iS
CHANNEL

Information Asymmetry and the Accountability Gap in Autonomous Vehicle Safety Governance

Claudia Voysest[1]

MSc Management of Information Systems and Digital Innovation
Department of Management
London School of Economics and Political Science

KEYWORDS

Autonomous vehicles
Accountability gap
Algorithmic accountability
Information asymmetry
Safety governance

ABSTRACT

This essay examines the governance of simulation based safety validation in the autonomous vehicle (AV) industry through an Information Systems (IS) lens. Because real world testing alone cannot demonstrate AV safety within any practical timeframe, simulation has become the primary evidence base for deployment decisions. Yet in the United States no binding federal framework mandates independent verification of that evidence. Drawing on Waymo as an illustrative case and the IS literature on algorithmic accountability, the essay asks how the institutional arrangement governing AV simulation evidence produces an accountability gap, and what kind of governance failure this constitutes. It argues that the failure is not merely regulatory absence but a structural information asymmetry: firms hold private knowledge of scenario coverage, embedded assumptions, and success criteria that no external party can observe. Extending proposals for architectural observability, standardised test variation, and adaptive governance, the essay proposes a federal auditing body empowered to define simulation baselines and commission independent verification. It shows why voluntary disclosure alone cannot hold AV firms accountable for the evidence behind their deployment decisions and argues for independent verification as the missing institutional safeguard.

1. Introduction

Autonomous vehicle (AV) developers face a statistical paradox. Demonstrating through real world driving alone that an AV is merely 20% safer than a human driver would require more than 11 billion miles of testing, which at plausible fleet sizes could take centuries (Kalra & Paddock, 2016). Simulation has therefore become the primary evidence base for safety validation. The governance of that evidence base, however, remains largely voluntary and unverified: firms design the scenarios, run the tests, evaluate the results, and disclose only what they choose.

This essay treats that arrangement as an Information Systems (IS) governance problem rather than a regulatory one. It asks: how does the institutional arrangement governing AV simulation evidence produce an accountability gap, and what kind of governance failure does this constitute? The argument advanced here is that the governance failure does not stem merely from the absence of a regulator but from a deeper structural information asymmetry. The firm holds private knowledge of scenario coverage, embedded modelling assumptions, and the success criteria against which its system is judged; no external party can observe these. Under such conditions, voluntary disclosure produces transparency without accountability

[1] Corresponding Author
Email Address: c.voysest-hurtado@lse.ac.uk

(Ananny & Crawford, 2018), because accountability requires an external actor with both the standing and the capacity to interrogate the evidence and act on it. The essay proposes a federal auditing body empowered to define simulation baselines and mandate independent verification.

Waymo serves as the illustrative case. As one of the first firms to operate a commercial driverless service and among the most forthcoming in publishing safety research, it is a demanding test: if independent verification is absent even for a highly transparent firm, the gap is systemic rather than a matter of individual firm conduct. The sections that follow conceptualise the failure, describe how the validation process works, examine it through Waymo, propose an auditing framework, and consider its limitations.

2. Conceptualising the Governance Failure

Characterising the AV simulation problem as the mere absence of a regulator is accurate but arguably incomplete. A useful way to understand the deeper issue is as a problem of delegation. By allowing firms to validate their own systems rather than requiring independent checks, regulators leave responsibility for safety assurance with the same companies whose products are being assessed. Baird and Maruping (2021) define delegation as "transferring rights and responsibilities for task execution and outcomes to another" (p. 316) and suggest that a delegated task tends to require accountability, some allocation of who is answerable for outcomes, which they argue must be actively established rather than assumed. Although their framework concerns delegation to technological artifacts, a similar logic may apply here: the firm appears to receive the responsibility to produce safety evidence while no external party clearly holds the right to verify it. On this reading, the problem is less that no regulator exists than that safety assurance has been left to the firm without a corresponding mechanism of accountability.

This connects to work on algorithmic accountability. Ananny and Crawford (2018) argue that transparency, the ability to see inside a system, does not by itself yield accountability. Seeing is not the same as being able to interrogate, contest, or act; accountability is a relational property that depends on an external party positioned to do so. The distinction matters here. Waymo's published safety research makes its methodology visible, but visibility of a described method is not verifiability of the proprietary tool that produced the evidence.

Underlying this is a structural information asymmetry. The firm is the informed party; the regulator, the public, and even peer reviewers are uninformed parties. AV developers rely on proprietary validation approaches (Koopman & Wagner, 2018) that no independent assessor is positioned to check. In markets with such asymmetry, the informed party can disclose selectively, and self serving disclosure is difficult to distinguish from candour without an external verification mechanism. The concern is not merely that a firm evaluates itself, but that the governance regime supplies no institution capable of converting disclosure into accountability. The failure is therefore an accountability-regime failure of a specific type: one in which the producer of algorithmic evidence also controls the conditions under which that evidence can be scrutinised.

This framing also clarifies why the problem is distinctively an IS governance issue rather than a generic safety regulation one. The object being governed is not a physical artefact whose properties can be inspected directly, but a machine learning system whose behaviour in novel situations is emergent and difficult to predict (Rahwan et al., 2019; Rajabli et al., 2021). The evidence for its safety is itself a software artefact, the simulation, which encodes the firm's own assumptions. Governing such systems therefore means governing the tools used to produce evidence about them, a recursive problem that conventional inspection based regulation is poorly equipped to address.

3. How the Accountability Gap Emerged

Rather than validating safety on roads, AV firms rely on simulation: virtual environments that

replicate road conditions, traffic, and hazards, running the software through scenarios at a speed and scale road testing cannot match. This is necessary because AV systems depend on machine learning for core functions such as object identification and prediction of other road users' behaviour. Unlike rule based software, these components learn patterns from training data rather than following explicit instructions (Rajabli et al., 2021), so their behaviour in novel situations cannot be verified by inspecting code alone. As Kalra and Paddock (2016) conclude, real world driving cannot demonstrate safety within a practical timeframe; simulation is the primary alternative.

Koopman and Wagner (2018) note that AV developers rely on proprietary validation approaches with no agreed governing standard. Unlike other safety critical industries, where validation can be evaluated by an independent assessor, the AV sector lacks an assessor to verify any stage of the pipeline. The authors also caution that simulation models carry residual risk because they rely on assumptions and simplifications, such as simplified pedestrian representations that omit real world variation in appearance, posture, and carried objects. The firm therefore constructs its own scenarios, runs its system through them, and evaluates performance against its own criteria: it sets the standards and tests itself against them, with no external party able to observe whether the simplifications are benign or consequential.

This governance vacuum was filled by voluntary programmes. As Gutierrez (2023) documents, absent binding regulation, organisations that implement such programmes are effectively answerable only to themselves. In the United States, the National Highway Traffic Safety Administration adopted what Koopman and Wagner (2018, p.1) describe as a non regulatory approach to automated vehicle technology safety effectively permitting firms to deploy when they judge themselves ready. Although some states regulate road testing and data reporting, there appears to be no general federal requirement for independent verification of simulation evidence before deployment.

This is not to dismiss voluntary governance wholesale. Gutierrez (2023) defends soft law as a flexible response to the lag of formal regulation, and suggests that its central weakness, that firms are "accountable to themselves" (p. 792), might be mitigated through better designed incentives and disclosure requirements. In the AV simulation case, however, this remedy may be harder to apply than it appears: incentives and disclosure can generally only discipline conduct that an external party is able to observe, whereas here much of the decisive evidence, the proprietary simulation and its embedded assumptions, tends to remain unobservable. To the extent that this holds, soft law's flexibility may be less of an advantage where the informed party also controls what disclosure reveals.

Competitive dynamics harden this arrangement. Iansiti and Lakhani (2020) argue that AI driven markets tend to favour scale and network effects, pointing to autonomous vehicles as a case where real time data can build barriers that limit rivals. As this advantage compounds with accumulated data, firms that establish an early lead may be difficult to displace. Shollo et al. (2022), studying how firms create value from machine learning, find that fully autonomous deployment tends to occur only where legal and ethical constraints are few. When that pressure operates inside an audit gap where no external authority verifies simulation quality, the informed party has both the means and the incentive to disclose selectively, and the public has no independent assurance that the evidence justifying deployment is adequate.

4. Evidence: The Waymo Case

Waymo, the AV subsidiary of Alphabet, operates one of the most established commercial driverless ride hailing service in the United States, having served over 20 million rides to date (Waymo, 2026). It publishes safety reports and submits methodological work to peer review, making it more transparent than most competitors. This combination of operational scale and relative openness makes it an especially informative case for examining how AV firms validate their own simulation evidence,

and, following the logic above, the hardest case for the argument that voluntary disclosure is insufficient.

Rahwan et al. (2019) observe that the researchers who study AI systems are predominantly the same researchers who built them. The Waymo case illustrates this directly. Scanlon et al. (2021), a study authored entirely by Waymo employees, used the company's own simulation environment to evaluate the Waymo Driver across scenarios drawn from its operational domain. The firm selected the scenarios, built the simulation, ran the tests, evaluated the results, and published the conclusions.

A charitable reading holds that voluntary disclosure is working: Waymo published this methodology in a peer reviewed journal (Scanlon et al., 2021) with no regulatory obligation to do so, discussed limitations openly, and submitted the work to academic review. If every firm behaved this way, the case for mandatory external governance would weaken. The information asymmetry framing shows why this reading is incomplete. Peer reviewers assessed the methodology *as described*; no independent party selected the scenarios, executed the simulation, or verified the assumptions embedded in the proprietary system. The paper reports 'no known competing financial interests' (Scanlon et al., 2021, p. 15), yet all authors were Waymo employees, so the study lacks institutional independence. The point is not that Waymo's evidence is unreliable, but that peer review validates the method while leaving the tool unverified: scenario coverage, embedded assumptions, and success criteria remain private.

This exposes a second order problem beyond self evaluation. Even maximal voluntary transparency cannot close the gap, because what is disclosed is chosen by the informed party and what is withheld, the proprietary simulation and its assumptions, is precisely the information an external assessor would need. Transparency here is not a partial solution that fuller disclosure would complete; it is structurally incapable of producing accountability without an independent verifier. This is why the Waymo case has broader significance: even one of the most transparent firms in the sector cannot make its central safety claims independently verifiable.

5. What Governance Should Look Like

The engineering literature already contains proposals that could close the audit gap, though none is currently mandatory. Koopman and Wagner (2018) call for observability points in AV architectures, the built in capability for a tester to observe system state and determine whether a test passed or failed, and stress that this must be designed in from the start. Lundgren (2021) adds that tests must incorporate sufficient variation to stop manufacturers optimising for the test rather than for real world conditions. Both are framed as voluntary best practice. This essay argues they must become mandatory, because, as the preceding analysis shows, voluntary adoption is structurally unreliable under information asymmetry.

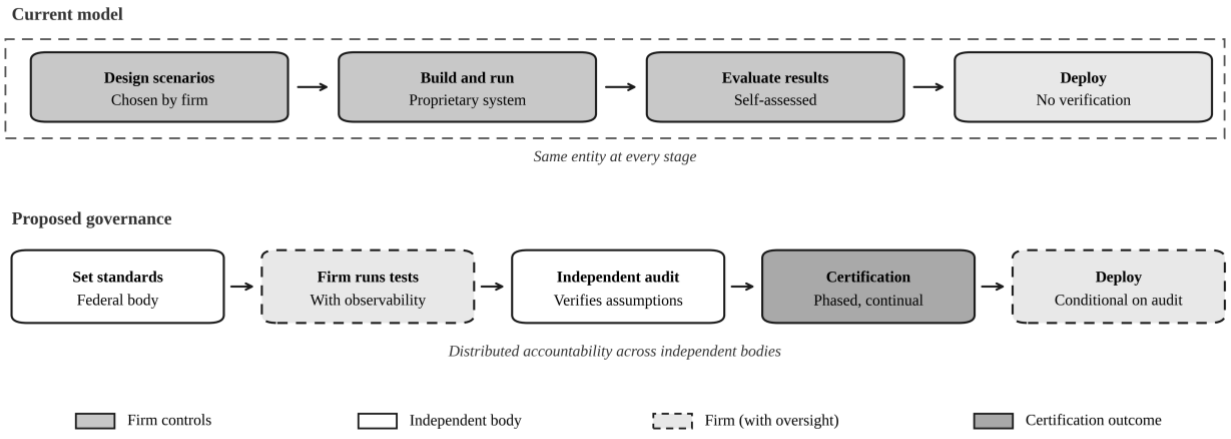
One reason no binding framework has emerged in the United States is that regulation struggles to keep pace with emerging technology; Gutierrez (2023) documents that the lag between an AI governance issue emerging and government acting can run to years. Yet a binding framework is achievable. Tan and Taeihagh (2021) describe how Singapore's Land Transport Authority developed a tiered milestone system for AV trials incorporating simulation and safety document review, with requirements for authorisation, data sharing, and liability insurance, an adaptive model that adjusts as the technology matures.

What this essay proposes is what Fisher et al. (2021) call certification: the process by which an independent body determines whether a system complies with the relevant regulations. The key advantage of this approach is that it addresses the information asymmetry by moving responsibility for verification away from the company being assessed. Drawing on expertise from different disciplines (Rahwan et al., 2019), the independent body would set common simulation standards, define the scenarios that must be tested, and commission external audits. In doing so, it would turn transparency into accountability by providing the independent oversight that Ananny and Crawford (2018) argue is essential.

Rather than approving a system once and granting permanent certification, the framework would require regular reassessments as the technology evolves (Tan & Taeihagh, 2021). Taken together, these measures mean that no single firm would control every stage of validating its own system. Figure 1 illustrates the proposed framework.

Figure 1. AV simulation governance: the current firm-controlled model versus a proposed framework distributing accountability across independent bodies

Figure 1: AV simulation governance



Sources: Koopman and Wagner (2018), Lundgren (2021), Tan and Taeihagh (2021). The current-model characterisation is the author's own synthesis.

Figure generated with the assistance of generative AI (Claude, Anthropic).

6. Limitations and Conclusion

Two limitations qualify the argument. First, the framework depends on a regulatory culture willing to enforce binding requirements; regulation is only as strong as the political context that sustains it. Singapore illustrates the point precisely because it has institutions with both the administrative capacity and the political will to regulate industry directly. Jurisdictions lacking either could not transplant the framework without first changing the underlying regulatory tradition. A federal auditing body without legitimacy or enforcement capacity is paperwork rather than governance.

Second, an auditing body works only if its auditors understand the industry they oversee and building that capacity is difficult. Fisher et al. (2021) note that certifying autonomous systems demands engineering and verification knowledge regulators have not historically needed. Acquiring it takes time, and the most direct route, recruiting from industry, would compromise the very independence the body exists to provide. The framework therefore depends on regulators developing genuine technical expertise not drawn from the firms they oversee, a tension the design must manage rather than assume away.

These limitations do not remove the underlying problem. The essay has traced how statistical necessity created a dependence on simulation, how voluntary governance appears to have left that simulation largely unverified, and how competitive pressure may entrench the status quo, with the same firms that benefit most from

deployment often producing, evaluating, and disclosing the evidence that justifies it. Understanding this as an accountability failure driven by information asymmetry helps clarify what a solution might need to do: not simply mandate more disclosure but establish an external body able to verify the evidence and act on it.

Future work could test the framing comparatively, examining whether sectors that resolved analogous asymmetries, such as aviation certification, did so through the institutional mechanisms proposed here, and could specify the audit standards, data access rights, and liability rule an AV certification body would require. More broadly, the essay suggests why disclosure alone struggles to hold the producers of algorithmic evidence accountable, and points to independent verification as the institutional safeguard that voluntary governance lacks. Peer review validates the method; until an independent body validates the tool, the rules governing how AI systems operate on public roads remain a privatised decision.

Notes

Adjacent concerns, such as the ethics of crash avoidance algorithms and algorithmic transparency in the broadest sense (e.g., Ananny & Crawford, 2018, on the limits of the transparency ideal), inform the theoretical framing but fall outside the specific institutional gap addressed here.

Data Accessibility

No data, code, or structured methods were produced or used in this research beyond the published sources cited in the reference list. No supplementary files are associated with this submission.

Competing Interests

The author has no competing interests to declare.

Authors' Contributions

This article was researched and written solely by the named author, who takes full responsibility for its content.

Acknowledgements

I would like to thank Dr. Aaron Cheng for his valuable guidance and insightful feedback throughout the development of this paper. I am also grateful to the anonymous reviewers and the iSCHANNEL editorial team for their thoughtful comments and constructive suggestions, which greatly improved the manuscript.

Generative AI Use acknowledgement and reflection

I acknowledge the use of Generative AI (Gen AI) in preparing this paper. I used it to brainstorm potential research topics, and to suggest an initial set of sources, each of which I then independently verified existed and read in full before using, continuing the literature search myself from there. Additionally, I used AI to render Figure 1 from a structure and content I specified myself. For proofreading I used Grammarly for spelling and grammar and Claude for general proofreading, re-editing the flagged passages myself. All AI outputs were critically reviewed, and the research direction, argument, and final wording are my own.

References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. ***New Media & Society***, *20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Baird, A., & Maruping, L. M. (2021). The Next Generation of Research on IS Use: A Theoretical Framework of Delegation to and from Agentic IS Artifacts. ***MIS Quarterly***,

45(1), 315–341. https://doi.org/10.25300/MISQ/2021/15882

Fisher, M., Mascardi, V., Rozier, K. Y., Schlingloff, B.-H., Winikoff, M., & Yorke-Smith, N. (2021). Towards a framework for certification of reliable autonomous systems. ***Autonomous Agents and Multi-Agent Systems***, *35*(1), Article 8. https://doi.org/10.1007/s10458-020-09487-2

Gutierrez, C. I. (2023). Uncovering Incentives for Implementing AI Governance Programs: Evidence From the Field. ***IEEE Transactions on Artificial Intelligence***, *4*(4), 792–798. https://doi.org/10.1109/TAI.2022.3171748

Iansiti, M., & Lakhani, K. R. (2020). Competing in the age of AI: Strategy and leadership when algorithms and networks run the world. **Harvard Business Review Press.**

Kalra, N., & Paddock, S. M. (2016). Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? ***Transportation Research. Part A, Policy and Practice***, *94*, 182–193. https://doi.org/10.1016/j.tra.2016.09.010

Koopman, P., & Wagner, M. (2018). Toward a framework for highly automated vehicle safety validation (SAE Technical Paper 2018-01-1071). **SAE International**. https://doi.org/10.4271/2018-01-1071

Lundgren, B. (2021). Safety requirements vs. crashing ethically: what matters most for policies on autonomous vehicles. ***AI & Society***, *36*(2), 405–415. https://doi.org/10.1007/s00146-020-00964-6

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., Jackson, M. O., Jennings, N. R., Kamar, E., Kloumann, I. M., Larochelle, H., Lazer, D., McElreath, R., Mislove, A., Parkes, D. C., Pentland, A. 'Sandy,' … Wellman, M. (2019). Machine behaviour. ***Nature (London)***, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Rajabli, N., Flammini, F., Nardone, R., & Vittorini, V. (2021). Software Verification and Validation of Safe Autonomous Cars: A Systematic Literature Review. ***IEEE Access***, *9*, 4797–4819. https://doi.org/10.1109/ACCESS.2020.3048047

Scanlon, J. M., Kusano, K. D., Daniel, T., Alderson, C., Ogle, A., & Victor, T. (2021). Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain. ***Accident Analysis and Prevention***, *163*, Article 106454. https://doi.org/10.1016/j.aap.2021.106454

Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. ***The Journal of Strategic Information Systems***, *31*(3), Article 101734. https://doi.org/10.1016/j.jsis.2022.101734

Tan, S. Y., & Taeihagh, A. (2021). Adaptive governance of autonomous vehicles: Accelerating the adoption of disruptive technologies in Singapore. ***Government Information Quarterly***, *38*(2), Article 101546. https://doi.org/10.1016/j.giq.2020.101546

Waymo. 2026. ***Waymo: Self-driving cars, autonomous vehicles, ride-hail***. https://waymo.com/