

When Intermediaries Become Participants: Classification, Institutional Authority, and AI Governance

Arthur D. Sidney

The Sidney Group PLLC, Washington, DC, USA
arthur.dennis.sidney@gmail.com

Abstract. Artificial intelligence unsettles a premise embedded in internet law and organizational governance: that an intermediary principally carries, hosts, or organizes the choices of others. Generative, retrieval-augmented, and agentic systems increasingly construct representations, shape consequential decisions, and execute actions. This paper develops a function-specific framework for determining when an AI-enabled intermediary becomes a participant in an outcome. It offers five indicators: content construction, decisional influence, action execution, control over material inputs, and institutional integration. The test is outcome-relative, control-sensitive, and change-sensitive; it classifies the disputed function rather than assigning a permanent label to an entire model or firm. The paper then connects external classification to internal governance. When participation is found, the deploying organization should map the relevant function to information access, approval rights, constraint-setting power, intervention authority, and accountability. Applications to AI overviews, enterprise decision support, vendor chains, and agentic transactions show how the framework distinguishes protected intermediation from active contribution without treating participation as strict liability. The paper concludes that classification is governance architecture: it determines which duties are implicated and whether any human or institution has the practical power to review, constrain, pause, override, or terminate the function that creates risk.

Keywords: AI governance, platform governance, intermediary liability, institutional authority, accountability, agentic AI.

1 Introduction

Section 230 of the Communications Decency Act was enacted for an internet organized around a recognizable distinction between an interactive computer service and information supplied by another content provider [1]. That distinction supported services that hosted, transmitted, organized, and moderated third-party speech without automatically becoming the publisher or speaker of every item they carried. The boundary has never been absolute. Courts have examined whether a service merely provided neutral tools or materially contributed to the unlawful feature of the

information at issue [2, 3]. Still, the legal starting point remained provenance: who supplied the relevant information, what role did the service perform, and does the claim seek to treat the service as the publisher or speaker of content supplied by another?

Artificial intelligence complicates that starting point. A synthesized search service retrieves sources, assigns weight, resolves or obscures disagreement, and delivers a newly composed answer. An enterprise copilot may transform internal records into a recommendation that shapes hiring, procurement, compliance, or financial decisions. An agentic system may select tools, communicate with external services, and complete a multi-step transaction. In each example the system is positioned between information and action, but its role cannot be described adequately by transmission or storage alone.

The central question is functional: when does an intermediary become a participant? This paper argues that participation begins when a system or the institution deploying it materially shapes the content, decision, or action at issue through functions that cannot fairly be reduced to carrying or storing another person's information. Participation does not imply legal personhood, publisher status for every purpose, or categorical liability for generated output. It identifies an active contribution that should not disappear behind a system-wide intermediary label.

The classification question precedes, but does not decide, liability. It identifies potentially relevant doctrines, regulators, controls, and professional duties. It also structures internal governance. An institution cannot assign meaningful responsibility until it identifies the function being governed, the way that function influences people or resources, and the actor capable of intervening. Misclassification can produce a fully staffed governance program aimed at the wrong object.

The paper makes four contributions. First, it defines participation as an outcome-relative continuum rather than a binary status. Second, it develops a five-indicator Functional Participation Test and a disciplined decision rule. Third, it connects external classification to an internal authority map. Fourth, it applies the combined framework to AI overviews and other deployments in which platform and organizational choices can produce online and offline harms.

2 Method and Scope

This is a conceptual and policy-design paper. It uses doctrinal-functional synthesis: boundary principles from intermediary law are placed alongside lifecycle governance frameworks and examined through recurring deployment scenarios. The aim is not to predict the outcome of a particular lawsuit. Section 230 supplies an especially useful example because its application turns on the relationship among third-party information, the defendant's own contribution, and the theory of the claim. The framework is also relevant outside Section 230, including product safety, administrative process, employment, professional services, contract, agency, and consumer protection.

The unit of analysis is the function connected to the disputed outcome, not the model, application, vendor, or enterprise as a whole. A single service may host third-party material, synthesize an answer, recommend an action, and execute a transaction. Each function can implicate different obligations and different actors. The approach is

therefore disaggregated and claim-relative. It asks what the system did with respect to the feature that caused harm or triggered governance concern.

The framework is designed for ex ante governance as well as ex post legal analysis. Before deployment, it can identify controls, contract rights, and decision owners that should accompany an active function. After an incident, it can help reconstruct where content, influence, action, and control entered the causal chain. Its validity is analytic rather than empirical at this stage; Section 7 identifies a research agenda for reliability testing and sector-specific calibration.

3 From Intermediation to Participation

3.1 The intermediary baseline

The traditional intermediary model remains necessary. Many services still transmit or host information created by users, and publisher liability based solely on carrying third-party speech would disrupt the allocation Congress adopted in Section 230 [1]. The Supreme Court declined in *Gonzalez v. Google* to resolve the broader application of Section 230 to recommendation systems [4]. Lower-court doctrine nevertheless offers a useful functional insight: a service can lose its intermediary character with respect to a particular feature when it materially contributes to what makes that feature unlawful [2, 3].

Algorithmic activity alone is not participation. Search indexing, spam filtering, caching, ranking, and moderation all involve computation. Nor does technical sophistication supply the boundary. The relevant distinction is between processing that enables another person's information to be carried or found and processing that materially constructs the representation, structures the decision, or performs the action at issue. The same ranking function may be ordinary curation in one claim and a consequential screening mechanism in another.

Platform scholarship has long shown that private systems govern through design, ranking, visibility, and access [5]. Content moderation research similarly emphasizes that harms emerge from interacting systems rather than isolated items of content [6, 7]. Generative and agentic systems intensify this governance role because they may supply both the representation through which a user understands an issue and the mechanism through which the user or institution responds.

3.2 A four-position continuum

Participation is best understood as a continuum with four recurring positions. A conduit transports information without materially changing the relevant feature. A curator selects, ranks, or organizes identifiable information. A synthesizer constructs a new representation from multiple inputs. An actor uses information and tools to change an external state. The positions are not corporate identities. They describe functions, and one service may occupy several positions at once.

Movement along the continuum is not automatically movement toward liability. It is movement toward a stronger need to identify the service's own contribution and the controls attached to it. A synthesized answer can be careful and accurate; an executed action can be authorized and beneficial. The continuum marks the point at which analysis must extend beyond third-party provenance to the design, deployment, and institutional use of the function itself.

4 The Functional Participation Test

The test contains five indicators. No arithmetic threshold is proposed because the significance of a function depends on the asserted harm and the governing regime. Instead, the decision rule is qualitative: participation is strongly indicated when action execution or content construction directly supplies the disputed feature; it may also be established by the convergence of decisional influence, control over material inputs, and institutional integration. Evidence that a function merely transports identifiable third-party material, leaves meaningful choices visible, and lacks control over the disputed feature weighs against participation.

Table 1. The Functional Participation Test

Indicator	Observable evidence	Governance consequence
Content construction	The system composes the material representation, reconciles sources, fills gaps, or states an inference as an answer.	Preserve provenance, test accuracy and uncertainty, and assign correction authority.
Decisional influence	The system scores, ranks, recommends, filters, or defines the options that receive meaningful consideration.	Audit actual reliance, thresholds, appeals, and the reviewer's power to reject the output.
Action execution	The system communicates, purchases, transfers, schedules, approves, denies, or changes an external state.	Define authorization limits, revocation, monitoring, and emergency intervention rights.
Control over material inputs	A provider or deployer controls retrieval, ranking, memory, tools, constraints, or data that materially shape the result.	Allocate duties to the actor able to inspect and change the relevant control.
Institutional integration	The output is embedded in an authoritative workflow, official record, policy, or service-delivery process.	Treat practical reliance as institutional action and assign an accountable decision owner.

The indicators should be applied according to four constraints. First, the inquiry is outcome-relative: the analyst identifies the specific representation, decision, or action in dispute. Second, it is function-specific: protection for hosting does not necessarily extend to synthesis or execution performed by the same service. Third, it is control-sensitive: duties should follow practical control over the feature, not model ownership alone. Fourth, it is change-sensitive: retrieval sources, tools, autonomy, workflow position, and actual reliance can change after initial approval.

The test separates provenance from contribution. Third-party material may remain part of the causal chain, but its presence does not end the inquiry when the system constructed, selected, or operationalized the feature that matters. Conversely, a component does not become a participant merely because its output is novel or probabilistic. The question is whether the defendant's or deploying institution's function

materially shaped the relevant outcome. Research on harmful AI speech reaches a compatible conclusion: legal analysis must distinguish harms created by model outputs from those attributable to users, data, deployment choices, or downstream use [8].

5 Classification as Governance Architecture

5.1 Classification allocates duties

Classification determines more than exposure to damages. It helps define jurisdiction, standards of care, documentation, audit, procurement terms, and the actor empowered to intervene. If a function is classified only as communications intermediation, governance may center on moderation and speech protections. If it operates as decision-support infrastructure, auditability, procedural fairness, professional judgment, and appeal become central. If it executes transactions, authorization, cybersecurity, agency, and revocation become unavoidable.

Existing frameworks recognize parts of this lifecycle problem. The NIST AI Risk Management Framework organizes practice around Govern, Map, Measure, and Manage and treats governance as a cross-cutting lifecycle function [9]. Its Generative AI Profile emphasizes risks that arise from confabulation, information integrity, human-AI configuration, and value-chain integration [10]. The European Union AI Act allocates obligations among providers, deployers, importers, distributors, and other actors according to role and risk [11]. Yet stable actor labels can still obscure functional change. A provider may add retrieval or tool use; a deployer may integrate an advisory output into a binding workflow. Continuing functional reassessment is therefore necessary.

5.2 Authority must follow participation

Once a system or institution is recognized as participating in an outcome, someone must possess authority commensurate with that participation. Responsibility without authority produces ceremonial oversight. A named human reviewer may receive an output yet lack access to sources, power to alter the workflow, or permission to halt use. Such a person is in the loop but does not govern it [12]. Sociotechnical research warns against abstracting an algorithm from the institution in which it operates [13]; internal auditing and reviewable decision-system research likewise require end-to-end attention to organizational processes [14, 15].

For each consequential function, an authority map should identify five powers: information access, approval or rejection, constraint setting, intervention, and accountability. Information access includes evidence, logs, provenance, uncertainty, and vendor disclosures. Approval authority determines whether the function may be used for the relevant purpose. Constraint-setting authority covers retrieval, tools, thresholds, permissions, and workflow rules. Intervention authority includes pause, override, rollback, termination, and restart. Accountability identifies the institution and decision owner answerable for the outcome.

These powers need not reside in one person, but the allocation must be complete and coordinated. A security team may disable a service without authority to decide whether its outputs remain legally usable. A procurement office may terminate a contract without access to incident evidence. A business owner may be accountable for a decision while the vendor alone can change the material control. The map exposes these separations before an incident converts them into governance failure.

6 Applications

6.1 AI overviews and synthesized search

A conventional results page ordinarily identifies links and leaves synthesis to the user. An AI overview selects sources and produces a single narrative answer. Content construction is therefore substantial when the alleged harm lies in the answer's wording, omission, confidence, or reconciliation of conflicting claims. Control over material inputs is also relevant because retrieval design, source eligibility, ranking, prompts, and safety rules shape the representation. This does not mean the service originated every factual claim. It means that third-party provenance does not fully describe the contribution at issue.

The classification produces specific governance consequences. The service should preserve source visibility, distinguish quotation from synthesis, communicate uncertainty, support correction, and define domains in which synthesis should be constrained or disabled. The authority map should identify who can inspect retrieval evidence, alter source rules, suspend the overview, approve remediation, and authorize restart. These controls directly address the conference's concern with how information is produced, distributed, consumed, and connected to offline consequences.

6.2 Enterprise decision support and public administration

An enterprise copilot may begin as a drafting aid and evolve into a recommender embedded in hiring, benefits, procurement, legal review, or compliance. Decisional influence and institutional integration are decisive when the output determines which candidates, claims, or options receive meaningful attention. A formal statement that a human makes the final decision is insufficient if employees presumptively follow the recommendation, cannot see alternatives, or face organizational penalties for rejecting it.

The governance object includes the model, but also thresholds, queues, evidence access, appeal procedures, data quality, and workflow rules. The designated reviewer must be able to understand and reject the recommendation, and the institution must monitor actual reliance rather than the workflow described at approval. This is where classification and authority become inseparable: a system treated as merely advisory can structure the field of choices so completely that the institution has delegated practical decision power without acknowledging it.

6.3 Procurement and vendor chains

AI capabilities are commonly assembled through cloud services, application providers, model vendors, data sources, and integration partners. Each party may control a different material input while the deploying institution remains publicly or corporately accountable. Procurement therefore allocates governance authority before the system operates [16]. Contracts determine audit access, logging, change notice, incident cooperation, data rights, model substitution, suspension, and termination.

The participation test prevents ownership from becoming a proxy for responsibility. The foundation-model provider may control model behavior; the application vendor may control retrieval and tools; the deployer may control workflow integration and reliance. Duties should attach to the function and the actor with practical control. Where accountability remains with the deployer but intervention power rests exclusively upstream, the mismatch is itself a procurement and governance risk.

6.4 Agentic transactions

Agentic systems move most clearly toward participation because they choose intermediate steps and execute actions. Yet the analysis should remain disaggregated. A model may plan, an orchestrator may choose tools, a platform may limit permissions, a vendor may retain logs, and an organization may define the objective. The relevant authority map must follow delegation and revocation across that chain.

A high-quality control system should define the action envelope, transaction limits, prohibited tools, confirmation points, monitoring, revocation, and restart authority. Legal remedies designed to deter autonomous software directly are unlikely to work unless they operate through the people and institutions capable of changing the system [17]. Participation classification identifies those control points without pretending the system itself is a responsible legal subject.

7 Objections, Boundary Conditions, and Research Agenda

A first objection is uncertainty. A categorical label is easier to administer than a multi-indicator test. But one label already fails to describe services that host, generate, recommend, and act. The framework channels uncertainty by tying analysis to observable functions, practical control, and the disputed outcome. Written findings under each indicator can make disagreement visible and reviewable.

A second objection is that participation may be mistaken for strict liability. It should not be. Classification identifies the relevant role and possible obligations; separate doctrine determines duty, breach, causation, defenses, immunity, and remedy. A participant may act reasonably and incur no liability. The framework's narrower purpose is to prevent an active contribution from disappearing behind an intermediary label.

A third objection concerns innovation and speech. Overbroad liability can discourage beneficial services and lawful expression, concerns that animate Section 230 and critiques of expansive platform liability [1, 18]. Function-specific analysis

offers a more tailored alternative. It preserves protection for hosting and transmission while recognizing synthesis, decisional integration, or execution when those functions materially contribute to the harm. Safeguards should target the active function and the actor able to control it.

The test also has boundary conditions. It is not a substitute for sector-specific definitions, constitutional analysis, evidentiary rules, or causal proof. The indicators may interact differently in medicine, employment, national security, search, or consumer transactions. Institutional integration may dominate in public administration, while content construction may dominate in synthesized search. Weighting should be explicit rather than assumed.

Future research should test inter-rater reliability by asking courts, regulators, vendors, deployers, auditors, and affected users to classify common scenarios. Comparative studies should measure whether the indicators predict existing doctrinal outcomes and whether authority mapping improves incident response time, appeal quality, correction rates, and risk detection. Longitudinal studies should examine classification drift after model updates, new tools, or changes in organizational reliance. This work could support sector-specific rubrics without converting the framework into a rigid universal score.

8 Policy and Organizational Implications

Policymakers should define AI roles by function and control, not corporate position or model ownership alone. A party that controls retrieval, ranking, tool access, or workflow integration may bear governance duties even when it did not train the underlying model. Material additions such as autonomous execution, a new retrieval source, or integration into a consequential workflow should trigger reassessment. Regulatory guidance can require a short functional analysis without preempting the elements of the underlying legal regime.

Organizations should maintain a functional system register rather than a model inventory alone. For each consequential use, the register should identify the outcome, participation indicators, relevant law, upstream dependencies, decision owner, information rights, intervention authority, and escalation path. Procurement should secure the evidence and contractual rights needed to exercise those powers. Audit should test whether formal rights remain usable in practice, including during vendor failure or urgent harm.

Courts confronting AI-related claims should separate third-party provenance from the defendant's own contribution. External information does not resolve whether a system constructed the allegedly harmful representation, materially structured a decision, or executed the relevant act. At the same time, a finding of participation should not short-circuit the statutory elements, defenses, or constitutional protections applicable to the claim.

9 Conclusion

Artificial intelligence is changing the institutional meaning of intermediation. Systems that once appeared to carry or organize the choices of others can now construct answers, influence decisions, and execute actions. Law and organizations need not choose between pretending those systems are passive and treating them as autonomous legal persons. They can classify functions.

The Functional Participation Test offers five indicators: content construction, decisional influence, action execution, control over material inputs, and institutional integration. Applied to the relevant outcome and practical distribution of control, the indicators distinguish intermediation from active contribution. They also expose when a service changes role after deployment.

When an intermediary becomes a participant, governance must follow the function. Someone must have access to the necessary information and authority to approve, constrain, pause, override, terminate, and, when appropriate, restart it. Classification is therefore not semantics. It is the architecture through which law and institutions allocate duties, evidence, decision rights, and intervention power.

Disclosure of Interests. The author is the founder of The Sidney Group PLLC. No external funding was received for this research, and the author declares no other competing interests.

References

1. 47 U.S.C. § 230 (2024)
2. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157 (9th Cir. 2008) (en banc)
3. FTC v. Accusearch Inc., 570 F.3d 1187 (10th Cir. 2009)
4. Gonzalez v. Google LLC, 598 U.S. 617 (2023)
5. Balkin, J.M.: Free speech in the algorithmic society: Big data, private governance, and new school speech regulation. *U.C. Davis Law Review* 51, 1149–1210 (2018)
6. Douek, E.: Content moderation as systems thinking. *Harvard Law Review* 136, 526–607 (2022)
7. Gorwa, R., Binns, R., Katzenbach, C.: Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7(1), 1–15 (2020)
8. Henderson, P., Li, X., Jurafsky, D., Hashimoto, T., Lemley, M.A., Liang, P.: Where's the liability in harmful AI speech? *Journal of Legal Analysis* 15, 559–608 (2023)
9. National Institute of Standards and Technology: Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1 (2023)
10. National Institute of Standards and Technology: Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1 (2024)
11. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence. *Official Journal of the European Union* (2024)
12. Sidney, A.D.: Human-in-the-loop is not AI governance. *FedScoop* (25 June 2026)

13. Selbst, A.D., Boyd, D., Friedler, S.A., Venkatasubramanian, S., Vertesi, J.: Fairness and abstraction in sociotechnical systems. In: Proceedings of FAT* 2019, pp. 59–68. ACM, New York (2019)
14. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P.: Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of FAT* 2020, pp. 33–44. ACM, New York (2020)
15. Cobbe, J., Lee, M.S.A., Singh, J.: Reviewable automated decision-making: A framework for accountable algorithmic systems. In: Proceedings of FAccT 2021, pp. 598–609. ACM, New York (2021)
16. Sidney, A.D.: AI procurement is quietly becoming a national security gatekeeper. RealClearDefense (21 April 2026)
17. Lemley, M.A., Casey, B.: Remedies for robots. *University of Chicago Law Review* 86, 1311–1396 (2019)
18. Citron, D.K., Wittes, B.: The internet will not break: Denying bad Samaritans § 230 immunity. *Fordham Law Review* 86, 401–423 (2017)