

Coordinating Fragmented Global AI Governance: Control, Rights, and Safety as a Regime Complex

hana kim

tea112@snu.ac.kr

Korea Advanced Institute of Science and Technology

Research Article

Keywords: artificial intelligence, policy coordination, global governance, policy diffusion, regulatory governance, regime complexity

Posted Date: May 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9374171/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Policy Sciences on September 23rd, 2026.
See the published version at <https://doi.org/10.1007/s11077-026-09632-w>.

Abstract

Artificial intelligence governance is often discussed as if it were a single global policy domain. In practice, however, governments confront a coordination problem across partially overlapping agendas pursued through different instruments by different actors. This research note argues that global AI governance is better understood as a regime complex composed of three partially overlapping regimes: Control, Rights, and Safety. Using an Overton-derived corpus of 198,823 AI-related policy documents and a monthly actor-by-regime panel for 1994–2025, I compare regime trajectories, leadership shares, temporal alignment, and short-window state uptake around selected policy moments. Three findings stand out. Rights is the clearest IGO-linked regime and shows the strongest short-window uptake signal after the OECD AI Principles, though not robust evidence of durable convergence. Safety diffuses through a hybrid public-private infrastructure linking states, firms, and standards bodies. Control is the weakest empirical case and is not predominantly state-led in the expected way. The implication is straightforward: effective AI governance requires regime-specific coordination strategies rather than a single omnibus approach.

Introduction

Artificial intelligence (AI) governance is expanding rapidly, but not as a single coherent policy field. Governments, intergovernmental organizations (IGOs), standards bodies, firms, and expert networks now produce strategies, principles, safety frameworks, technical standards, audit guidance, and strategic control measures under the shared label of "AI governance" (Jobin et al., 2019; Fjeld et al., 2020; Ulnicane & Erkkila, 2023). The result is not simply growth. It is a coordination problem across overlapping agendas that are only partly aligned.

Recent research has mapped important parts of this landscape, including agenda setting, national AI strategies, global indicators, standards, and institutional fragmentation (Schiff, 2023; Erkkila, 2023; Kim, 2023; Cihon et al., 2020; Tallberg et al., 2023). Yet these debates still often treat AI governance as one domain. That makes it difficult to see which venues matter for which policy tasks, where templates travel, and why document growth can mask divergent governance pathways.

In policy-sciences terms, the issue is not fragmentation alone but coordination, coherence, and partial integration across heterogeneous instruments and organizations (Peters, 1998; Howlett & Rayner, 2007; Cejudo & Michel, 2017). This note argues that global AI governance is better understood as a regime complex composed of three partially overlapping regimes: Control, Rights, and Safety. Control centers on strategic restriction, security, and capacity. Rights centers on legality, due process, privacy, and nondiscrimination. Safety centers on testing, evaluation, assurance, and risk management.

Distinguishing these regimes is useful because each mobilizes different actor coalitions, favors different instruments, and creates different coordination burdens. Rather than ask who leads "AI governance" in the aggregate, the analysis asks which actors lead which regime, through which policy objects, and with

what kind of downstream uptake. The aim is therefore not simply to describe fragmentation, but to show that fragmentation in global AI governance is structured.

To make that argument concrete, I use a large Overton-derived corpus of AI-related policy documents. Overton is a large-scale database that aggregates policy documents produced by governments, IGOs, think tanks, and other policy-relevant organizations across countries. It is especially useful here because it captures not only formal law, but also the broader documentary infrastructure through which governance travels: strategies, recommendations, standards, guidance, and reports. Three findings organize the note. Rights is the clearest IGO-linked regime and shows the strongest short-window uptake signal after the OECD AI Principles, though not robust evidence of durable convergence. Safety diffuses through a hybrid public-private infrastructure linking states, firms, and standards bodies. Control is the weakest empirical case and is not predominantly state-led in the expected way.

AI governance as a regime complex

Policy sciences has long treated complex public problems as coordination problems rather than as questions solved by one institution or one statute (Peters, 1998; Howlett & Rayner, 2007). AI governance fits this pattern. The field expands through a heterogeneous mix of strategies, standards, principles, audits, rights-based rules, and strategic restrictions issued by multiple public and private actors. In such settings, policy activity alone is not evidence of coherence (Cejudo & Michel, 2017).

A useful way to organize this complexity is to conceptualize global AI governance as a regime complex. Here, a regime is a relatively coherent bundle of problem definitions, normative priorities, favored instruments, and recurrent sponsor coalitions. A regime complex exists when several such regimes overlap without a clear hierarchy (Raustiala & Victor, 2004; Alter & Meunier, 2009; Keohane & Victor, 2011; Orsini et al., 2013). This framing helps make fragmentation more diagnosable: not whether AI governance is fragmented in the abstract, but how different governance projects are assembled and how they interact.

The three regimes differ in ways that matter for policy design. Control frames AI as a matter of strategic capacity, security, and restriction. Rights frames AI as a question of legality, accountability, privacy, and nondiscrimination. Safety frames AI as a challenge of testing, assurance, and risk management. These regimes are not mutually exclusive, but they are not interchangeable. Their governance infrastructures also differ. Rights travels heavily through principles, accountability language, and legal templates, while Safety relies more on standards, benchmarks, and assurance routines (Black, 2002a, 2002b; Abbott & Snidal, 2010; Buthe & Mattli, 2011; Tabassi, 2023; von Ingersleben-Seip, 2023).

This distinction also helps clarify what diffuses in contemporary AI governance. In many domains, what travels first is not a statute, but a reusable policy object: a reporting category, a testing routine, a governance template, or a problem definition (Simmons et al., 2006; Shipan & Volden, 2008; Gilardi, 2010). In AI governance, the portability of principles, ethics frameworks, rights vocabularies, and model

risk-management routines is already well documented (Jobin et al., 2019; Fjeld et al., 2020; UNESCO, 2021; Veale & Borgesius, 2021).

Recent work on fragmented architectures and generative-AI governance likewise suggests that power, venue choice, and policy scripts do not align neatly across issue areas (Cihon et al., 2020; Tallberg et al., 2023; Taeihagh, 2025; Ulnicane, 2025). A regime-complex lens therefore helps separate aggregate document growth from more specific coordination pathways.

A brief empirical probe

The evidence presented here is documentary rather than implementation-based. The corpus begins with 198,823 unique AI-related policy documents derived from Overton and uses a consistent metadata text blob built from titles and indexed topical fields. Using the same metadata representation across the full corpus improves comparability across source types, even if it is thinner than full text (Grimmer & Stewart, 2013).

Table 1. Control, Rights, and Safety as distinct coordination logics in global AI governance.

Regime	Core problem frame	Favored instruments	Typical leading actor families
Control	State capacity, security, strategic control of AI capabilities	Restrictions and controls; centralized oversight; surveillance/security governance; export/compute/data controls	National governments; security agencies; some strategic industry alliances
Rights	Legality, due process, privacy, non-discrimination, fundamental rights	Rights-based requirements; transparency and accountability duties; impact assessments; enforcement and redress mechanisms	IGOs; national regulators; civil society; legal communities
Safety	Reliability, robustness, alignment, systemic risk management	Risk-management frameworks; evaluation and assurance; audits; incident reporting; technical standards	Standards bodies; technical communities; firms; multi-stakeholder initiatives

The corpus begins with 198,823 unique AI-related policy documents derived from Overton. For the regime diagnostics, I use a filtered 1994-2025 analytic sample of 75,053 documents with valid dates, usable metadata, and at least one regime hit. Each document is represented by a standardized metadata text blob combining title and indexed topical fields. Using the same metadata representation across the full corpus improves comparability across source types, even if it is thinner than full text.

Documents are grouped into five broad actor families: State, IGO, Industry, Standards/technical bodies, and Civil society/Experts. Regime membership is assigned through transparent metadata-based lexicons for Control, Rights, and Safety. The coding is intentionally conservative. A regime hit should

therefore be read as evidence that a governance logic is salient enough to appear in the document's title or indexing fields, not as proof that it dominates the full text.

For each actor family and month, regime salience is measured as the share of that actor family's documents that register a regime hit. From these actor-by-month series I derive three descriptive indicators: trajectories, leadership shares, and lead-lag alignments. These measures are diagnostic rather than causal. They help distinguish document growth from within-actor emphasis, and simultaneous movement from patterned precedence.

To complement this descriptive mapping, I also estimate simple jurisdiction-month panel probes on state documents around selected policy moments: the OECD AI Principles for Rights, a 2022 export-control shock proxy for Control, and a 2023 frontier-safety coordination proxy for Safety. These probes are intended to detect short-window uptake, not to establish dyadic copying or durable convergence.

Detailed lexicons, actor-coding rules, network-construction steps, and robustness checks are reported in the appendix. The main text keeps the empirical strategy intentionally compact, in keeping with the purpose of a research note.

Regime patterns

The evidence is asymmetric across regimes. The four sets of results below ask how the regimes evolve over time, which actor families most often lead them, where temporal alignment appears, and whether selected policy moments are followed by short-window uptake in state documents.

Regime trajectories

Figure 1 shows that the three regimes accumulate on top of one another rather than appearing as a simple sequence. Control is the most expansive regime in the analytic sample, matching 65.8% of regime-bearing documents. Rights and Safety each account for roughly one quarter of the sample, but their temporal profiles differ markedly.

The key point is that aggregate growth does not imply a single pathway of coordination. Rights rises earlier and then stabilizes. Safety accelerates after 2020. Control grows throughout the period and becomes especially prominent after 2016 and again after 2022. The field therefore looks layered rather than sequential.

Leadership across actor families

Leadership is clearly regime-contingent. Rights is the most consistently IGO-led regime. Safety shows a more hybrid pattern, with states remaining central but industry and standards/technical bodies becoming increasingly important after 2016. Control is the most surprising case: leadership is mixed

and shifts away from a predominantly state-centered pattern. These contrasts matter because coordination capacity depends less on absolute volume than on who supplies reusable templates and interpretive authority within each regime.

Temporal alignment

Lead-lag diagnostics point to different coordination pathways across the three regimes. Control does not display a clean state-to-state cascade; the strongest patterned alignments cluster instead among standards/technical bodies, industry, and expert intermediaries. Rights combines IGO-linked agenda setting with broader downstream reuse across actor types. Safety is the clearest hybrid configuration, with repeated alignment among states, firms, and standards bodies. These diagnostics are descriptive rather than causal, but they indicate that the three regimes rely on different coordination infrastructures.

Cells report the OECD x Post coefficient, p-value, and pretrend p-value for each event window. Shading summarizes the substantive reading used in the text.

Rights provides the clearest short-window uptake signal, but not robust evidence of durable convergence. In the 12-month window around the May 2019 OECD AI Principles, the OECD x Post coefficient is positive and statistically significant, with acceptable pretrends. The signal weakens at 18 months and becomes harder to interpret at 24 months because pretrend problems emerge.

Control and Safety are mixed in different ways, but Control is clearly the weaker empirical case. The Control probe shows no clear short-window differential and becomes only suggestive in wider windows. Safety shows a positive 12-month differential around the 2023 safety shock proxy, but the estimate weakens across broader windows. The panel evidence therefore reinforces the broader descriptive pattern: Rights is the strongest short-window case, Safety is suggestive but unstable, and Control remains comparatively unsettled.

Appendix Table A3 narrows the Rights claim further. The positive result is concentrated in the short-window log-count specification, while composition-based outcomes, matched-sample estimates, and placebo timings do not support stronger claims. The finding is therefore best read as evidence of short-window uptake, not definitive convergence.

Volume and emphasis

Absolute counts and within-family shares help separate volume from emphasis. Because state documents constitute 60.8% of the deduplicated corpus, states contribute the largest absolute number of regime-hit documents across all three regimes. Yet within-family emphasis tells a different story. Control language is especially concentrated in industry, civil-society/expert, and standards documents rather than in state documents alone.

Rights vocabulary is most concentrated in IGO documents, while Safety vocabulary is comparatively concentrated in state, IGO, and standards/technical documents. These contrasts reinforce the broader point of the note: document volume, agenda leadership, and within-family emphasis are related but not identical dimensions of governance influence.

What the evidence supports

The combined evidence is asymmetric across regimes. Rights is the strongest empirical case: it shows the clearest IGO-linked agenda leadership and the clearest short-window uptake signal, though not robust evidence of durable convergence. Safety is best understood as a hybrid public-private technical regime. Control is the weakest case: it is active, but its leadership and uptake patterns do not support a simple state-led diffusion story.

Implications for coordination

These findings recast a central policy-science question: how should policymakers structure and coordinate problems that exceed any one organization, instrument, or venue? In AI governance, the answer is not to search for one master regime. It is to identify where different governance problems cluster, which actor ecologies carry them, and how interfaces can be built across partially overlapping regimes (Peters, 1998; Howlett & Rayner, 2007; Cejudo & Michel, 2017).

The main value of regime decomposition is diagnostic. Before asking whether AI governance is converging, policymakers need to ask which governance problem they are actually trying to coordinate. Rights, Safety, and Control do not move through the same infrastructures, and they do not create the same coordination burdens.

Rights is the clearest case of template-mediated governance. Its IGO-linked leadership and short-window uptake pattern suggest that multilateral principle production and interpretive guidance remain especially important when the policy object is portable language about legality, accountability, and non-discrimination (Jobin et al., 2019; Fjeld et al., 2020; UNESCO, 2021).

Safety looks different. The recurring role of firms and standards bodies shows that implementable governance often depends on assurance infrastructures - benchmarks, testing routines, reporting formats, and evaluative standards - that no single public actor fully controls (Black, 2002b; Buthe & Mattli, 2011; Tabassi, 2023; von Ingersleben-Seip, 2023). For policy design, that implies a need for interfaces between regulators, standards bodies, auditors, and firms rather than a purely state-centric model.

Control is the most cautionary regime. Strategic-control discourse is politically salient, but the documentary evidence does not show a clean or stable state-led cascade. That caution is consistent

with broader observations that AI governance remains fragmented across institutions, coalitions, and geopolitical agendas (Cihon et al., 2020; Tallberg et al., 2023; Taeihagh, 2025; Ulnicane, 2025).

More broadly, the results suggest three design lessons. First, separate document growth from coordination: more texts do not necessarily mean greater coherence. Second, separate agenda leadership from implementation uptake: the actors setting vocabularies are not always the actors converting them into state practice. Third, build explicit boundary-management mechanisms across regimes, because tensions among control, rights, and safety are substantive rather than merely administrative.

Limitations

This note favors breadth over close process tracing, and several limitations follow. Corpus coverage depends on document availability and Overton indexing, so some regions and language communities are likely better represented than others. The analysis is document-centered: it captures what actors publish and circulate, not necessarily what they implement or enforce. These are common tradeoffs in large-scale text-based policy analysis (Grimmer & Stewart, 2013).

Event-based uptake

Figure 2 summarizes simple jurisdiction-month event-based probes on state documents. These estimates should be read as shock-responsive uptake tests rather than as definitive causal diffusion coefficients.

Conclusion

This research note has argued that the policy problem in global AI governance is not simply proliferation but coordination across overlapping regimes. Treating AI governance as a regime complex composed of Control, Rights, and Safety makes that problem more tractable. It reveals that Rights is the clearest IGO-linked regime with the strongest short-window uptake signal, Safety is a hybrid public-private technical regime, and Control is the weakest and least consolidated empirical case.

The practical implication is straightforward. Effective AI governance is unlikely to come from a single omnibus strategy. It requires regime-specific coordination strategies, attention to the different infrastructures through which governance travels, and more explicit management of the boundaries between control, rights, and safety.

Declarations

Funding :

not applicable

Author Contribution

Hana Kim wrote the whole paper.

References

1. Abbott, K. W., & Snidal, D. (2010). International regulation without international government: Improving IO performance through orchestration. *The Review of International Organizations*, 5(3), 315–344. <https://doi.org/10.1007/s11558-010-9092-3>
2. Alter, K. J., & Meunier, S. (2009). The politics of international regime complexity. *Perspectives on Politics*, 7(1), 13–24. <https://doi.org/10.1017/S1537592709090033>
3. Black, J. (2002a). Critical reflections on regulation. *Australian Journal of Legal Philosophy*, 27, 1–35.
4. Black, J. (2002b). Regulatory conversations. *Journal of Law and Society*, 29(1), 163–196. <https://doi.org/10.1111/1467-6478.00215>
5. Bütte, T., & Mattli, W. (2011). *The new global rulers: The privatization of regulation in the world economy*. Princeton University Press.
6. Cejudo, G. M., & Michel, C. L. (2017). Addressing fragmented government action: Coordination, coherence, and integration. *Policy Sciences*, 50(4), 745–767. <https://doi.org/10.1007/s11077-017-9281-5>
7. Cihon, P., Maas, M. M., & Kemp, L. (2020). Fragmentation and the future: Investigating architectures for international AI governance. *Global Policy*, 11(4), 455–461.
8. Erkkilä, T. (2023). Global indicators and AI policy: Metrics, policy scripts, and narratives. *Review of Policy Research*, 40(5), 811–839. <https://doi.org/10.1111/ropr.12556>
9. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication No. 2020-1. <https://cyber.harvard.edu/publication/2020/principled-ai>
10. Gilardi, F. (2010). Who learns from what in policy diffusion processes? *American Journal of Political Science*, 54(3), 650–666. <https://doi.org/10.1111/j.1540-5907.2010.00452.x>
11. Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297. <https://doi.org/10.1093/pan/mps028>
12. Howlett, M., & Rayner, J. (2007). Design principles for policy mixes: Cohesion and coherence in 'new governance arrangements'. *Policy and Society*, 26(4), 1–18. [https://doi.org/10.1016/S1449-4035\(07\)70118-2](https://doi.org/10.1016/S1449-4035(07)70118-2)
13. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. <https://doi.org/10.1038/s42256-019-0088-2>

14. Keohane, R. O., & Victor, D. G. (2011). The regime complex for climate change. *Perspectives on Politics*, 9(1), 7–23. <https://doi.org/10.1017/S1537592710004068>
15. Kim, J. (2023). Traveling AI-essentialism and national AI strategies: A comparison between South Korea and France. *Review of Policy Research*, 40(5), 705–728. <https://doi.org/10.1111/ropr.12552>
16. Orsini, A., Morin, J.-F., & Young, O. R. (2013). Regime complexes: A buzz, a boom, or a boost for global governance? *Global Governance*, 19(1), 27–39.
17. Peters, B. G. (1998). Managing horizontal government: The politics of co-ordination. *Public Administration*, 76(2), 295–311. <https://doi.org/10.1111/1467-9299.00102>
18. Raustiala, K., & Victor, D. G. (2004). The regime complex for plant genetic resources. *International Organization*, 58(2), 277–309. <https://doi.org/10.1017/S0020818304582036>
19. Schiff, D. (2023). Looking through a policy window with tinted glasses: Setting the agenda for U.S. AI policy. *Review of Policy Research*, 40(5), 729–756. <https://doi.org/10.1111/ropr.12535>
20. Shipan, C. R., & Volden, C. (2008). The mechanisms of policy diffusion. *American Journal of Political Science*, 52(4), 840–857. <https://doi.org/10.1111/j.1540-5907.2008.00346.x>
21. Simmons, B. A., Dobbin, F., & Garrett, G. (2006). Introduction: The International Diffusion of Liberalism. *International Organization*, 60(4), 781–810. <https://doi.org/10.1017/S0020818306060267>
22. Tabassi, E. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>
23. Taeihagh, A. (2025). Governance of generative AI. *Policy and Society*, 44(1), 1–22. <https://doi.org/10.1093/polsoc/puaf001>
24. Tallberg, J., Erman, E., Furendal, M., Geith, J., Klamberg, M., & Lundgren, M. (2023). The global governance of artificial intelligence: Next steps for empirical and normative research. *International Studies Review*, 25(3), viad040. <https://doi.org/10.1093/isr/viad040>
25. Ulnicane, I. (2025). Governance fix? Power and politics in controversies about governing generative AI. *Policy and Society*, 44(1), 70–84. <https://doi.org/10.1093/polsoc/puae022>
26. Ulnicane, I., & Erkkilä, T. (2023). Politics and policy of artificial intelligence. *Review of Policy Research*, 40(5), 612–625. <https://doi.org/10.1111/ropr.12574>
27. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. UNESCO. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>
28. Veale, M., & Borgesius, F. J. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act-Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97–112.
29. von Ingersleben-Seip, N. (2023). Competition and cooperation in artificial intelligence standard setting: Explaining emergent patterns. *Review of Policy Research*, 40(5), 781–810. <https://doi.org/10.1111/ropr.12538>

Figures

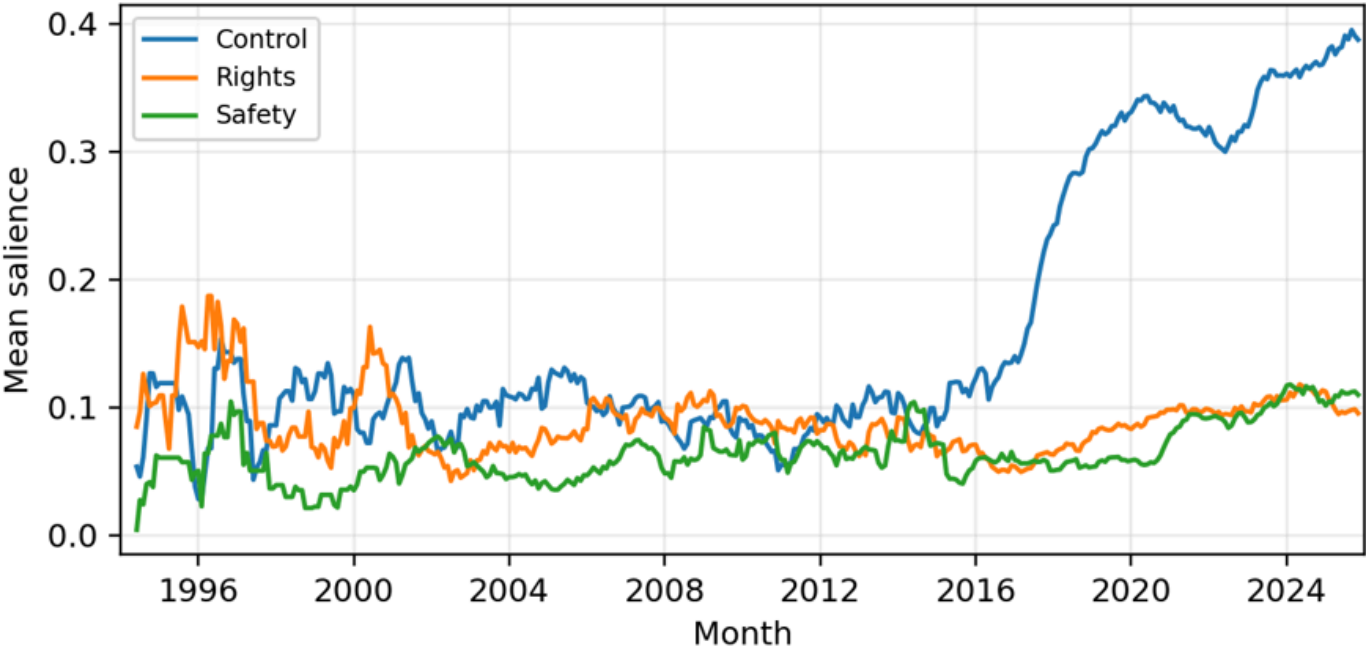


Figure 1

Regime salience over time (metadata-based analytic corpus; 1994–2025).
Higher salience corresponds to denser articulation of implementable governance tools within each regime.

	$\pm 12m$	$\pm 18m$	$\pm 24m$
Rights	$\beta = 0.056$ $p = 0.047$ pre = 0.188 Positive signal	$\beta = 0.067$ $p = 0.055$ pre = 0.070 Marginal	$\beta = 0.124$ $p = 0.007$ pre = 0.007 Pretrend problem
Control	$\beta = 0.049$ $p = 0.368$ pre = 0.321 No clear differential	$\beta = 0.100$ $p = 0.060$ pre = 0.100 Suggestive only	$\beta = 0.098$ $p = 0.080$ pre = 0.013 Pretrend problem
Safety	$\beta = 0.114$ $p = 0.037$ pre = 0.143 Positive signal	$\beta = 0.098$ $p = 0.054$ pre = 0.096 Suggestive only	$\beta = 0.078$ $p = 0.113$ pre = 0.135 Not stable

Figure 2

Event-window uptake probes across regimes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendixs.docx](#)