Journal of Engineering Innovation and Global Impact (JEIGI)
Volume 1, Issue 2 | PP 14-27 | www.jeigionline.com |

Submitted : 14-08-2026 | Revised : 25-09-2026 | Accepted : 03-10-2026 | Published : 13-10-2026

Governing Artificial Intelligence Through Public–Private Partnerships: A Conceptual Framework for Balancing Innovation and Accountability

Dr. Aparna Marwah

[2] Associate Professor, Department of Management Studies, Bharati Vidyapeeth (Deemed to be University), Institute of Management & Research (BVIMR), New Delhi, India.

Corresponding Author: Dr. Aparna Marwah
Email: aparna.marwah@bharatividyapeeth.edu

Abstract

Public–private partnerships have emerged as the operating tool for the frontier artificial intelligence governance. Some states that don't have the computing power, talent or model access to assess advanced systems are contracting out that need to the companies that develop them, and the deal is often called a "tradeoff between innovation and accountability." This concept paper does not accurately reflect what is being balanced. Classical partnership theory assumes that the public principal is competent, and he/she can define the service, check for its delivery, and renegotiate with an alternative supplier. The three assumptions of frontier AI governance are reversed in the public partner: They are unlikely to be able to define the safety themselves, they have no way to validate claims without the private partner's help, and they have no alternative source for the capability they need. The nature of the exchange is therefore not about risk in exchange for efficiency, as in an infrastructure partnership. Nonetheless, the responsibility for capability (not money) is what makes it show up on no balance sheet. The paper builds upon three constructs: capability asymmetry, verification dependence, and exit cost; and draws out its key dynamic: the high level of cooperation is voluntary and can be withdrawn at any time, for this reason, the level of public leverage is weakest at the times of highest suspicion, and the accountability is most available when most needed. There is a typology of four regimes between courtesy governance and sovereign verification and it situates current practice. For one reason alone, the two developments of 2025–2026, which go in opposite directions, are calibrated with the EU's binding high-risk regime being deferred, and the voluntary pre-deployment testing partnerships growing. The propositions and institutional design agenda are then presented. This paper is conceptual and does not contain any empirical data.

Keywords: artificial intelligence governance; public–private partnerships; accountability; regulatory capacity; frontier AI; verification

1. Introduction

Governance of frontier artificial intelligence has quietly come to rest on an instrument rather than a theory. States can learn, by contract with the manufacturer, whether a more advanced model is unsafe. The models are tested under voluntary arrangements by national evaluation bodies, firms provide pre-deployment access, researchers are seconded, compute is offered, and standards are co-drafted. It's a

public–private partnership in all but jargon, and it's being touted as "what we can do together that we can't do individually," "balancing innovation with accountability," and "pooling complementary strengths.

The vocabulary that is inherited from a different problem. The allocation of risk between the parties under a contract to which the public can write and enforce is the basis of PPP scholarship, with respect to infrastructure and public services, roads, hospitals, prisons, IT systems, and its analytical core (Grimsey & Lewis, 2004; Hodge & Greve, 2007). One recurring concern in the literature is that the accountability of partnerships is difficult to find: if many hands are involved in the execution of a public purpose, accountability can get lost, as is often reported as the accountability paradox of collaborative arrangements (Willems & Van Dooren, 2011; Bovens, 2007). Its corollary, however, is that there is a very competent public partner in the principal who can say what it is looking for, check to see if it has it, and seek out other possibilities if it has not. All the mechanisms, ranging from output specification to performance regime to the end of the contract, are built on that premise.

Frontier AI governance is guilty of three breaches of it. Because the public partner usually has no influence on what constitutes adequate safety evaluation for a system whose capabilities are only discovered, this is also a question of research (Anderljung et al., 2023; Bommasani et al., 2021). It can't confirm model weights, training data, internal testing, and compute required to perform testing at scale, matters that reside within the firm, which is why the state's picture of this system is to a large extent the picture that the firm provides (Whittlestone & Clark, 2021; Raji et al., 2020). Furthermore, it can't exist because there's no other source of access to this company's frontier model, and terminating them is a threat to the state's information supply. PPP theory analysis is not about a partnership where the agent only has the answer to determining if the principal's goals are being achieved. It's more like the inverse.

The odds are real in two developments in 2025-2026, which are often interpreted separately. The binding obligations of the AI Act, the main feature of the world's most comprehensive statutory framework, were delayed by the Digital Omnibus on AI, which was finally approved by the Council on 29 June 2026 (Council of the European Union, 2026), putting them off until 2 December 2027 for “stand-alone” high-risk systems and 2 August 2028 for “product-embedded” high-risk systems. The reason given was not one of lobbying by the industry but of institutional unreadiness (harmonised standards were not complete, national competent authorities were still not designated). In the voluntary space, government evaluation bodies grew: the United States' AI Safety Institute became the Center for AI Standards and Innovation (CAISI), a government agency focused on national security, in 2025, while the United Kingdom's institute was renamed the AI Security Institute, and in May 2026, pre-deployment testing agreements were expanded to other frontier developers, with the companies describing the arrangements as voluntary collaborations (Microsoft, 2026).

When read together these movements are mutually explanatory. Where the state is required to prove by the oath of a jury, it was sloped, for the machinery of proof was not in existence. It progresses where it is thus granted by permission, which is readily obtainable and easily arranged. The public comes to learn that there is an accountability of AI, but it is the accountability that the state is able.The public learns that there is accountability for AI, but it is the accountability that the state is able. This paper theorises that is the phenomenon.

The paper asks:

RQ1. What happens to the mechanisms of partnership theory when the public principal cannot specify, verify, or exit?

RQ2. Through what dynamic does the capability–accountability exchange evolve over a partnership's life, and when is public leverage greatest?

RQ3. What institutional designs preserve accountability while acquiring capability, and what do they cost?

Contributions: a structural diagnosis of the principal–agent inversion in AI governance partnerships (Section 4.1); the Verification Dependence Framework, built on capability asymmetry, verification dependence, and exit cost (Section 4.2); the inverse-leverage dynamiccooperation is most forthcoming

when least needed (Section 4.3); a four-regime typology locating current practice (Section 4.4); and seven propositions with a design agenda (Sections 4.5, 5.2). The paper is conceptual and reports no empirical data.

2. Literature Review

2.1 Partnership Theory and Its Competent Principal

In the evolution of PPP research, it was the shift in the state's role from provider to purchaser as part of the reorganisation of the state under new public management (NPM) that has led to the development of a contract and performance literature under contract and risk transfer (Grimsey & Lewis, 2004; Osborne, 2000). It has been a controversial and mixed success in terms of its empirical conclusions and conceptual scope. Accountability is further blurred because the word "partnership" is used to refer to project finance, policy networks and more. It is also known that partnerships have pathologies that are not technical but institutional and strategic (Bovaird, 2004; Hodge & Greve, 2007). Diffusion, the dispersion of responsibility across multiple hands so that no single hand has it (Willems & Van Dooren, 2011), is a major problem in the literature, the key to this paper. Diffusion assumes a requirement for accountability to be spread, that the public side could, in theory, have known what it was to ask for. It's the assumption of the competent principal and that's what frontier AI takes away.

2.2 Agency, Capture, and Regulatory Capacity

The vocabulary for agency theory is provided by information asymmetry, monitoring cost, and moral hazard (Eisenhardt, 1989). Regulation scholarship focuses on the failure modes: when the regulator becomes the regulated (Stigler, 1971), in a refined form as a set of mechanisms and remedies rather than an inevitability (Carpenter & Moss, 2014). Responsive regulation revealed that enforcement can be pyramid and that the modes of enforcement that are not cooperative can be reserved for the last resort, but that the central element of cooperation is the credibility of enforcement escalation (Ayres & Braithwaite, 1992). Polycentric-regime scholarship has shown that legitimacy and accountability need to be built between public and private nodes and not taken for granted (Black, 2008; Baldwin et al., 2012). The temporal dilemma this paper formalises, is captured by the pacing problem and technology outrunning oversight (Marchant et al., 2011) and Collingridge's (1980) dilemma, where control is easy but impact is unknown and hard once known.

2.3 AI Governance: Instruments Without a Partnership Theory

There are principles in AI governance research (Jobin et al., 2019), AI architectures (Cihon et al., 2020), and a research on the auditing practice (Raji et al., 2020). Specific to the frontier, regulatory proposal spree-pre-deployment evaluation, licensing, compute thresholds (Anderljung et al., 2023) and the case for governments to develop their own monitoring of the frontier (Whittlestone & Clark, 2021). EU's strategy and its trust in standards bodies and self-assessment have been subject of legal analysis (Veale & Borgesius Zuiderveen, 2021; Kaminski, 2019). This literature has not considered the partnership itself as the unit of analysis; it has been treated as a mere expedient, not as a form of organisation with a structure, a life cycle and predictable failure modes. The resulting gap is given in Table 1.

Table 1. *Partnership theory's assumptions against frontier AI conditions*

Assumption in classical PPP theory	Mechanism it enables	Condition in frontier AI governance	Consequence

Assumption in classical PPP theory	Mechanism it enables	Condition in frontier AI governance	Consequence
The principal can specify the service.	Output specification; contractual completeness	Adequate evaluation is an open research question	The agent co-writes the specification
The principal can verify delivery.	Performance regime; audit; penalties	Weights, data, internal evals, and compute sit inside the firm	Verification runs on the agent's disclosure
The principal can resort to alternatives.	Termination; competitive tendering	No alternative supplier of access to a given model	Exit threatens the principal's own information supply
Risk is transferable by contract.	Risk allocation as the core bargain	The salient risk (catastrophic capability) is not insurable or transferable.	The bargain's object changes from risk to access.
Accountability failure is diffusion.	Locate the many hands	Accountability may never have been constituted	Diffusion analysis misses the prior problem

3. Research Methodology

3.1 Design

It is a conceptual paper that involves theory transport/transformation, a well-known theory of public–private partnership is imported into a space outside of its scope conditions, and the theory is reconstructed around the challenge (Jaakkola, 2020). The mode of adaptation is appropriate here because the challenge is not one of the absence of PPP theory in AI governance, but rather one of its use without awareness of the theory's assumptions that have proved incorrect. Adequacy is measured by standard criteria- constructs, relations, justification, boundary conditions (Whetten, 1989) and by the need to predict when governance partnerships account for accountability and when they consume it (MacInnis, 2011).

3.2 Procedure

Four steps. First, the enabling assumptions of the PPP theory were made explicit and validated in the context of conditions at the edge of the field of AI, yielding Table 1 and the paper's diagnostic claim. Second, constructs were set as to capture what the inversion turns on asymmetry, dependence, exit, and defined to be, in principle, observable by third parties. Third, the framework's dynamics were found by asking when a voluntary partner's incentive to cooperate is highest, which resulted in the inverse-leverage result. Fourth, the typology was developed, and different propositions were formulated for each relation, with reference to actual arrangements.

3.3 Calibration Sources and Their Reflexivity

The paper is conceptual, but its claims are based on public descriptions of institutional record: EU AI Act and amended in 2026 (Council of the European Union, 2026) and public descriptions of pre-deployment testing arrangements (Microsoft, 2026). One aspect of that record is worthy of note, being a case of the paper's argument. Much of the public knowledge about frontier evaluation partnerships comes from the companies' announcements of the extent of access, the models included, and where the findings will be put. The evidence base for the analysis of verification dependence in turn is verification-dependent.What is studied as evidence of verification dependence is itself verification dependent. This isn't a methodological critique, it's data.

3.4 Scope and Boundary Conditions

The framework includes provisions on partnerships that allow states to become capable of governing the evaluation, standard-setting, monitoring, and compute and expertise needed. It does not address procurement of AI for public services, which is a typical contracting challenge with traditional solutions, unless otherwise indicated in Table 3. It does not model jurisdictions where the leverage comes only from market access, which is another, important case. No empirical data were gathered, quantities are not claimed.4. The Verification Dependence Framework

4.1 The Principal–Agent Inversion

Figure 1 is the claim about the structure. The classic model of partnership sees the state set the service's specifications, the private partner provide the service, and a principal oversee the service to be able to evaluate it, with accountability from the private partner to the public partner, and from the public partner to the citizens. In the case of AI governance partnership, the state has the knowledge of the system as a result of the articulation of the private partner, the private partner possesses the assessment capability and the artefact, while the accountability of the state towards citizenry is shaped by the disclosure options of the entity governed. The partnership isn't just a service, it's the epistemic infrastructure of accountability.

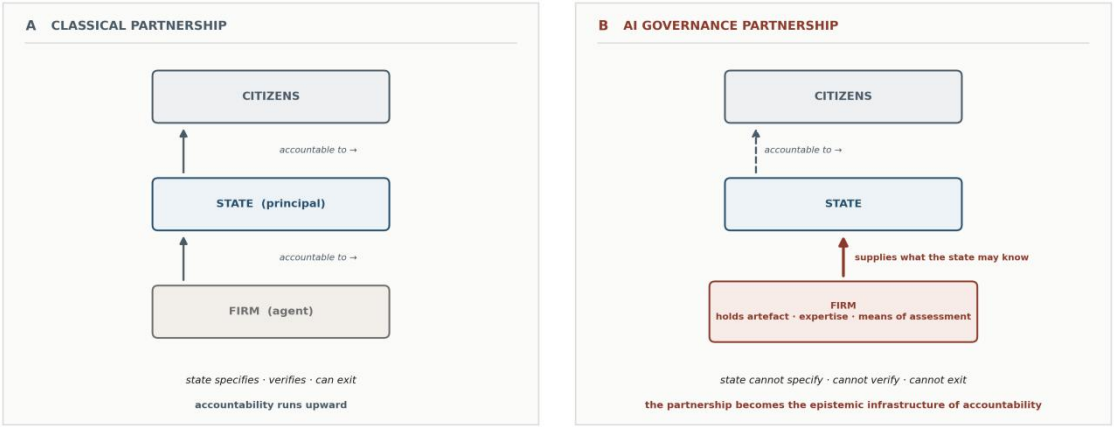


Figure 1. *The principal–agent inversion: accountability geometry in classical partnership and in AI governance partnership.*

This reframes the literature's central worry. The accountability paradox of collaboration holds that partnership disperses responsibility (Willems & Van Dooren, 2011). In frontier AI, the prior problem

is constitution rather than dispersion: the state cannot disperse an accountability it has not yet assembled, because assembling it requires a capability it lacks. Diffusion is a problem of governance's second act; AI governance is stuck in the first.

4.2 Constructs

Table 2. *Constructs of the Verification Dependence Framework*

Construct	Definition	Indicative observation
Capability asymmetry (CA)	The gap between private and public capacity to evaluate a system, compute, talent, tooling, model access	Public-body compute and specialist headcount relative to the frontier developer's; provenance of evaluation methods.
Verification dependence (VD)	The degree to which a public assessment cannot be reproduced without the private partner's continuing cooperation	Share of assessment steps requiring firm-granted access; reproducibility of findings by third parties
Exit cost (EC)	The cost to the public partner of terminating the arrangement, including loss of visibility and of accumulated expertise	Availability of alternative access routes; sunk institutional investment in firm-specific tooling
Cooperation incentive (CI)	The private partner's expected net benefit from cooperating in a given period	Reputational, market-access, and pre-emptive-regulation returns to cooperation, net of disclosure risk.
Effective accountability	The public partner's capacity to reach and enforce a finding contrary to the private partner's interest	Existence of adverse findings; enforceability of consequences following them

Effective accountability is the one thing that is clearly outlined in the adverse case. A relationship that's fruitful in agreeable findings tells nothing about accountability, as if the findings were agreeable, it must be true either that there was some genuine safety or that there was some managed disclosure. The differentiating question is, the arrangement can it create, and execute, a finding that the company does not wish to have. This definition is what makes the propositions of the framework testable, and is also why engagement style indicators (how many evaluations, which models covered, which agreements signed) are not evidence of accountability, although they may be evidence of capability.

4.3 The Inverse-Leverage Dynamic

The main outcome of the framework is related to timing, and it can be seen in Figure 2. Think about the cooperation incentive that the private partner can gain between the states of the world. If the model is not particularly interesting, then cooperation is more or less free, and its benefits are more or less positive: reputation, goodwill in front of regulators, a pre-emption of more difficult rules, etc. The firm cooperates and the partnership is good. When there is a serious issue with a model, the calculus is reversed: disclosure now has market, legal and competitive costs, and the state's instruments for demanding it are those that they are, a voluntary agreement. Cooperation is therefore most forthcoming when it is least needed, and least forthcoming when most it is needed. The most successful voluntary verification systems are in countries where they are not mandated.

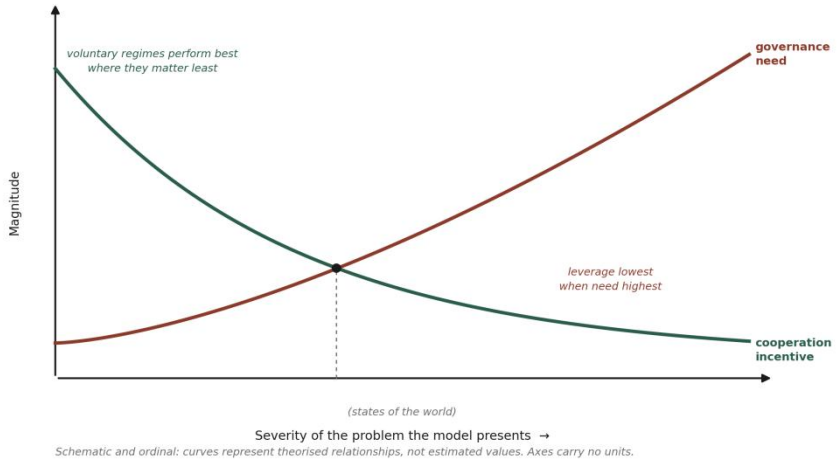


Figure 2. *The inverse-leverage dynamic: cooperation incentive and governance need to move in opposite directions across states of the world. Curves are schematic and ordinal, not estimated values.*

This isn't a statement of bad faith, and it's important. A firm can cooperate sincerely in each state; but it is a structural situation: the evidence of cooperator includes no way for an outsider to detect that he is an outside of a sincere cooperator. In these circumstances, the good behaviour of good actors is unprovable, and the cost of good behaviour goes to the good actors. Cooperative modes are effective when supported by credible escalation (Ayres & Braithwaite, 1992) and the importance of insight applies here. What a voluntary agreement doesn't have is escalation, and that's because it must not have to be able to escalate without cooperation, that is, if it has to make the capability rather than rent it.

The second dynamic occurs slower and is compounded by the first. In a more mature partnership, the state's tooling, knowledge and institutional routines become more specific to the partner's and exit costs increase, while rented capability does not become part of the public's capacity, since it is transferred back to the state when the secondee leaves and the access agreement ends. Both dependence and spending increase, and that of the sovereign, in particular, decreases; the most enduring product of the arrangement is the difficulty of getting out. This is the Collingridge (1980) scenario made institutional: when a state can easily have control systems that independently validate it is when it doesn't need it; when it needs it, the cost has increased.

4.4 A Typology of Governance Regimes

Two design choices are available to a state: whether capability is rented or built, and whether access rests on permission or statute. Table 3 crosses them.

Table 3. *Four regimes of AI governance partnership*

Regime	Capability × access	Description	Speed/cost	Characteristic weakness
Courtesy governance	Rented × voluntary	State evaluates by agreement, using access and often expertise supplied by the firm.	Fast, cheap	Inverse leverage in full force; nothing survives withdrawal
Compelled access	Rented × statutory	The state can demand artefacts but lacks the capacity to interpret them independently.	Moderate	Documents without comprehension; compliance theatre
Capable courtesy	Built × voluntary	The state has independent capability but no mandate; it proceeds by invitation	Slow, expensive	Capability idles where consent is withheld; strongest actors opt out
Sovereign verification	Built × statutory	The state can compel and independently interpret	Slowest, costliest	Political and fiscal durability; risk of ossification

The typology locates current practice. Frontier evaluation partnerships of the CAISI and AISI kind sit in courtesy governance: they are explicitly voluntary, and their expansion through 2026 to additional developers reflects the regime's genuine virtues- speed, technical seriousness, and access to unreleased systems that no statutory regime currently reaches (Microsoft, 2026). The EU's high-risk regime aims at sovereign verification and was deferred to 2027–2028 in part because the apparatus that regime requires- harmonised standards, designated competent authorities- was not ready (Council of the European Union, 2026). Figure 3 places these movements on a common timeline, and their opposition is the framework's calibration: the binding regime slipped for want of public capability. In contrast, the voluntary regime advanced because it needs none. Both facts are consequences of the same shortage.

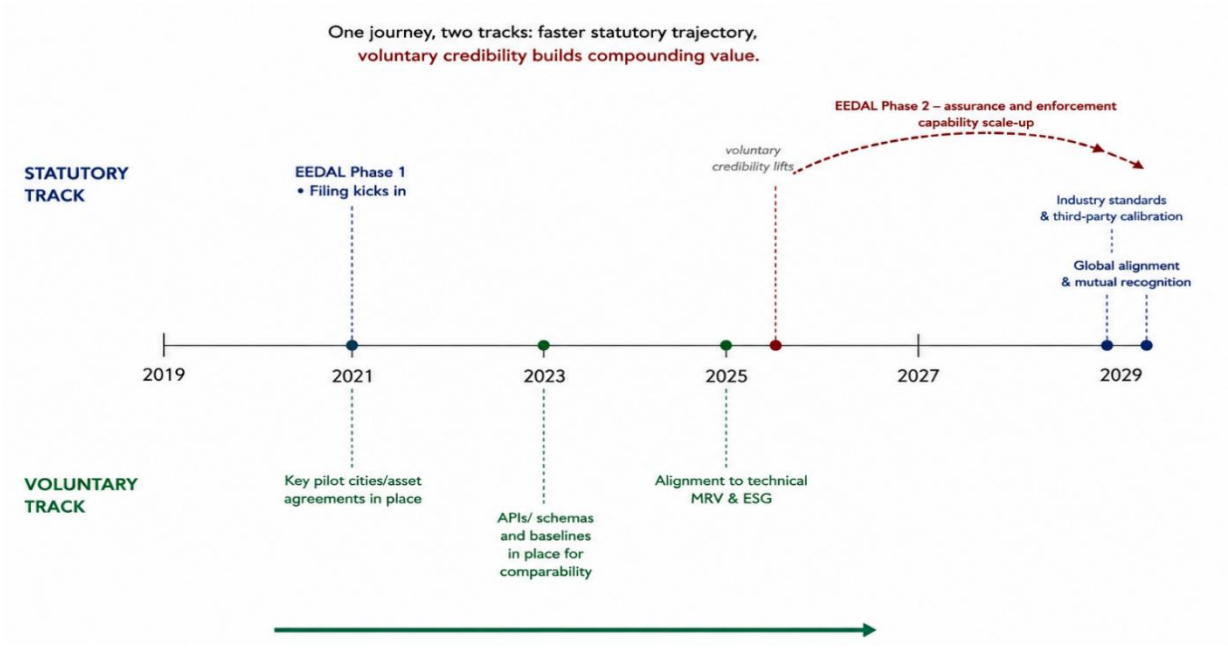


Figure 3. *Two regimes moving in opposite directions, 2024–2028: binding obligations deferred while voluntary evaluation partnerships expand. Dates from the public institutional record; placement of regimes is the authors' schematic reading.*

The typology also explains why courtesy governance is not a way-station. Rented capability doesn't build up, time in the courtesy cell does not advance a state towards sovereign verification, it pushes it back since there's no increase in public capacity and there's an increase in exit cost. Wait is only part of the institutional design challenge for transition, which needs conscious investment to run against the incentives of the arrangement itself.

Table 4. *Partnership forms and their accountability geometry*

Form	What the state obtains	What it surrenders	Where accountability lodges
Evaluation partnership	Model access; pre-deployment findings	Independence of assessment; timing control	With the firm's disclosure decisions
Capability secondment	Expertise, temporarily	Institutional memory; conflict-free staffing	Ambiguous evaluators return to industry
Compute and infrastructure partnership.	Research capacity	Path dependence on provider stacks	With procurement, weakly
Standard co-production	Legitimate, workable rules	Agenda control over the rule's content	Diffused across the standards body
Public-service AI procurement	Delivered systems	Little, if contracting is competent	Conventional, contract, and

Form	What the state obtains	What it surrenders	Where accountability lodges
			audit suffice

4.5 Propositions

Table 5. *Propositions of the framework*

ID	Proposition
P1	Where verification dependence is high, the probability of a public finding adverse to the private partner declines, independent of the partner's actual conduct.
P2	Cooperation incentive and governance need are inversely related across states of the world (inverse leverage); voluntary regimes perform best where they matter least.
P3	Exit cost rises monotonically with partnership tenure, while sovereign capability under rented arrangements does not; dependence and capacity diverge over time.
P4	Statutory access without built capability (compelled access) produces disclosure volume without adverse findings.
P5	The credibility of escalation, not the Intensity of cooperation, determines whether a partnership yields effective accountability.
P6	States that build capability during low-need periods reach sovereign verification at a lower total cost than states that transition under crisis.
P7	Public capability investment and binding-regime feasibility are complements: statutory regimes are deferred in proportion to the public verification capacity gap at the compliance date.

P7 is the framework's most directly testable claim against the record: it predicts that binding-regime slippage tracks the public sector's readiness rather than the technology's difficulty, and the stated grounds of the 2026 deferral standards and authority designation are consistent with it (Council of the European Union, 2026). A stronger test would compare deferral across jurisdictions with differing capacity investment.

5. Discussion

5.1 Theoretical Implications

Three literatures are revised in the framework. In PPP scholarship, it points to a class of partnership where the canonical accountability issue is not diffusion but constitution, and where the central bargain of the field, risk transfer, is replaced by an exchange of access; theory that has never needed to rely upon a scope condition. In competition regulation, it is suggested that the base of the pyramid of responsive regulation is not regulation power, but ability. Capability is what is at the bottom of the pyramid since in competition law, the base of the pyramid is the capacity to act without the assistance

of the regulated party (Ayres and Braithwaite, 1992). Furthermore, on the issue of AI governance, it takes the go/no-go dichotomy of the field and changes it into another one: rented versus built – a binding governance without public verification capability is forced access, and the framework suggests that this will result in documents instead of findings (P4), a finding that directly impacts the ordering of statutory regimes.

5.2 Practical and Institutional Design Implications

The constructs lead to the design agenda. Since exit cost is the "binding" pathology, it is important to design arrangements so as to be also "leavable": fixed terms, lack of any specific tooling monopoly, published evaluation methods, so that results can be reproduced by third parties, making exit cost a measurable one rather than a structural one. Secondment should always be accompanied by retention as there is no accumulation of rented capability. Voluntary arrangements should be linked to standing statutory reserve powers, the apex of the pyramid should be kept visible, though unused, and capability should be the base of the pyramid with powers being useless if public compute and staff are not invested concurrent. The framework acknowledges the inverse relationship between leverage and opportunity (P3), as well as between need and opportunity (Collingridge, 1980), and the quiet time is the time to build, but it's exactly when the political economy of building isn't easy.
None of this makes any sense as a criticism of the existing partnerships. Courtesy governance is providing the information that matters from real systems – even those not yet released – more quickly than any statutory system, and the idea is that the good will of the governed is the basis on which its virtues rest, and hence that it is an interim arrangement financed by an accountability debt that will eventually come due.

5.3 Limitations

The framework is conceptual and the constructs are demanding: one has to know what an assessment would have found in the absence of cooperation (a counterfactual) and adverse findings are what that defines effective accountability, but if the framework is correct, they will be few and difficult to observe. Market-access based leverage is not included in the scope conditions, and could greatly change the analysis of large jurisdictions, which can regulate without evaluating. The framework assumes that the private partner is a single strategic actor, thus understating the internal diversification – the safety staff within firms is often the same group as the one providing the "strategically supplied" disclosures. Furthermore, its calibration is based on two jurisdictions, and a limited temporal span; the inverse-leverage dynamic is a structural speculation, rather than any kind of regularity, because the cases where such an inverse-leverage effect would be felt, are precisely the ones the framework says we wouldn't see.

5.4 Future Research

Three priorities. First, capacity metrology: similar metrics of public verification capability, compute capacity, specialist inventory, method provenance, and repeatability of results that would otherwise prevent P7 from being tested and policy debate to be decided. Second, the adverse finding record: systematic investigation of the cases when governance partnerships have found against partner interest, what has followed, and a null result would be a confirmation of the framework's strongest and an important finding in the field. Third, comparative regime studies taking advantage of the natural variation now becoming available among jurisdictions that are statutory and voluntary and comparing their P7 and P6 records with their deferral and enforcement records. A fourth, less speculative, direction relates to internal heterogeneity: the firm as a coalition instead of an actor may offer an explanation for the occurrence of disclosure when it is not in the best interest of the firm.

6. Conclusion

The public–private partnership (P-PP) kind of entered AI governance as a response to a crisis: States needed to gain an insight to systems they were not able to assess and companies were willing to provide it. On the one hand the offer was taken up in the language of balance and innovation, on the other in the language of accountability, and in the language of partnership on the scales. This paper has claimed that the metaphor hides this transaction. The ability to supply a partnership is the ability to provide independence in the assessment, not the money it costs, and the money component of the assessment never appears in a budget. The resulting bind: leverage comes before dependence, generosity of cooperation comes when it's least expected, and rented capability (when it leaves) leaves nothing behind. The shape of the 2025-2026 record is a binding regime withheld due to a lack of public apparatus and voluntary testing growing, two facts with one cause. None of this is a mistake; none of this is an in-terim; and in-temporary is the same as in-permanent, except that nobody prices in-temporary. The only question a state should ask of any AI governance partnership is whether it is working, and in the states in the world where it's easy, it will seem to work. Rather, the question the state should ask is what does the state know if the partner doesn't respond, and the answer is that it doesn't know anything.

References

1. Anderljung, M., Barnhart, J., Korinek, A., Leung, J., O'Keefe, C., Whittlestone, J., Avin, S., Brundage, M., Bullock, J., Cass-Beggs, D., Chang, B., Collins, T., Fist, T., Hadfield, G., Hayes, A., Ho, L., Hooker, S., Horvitz, E., Kolt, N., … Wolf, K. (2023). Frontier AI regulation: Managing emerging risks to public safety (arXiv:2307.03718). arXiv. https://doi.org/10.48550/arXiv.2307.03718
2. Ayres, I., & Braithwaite, J. (1992). Responsive regulation: Transcending the deregulation debate. Oxford University Press.
3. Baldwin, R., Cave, M., & Lodge, M. (2012). Understanding regulation: Theory, strategy, and practice (2nd ed.). Oxford University Press.
4. Black, J. (2008). Constructing and contesting legitimacy and accountability in polycentric regulatory regimes. Regulation & Governance, 2(2), 137–164. https://doi.org/10.1111/j.1748-5991.2008.00034.x
5. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021). On the opportunities and risks of foundation models (arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258
6. Bovaird, T. (2004). Public–private partnerships: From contested concepts to prevalent practice. International Review of Administrative Sciences, 70(2), 199–215. https://doi.org/10.1177/0020852304044250
7. Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. European Law Journal, 13(4), 447–468. https://doi.org/10.1111/j.1468-0386.2007.00378.x
8. Carpenter, D., & Moss, D. A. (Eds.). (2014). Preventing regulatory capture: Special interest influence and how to limit it. Cambridge University Press.
9. Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the future: Investigating architectures for international AI governance. Global Policy, 11(5), 545–556. https://doi.org/10.1111/1758-5899.12890
10. Collingridge, D. (1980). The social control of technology. Frances Pinter.
11. Council of the European Union. (2026, June 29). Artificial intelligence: Council gives final green light to simplify and streamline rules [Press release]. https://www.consilium.europa.eu/en/press/press-releases/2026/06/29/artificial-intelligence-council-gives-final-green-light-to-simplify-and-streamline-rules/

12. Eisenhardt, K. M. (1989). Agency theory: An assessment and review. Academy of Management Review, 14(1), 57–74. https://doi.org/10.5465/amr.1989.4279003
13. Grimsey, D., & Lewis, M. K. (2004). Public-private partnerships: The worldwide revolution in infrastructure provision and project finance. Edward Elgar.
14. Hodge, G. A., & Greve, C. (2007). Public–private partnerships: An international performance review. Public Administration Review, 67(3), 545–558. https://doi.org/10.1111/j.1540-6210.2007.00736.x
15. Hood, C. (1991). A public management for all seasons? Public Administration, 69(1), 3–19. https://doi.org/10.1111/j.1467-9299.1991.tb00779.x
16. Jaakkola, E. (2020). Designing conceptual articles: Four approaches. AMS Review, 10(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0
17. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2
18. Kaminski, M. E. (2019). Binary governance: Lessons from the GDPR's approach to algorithmic accountability. Southern California Law Review, 92(6), 1529–1616.
19. Klijn, E.-H., & Teisman, G. R. (2003). Institutional and strategic barriers to public–private partnership: An analysis of Dutch cases. Public Money & Management, 23(3), 137–146. https://doi.org/10.1111/1467-9302.00361
20. MacInnis, D. J. (2011). A framework for conceptual contributions in marketing. Journal of Marketing, 75(4), 136–154. https://doi.org/10.1509/jmkg.75.4.136
21. Marchant, G. E., Allenby, B. R., & Herkert, J. R. (Eds.). (2011). The growing gap between emerging technologies and legal-ethical oversight: The pacing problem. Springer. https://doi.org/10.1007/978-94-007-1356-7
22. Microsoft. (2026, May 5). Advancing AI evaluation with the Center for AI Standards and Innovation (US) and the AI Security Institute (UK). Microsoft On the Issues. https://blogs.microsoft.com/on-the-issues/2026/05/05/advancing-ai-evaluation-with-the-center-for-ai-standards-us-and-innovation-and-the-ai-security-institute-uk/
23. Osborne, S. P. (Ed.). (2000). Public–private partnerships: Theory and practice in international perspective. Routledge.
24. Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 33–44). ACM. https://doi.org/10.1145/3351095.3372873
25. Stigler, G. J. (1971). The theory of economic regulation. The Bell Journal of Economics and Management Science, 2(1), 3–21. https://doi.org/10.2307/3003160
26. Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402
27. Whetten, D. A. (1989). What constitutes a theoretical contribution? Academy of Management Review, 14(4), 490–495. https://doi.org/10.5465/amr.1989.4308371
28. Whittlestone, J., & Clark, J. (2021). Why and how governments should monitor AI development (arXiv:2108.12427). arXiv. https://doi.org/10.48550/arXiv.2108.12427
29. Willems, T., & Van Dooren, W. (2011). Lost in diffusion? How collaborative arrangements lead to an accountability paradox. International Review of Administrative Sciences, 77(3), 505–530. https://doi.org/10.1177/0020852311408648

Declarations

Conflict of Interest

The authors declare no conflicts of interest related to this study.

Funding Statement



This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical Approval

This study did not involve any human or animal subjects and, therefore, did not require ethical approval.