Phase gates, tranche boundaries, and exit ramps are how an incomplete contract over uncertain science handles the states of the world it could not enumerate at signing. They convert the inevitable surprises from breaches into scheduled decisions. An R&D contract without them is pretending to a completeness it does not have, and the pretence is expensive.

One more practical note for anyone redlining a draft that arrives as a result guarantee. The change from a result obligation to an effort obligation is usually a small edit in wording and a large change in meaning, and it tends to reassure clients rather than alarm them once it is explained. The sentence that says the provider is not responsible for achieving a specific result reads, at first, like the vendor backing away from accountability. It is the opposite. It is the vendor refusing to sell a promise it cannot keep, and offering instead the thing it can stand behind: documented, reproducible, professionally competent work, gated so the client can stop at any boundary, capped so a downstream surprise cannot bankrupt

anyone, and priced so the science stays fundable while it is still uncertain. A client who understands that they were about to buy a bet, and are instead buying research done properly, is a client who signs with more confidence, not less.

The programme this contract governed ran its course, produced the accuracy curve that no one could have guaranteed at the start, and ended without a commercial dispute. That absence of dispute is the quiet evidence that the structure worked. The science was uncertain from the first day to nearly the last, exactly as research is supposed to be, and the paper it was written on was built to stay honest through the uncertainty rather than to deny it.

Limitations

This is a synthesis of one engagement's contract architecture, anonymised, with no amounts disclosed; the tranche shares in the instruments are illustrative proportions rather than the real split, and the delay-cost and committed-fee figures are normalised indices, not currency. The legal terms described are from a draft service agreement reviewed during negotiation, and a specific engagement's final executed terms and governing law may differ; nothing here is legal advice, and any real contract should be drafted and reviewed by qualified counsel for the relevant jurisdictions. The perverse-incentive argument is grounded in established contract theory [1][2] and in this programme's measured accuracy trajectory [3], but the quantitative payoff model is a stylised illustration, not an estimated structural model, and the probabilities it references were not elicited formally at signing. The claim that data availability, not modelling, dominated the timeline reflects this engagement and the broader field evidence [4]; it will not hold for every AI programme, and engagements where the vendor controls the data supply should allocate the timeliness risk differently. Finally, the accuracy figures are recall and F1 at stated depth tolerances on a specific fractured-carbonate dataset; they are evidence that this outcome was unguaranteeable, not a general benchmark for what such models achieve.

References

- 1 Hart, O., Moore, J. (1988). *Incomplete Contracts and Renegotiation*. *Econometrica*. <https://www.jstor.org/stable/1911259>

- 2 Holmstrom, B., Milgrom, P. (1991). *Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design*. Journal of Law, Economics, and Organization. <https://www.jstor.org/stable/764957>
- 3 Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S. (2020). *End-to-End Object Detection with Transformers*. ECCV. <https://arxiv.org/abs/2005.12872>
- 4 Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., Zimmermann, T. (2019). *Software Engineering for Machine Learning: A Case Study*. ICSE-SEIP. <https://ieeexplore.ieee.org/document/8804457>

Get the full whitepaper

This page is the long-form summary. The complete whitepaper adds the clause-by-clause annotated structure, the full phase-boundary schedule with the tranche logic, the data-supply timeline that made the timeliness clause load-bearing, and a redlining checklist for turning a result-guarantee draft into an effort obligation without losing the client's confidence.

Contracting for Uncertainty: How to Paper an AI R&D Engagement

Published February 2026. Generated 2026-07-22.

<https://www.earthscan.io/whitepapers/contracting-for-uncertainty-how-to-paper-an-ai-rd-engagement>

Copyright EarthScan. All rights reserved.