Contractual AI Governance as an Architecture of Instrumental Rationality

Di Alberto Stazzone, Senior Advisor, Pollicino AIdvisory

Executive Summary

Artificial intelligence governance is increasingly framed through ethical principles and regulatory obligations; a recurring practical question, however, concerns how these expectations are translated into effective governance mechanisms within operational and contractual settings.

In complex AI value chains, where control and information are unevenly distributed, risk is often allocated implicitly through contractual arrangements. When governance is not deliberately structured, responsibilities may shift by inertia rather than by design, gradually reducing organisational visibility and control. This challenge is time-sensitive, too: decisions on procurement, data use and system integration taken today shape durable assets, dependencies and risk profiles. Deferring governance choices until full regulatory application does not neutralise their effects.

This Policy Brief examines contracts as an enabling layer of AI governance. Rather than treating contractual arrangements as residual instruments of compliance or liability allocation, it conceptualises them as governance devices capable of structuring control, information flows and decision rights *ex ante*. In this perspective, contractual design determines whether regulatory expectations can be exercised in practice, by aligning responsibility with effective control and by making governance mechanisms enforceable within operational relationships.

To this end, the Brief proposes a contractual governance logic articulated around five stages: risk identification, allocation of control, verifiability, enforceable remedies and adaptive capacity over time. The purpose of this framework is not to standardise clauses, but to provide a coherent logical architecture through which AI-related risk can be intentionally governed across different contractual settings; the logic is applied across deployers, providers and organisations, illustrating how the same contractual structure can support AI risk management throughout the value chain.

1. Introduction: the Operational Gap in AI Governance

Academic and institutional discourse on artificial intelligence governance has developed primarily around the protection of fundamental rights, ethical principles and societal impacts. This body of literature has been essential in framing AI as a matter of public interest, legitimacy and accountability, and has significantly informed the architecture of recent regulatory initiatives; in parallel, a growing corpus of standards, ethical frameworks and sector-specific guidance - particularly in public procurement and compliance-oriented settings - has sought to translate these principles into governance requirements. Taken together, these contributions have shaped a predominantly normative understanding of AI governance, cantered on values, safeguards and formal obligations[1].

Despite this rich normative landscape, organizations increasingly lack practical reference points to orient themselves in the day-to-day governance of AI systems. Ethical principles and regulatory obligations articulate what ought to be protected, but offer limited guidance on how risks should be concretely identified, prioritised and managed across complex technical, contractual and organizational environments[2]. As a result, enterprises operate within a fragmented governance landscape, where obligations are known in abstract terms but operational pathways remain unclear. At the moment this disjunction becomes particularly evident when concrete decisions concerning system acquisition, deployment, modification and oversight must be taken, often before formal compliance or enforcement mechanisms are activated.

2. AI Systems Risk Dynamics and Governance Gaps

AI-related risks can be logically disaggregated into distinct regulatory categories: a first category comprises risks that remain largely outside the explicit scope of the AI Act, including intellectual property exposure, secondary use of data for training purposes, and interoperability constraints resulting from technical or contractual lock-in; a second category includes risks that are formally addressed by the regulatory framework (such as accountability, auditability and liability) but whose effective management depends on organizational and legal arrangements that are not prescribed in operational terms by the regulation itself.

[1] Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., ... & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns, 4(10).

[2] Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: a systematic literature review. *AI and Ethics*, 3265-3279.

The distinction is not merely analytical: it determines whether risks can be actively governed, remain confined to abstract regulatory recognition or fail to be explicitly identified at all. With respect to the first category of AI-related risks, allocation often fails to occur altogether. Because these risks fall outside explicit regulatory prescriptions, they are not framed as governance-relevant variables at the moment of procurement or deployment, and no actor is structurally incentivised to identify or manage them ex ante. As a result, exposure remains embedded within technical architectures and contractual relationships, accumulating through default configurations, asymmetries of control and progressive dependency.

In such settings, risk does not materialise as a breach or compliance failure, but as a structural effect of non-recognition, becoming visible only once strategic or economic constraints have already crystallised; in the absence of explicit and structured risk allocation, risks belonging to the second category tend to shift by default from AI systems providers onto deployers, as regulatory obligations cascade downstream, while the informational, technical and organizational means required to discharge them remain upstream with the former.

Beyond these regulatory categories, AI systems introduce technology-intrinsic risks structurally distinct from traditional ICT hazards. The probabilistic nature of model outputs generates stochastic risk, while opacity, model drift and emergent behaviours produce algorithmic risk difficult to anticipate or attribute. When combined with complex data flows, cross-border processing and fragmented jurisdictions, these technical uncertainties compound into layered exposure profiles that generate governance imperatives upstream of harm or breach[3]. Organizations must establish monitoring protocols, output validation, change management and accountability structures before liability mechanisms activate: these ex-ante decisions shape exposure regardless of whether violations materialize, as governance shall precede enforcement and cannot be retrofitted through compliance alone.

The combined effect of unallocated regulatory risks, downstream-cascading obligations and technology-intrinsic uncertainties produces a silent transfer of risk: exposure accumulates not through deliberate allocation, but through organizational and contractual inertia. In such settings, risk is neither negotiated nor governed ex ante, but absorbed implicitly once operational, legal or economic consequences materialise.

[3] For a technical and quantitative perspective with respect to technology-intrinsic uncertainty and stochastic risk in AI systems: Wang, T., Wang, Y., Zhou, J., Peng, B., Song, X., Zhang, C., *et al.* (2025). From aleatoric to epistemic: Exploring uncertainty quantification techniques in artificial intelligence. *arXiv preprint arXiv:2501.03282*.

This misalignment manifests in recurring operational scenarios: organizations may deploy AI systems without a shared or stable classification of their regulatory risk profile, leaving open questions as to who bears responsibility when system characteristics evolve over time; data governance is frequently left undefined, with providers lacking clear policies on data reuse, retention or secondary training, exposing deployers to downstream legal and strategic risks[4]; moreover, where internal AI governance frameworks are incomplete or still under development, risk allocation often occurs implicitly through fragmented practices, resulting in inconsistent and fragile governance outcomes across comparable deployments.

The problem is further aggravated by a structural temporal asymmetry inherent to AI systems: on the one hand, datasets used today generate derived data, model artefacts and dependencies that persist over time and shape future capabilities; on the other, deferring governance until the full application of regulatory frameworks does not preserve optionality, as it entrenches risk within operational and relational structures that are difficult to unwind ex post. In this sense, inaction translates directly into the loss of control over strategic assets and the accumulation of exposure that is already material.[5]

In this context, the European Artificial Intelligence Act represents a fundamental regulatory milestone, yet its architecture reveals implementation gaps, as many obligations presuppose organizational arrangements, processes and tools that are

[4] Empirical evidence corroborates this structural misalignment in current AI contracting practices. An analysis of AI vendor contracts conducted by Stanford CodeX and TermScout found that 92% of AI vendors claim data usage rights beyond service delivery needs, while only 17% commit to full regulatory compliance and merely 33% provide indemnification for third-party intellectual property claims — figures that diverge sharply from broader SaaS market norms and confirm that default contractual configurations systematically favour provider control over deployer governance capacity: CodeX — Stanford Law School (2025). *Navigating AI Vendor Contracts and the Future of Law: A Guide for Legal Tech Innovators*. Available at: https://law.stanford.edu/2025/03/21/navigating-ai-vendor-contracts-and-the-future-of-law/. On the broader scaling gap between AI experimentation and operational governance maturity, see Deloitte AI Institute (2024). The State of Generative AI in the Enterprise: Now decides next. Q4 Survey (July–September 2024); as cited in OECD (2025). *Governing with Artificial Intelligence: The State of Play and Way Forward in Core Government Functions*, where the majority of surveyed organisations anticipated scaling fewer than one-third of their generative AI experiments into full production within 3-6 months, suggesting that governance frameworks remain structurally underdeveloped relative to deployment ambitions.

[5] Hausenloy, J., McClements, D., & Thakur, M. (2024). Towards Data Governance of Frontier AI Models. *arXiv preprint arXiv:2412.03824*, p. 1 «The development of today's frontier artificial intelligence (AI) models, highly capable foundation models, is inextricably linked to data, so much so that the systems are regularly defined by their training on "broad data at scale". There is a growing scientific consensus that, as well as tremendous benefit, such models may pose risks to public safety. Yet, because of the rapid pace of AI development and the growing secrecy surrounding frontier model training, the production, aggregation, and processing of the datasets used by frontier models has thus far received little regulatory and public attention».

not self-executing. Accountability, auditability and compliance capacity remain unevenly distributed along AI supply chains, creating asymmetries between providers and deployers. These normative frameworks must therefore be translated into operational mechanisms[6] that determine how risks are allocated, monitored and governed in practice. Strategic contract design counters this dynamic by making risk allocation explicit, enforceable and aligned with actual capacities.

3. *Zweckrationalität* as Governance: The Contract as Rational Instrument

Given this operational gap, contracts can operate as an immediately actionable layer of AI governance: beyond allocating liability ex post, they structure control, information flows and decision rights *ex ante*, translating governance expectations into binding operational commitments between providers and deployers: by embedding procedural obligations (as oversight mechanisms, documentation duties, monitoring requirements, escalation paths) into enforceable terms, contracts render governance observable and exercisable in practice[7].

Importantly, this contractual architecture produces governance effects independently of regulatory enforcement timelines. On the one hand, while regulatory frameworks articulate obligations, contracts determine *whether* and *how* those obligations can be operationalised within real organisational and technical constraints. On the other hand, where governance processes are embedded in contractual relationships, risk allocation becomes explicit, capacity-aligned and verifiable; where absent, allocation occurs implicitly through organisational inertia, resulting in the silent risk transfer along the value chain mentioned above.[8]

---

[6] Finch, W. W., & Butt, M. (2025). Gaps in AI-Compliant Complementary Governance Frameworks' Suitability (for Low-Capacity Actors), and Structural Asymmetries (in the Compliance Ecosystem)—A Systematic Review. *Journal of Cybersecurity and Privacy*, *5*(4), 101; Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885;

[7] On the theoretical framework of contracts as governance instruments structuring control, compliance and regulatory implementation within complex value chains: Cafaggi, F. (2013). The Regulatory Functions of Transnational Commercial Contracts: New Architectures. *Fordham International Law Journal*, 36(6), 1557-1618; for the foundational concept of contract governance as a distinct analytical dimension in private law: Grundmann, S., Möslein, F., & Riesenhuber, K. (2015). Contract Governance: Dimensions in Law and Interdisciplinary Research. In S. Grundmann, F. Möslein, & K. Riesenhuber (Eds.), *Contract Governance: Dimensions in Law and Interdisciplinary Research* (pp. 3-57). Oxford University.

[8] A pertinent and broader note on governance objectives at a regulatory and organisational level: Fabiano, N. (2025). Subject Roles in the EU AI Act: Mapping and Regulatory Implications. *arXiv preprint arXiv:2510.13591*; Montinaro, R. (2023, December). Responsible data sharing for AI: a test bench for EU Data Law. In *Global Data Law Conference Series* (pp. 81-99). Cham: Springer Nature Switzerland.

This proposed contractual governance approach operates through what Weber termed *Zweckrationalität* - that is, *instrumental rationality oriented exclusively to means conceived as adequate for precisely defined ends*:[9] at higher levels of technical complexity, governance mechanisms increasingly converge on technical risk control, risk allocation and regulatory anticipation, rather than ethical balancing exercises. The operational challenge is not formulating moral judgments, but rather translating governance objectives - however derived - into binding, executable configurations of control within procurement, deployment and integration processes.

The governance capacity of contracts lies in their ability to structure these means under real constraints,[10] as many AI-related risk areas (data governance, classification stability, liability alignment, regulatory adaptability) can already be addressed through existing legal instruments: *ad hoc* clauses, atypical structures, and emerging practices. The critical limitation is not the availability of legal tools, but the absence of systematic logic guiding their deliberate design and coordinated deployment across the AI value chain.

4. Contractual Governance as Instrumental and Rational Control Design

Having established contracts as rational instruments of governance, the methodological question becomes operational: how are governance objectives translated into binding contractual architectures? This section articulates the logical framework through which contractual risk governance can be structured.

In this case, contractual governance can operate through a constant five-stage logic: (1) identify the governance risk, (2) allocate responsibility to the party with capacity and control, (3) render governance verifiable through observation mechanisms, (4) define remedies enabling intervention before failures materialise, (5) preserve governance capacity as systems and regulations evolve. This logic reflects the instrumental rationality described earlier: it focuses on means (control mechanisms, verification pathways, intervention rights) rather than normative ends and each stage must produce observable, verifiable and enforceable commitments that translate governance expectations into operational reality.

Translating this sequential logic into enforceable contractual architecture requires designing four interconnected components, each of which operationalises one or more of the stages identified above across the lifecycle of the contractual

[9] Weber, M. (2019). *Economy and society: A new translation*. Harvard University Press.

[10] Zumbansen, P. (2007). The law of society: governance through contract. *Ind. J. Global Legal Stud.*, *14*, 191.

relationship: first, *control allocation mechanisms* that assign decision rights, oversight authority and technical access based on operational capacity rather than negotiating power; second, *verification infrastructure* through audit rights, reporting obligations and documentation duties that render governance processes observable; third, *adaptation mechanisms* that preserve contractual alignment as systems, regulations and use cases evolve (including notification duties, renegotiation triggers and termination pathways); fourth, *remedial structures* that graduate responses from technical adjustment to contractual termination, aligned with severity and reversibility of governance failures.

Central to this framework is the principle that contractual risk allocation must follow effective control. This constitutes the decisive departure from prevailing contracting practice, where AI-related obligations are typically distributed according to negotiating leverage, default liability templates, or regulatory categorisation alone. In such settings, deployers may formally bear accountability for system behaviour they cannot observe, while providers retain de facto control over data, architecture and update cycles without corresponding governance obligations. The resulting misalignment does not merely produce contractual inefficiency: it renders governance structurally inoperable, as the party charged with oversight lacks the informational access, technical capability or intervention authority to exercise it. The design question therefore precedes - and must discipline - negotiation: *which party can observe system behaviour, intervene when necessary, and verify compliance?* Contractual clauses become governance mechanisms only when they operationalise this capacity-aligned logic, structuring rights and obligations around actual technical and organisational configurations rather than inherited allocation patterns. Where this principle is not observed, contracts reproduce the same asymmetry they purport to govern[11].

Given these premises, the proposed framework seeks to instil a systematic logic into the often-fragmented landscape of current AI contracting. Where AI contracts rely on generic terms or sector-agnostic templates, governance processes remain unstructured and default allocation prevails (Section 3). The following chapter shows how this framework applies to recurring risk areas, translating the abstract design logic into concrete contractual mechanisms that function as operational governance infrastructure.

[11] Williamson, O. E. (1979). Transaction-Cost Economics: The Governance of Contractual Relations. *Journal of Law and Economics*, 22(2), 233-261; Klausner, M. (2015). Governance Mechanism in Long-term Contracts. In Grundmann, Möslein & Riesenhuber (Eds.), *Contract Governance* (pp. 237 ss.)

5. Operationalizing Contractual AI Governance: a Practical Framework

The contractual design framework articulated in the previous section now requires concrete demonstration. This section proposes a five-stage governance scheme based on the above-described logic to three distinct practical cases within the AI value chain: deployers managing regulatory compliance under asymmetric control; providers seeking to stabilise governance commitments at scale; and organisations building internal capacity to operationalise contracts as governance infrastructure. Each scenario is structured around the same analytical sequence, illustrating how abstract design principles translate into binding operational mechanisms.

Deployers face structural governance asymmetry: they bear primary AI Act obligations while critical control mechanisms-data usage, system updates, performance monitoring-remain upstream with providers. Governance failures in this context stem not from regulatory disregard, but from insufficient access to decision rights, information flows and intervention mechanisms. Contracts must therefore create the missing infrastructure enabling deployers to exercise oversight, verify compliance and intervene proportionally, as described in the table below.

Logical stage	(Possible) Contractual mechanisms	AI-related risk governed
1. Risk identification	• Defines permitted use cases in specification annex • Excludes unassessed uses through explicit prohibition clauses • Requires provider to furnish technical documentation for regulatory classification	Inability to identify or classify AI-related legal, technical and operational risks
2. Control allocation	• Establishes human oversight points with named accountable roles • Defines validation gates before deployment and material changes • Creates escalation protocols linking technical alerts to decision authority	Regulatory responsibility without the capacity to influence system behaviour
3. Verification	• Grants audit rights with defined scope, notice periods and access to facilities	Inability to verify compliance and safeguards over time

Pollicino AIdvisory
Via Visconti di Modrone 8
20122, Milano, Italy
info@pollicinoaidvisory.eu

	• Requires access to model cards, training data lineage, performance logs • Mandates incident notification within specified timeframes	
4. Remedy	• Graduated response framework: correction plan → suspension → termination • Suspension rights exercisable pending provider remediation • Managed disengagement protocol including data return and transition support	Lack of proportional intervention mechanisms before harm or non-compliance
5. Adaptation (recursive)	• Notification duty when system updates affect risk classification • Reclassification triggers requiring joint assessment • Renegotiation pathway or termination right if alignment cannot be restored	Governance obsolescence as risks and obligations change

Providers control data, model design and technical documentation, but face escalating transaction costs as deployers seek AI Act assurances through fragmented, bespoke negotiations. Without anticipatory contractual positioning, governance becomes reactive: repeated due diligence requests, ad hoc documentation demands, and case-by-case renegotiations undermine scalability. Contracts must function as governance-enabling infrastructure, stabilising expectations and managing regulatory evolution without sacrificing commercial viability:

Logical stage	(Possible) Contractual mechanisms	AI-related risk governed
1. Risk identification	• References standardised data governance policy document • Specifies data retention periods, training reuse restrictions, derived data handling • Provides transparency on model architecture and limitations via model cards	Uncertainty and inconsistency in how data-related risks are assessed across deployments

Pollicino AIdvisory
Via Visconti di Modrone 8
20122, Milano, Italy
info@pollicinoaidvisory.eu

2. Control allocation	• Anchors responsibility for model design, updates and versioning to provider • Defines provider obligations for documentation maintenance and accuracy • Clarifies boundaries between provider-controlled and deployer-controlled risk areas	Diffuse or contested responsibility for risks arising from upstream technical choices
3. Verification	• Links contractual commitments to reusable technical documentation packages • Standardises transparency artifacts: model cards, test reports, classification rationale • Enables deployer verification without duplicative bespoke due diligence	Repeated due diligence requests and inability to demonstrate governance maturity
4. Remedy	• Defines correction obligations with specified timelines for governance deviations • Proportional escalation: technical fix → service credit → managed exit • Structured off-boarding procedures preserving continuity during disputes	Escalation of disputes or termination driven by unstructured compliance failures
5. Adaptation (recursive)	• Notification protocols when regulatory landscape or system capabilities change • Reclassification procedures triggered by material system modifications • Amendment framework balancing regulatory adaptation with commercial stability	Contractual rigidity that crystallises regulatory risk as norms evolve

Organizations deploying AI systems often treat contracts as administrative artefacts managed in isolation by legal or procurement functions, disconnected

from risk management and technical operations. Governance choices occur downstream (after vendor selection and system integration) when the scope for shaping risk has narrowed. Without coordinated review processes, shared standards and cross-functional ownership, contracts remain static documents incapable of supporting systematic AI risk governance.

The challenge is organizational: building internal structures that activate, monitor and adapt contractual commitments as governance instruments:

Logical stage	(Possible) Contractual mechanisms	AI-related risk governed
1. Risk identification	• Mandatory pre-contract assessment involving legal, procurement, risk, technical functions • Standardised AI risk questionnaire completed before vendor selection • Cross-functional sign-off required before contractual commitment	Fragmented or incomplete identification of AI risks before vendor selection
2. Control allocation	• Internal ownership matrix assigning contractual obligations to specific roles • Decision rights protocol defining who can approve, suspend, or terminate • Coordination mechanism ensuring accountability spans legal, technical and business functions	Diffuse responsibility and lack of accountability for contractual governance
3. Verification	• Periodic contract portfolio review by cross-functional governance committee • Incident tracking system linking contractual terms to operational failures • Regular scanning of regulatory developments affecting existing agreements	Inability to monitor whether contractual safeguards remain effective over time

4. Remedy	• Predefined escalation paths: operational → management → executive • Corrective action playbook for common governance failures • Authority matrix enabling timely suspension or redesign decisions	Delayed or uncoordinated responses to governance failures
5. Adaptation (recursive)	• Scheduled template updates reflecting regulatory changes and lessons learned • Contract renewal triggers for legacy agreements predating governance standards • Organizational learning process feeding operational experience into policy evolution	Organizational rigidity and governance obsolescence as AI practices evolve

6. Policy Recommendations: Embedding Contractual Design in AI Governance

As this Policy Brief has argued, the core policy recommendation is that AI governance strategies, whether developed by regulators, enterprises, or industry bodies, should explicitly structure the use of contracts as instruments of risk governance rather than treating them as residual tools of liability allocation. In particular:

- Regulators and policymakers should embed contractual governance in AI Act implementation guidance. Compliance depends on operational mechanisms that regulations cannot directly enforce. Guidance should provide model contract clauses for high-risk systems and explicitly recognise contractual risk allocation as a primary compliance instrument[12]. Regulatory frameworks should incentivise providers who adopt standardised governance documentation and transparent data policies, reducing transaction costs and enabling systematic rather than reactive compliance;
- Enterprises deploying AI systems must coordinate legal, procurement, technical and risk functions in contract governance. AI-related risk cannot be managed by legal teams alone. Deployers should demand contracts that grant audit rights, technical access and oversight mechanisms aligned with

[12] As directly implied by Art. 25(4); Rec. 88, 89, 90 Reg. (EU) 2024/1689.

their actual control capacity. Contractual arrangements that assign regulatory obligations without corresponding control mechanisms make governance unmanageable, regardless of formal compliance;

- AI providers should treat governance infrastructure as a competitive advantage. Standardised documentation, transparent data policies and clear technical specifications reduce negotiation friction and enable scalable deployment. Provider contracts should explicitly define which responsibilities remain under provider control and which transfer to deployers. Clear allocation prevents disputes and enables both parties to fulfil their obligations effectively.

Across all stakeholder positions, effective contractual governance requires recursiveness. AI systems, regulatory requirements, and organisational capabilities evolve; governance arrangements must adapt accordingly. Contractual frameworks should incorporate notification duties, reclassification triggers, amendment procedures, and managed termination pathways that preserve alignment over time. Static contracts crystallise governance risk. Adaptive contracts enable sustained compliance and control.

7. Conclusions: Contracts as the Operational Layer of AI Governance

This Policy Brief has argued that AI governance cannot be achieved through regulatory compliance alone. While the AI Act establishes a necessary normative framework, its obligations presuppose operational capacities that remain unevenly distributed along the AI value chain. Governance is not determined by formal allocation of duties, but by who controls information, system behaviour and intervention mechanisms in practice.

Contracts emerge as the operational layer where governance becomes executable. By allocating control, defining verification mechanisms and enabling enforceable remedies, contractual architecture translates regulatory expectations into binding structures capable of operating ex ante and independently of enforcement timelines. Where contractual governance is not deliberately structured, risk is still allocated through inertia rather than design. The policy challenge is therefore not to invent new tools, but to structure their deliberate, systematic and capacity-aligned deployment across the AI value chain.