

The Responsibility Frontier

Why hyperscalers are descending toward the result, and why the return waits for them where they control only a subset of the factors

For forty years the software industry answered for the availability of a means, and the customer answered for the result. That line is moving.

On July 2, 2026, Microsoft announced Frontier Company, a 2.5 billion dollar operating unit that embeds roughly six thousand engineers and industry specialists inside enterprise customers to design, build, and operate their AI systems. The announcement is not the movement. It is its first observable case. The thesis of this article is that the value of an enterprise AI system is not an additive but a multiplicative function of complementary factors, that a single factor reduced toward zero collapses the product, and that the factor which most often sits near zero, the customer's institutional capacity, is precisely the one no vendor can deliver from the outside, because an asset is deployed and a capability is instituted.

Read that way, the interesting question about Frontier is not whether Microsoft misread the problem. It is what happens to a software company when the rent it lives on has moved to a place it does not control, and it follows the rent there anyway. This piece extends an earlier and more immediate reading of the same announcement, and moves deliberately from reaction to the structure beneath it.

From publisher to operator

The vocabulary of the announcement is the tell. Microsoft describes the unit as outcome-driven, promises measurable business outcomes, and grants it its own financial accountability, with engagements judged against business results rather than software delivered. GeekWire, reading the same material, concluded that the initiative is less a new company than a larger and better-branded push behind work Microsoft already did. Both descriptions point at the same displacement. A vendor that proposes to be judged against the customer's result is a vendor proposing to answer for the result.

Situate this in the long line of the industry. The software house sold a licensed artifact and answered for its conformance. The platform sold a substrate and answered for its interfaces. The cloud sold a resource and answered for its availability. The copilot sold an assistant and answered for its suggestions. Frontier proposes the next station: partial accountability for the outcome itself. The displacement is not technological. It is juridical, economic, and organizational, and it is the reason the story matters beyond one quarter.

The human composition of the unit deserves a paragraph and no more, because in five years no one will remember the org chart while everyone will still argue about accountability. The six thousand are drawn primarily from existing engineering and forward-deployed teams. A significant share descends from Industry Solutions Delivery, the group that had itself absorbed the former Microsoft Consulting Services. Of the 2.5 billion dollars, roughly 1.2 billion is reported to cover personnel costs, including relocation and AI-readiness training, which is the language of reskilling existing staff rather than hiring a new army. And the announcement arrived within days of a reduction of about nine thousand roles concentrated in sales and consulting, timed to the opening of the fiscal year. Purged of intention, the pattern is consistent with a reclassification of delivery labor into a higher-accountability unit rather than with the creation of net capacity. Whether the 2.5 billion is new capital or a reallocation, and over what period, has not been disclosed. The gap is worth noting and not worth filling.

Software used to answer for a means. It now proposes to answer for an end. The question is whether the end is a thing a vendor can supply.

The rent migrates because the model commoditizes, and the enterprise is under tension

The migration is probably inevitable, and the reason is economic rather than strategic. The proximate cause is the commoditization of the layer that briefly held the rent, the model itself. The Stanford AI Index reports that the cost of querying a model to reach GPT-3.5-level performance on the MMLU benchmark fell from about twenty dollars per million tokens in late 2022 to roughly seven cents by late 2024, close to two orders of magnitude in eighteen months. Epoch AI finds inference prices falling at a median rate near two hundred times a year, for a fixed level of capability, since the start of 2024. When the architectural spread between the leading models narrows, when APIs converge, and when the price of invoking a given level of intelligence collapses, the model stops being a durable source of economic rent. The rent has to move to the layer that does not commoditize.

That layer is integration, governance, and the customer's own outcome. Across four decades the rent has already moved down a chain: hardware, then the operating system, then the office suite, then the server, then the cloud, then the data, then the model, and now the result. Each displacement stranded the previous rent-holder in a commoditized layer and pushed value one station further toward the customer's outcome. The convergence of the vendors is the evidence, not imitation. Two days before Frontier, Amazon committed a billion dollars to a comparable embedded-engineering initiative. Anthropic and OpenAI had each stood up forward-deployed groups in May. When four capable actors make the same move within weeks, the parsimonious reading is a shared

displacement of rent, not a fashion. Frontier can be read as an adaptation to the migration of value, which is a weaker and more defensible claim than any account of Redmond's intentions.

The symmetry that most analyses miss sits on the demand side. The customer accepts the transfer of responsibility because the customer is also in distress. ManpowerGroup's 2026 talent survey found that AI skills had become, for the first time on record, the hardest category of role to fill, with roughly seventy-two percent of some thirty-nine thousand employers reporting difficulty and France near the top at seventy-four percent. McKinsey has described AI innovation as constrained by the talent shortage, and finds that a large majority of organizations now use AI while fewer than two in five report enterprise-level EBIT impact. BCG apportions about seventy percent of AI success to people, process, and change, and only the remaining thirty to algorithms, data, and infrastructure combined. These are consultancies with an interest in the conclusion that the binding constraint is organizational rather than technical, which is also the argument advanced here, and the figures should be weighed with that alignment in mind. The World Economic Forum, on firmer ground, expects around forty percent of workers' core skills to change by 2030, which compresses the half-life of any competence a firm manages to build. Beneath the talent crisis sits a documentary one. IDC's Global DataSphere grew from thirty-three zettabytes in 2018 toward roughly one hundred seventy-five in 2025, the large majority of it unstructured, even as governed and usable data stayed scarce. The customer is drowning in raw material and starved of the version of it a system can use.

So the encounter is not a vendor descending on a passive, immature client. It is one party in distress meeting another. The hyperscaler arrives with its stock down about twenty-one percent on the year and its capital expenditure under investor scrutiny. The enterprise receives it with a talent shortage it cannot hire its way out of and a data estate it cannot govern at the rate it grows. *Two distresses meet. Neither is a position of strength.* That is a more honest frame than immaturity, and it sets up the real question, which is where the binding factor actually sits.

From constraint to complementarity

The tempting reading, and the one an earlier version of this analysis reached for, is that Microsoft has correctly diagnosed the disease and is acting on the wrong side of a single bottleneck. That reading is wrong. It is not unkind, it is bad theory, and it fails for a specific reason: there is no single bottleneck.

The correct object is not a constraint but a production function, and it is worth being exact about which one, because the precision is the argument. The value of an enterprise AI system is the product, not the sum, of its complementary factors: infrastructure, data, governance, skills, process, adoption, measurement, and incentives. A strict Leontief function of fixed proportions would make those factors perfectly non-substitutable and

set the value at the minimum of the inputs. That is too strong, and it is not the claim here. The right form is the one Michael Kremer formalized as the O-Ring theory of economic development (*Quarterly Journal of Economics*, 108(3), 1993, pages 551 to 575), in which output is the product of task qualities, limited substitution between tasks is possible, and yet the weakest task still dominates the outcome, so that quantity cannot be substituted for quality. The name is owed to the Challenger, lost to a single failed seal while every other component performed to specification. Paul Milgrom and John Roberts gave the organizational counterpart in their analysis of modern manufacturing (*American Economic Review*, 80(3), 1990, pages 511 to 528), where the return to raising one activity appears only when its complements are raised alongside it, and where raising one while the others lag yields little. The distinction between Leontief and O-Ring is not pedantry. It concedes that a vendor can lift a low factor somewhat, and denies that lifting the factors it controls can compensate for a factor near zero that it does not.

Write the value of an AI system as a product across those factors and the consequence follows immediately. A single factor near zero collapses the product, however large the others. Microsoft controls a subset. It can raise infrastructure, and it can raise some of the skills factor by placing capable engineers on site. It does not control governance, it does not control adoption, and it does not control the customer's institutional capacity to run what is built. The inclusion of incentives as an explicit factor marks the outer boundary of the vendor's reach. A firm may hold clean data, sound processes, and trained people, and still extract nothing if its internal incentives contradict one another, when the technology function optimizes for cost containment, the business units for deployment velocity, finance for a near-term profit and loss effect, and compliance for the minimization of risk. No amount of external engineering resolves a misalignment of internal incentives, because the vendor has no standing to redistribute a customer's decision rights. Over-supplying engineering against a near-zero governance or incentive factor does not move the product, because the product is multiplicative and the missing factor is not Microsoft's to raise.

This is why the now-famous statistic, the MIT NANDA finding that ninety-five percent of enterprise generative AI pilots produce no measurable impact on profit and loss, should be read not as a deficiency of the models but as the statistical signature of multiplicative collapse. The models work in the demonstration. The product fails in the business because at least one client-held factor is near zero.

The scarce profile everyone is bidding for is itself one of these factors, not a way around them. The engineer who simultaneously commands LLM operations, the profit-and-loss structure of a regulated sector, and the constraints of the AI Act and the GDPR is rare because the joint distribution of three uncommon skills is rare, and capital does not compress the maturation of a joint distribution. In McKinsey's 2024 estimate, recruitment cycles in regulated industries run roughly seventy percent longer for exactly these

profiles. Amazon, Anthropic, and OpenAI can bid the price of that profile up. None of them can manufacture it in a year. The rent, in the end, accrues to whoever masters the geometry of the complementary factors, not to whoever aligns the largest number of engineers. *You cannot buy your way across a multiplicative production function by over-supplying one input.*

Asset, capability, capacity, institution

What, precisely, does the vendor deliver, and what does that delivery leave untouched? The friction in outcome-based software has an ontological source, and it is clearer once four classes of resource are separated rather than two. An asset becomes a capability becomes a capacity becomes an institution, and a vendor can sell only the first and briefly rent the second.

An asset is a thing. Azure is an asset. Copilot is an asset. A retrieval pipeline is an asset. An agent is an asset. An asset has three properties that make it saleable: it is transferable, it is financeable, and it is deliverable. A vendor can package it, price it, ship it, and book the revenue. A capability is the localized competence to operate and maintain that asset: an engineer's fluency in LLM operations, a team's routine for evaluating a model's output. A capability can be trained, and it can be scaffolded from outside for a time. A capacity is the broader institutionalized state, endogenous to the firm, cumulative over years, and embedded in political and social structure, which is to say non-transferable by contract. It is the organization's collective ability to absorb, govern, and extract value from its capabilities and thereby change its own performance. An institution is the layer above even that: the socio-technical matrix that distributes decision rights, assigns legal accountability, arbitrates internal conflict, absorbs algorithmic error, and manufactures the baseline trust without which none of the rest operates.

A vendor can sell an asset. It can scaffold a capability for the length of an engagement. It cannot sell a capacity, and it certainly cannot sell an institution, for the same reason a gym cannot sell fitness. It can sell access to the equipment and the presence of a coach. The capacity is what the member builds, or fails to build, after the coach goes home. The binding factor of the previous section is a capacity, and therefore it is exactly the thing that cannot be delivered. This is the paradox at the center of the Frontier proposition, stated without malice: Frontier deploys an asset and bills for the institution of a capacity that it cannot itself institute.

The strongest objection must be stated at full strength before it is answered. The forward-deployed model, invented at Palantir and now copied across the industry, rests on the claim that technical immersion is how you build the maturity that is missing. On this view the customer's immaturity is the target, not a mismatch, and to complain that the client is not ready is to blame a fire for the existence of firefighters. A sharper version turns the complementarities argument against itself: if value is multiplicative and Microsoft

understands that, then embedding engineers who also touch the data, the governance, and the process is precisely an attempt to raise several factors at once, so the claim that Microsoft controls only a subset is too strong.

The answer is a distinction inside the objection. A vendor can raise the technical factors, and it can scaffold the others for the duration of the engagement. Scaffolding is not institution. Governance, adoption, measurement, and the arbitration of incentives are organizational and political acts, cumulative and non-transferable, and they decay when the vendor's people leave the building. The cadence makes the point concrete. Frontier engagements are described as forty-five day sprints staffed by five or six people, a rhythm calibrated for the demo-to-deployment gap of lightly regulated workflows. Microsoft's own showcase customers include Novo Nordisk and the London Stock Exchange Group, where qualification, change control, and audit trails do not compress to forty-five days. The velocity advantage evaporates in exactly the regulated terrains the vendor chooses to display. *A capability that lasts only while the vendor remains is not a capability. It is a dependency.* The vendor can install the machinery of reliability. It cannot install the person on the customer's side who would run it.

The stewardship hierarchy and the governance of measurement

Microsoft ships what it calls a trust platform, responsible for governance, security, and the measurement of return. A platform of that kind exposes a set of controls, and a control is only as useful as the role on the other side that operates it. Those roles are human, and most customer organizations are missing them. The proposal that follows is doctrinal rather than empirical, and it should be read as a framework and not a finding.

Five forms of human accountability have to exist on the client side for the delivered platform to connect to anything, and they form a hierarchy rather than a list. A Data Steward answers for the truth, lineage, and compliance of the source data. A Knowledge Steward answers for the truth and currency of the documentary corpus, the difference between a document that exists and a document that is still correct, which is the role the earlier documentary sprawl most directly defeats. An AI Steward answers for the algorithmic layer, the control of drift, the boundary of refusal, and the stewardship of inference cost, which is the meter that runs whether or not the value arrives. A Process Owner answers for the operational and legal liability of the workflow the system informs. Crowning these four is the Decision Owner, who governs the threshold of automated reasoning itself: what level of algorithmic confidence is sufficient to act, when a human is placed back in the loop, and how the cost of an automated error is weighed against the velocity of autonomy. The first four keep the plumbing honest. The Decision Owner decides how much of the enterprise's judgment is ceded to the machine, which in a

regulated setting is the only question that finally matters. The Data Steward is at least a named role, occasionally staffed. The Decision Owner is almost nowhere formalized, which is a precise way of saying that the platform's most consequential control connects, on the customer's side, to no one.

This reverses the usual causal story. The absent steward is not the cause of failure. It is a symptom. The chain runs from the absence of institutional capacity to the absence of the stewards, from the absence of stewards to the absence of governance, from the absence of governance to the absence of measurement, and from the absence of measurement to the absence of any return that can be verified. Diagnosing the missing steward and then hiring one treats the symptom, because the role cannot hold without the capacity underneath it.

Which brings the measurement question to its sharper form. The complaint that Microsoft measures its own return is true but shallow. The consequential fact is that the vendor progressively defines the instrument of measurement itself: the metrics, the observables, the dashboards, the traces. In the economics of organizations, whoever defines the metric orients the behavior that is measured against it. The issue is not self-notation, which is the crude version, but the governance of measurement, which is the consequential one. Marking your own homework is embarrassing. Designing the grading rubric that everyone else will use is structural power, and it is the quieter of the two.

Systemic blind spots, and what we actually know about each

Several second-order effects follow, and intellectual honesty requires attaching a level of certainty to each rather than presenting them at a uniform confidence.

1. The first is channel preemption, and it is documented. Microsoft names a partner ecosystem, Accenture, Capgemini, EY under a five-year alliance, KPMG, and PwC, while embedding its own engineers directly in marquee regulated accounts such as the London Stock Exchange Group and Novo Nordisk. The tension is real but its shape is specific. This is not an industrial substitution of the global integrators, which would be impossible at that scale, but something narrower: capturing the upstream architecture decision on Azure at board level while leaving the heavy industrialization to partners. What is taken is the decision, not the delivery.
2. The second is an asymmetry of intellectual property, and it is structural. The protection of the customer's training data is real and stated at the level of principle, with Microsoft's leadership arguing that there is no societal license for an AI future that consumes the intelligence of the firms it is deployed inside. But the protection is scoped to training data. The methodological learning, the reusable abstractions, and the sector playbooks distilled from each sprint accumulate in the vendor's

cognitive estate across engagements. Each deployment lowers the vendor's own uncertainty-reduction cost on the next one, while the individual customer stays sealed inside its own data walls. The data stays with the client. The method compounds with the vendor.

3. The third is a reconfiguration of the software market, offered here as sectoral analysis rather than fact. By descending through autonomous agents and architecture advice, the hyperscaler threatens the relevance of vertical SaaS publishers whose features are increasingly liable to be absorbed into the Azure AI substrate. The application layer that once sat above the platform is precisely the layer the platform is now learning to generate.
4. The fourth is the loop, and it must be marked as a structural tendency and an explicit hypothesis, not an imputed intent. The deep economic model of the hyperscaler rests on the consumption of compute, which creates an objective divergence between the vendor's function, the generation of consumption, and the customer's, the pursuit of operational efficiency and lower fixed cost. The token economics sharpen this rather than soften it. The per-token price falls, but the total bill rises, because reasoning models and agentic workflows multiply the number of tokens a task consumes, a pattern industry analysts have begun to call the cost paradox. Cheaper units and larger invoices is not an accident for a vendor whose base is consumption; it is the engine. Once an agent reaches production, the chain becomes a loop: usage produces data, data improves the agents, better agents deepen dependence, dependence raises consumption, and consumption produces more data. This is a tendency rather than a demonstrated compounding, and it is a feature of the architecture rather than a hidden motive.
5. The fifth belongs in this section but must be stated as continuity rather than monopoly, and as transitory rather than permanent. The unified graph across SharePoint, Teams, Exchange, OneDrive, Entra, and the Microsoft Graph is what makes enterprise agents tractable at all, because the documentary substrate already exists as a connected whole, and it is difficult to reconstruct quickly. It is not, however, a durable moat. Three forces erode it. European data sovereignty rules and the growth of industrial data spaces restrict the unconstrained centralization of corporate data in a single vendor's cloud. The Model Context Protocol, released by Anthropic in late 2024, adopted within months by OpenAI, Google, Microsoft, and AWS, and donated to the Agentic AI Foundation under the Linux Foundation in December 2025, decouples the integration layer from any single model or data plane, so that the investment in connecting a system is no longer bound to the vendor of the model. And large enterprises are building vendor-neutral semantic layers and graph databases precisely to avoid this class of lock. Microsoft itself adopted the Model Context Protocol into Copilot Studio, which is the quiet admission that the standard eroding the moat is one the incumbent cannot refuse.

Two hypotheses: capacity-as-a-service against the internal factory

The claim that a vendor cannot deliver the outcome from the outside assumes the customer wants, and is able, to build an internal institution. That assumption is where the argument is most exposed, and two futures test it. They are not decorative alternatives. They are the two ways this thesis could turn out wrong, and stating them is the discipline that separates a reading grid from a prophecy.

1. The first future is capacity-as-a-service. A large share of the market may decline to institutionalize at all, and choose instead to rent its cognitive production permanently, treating organizational intelligence as an externally managed utility and converting the accountability deficit into a standing fee. Under this future the distinction between deploying an asset and instituting a capacity dissolves, because the capacity never returns to the customer. It becomes a leased appendage, run by the vendor, priced as a cognitive utility. The talent shortage and the absent stewards are not solved on the client side; they are absorbed by the supplier and billed back. If this is where the market settles, Frontier is not overreaching. It is early.
2. The second future is the internal AI factory. Enterprises may treat cognitive production as a core competence, stand up their own AI platform teams, centers of excellence, and program offices, and reduce the hyperscaler to an infrastructure layer beneath a sovereign orchestration they own. This future is not hypothetical. Gartner projects that by 2026 about eighty percent of large enterprise finance teams will run internally managed and owned generative AI platforms trained on proprietary data, and the center-of-excellence model has become the default organizational answer to scaling AI across a firm. For these enterprises the hyperscaler's intensive intervention is a transient accelerator, not a permanent institutional replacement. The irony sits inside Microsoft's own material: its Cloud Adoption Framework instructs customers to build an internal AI center of excellence, while Frontier Company offers to be that center on their behalf. It is arguing both sides at once, and it cannot know in advance which the market will buy.

The two futures are not mutually exclusive across the market, and the likely settlement is a split by maturity. The mature enterprise, already building its own factory, will use Frontier as an accelerator and repatriate the capability. The immature one, unable to staff the stewards, will slide into permanent rental. Which of the two dominates is precisely what the next eighteen months will reveal.

The window of falsification

Raise the whole matter one level. For forty years software sold a means, and the customer supplied the end. The hyperscalers are now trying to sell the end itself. Software is ceasing to be sold as a product, and even as a service, and is beginning to be sold as a factor of cognitive production, defined here as the institutional capacity to turn organizational knowledge into economically valuable decisions through algorithmically mediated reasoning. If that reading holds, Frontier Company is not the subject of this analysis. It is the first observable case study of it, and the responsibility frontier, having moved from availability toward outcome, will keep moving.

A grid that cannot be refuted is not worth publishing, so here is the condition of refutation, with a date, and along two distinct vectors. The window is twelve to eighteen months.

The first is the autonomy vector. The claim that the binding factor is a non-deliverable capacity will be falsified if Frontier demonstrates, in a public and documented way, a durable impact on profit and loss inside customers that began data-immature and without internal stewardship, and if the human accountability structures, the several stewards up to the Decision Owner, survive and become autonomous after Redmond's engineers withdraw. Two observations govern this vector, and both are visible. One is whether the customers institute a capacity or retain a dependency that decays at exit. The other is whether the customers come to define their own metrics, observables, and traces, or continue to be measured by the instrument the vendor supplied.

The second is the institutionalization vector, and it is the concession the earlier drafts of this argument did not make. The thesis will also be falsified if the market embraces capacity-as-a-service in durable form, that is, if an infrastructure vendor proves able to manage the cognitive outcomes and decision architecture of a firm from the outside, indefinitely, without the client ever cultivating internal maturity. In that case the distinction between asset and capacity has not been vindicated. It has been dissolved by a market that chose permanent dependency and was satisfied with it.

If neither vector resolves in Microsoft's favor, if returns appear only inside firms that were already mature, and if the departure of the engineers is followed by the operational decay of the agents, the initiative will stand as the empirical demonstration that an organizational accountability deficit cannot be closed by over-supplying technical power against a single factor of a multiplicative function. The grid is worth its length only if it generalizes, and it does. It explains Microsoft in July of 2026, and it anticipates the same move at Amazon Web Services, Google Cloud, Oracle, and Salesforce, because none of them controls more than a subset of the factors either, and all of them are following the same rent to the same place. The rent has moved to the result, and the vendors are moving to the result behind it. The difficulty they will all meet there is the same one, and it is not

a difficulty of engineering. The result is the one factor a vendor cannot deliver, because it is not deployed. It is instituted.