

From System Delivered to Risk Allocated: Why AI Insurance Covers Only Fragments of the Business Case

The vendor delivers a system, the business expects an outcome, the insurer covers a peril.

The shift of the problem

For decades, deploying an information system meant transferring an artifact: a piece of software, a database, a data flow. Deploying an AI means transferring a decision-making capability, and that shift reconfigures at once the governance, the liability, the insurance, the financing and the economics of deployment. The previous article isolated one of its faces, the deferred break-even: between the moment a system goes live and the moment its cumulative value exceeds its cumulative cost lies an interval that customer-facing personalization tends to lengthen rather than shorten. A question follows: can this delay, this uncertainty over profitability, be transferred to a third party? A market for AI-specific insurance now exists, it is growing, and it would be tempting to read it as the answer.

The thesis of this note is that it is not, and that the reason for this failure is not the market's immaturity. The offerings that are emerging cover liabilities, failures and performance thresholds, not the client's economic result. This difference does not reveal a lag that time will close: it reveals a *defect of translation* between three units that do not overlap. The vendor delivers a system, the business expects an outcome, the insurer covers a peril, and nothing guarantees that an object expressed in one of these units is expressible in the other two. The market can cover a defined AI failure; it cannot turn every economic disappointment into a claim. A single question therefore carries this note: why does the residual economic risk remain on the client's balance sheet? Insurance is its entry point, not its exclusive object. Answering it requires crossing proof, norm and cost, but always for the same reason, and it is that reason, not the inventory of mechanisms, that holds the thread.

Scope. This note addresses the insurability status of the residual economic result that weighs on the business case after delivery. It reasons at the level of public product descriptions: carrier pages, broker notes, academic work on the insurability frontier. Contractual wordings are not public, and nothing here claims to interpret them as enforceable. It describes a young market; it predicts neither its size nor the sustainability of its premiums.

A vocabulary that must be irreproachable

The argument requires a lexical discipline that ambient discourse neglects. Six terms must stay distinct.

- The **peril** is the mechanism that generates damage.
- The **exposure** is the situation in which the insured may suffer it.
- The **damage** is the economic or legal consequence.
- The **risk** is the probabilistic distribution combining peril, exposure and vulnerability.
- The **trigger** is the contractual condition that activates the cover.
- The **object of the cover** is the loss the policy agrees to indemnify.

One chain is enough to hold them together:

- A hallucination is a mechanism.
- Its use in financial advice is an exposure.
- The mistaken decision of a third party who relies on it is the causal event.
- That third party's loss is the damage.
- The lawsuit is the potential claim.
- The defense costs are the object of the cover.

To confuse these levels is to believe a hallucination is a claim. It is not: it is, at best, the distant upstream of a claim that remains to be established.

Break-even plays a particular role here, and it must be framed before it is used. The business case is the forward-looking architecture of a deployment's value. Break-even is merely one of its temporal variables. Risks do not destroy the break-even: they modify its distribution. The instruments of allocation, insurance among them, absorb or redistribute certain components of those risks. The insurer never prices the break-even itself, and a reader who imagined otherwise would fall back into precisely the confusion this note fights. One convention will hold: break-even here means the first cumulative crossing, not the guarantee of value creation that would persist thereafter.

The ROI is not a peril

The central difficulty lies in the nature of the economic result. The objects the market knows how to handle (harm to a third party, a security incident, an intellectual-property violation, a deviation from a performance threshold) are relatively bounded events or

states. The net economic result of a deployment is not of that nature. It is a composite function in which technical performance, organizational adoption, real usage, operating and supervision costs, change management, the quality of management, market conditions and external shocks unrelated to the AI all intervene. The ROI is not a poorly documented peril. It is a causal aggregate endogenous to the client.

Four properties make it refractory to classical indemnity. Causality is diffuse: a firm can miss its ROI with a technically compliant system, or produce ROI despite a mediocre model, thanks to a reorganization or a price increase. Control belongs largely to the insured, who can under-invest in adoption, change the data, or ignore recommendations, which installs a moral hazard of an altogether different scale than on a technical threshold. Measurement is counterfactual: indemnifying a shortfall requires comparing the observed world with a world without failure, a counterfactual rarely observable. Volatility, finally, is partly external. An object endowed with these four properties is not waiting to be better reconstructed. It is waiting, perhaps, to be decomposed.

The market covers fragments, not the result

An examination of the public offerings confirms this shift, provided one reads them at the level of proof at which they present themselves. Armilla, a Lloyd's coverholder, structured from launch a standalone policy whose overall capacity was raised to 25 million dollars in early 2026¹. Its described objects fall under liability for model error, defined as the loss suffered by third parties from underperformance, liability for outputs (defamation, confidentiality), material damage to third parties, defense costs and regulatory fines. The described trigger is not an isolated error but performance judged below initial expectations. This is not the destruction of the business case: it is a measured deviation from a technical promise. Munich Re, with aiSure, has since 2018 built guarantees around defined and measurable indicators, the trigger being the crossing of a performance threshold, the analysis bearing on the model.

A third case forbids any hasty conclusion. Testudo describes a standalone third-party liability policy that claims to require no integration into the system and no extended technical audit, priced notably on litigation and exposure data². This counter-example matters. It forbids asserting that AI insurance converges uniformly toward technical verification of the system. Reality is graded: the more the trigger depends on the model's own behavior, the more central its technical characterization becomes; the more it

¹ Armilla AI, Lloyd's Coverholder Facility Underwriting Report (2025-2026). The public product descriptions target model performance liability (Model Performance Liability) and regulatory defense costs, backed by the capacities of the partner syndicates

² Testudo Risk, Technical Underwriting Specifications (2025). The documented methodology favors the analysis of legal exposure vectors and sectoral litigation history over algorithmic-engineering audits

depends on generic legal liability, the more underwriting can rely on exposure proxies without auditing the model.

This gradation requires separating instruments that the word "insurance" conflates. Indemnity insurance requires a demonstrated loss and a causal link. A performance guarantee requires a contractual threshold and a measurement method. Parametric cover requires an objective index and accepts a basis risk. A service-level agreement does not necessarily fall under insurance and often transfers a capped compensation far below the actual damage. Basis risk is the blind spot here: a parameter can be met while the client loses money, or breached while the project stays profitable. Insuring the proxy is not the same as insuring the outcome. One cannot therefore file under a single heading an indemnity for harm and a guarantee indexed on a metric: they do not cover the same object, and they do not leave the client the same remainder.

Decompose to allocate, not to insure everything

If the economic result is not a peril, the work required is not to reconstruct it better but to cut it up. This work is nothing new: insurance has decomposed complex exposures for three centuries, and to deny it would offer any underwriter an immediate rebuttal. The contribution is not decomposition; it is the observation that AI deployment and its business case remain structured along units the insurer cannot use as they stand.

A first layer decomposes causality. Each sub-risk reads as a triplet: a source (the model, the data, the integration, the infrastructure, human use, governance, a third-party supplier, the regulatory environment), a manifestation (underperformance, error, drift, unavailability, violation, unauthorized behavior, loss of control, security incident), a consequence (own damage, third-party damage, regulatory liability, remediation cost, interruption, revenue loss, break-even delay). This matrix does not, however, suffice to produce an insurable peril. Two incidents identical on these three axes differ radically for the insurer according to their profile. A hallucination causing a 100,000-euro loss does not have the same insurance value if it strikes an isolated client or if an update to the same model reproduces it simultaneously across twenty thousand insureds. A second layer is therefore needed, one that characterizes the sub-risk without mechanically deciding its fate: a profile defined by controllability, observability, frequency, severity, correlation and boundability. This profile informs the arbitration between prevention, contractual allocation, insurance transfer and retention. It does not determine it: bargaining power, applicable law and financial capacity also intervene.

From this decomposition come three destinations, not one. Some perils are mutualizable with an external insurer. Others are allocable or shareable contractually between vendor and client, through a performance guarantee, a contract indexed on an agreed outcome, a sharing clause. Others still are retained or financed internally, possibly through a captive, not because they would be uninsurable but because they belong to execution,

strategy or entrepreneurial uncertainty. A part of the economic result is not a failure of the insurance market: it is ordinary capitalism, that institution in which the one who decides sometimes keeps the risk of the decision. To confuse a strategic bet with a retained deductible is to ask insurance to cover what the entrepreneur is precisely meant to bear.

The most important point remains, and the most original. Decomposition separates perils, but above all it reveals that three maps of the risk do not coincide. *Operational allocation* designates who controls the causes. *Contractual allocation* designates who legally accepts the consequences. *Economic incidence* designates who ultimately bears the loss. These three distributions are distinct, and their gap is structural. The vendor controls the model, contractually limits its liability, and leaves the client to bear the incidence. The insurer indemnifies a share, then turns back through subrogation. A risk may be controlled by one party, contractually imputed to another, and economically borne by a third. The least-cost risk-bearer doctrine, which would assign each risk to whoever observes or diversifies it best, describes an *efficient* allocation. It is not the *effective* allocation, where risk goes to whoever the contract manages to impute it to. The gap between efficient and effective allocation is an asymmetry of the fixed-fee deployment model, not a mere market friction.

Who holds the proof?

To attribute a loss, one must reconstruct it, and this is where the delivery model shows its hole. The proof required is *fragmented* among actors none of whom holds it alone. The vendor knows the system. The client knows real usage, the workarounds, the organizational decisions, the internal costs, the unreported incidents, the effective result. The integrator knows the dependencies. The insurer, for its part, needs a consolidated representation that the contractual structure charges no one with producing. In compression: the insurer needs an evidentiary chain that no single actor can provide alone. The formula must be bounded, on pain of falsehood: the more the guarantee depends on the system's real behavior and on causal attribution of the loss, the more this chain becomes necessary; a parametric guarantee makes do with an index, a liability policy awaits the legal establishment of fault.

Reconstructibility intervenes here, but not as the magic property one is tempted to see. Information asymmetry covers three problems that must be separated. Information is distributed: each holds a different link of the chain. Information is partly non-observable: workarounds, tacit decisions, users' actual behavior are not documented. Information is, finally, strategically withheld: the vendor protects its liability, the client seeks a better premium, the business masks insufficient adoption. Reconstructibility addresses the first two problems. It does not resolve the third, and it may even produce more data without guaranteeing their credibility. One will therefore not write that more proof reduces asymmetry, but that accessible, interpretable and sufficiently independent proof can

reduce a part of it. Reconstructibility turns a fraction of uncertainty into contractually usable information; it does not make the business case insurable.

Time aggravates the problem. At underwriting, the insurer makes do with an architecture, tests, a domain of use, a governance. At claim, it demands the exact model version, the logs, the changes that occurred, the decision chain, proof of the damage, the causal link, compliance with mitigation obligations. These two sets of proof do not coincide, hence a proposition that must stay exact so as not to slip into dramatic effect: a system may pass underwriting and still produce a claim that fails the conditions for indemnification. Not by technical mystery, but by contractual rupture. One must nonetheless refrain from universalizing this rupture. An undeclared material modification may lead to a repricing of the premium, a restriction of cover, a reduction of indemnity, even a loss of cover, depending on the contract and the applicable law. The notion of aggravation of the risk is legally loaded, and it is not deduced from an operational intuition. Likewise, the client's insurance does not deprive it of recourse against the vendor: it may on the contrary shift the conflict from the initial dispute to the subrogated claim. The evidence file is then not a purely defensive asset: it can establish the vendor's compliance as much as document its breach.

This file, precisely, is not an ordinary deliverable. The knowledge of a customer-facing deployment is contextual, tacit, distributed, evolving. Converting it into persistent proof, interrogable by a financial third party, requires work that the delivery contract does not pay for: it pays for going live, not for the constitution and maintenance of a reusable base. One will moreover refrain from assuming a uniform standard: between the internal risk-mastery file, the underwriting file configured for one insurer, and a base interrogable by several carriers, the passage is not acquired.

This base deserves better than a passive role. Reducing asymmetry is only its first function. By making an opaque risk observable, attributable and boundable, reconstructibility turns it into a negotiable financial object: this is the condition not only of insurance, but of refinancing, of reinsurance, of rating, of securitization, of captive provisioning. The insurer is here only a particular case of a *financial risk-bearer*: the bank that lends against the deployment, the investor who values it, the agency that rates it, the auditor who certifies it all ask the same thing, reliable information about the risk. Reconstructibility therefore opens not only a policy: it opens a market. And it is because it has this value that its absence is costly: the delivery contract that does not fund its production leaves the risk without the instrument that would make it transferable, and therefore on the balance sheet of whoever lacks the means to transfer it.

A selection consequence deserves to be stated without caricature. A guarantee close to the result may attract the most fragile projects, out of a need for protection, but the demand for proof may conversely retain only the organizations able to produce its pieces. The mechanism is not size: an SME with a narrow use case may be more insurable than a

large group unable to say who changed the system prompt last Tuesday. Exclusion strikes whoever cannot amortize the fixed cost of proof, not the small as such.

The norm that does not yet exist

The previous section showed that the system can drift beneath the policy: the version changes, usage changes, and the cover subscribed ceases to match the risk run. There is a second source of drift, harder to grasp because it does not touch the system. It is the norm itself that moves.

To be insurable, an AI risk must clear at least two independent gates, and the actuarial one is not the higher. The first is actuarial: frequency, severity, correlation, capital. The second is normative: the covered peril must be defined by a norm stable enough to be bounded, priced and litigated before a judge with a predictable outcome. Beyond actuarial acceptability, a *normative determination* is therefore needed. Yet a risk can be actuarially tractable and normatively indeterminate. It is not then uninsurable in principle, but it is in the sense that matters: the insurer does not know what it is promising to indemnify.

Two reasons make the norm applicable to AI resist. The first is the *orthogonality* of certain requirements, which cannot be maximized together. The explicability and human oversight expected by the AI Act pull against the data minimization of the GDPR; the right to explanation pulls against the protection of the trade secret that covers the model; the MDR's demand for a stable clinical evidence base enters into ontological collision with an AI Act that presupposes systems that keep learning. I have argued elsewhere that this MDR / AI Act collision is not a coordination defect but a structural property. When two requirements are orthogonal, compliance is not a state one reaches: it is an arbitration one renders, and an arbitration has no single correct answer until an authority has decided. The peril "non-compliance" is under-defined by construction.

The second reason follows. A large part of the operative norm will not be written by the legislator but forged by the courts. The text fixes the requirements; the judge will fix what satisfying them means when they conflict. The underwriter prices the written rule; it pays against the enforceable rule, forged after the fact. It will be objected that insurance has lived for a century with legal uncertainty, in product liability as in directors' and officers' liability, and that is exact. But it does so because decades of case law have stabilized the operative meaning of the norms: the insurer prices against a settled corpus. The AI regulatory corpus is at once recent and internally orthogonal, so that the stabilizing case law does not yet exist and will take years of litigation to form. During that formation, the meaning of the covered peril shifts beneath the insurer's feet without the insured system having changed.

This reopens, from the legal side, the gap the previous section opened from the technical side. A system may be compliant at underwriting and non-compliant at claim, not because it changed, but because the norm changed. It is a *jurisprudential* aggravation of the risk, and it does not fit the contractual categories of material modification, which all assume that what moves is the system. Here, what moves is the reading. It also introduces a correlation the technical audit does not see. A single landmark ruling can reclassify at a stroke the liability of every deployment built on the same orthogonal arbitration, across an entire portfolio. This normative correlation is distinct from the technical correlation of a shared foundation model: it lives in the legal layer, not the model layer, and no reconstructibility of the system detects it.

The consequence for cover is direct. What the insurer cannot bound because it is not yet written, it treats like any undefined peril: by exclusion, by sub-limit, by surcharge, or by a defense-cost cover that pays for the trial without guaranteeing its outcome. What the text leaves open, the court closes, often after the loss. The regulatory fines and defense costs that appear with such assurance in the product descriptions are, on this reading, less a covered peril than a bet on the pace and direction of a case law barely begun.

Insurance modifies the break-even distribution

The economic effect of the transfer remains, and this is where the loop with the previous article closes. One would be tempted to write that insurance reduces the variance of the break-even. That would assert more than the mechanism permits. A bounded, deductible-bearing and selective cover does not reduce global variance, and it may even increase certain financial volatilities. The exact proposition is more modest: insurance adds a certain cost to all scenarios and may reduce the severity of certain extreme scenarios that would delay or prevent the break-even. Even a covered claim leaves the client a deductible, a sub-limit, a settlement delay, a causation dispute, out-of-scope costs. One absorbs part of certain bounded shocks; one neutralizes nothing.

This precision commands a correct reading of cost. The total operating cost of a deployment breaks down: infrastructure, operations, governance (proof, control, compliance, validation), transfer (premium, brokerage, underwriting audit). The premium belongs to transfer, not to governance, and this separation is not an accounting coquetry. Without it, one would impute to insurance governance costs the client would have incurred anyway. With it appears the paradox that concerns the business case: the very infrastructure that enables risk transfer can defer the break-even it was meant to secure, once the cost of transfer, added to the governance cost it requires, grows faster than the risk it removes. The value of a policy is therefore not read in the improvement of the average break-even. It may lie elsewhere: in the reduction of certain extreme scenarios, in access to contracts that require it, in the reduction of the cost of capital. The word "may" is decisive here, for the policy is not always bought for its actuarial expectation. It

is sometimes bought because a principal, an investor or a regulator imposes it. Insurance can function less as expected indemnity than as a contractual entry ticket to certain markets.

By conditioning cover on the audit, the insurer becomes a private standard-setter, without the movement being virtuous for all that: it standardizes what is measurable, not what reduces risk, and gives rise to a market of auditors paid to produce a file compliant with a grid. The flaw is not the framework, it is *photographic compliance*: the exact snapshot of a system that keeps changing after the shot.

Nothing guarantees, moreover, that the market takes this route. Reconstructing costs, and a bearer may judge it cheaper not to reconstruct: to exclude, to cap, to raise deductibles, to fall back on the parametric, to return the risk to the contract. Reconstructibility is therefore not a destiny of the market but a conditional utility: the more a guarantee depends on the system's real behavior, the more useful it becomes economically; the less it depends on it, the more exclusion suffices. The thesis does not predict a convergence toward proof: it predicts a gradation, and leaves open the trajectory of a market that would choose priced ignorance.

Discussion, limits, and conditions of refutation

This note rests on public descriptions, not on contracts. That is its first limit, and it must be assumed rather than masked. The wordings, deductibles, caps and causal links required by each product are not accessible; what is described as a trigger is a public positioning, not an enforceable clause. The exact correspondence between the risks removed from general policies (the exclusions of third-party damage tied to generative AI, circulated around 1 January 2026) and the risks taken up by dedicated products remains to be established. One will note in passing that the coverage gap of the economic result does not stem from these exclusions: that result was, most often, not covered before. Any aggregate figure on the extent of the exclusions would be premature here as long as the precise metric is not named.

The thesis is falsifiable, and it is better to say under what conditions. The insurance value of reconstructibility, which should grow with the trigger's dependence on the system's behavior, would remain to be established on underwriting questionnaires, wordings, claim files and premium differentials, which public descriptions alone do not provide. The thesis would weaken if indemnity guarantees durably covered a global economic result without reducing it to bounded events, thresholds or indices; a cover indexed on revenue alone would not suffice, for it would remain parametric. It would weaken further if ordinary deployment deliverables regularly produced an evidence base simultaneously maintained, causally relevant and reusable by several insurers, which an isolated model card or security file does not constitute. It would strengthen, conversely, if premiums, deductibles and limits proved statistically associated with the quality of monitoring, logs,

change governance and independent verification. As for the cost of insurability, its economic compensation varies from project to project: it may come from a reduction of extreme losses, of the cost of capital or of contractual barriers, but it is neither automatic nor total, and it is project by project that it is demonstrated.

The normative argument calls for its own reservation. It depends on the jurisdiction: in civil-law systems, case law binds less formally than in common-law systems. But administrative interpretation and the decisions of authorities fix in practice the operative meaning, and a ruling of the Court of Justice or the practice of a European authority would play the same stabilizing role there. The normative thesis would weaken if determinate standards, endowed with non-orthogonal safe harbors, appeared early enough to close the gap before litigation forges it. It predicts the opposite: as long as the requirements remain orthogonal, stabilization will come from the courts, with the delay that characterizes them.

Conclusion

The market for AI-specific insurance is not behind the client's business case: it works on another object. It isolates liabilities and failures defined enough to be attributed, bounded and priced, and it leaves on the client's balance sheet what, in the economic result, belongs to execution, strategy and the counterfactual. Between the system delivered, the outcome expected and the peril covered, risk is not merely transferred: it must be translated, attributed and documented, and this translation has been contractually charged to no one. Reconstructibility reduces a part of it, conditionally, by turning uncertainty into usable information. It does not make the business case insurable, and it removes neither moral hazard, nor diffuse causality, nor the correlation of losses, nor the indeterminacy of the norm.

One must therefore give up the idea that decomposition would end up making everything insurable. It serves to decide what must be prevented, what must be allocated to the vendor, what can be mutualized, and what the client must keep bearing. The deployment produces an operational system. It does not produce the map of the risk that would allow one to say who controls, who answers and who pays. It is the same shift that the series observes from different angles: transferring a decision-making capability is not transferring a tool, it is redistributing a responsibility that neither the law, nor the contract, nor the policy has finished assigning. As long as this map is not drawn, the loss, for its part, always finds someone to bear it: it is simply not always the one who caused it.