

Integration is a deliverable. Conformity is an institution.

Why the forward deployed engineer lowers the cost of integration and can raise the cost of imputation.

Two earlier pieces read Microsoft's Frontier Company as one move in a larger displacement: first as a bid to occupy the layer where value is proven rather than built, then as the migration of rent toward the binding factor, the customer's institutional capacity, which no vendor delivers from outside. A parallel set of weekly notes said it from the buyer's chair: performance is not deployability, and the cost that breaks a deployment is the cost of its operating regime, a line the chief financial officer now reads on the income statement.

Grant all of it, and one question remains that Microsoft, Amazon and OpenAI leave unspoken. Once integration is solved, what is the binding factor, and can the instrument the vendors chose to solve it reach that far? The instrument is the forward deployed engineer, and the answer is compact enough to serve as the thesis. The forward deployed engineer lowers the cost of integration and can raise the cost of imputation. It ships the system faster, and it leaves the question of who answers for the system not only unsolved but heavier.

The mechanism does not appear everywhere. It appears where three conditions hold together:

1. Responsibility allocated before deployment,
2. Juridical roles named in advance,
3. And a system whose decisions carry high stakes.

Remove any one and the forward deployed engineer may well be enough.

Hold the three and the binding factor stops being technical.

Stated in full: under ex ante allocation of responsibility, once integration is solved, the binding factor becomes institutional, and the family of forward deployed engineering strategies, superb at building, is neither the authority that institutes nor the party that answers.

The European regime under Regulation (EU) 2024/1689 is the clearest laboratory for this, not because it is the only place the mechanism operates, but because it is the place the

mechanism is written down. Deployability, whose meaning has widened across this series from proof to responsibility to institution, is fixed here for the corpus: the capacity of a system to remain operable under technical, economic and institutional constraints across its entire service life.

Three roles the market runs together

The deployment turn rests on a correct observation and a hasty inference. The observation: the model is no longer the problem, deployment is. The inference: therefore deployment is integration. It is not. Deployment in a regulated regime is three roles, and the market prices only the first. To build is to produce the working system, the integrations, the harness, the evaluations, the observability. To institute is to establish the rule system under which the system may operate: the conformity file, the risk management system, the human oversight design, the allocation of who is accountable for what. To assume is to carry the responsibility for the decisions the system makes.

The third role splits, and this split is the load bearing move of the argument. Why two halves, and not one or three? Because the line that matters is economic, the line of transferability. Attesting produces a transferable object; answering produces a non transferable exposure. Everything that can be handed to a notified body, an insurer or a firm falls on one side; everything that stays with a person once those parties have done their work falls on the other. The cut is not moral, it is contractual: what can be sold, and what cannot. So attesting is producing the instrument of conformity, the declaration, the marking, the technical file, the formal risk acceptance. It is a deliverable, delegable, insurable, and partly codable as policy as code and evidence as code. Because it is transferable it makes a market, and because it makes a market it carries rent. Answering is the other half, and it must be stated in its technical sense and not as a moral flourish. To answer is to remain the duty holder who is exposed, in civil and in criminal law, once attestation, insurance and documentation have done all they can do. It is not insured, not delegated, not coded. It makes no market, so it carries no rent, only exposure.

Hence the strict definition that keeps the thesis falsifiable. An institution here is the rule system, formal, meaning law and contracts, or informal, meaning corporate governance, that allocates the responsibility of a decision act. A constraint is institutional when its lifting requires an authority, a third party or a formalized governance, rather than a performance metric such as a coverage rate or a calibration score. Data quality is not, in itself, an institutional constraint; it is lifted by better data. The obligation to attest the representativeness of that data, which Article 10 of the Regulation places on the provider of a high risk system, is institutional, because no metric lifts it, only a signature before an authority. That difference is the whole difference between a thesis and a platitude.

Three clean oppositions follow, and the article turns on them. A technical deliverable against an institution of responsibility, what one produces against what one commits. The

economy of acquisition against the economy of continuous operation, the cost of the pilot the engineer ships against the cost of the monitoring, the recalibration and the evidence retention the regime demands for the system's life. And the engineer diplomat against the authority that institutes, the one who aligns the teams against the one who signs the risk. It would be unfair, and therefore weak, to say the forward deployed engineer does not institute. He can document an institution, accelerate its artefacts, and reveal its gaps. What he cannot do is constitute the authority from which the institution draws its force.

The family of FDE strategies, and Microsoft as a rational actor

The forward deployed engineer is not a straw man. It is a dated, funded, documented object. In May 2026, within seventy two hours of each other, Anthropic announced a services venture with Blackstone, Hellman and Friedman and Goldman Sachs, and OpenAI announced a deployment company and the acquisition of the consultancy Tomoro. The figures reported at the time placed the first near one and a half billion dollars and the second near four billion, with Tomoro adding on the order of one hundred and fifty deployment engineers. The playbook is Palantir's, and not in dispute: embed a senior engineer inside the customer for a fixed period of weeks to months, ship production code early, refuse the systems integrator role. The creed is explicit, and correct as far as it reaches: the model is not the bottleneck, deployment is, a reading the vendors anchor in the MIT NANDA report State of AI in Business 2025, which found that ninety five percent of enterprise generative pilots show no measurable profit and loss impact.

Four positions order the whole field, and it helps to name them before the actors crowd in. A builder makes the model. An integrator embeds it. An attester certifies it. An answerer stands behind its decisions. The vendors occupy the first two and reach toward the third; the fourth is, by construction, the customer's. Every firm below takes its place on that line, and the family that runs the embed is not homogeneous. Palantir accepts the deepest operational involvement, Microsoft stays closer to accompaniment and distributes the OpenAI infrastructure, Anthropic sells integration acceleration, OpenAI is still building its model, Amazon supplies the container, Bedrock. One speaks of a family of strategies. And the order the critique will draw is not strictly sequential: integration and conformity co evolve, and in a real programme the further the integration advances, the more the conformity requirements surface. The two rise together, which is precisely why the party who ships is tempted to treat conformity as a downstream chore rather than a distinct role.

Microsoft is used across this series as a symptom, but it is also a firm making a choice, and the instructive fact is where the choice stops. Frontier Company will build and

integrate at scale. It will not attest, and it will not answer. The reason is structural, not psychological, and one need not attribute intentions to an org chart to read it. To attest a system one also supplies would collapse the independence that conformity assessment requires. To answer for it would put a supplier's balance sheet behind a customer's decision, and that is the non transferable exposure a supplier's economics cannot absorb. Microsoft stops precisely where its business model ends, which happens to be exactly where institution begins.

The FDE does not institute, and it may requalify

Consider a composite but unremarkable scene, a go to market committee in a European retail bank in mid 2026. A vendor's forward deployed team has integrated a graph completion model for small business credit scoring that automates most of the case preparation. The system is ready. The load tests pass. The connector ships code to production. Then the programme stalls, and on nothing technical. The chief compliance officer will not put her name to the fundamental rights impact assessment that Article 27 of the Regulation requires of a deployer of a credit scoring system, and that the bank's own risk governance already demands ahead of the deadline. Her reason is exact. The embedded engineers, having tuned the model's local weights to the bank's sectoral specifics, have done precisely what the vendor was paid to do, and in doing so have moved the bank toward the substantial modification that Article 25 treats as turning a deployer into a provider. The bank is no longer merely using an external tool. It is on its way to being the provider of the system, exposed to the heaviest regime in the text, for a decision, the refusal of credit, that carries a systemic discrimination risk. The engineer diplomat, whose mandate ends at the software deliverable, packs up. He has neither the contractual standing to bind his employer to the conformity of the final architecture, nor the mandate to co sign that risk. Integration is complete. Deployment is paralysed.

The argument does not depend on banking. Replace the credit scoring model with a clinical decision support tool, an autonomous manufacturing controller, or a public sector eligibility engine, and the structure is unchanged: the technical work completes, and the programme halts at the signature. What the scene shows is a paradox. The forward deployed model is built to modify substantially, it personalises, it fine tunes, it weaves the system into the customer's infrastructure, and that depth is the value proposition. It is also the trigger. The instrument sent to reduce the friction of deployment is, by the same technical gesture, the instrument that qualifies the customer for the regime it was least prepared to carry. The vendors see the friction coming and draft against it: their terms increasingly stipulate that fine tuning, alignment, or retrieval augmented adaptation remain within normal use and transfer no qualification. It helps them less than they hope, because requalification under Article 25 is a matter of law and fact, of what was actually done to the system, not of the clause by which a supplier

disclaims it. Substantial modification is not yet stabilised in guidance or case law, so the embed can lead to requalification, it does not mechanically produce it. The point is not a certainty. It is an unpriced risk that the go to market motion manufactures, and that no contract fully waives.

The second limit answers the objection that conformity will simply be automated. Continuous compliance, policy as code, evidence as code, automated conformity assessment: the market is already industrialising the evidentiary, documentary and procedural layers of conformity, and it will succeed at much of it. What the automation reaches is attesting, the transferable instrument. What it does not reach is answering. Here Bainbridge's ironies of automation, written for control rooms in 1983, transpose exactly, provided one applies them to the right residue. The more of building, integrating and even attesting is automated, the rarer, the more critical and the more concentrated the remaining act becomes, and the remaining act is not a signature on a form, which delegates and insures. It is answering for the decision, which does not. The compact reading: one codes the proof, one transfers the attestation, one does not automate the act of answering.

The deferral as proof, and the AI Act as a revealer of allocation

The regulatory calendar moved while this series was being written, and the move is the strongest evidence for the thesis, not against it. Precision matters here, because the passage invites a jurist's objection. Under the Digital Omnibus on AI, on which the Council and the Parliament reached political agreement on 7 May 2026 (Council of the EU, 7 May 2026), endorsed by both institutions in June and awaiting publication in the Official Journal at the time of writing, the high risk obligations for standalone Annex III systems, recruitment, credit scoring, education, border control, are set to be deferred from 2 August 2026 to 2 December 2027, and those for Annex I systems embedded in regulated products, medical devices, machinery, vehicles, to 2 August 2028. The transparency obligations of Article 50, including those on deployers, are set to hold at 2 August 2026.

The deferral was granted because the machinery that makes the obligations operable, the harmonised standards from CEN and CENELEC, the notified bodies, the national competent authorities, was not ready. The binding factor for deployment at scale is therefore not the model, which shipped, and not the integration, which the engineers ship. It is the infrastructure of institution, and the regulator concedes it does not yet exist by conceding the calendar. In May 2026 the industry commits billions to embedded engineering at the moment the regulator retreats for want of an institution to enforce. The technical infrastructure runs faster than the infrastructure of trust. And this is the point at which the article must be read against its own European frame: the AI Act does not create

the mechanism, it reveals it. The Act is a microscope, not the phenomenon. It does not invent responsibility, since manufacturers built AI enabled medical devices under the Medical Device Regulation and answered under product liability long before it existed. What it adds is legibility. It names the deployer as an autonomous juridical actor, in Article 26, and makes the allocation of responsibility visible. The thesis rests on that visibility, not on any inflation of duties. One guardrail follows: a deferred calendar is not a soft constraint. The Omnibus moves the date of attesting; it does not move answering. Sectoral law bites in 2026 regardless, and personal exposure does not defer with the annex.

Where the rent goes

This is where the series closes on itself. The frontier of responsibility named in the second piece now has a juridical name, the provider deployer boundary. The Decision Owner named there, the accountable role almost never formalised, is the one who answers, the holder of the non transferable share. The insistence, in the RAISE framework, that industrialising regulated AI is a matter of instituting and not only of integrating, is the same claim seen from the method side. The buyer's test from the weekly notes extends cleanly: for each system in production, who owns the operating line, who owns the incident line, and now the third, who owns the institution line, the attesting and the answering. In a CE marked medical device the distinction is already visible, the manufacturer builds, a notified body attests, and a designated responsible person answers, three parties for three roles, the middle one buyable and the last one not.

Now the consequence, and it needs one layer more than the earlier notes gave it. Rent does not migrate toward attestation. Rent migrates toward power, and the relevant power is the authority to authorise operation. That authority is the scarce asset. Attestation is only its marketable face, the instrument through which the authority is exercised and evidenced, which is why the money appears to flow to attestation when it flows, in truth, to whoever may say yes. This is old ground in the economics of organisation. Institutions, in North's sense, are the rules of the game, and the rules here confer a right to authorise; value accrues to the holder of a rule conferred right, not to the technician who prepares the file. The boundary the vendor cannot cross is Williamson's: attestation is contractible and therefore a market, answering is the residual that no contract completes and therefore stays inside the party who holds it. So the rent settles on whoever holds the power to authorise operation, a notified body in one sector, one of the large audit firms in another, an insurer, an internal governance function. The vendors will not stay passive before it. They will reach through AI assurance services, certified deployment offers, contractual guarantees, managed operation, and they will take the transferable part, from beside their balance sheet, through a separate assurance vehicle that monetises attestation without costing them their neutrality, precisely because it is separate. What

no vehicle and no contract capture is answering. One can buy the attestation; one cannot delegate the act of answering. The CTO frames the integration, the CFO budgets the operation, a third party is paid to attest, but the Decision Owner alone answers, and the vendor sent no one to hold that pen.

The transatlantic asymmetry, and the horizon

The boundary of the thesis is the engine that explains it, and it should not be read as too clean a line. The forward deployed model was born in the United States, and there responsibility is produced after the fact and from many sources, jurisprudence, class actions, insurers, contracts, procurement requirements, arbitrated at the level of operational use rather than of ex ante attestation. In that regime the embedded engineer is closer to a complete solution, because the institution forms later, around the use, so he can carry a deployment to the point where the courts and the insurers take over. But the asymmetry is one of mechanism and degree, not of presence. The American regime produces its own imputation cost upstream and informally: the fear of a class action or of reputational damage already stops many an American committee, with no requalification in sight. The same paralysis of answering is beginning to infuse there; the European regime merely names it in advance, and puts a door where the American one leaves an open field. Europe is not the only place the mechanism operates. It is the place it is legible.

The thesis is falsifiable twice over. It fails in principle the day answering becomes transferable, if a regime grants a leader personal immunity in exchange for a licensed third party's attestation. It fails in fact if firms with embedded forward deployed teams are shown to deploy regulated, high impact AI systematically faster, with no increase in institutional cost and no reallocation of responsibility. The first test is elegant and improbable, the second mundane and measurable, and the second is the one to watch. Two limits stay honest alongside it. The institutional factor is not always the one that binds first, a bank's own governance, a hospital's data obligations, a public buyer's procurement rules may each bind before the law, but each is a rule system allocating responsibility, so each confirms the level of the claim while displacing its content. And what the public theses omit is not proof that the private contracts omit it too; the allocation may already be negotiated where it cannot be seen.

The corpus points past this article, and past artificial intelligence. What holds for a credit model holds for a medical device, a power grid controller, an aircraft system, a foundation model: the technical layer industrialises, fast and with great capital, and then the work stops where authority begins. That is the general proposition the AI Act only happens to illustrate. Every technological revolution eventually halts at the same frontier, the one where a rule must confer a right and a person must stand behind a decision. Software builds. Rules institute. A person answers.