Fairfield UNIVERSITY

North East Journal of Legal Studies

Volume 47 Article 3

Fall 6-25-2026

The Cheapest Cost Avoider is Dead: Long Live the Best Algorithmic Risk Governor

Boaz Segal
UC Berkeley School of Law, boaz.segal41@gmail.com

The Cheapest Cost Avoider Is Dead—Long Live the Best Algorithmic Risk Governor

Boaz Segal*

Abstract

Artificial intelligence is destabilizing one of tort law's most influential organizing principles: the Cheapest Cost Avoider ("CCA"). Classical CCA analysis emerged in relatively bounded accident settings, where the actor best positioned to prevent harm was often also best positioned to evaluate the relevant cost–benefit tradeoffs. Algorithmic systems disrupt this alignment. In AI-driven environments, harms increasingly emerge from layered socio-technical ecosystems involving developers, platform providers, institutional deployers, and opaque model architectures. Under these conditions, the most visible human actor at the point of injury—the physician, driver, or loan officer—often appears to be the CCA, yet lacks meaningful control over the system-level risks that precipitated the harm.

This Article argues that tort law must move beyond the CCA paradigm and reorient liability around a new organizing concept: the Best Algorithmic Risk Governor ("BARG"). Building on Calabresi and Hirschoff's best decision maker ("BDM") framework and the economic logic of the Hand formula, the Article contends that liability should concentrate on the actor best positioned to observe, evaluate, and govern systemic algorithmic risk at scale. The Article develops functional markers for identifying BARGs, including control over training data, model architecture, monitoring infrastructure, and population-level risk observability.

Using medical AI, autonomous vehicles, and algorithmic credit systems as case studies, the Article demonstrates how contemporary tort doctrine systematically misallocates responsibility by focusing on downstream human discretion rather than upstream governance power. It concludes that tort law's efficacy in the algorithmic age lies not in assigning blame for isolated accidents, but in structuring incentives for continuous institutional risk governance.

I. Three Rooms, One Question—Tort Law at the Point of Algorithmic Harm

In a single mid-sized American city, three quiet rooms become sites of algorithmic fate.

In the first room, a radiologist stares at a screen in the city's main hospital. The AI diagnostic system flashes "low malignancy risk" next to a lung nodule the physician's instincts find troubling. A cursor blinks beside a green confidence score; a second monitor shows the hospital's risk-management dashboard, reassuringly clean. In the next room, a patient signs a consent form that mentions "decision-support software" in one buried line of boilerplate, indistinguishable from the noise of other clauses. Months later, the patient dies from a cancer the radiologist would likely have caught without the AI-generated reassurance.

Across town, on the same gray afternoon, a driver "supervises" an almost-autonomous vehicle gliding through familiar streets. The system promises to handle everything "except rare edge cases". The driver, succumbing to the predictable lull of automation bias after hundreds of uneventful miles and a flawless marketing campaign, glances at a text message just as the car's vision system misclassifies a pedestrian. The impact occurs faster than any human could reasonably react, faster than any ordinary negligence narrative can comfortably describe.

A few blocks away, in a glass-front bank branch on the city's main street, a loan officer clicks "approve" or "deny" based on a credit score generated by a proprietary machine-learning model—trained, tuned, and periodically updated by a vendor the bank barely understands. The officer has been instructed not to "second-guess the algorithm," and the bank's compliance manual treats the score as an objective, neutral fact. The applicant denied that afternoon will later default on high-cost informal credit, triggering a cascade of financial and emotional harms that no one in the room believes they "chose."

In each of these "rooms", tort law is about to be asked its oldest question in a radically new setting: *who should pay for these harms—and why?*

For over half a century, the canonical answer has leaned on the figure of the Cheapest Cost Avoider ("CCA"). But when harms emerge from multi-layered, opaque, and constantly updating socio-technical systems, that familiar figure begins to blur. This article argues that tort law must

be re-designed for the age of AI by shifting its focus from locating the CCA to identifying—and legally cultivating—the Best Algorithmic Risk Governor ("BARG").

Artificial intelligence now permeates high-stakes domains: *Medicine*—diagnostic and triage algorithms, radiology tools, robotic surgery, and, increasingly, large language models for clinical decision support;[1] *Mobility*—autonomous and semi-autonomous vehicles, driver assistance systems, and mixed traffic platooning;[2] *Finance and work*—credit scoring, fraud detection, underwriting, algorithmic hiring, and automated performance management;[3] *Generative AI*—large models that produce text, images, and code, with sophisticated but fragile guardrails.[4]

In each of these environments, the AI "decision" is the visible tip of a long chain of design choices, data curation, model training, deployment, updating, and institutional integration. Yet traditional tort reasoning often gravitates toward the last human in the loop – the doctor, the driver, the clerk, or the content poster – as the CCA or primary bearer of negligence-based duties.

The central question of this article is: *On which actor(s) in the AI ecosystem should tort law concentrate liability if it wants to actually minimize the social costs of accidents and promote safe design?*

The article advances three claims:

---

[*] Doctor of Law, Visiting Scholar UC Berkeley School of Law and Vice Dean, School of Law, Sapir Academic College.

[1] David O. Shumway & Hayes J. Hartman, *Medical Malpractice Liability in Large Language Model Artificial Intelligence: Legal Review and Policy Recommendations*, 124 J. OSTEOPATH. MED. 287, 287–288 (2024); Clara Cestonaro et al., *Defining Medical Liability When Artificial Intelligence Is Applied on Diagnostic Algorithms: A Systematic Review*, 10 FRONT. MED. 1, 1–3 (2023); George Maliha et al., *Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation*, 99 MILBANK Q. 629, 629–632 (2021).

[2] Shi Rui, *Research on Tort Liability of Autonomous Vehicles in Traffic Accidents*, 19 BCP SOC. SCI. & HUMAN. 157, 157–160 (2022); Muhammad Uzair, *Who Is Liable When a Driverless Car Crashes?*, 12 WORLD ELECTR. VEH. J. 1, 1–3 (2021); Xu Chen & Xuan Di, *Legal Framework for Rear-End Crashes in Mixed-Traffic Platooning: A Matrix Game Approach*, 3 FUTURE TRANSP. 417, 417–419 (2023).

[3] Xukang Wang et al., *Algorithmic Discrimination: Examining Its Types and Regulatory Measures with Emphasis on US Legal Practices*, 7 FRONT. ARTIF. INTELL. 1, 2–5 (2024); Lauri Kai, *Machine-Learning Credit Scores and Disparate Impact Theory* 2–8 (2018) (unpublished manuscript) (available at SSRN 3166562); Natalie Sheard, *Employment Discrimination by Algorithm: Can Anyone Be Held Accountable?*, 45 UNSW L.J. 617, 622–630 (2022); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 679–691 (2016).

[4] Justin D. Weisz et al., *Toward General Design Principles for Generative AI Applications*, ARXIV:2301.05578, 2–3, 8–10 (2023); Susan Hao et al., *Safety and Fairness for Content Moderation in Generative Models* 1, 1–3 (2023).

1. *In AI-driven environments, reliance on the classical CCA test alone systematically misallocates responsibility*—Economic logic that once worked tolerably well in simple, bilateral interactions breaks down in multi-layered algorithmic ecosystems.[5]

2. *Courts should pivot toward a "best cost–benefit decision maker" perspective, adapted to algorithmic systems*—This builds on Calabresi and Hirschoff's insight that the core question is who is best positioned to perform and act upon the cost–benefit calculus, not merely who can physically prevent a particular accident.[6]

3. *To operationalize this shift in AI, tort law needs a new organizing concept: the "Best Algorithmic Risk Governor"*—This is the actor best positioned to gather information about algorithmic risks, evaluate cost–benefit trade-offs, and implement system-level changes that affect many users at once.

The payoff is not merely theoretical. Properly used, tort law can function as an architect of algorithmic reality, aligning private incentives toward safer AI design and governance while avoiding the scapegoating of frontline human actors for systemic failures beyond their control.[7]

II. The Best Decision Maker Move—From CCA to BDM to BARG

This chapter develops the economic foundations of the argument by returning to three canonical frameworks in tort theory: Calabresi's CCA, the Calabresi–Hirschoff best decision-maker ("BDM") test, and Judge Learned Hand's negligence formula. It unpacks each of these in turn, showing how they conceptualize the allocation of accident costs, the locus of cost–benefit analysis, and the role of courts in specifying reasonable care. Together, these three theories provide the analytic vocabulary that the rest of the article adapts and extends to AI-driven, algorithmic environments.

---

[5] Roberto Pardolesi & Bruno Tassone, *Guido Calabresi on Torts: Italian Courts and the Cheapest Cost Avoider*, 1 ERASMUS L. REV. 7, 17–18, 34–35 (2008); See generally John C.P. Goldberg, *History, Theory, and Tort: Four Theses*, 11 J. Tort L. 17 (2018); Miriam Buiten, Alexandre de Streel & Martin Peitz, *The Law and Economics of AI Liability*, 48 COMPUT. L. & SEC. REV. 1, 8–12 (2023).
[6] John C.P. Goldberg, *Id*; Helmut Koziol ed., BASIC QUESTIONS OF TORT LAW FROM A COMPARATIVE PERSPECTIVE 453–456 (Jan Sramek Verlag 2015).
[7] George Maliha et al, *supra* note 1, at 634–639; Madalina Busuioc, *Accountable Artificial Intelligence: Holding Algorithms to Account*, 81 PUB. ADMIN. REV. 825, 832–834 (2021).

A. Move One: Avoidance—The CCA Logic

Calabresi's project in The Costs of Accidents is not only to define the CCA, but to embed that figure in a broader optimization of primary, secondary, and tertiary accident costs in a world of positive transaction costs. Primary costs are the expected losses from accidents plus the resources invested in avoiding them; secondary costs are the costs of spreading or insuring against those losses; tertiary costs are the administrative expenses of operating the liability system itself. The cheapest cost avoider is therefore the actor who, taking all three cost categories into account, can most efficiently combine precautions, activity-level adjustments, and risk-spreading.[8]

This is why Calabresi famously associates the CCA idea with targeted strict liability rather than with negligence: concentrating liability on the CCA is supposed to induce optimal investments in safety while also exploiting that actor's superior capacity to buy insurance and pass residual accident costs along in prices. Subsequent law and economics work has elaborated and refined this insight, for example by asking when loss-sharing or mixed liability rules outperform all-or-nothing allocation, and by testing whether real courts actually succeed in identifying CCAs in complex multi-party accidents. These developments reinforce the core lesson for AI: if courts keep focusing liability on the last human actor, they risk ignoring those institutions that are in fact the cheapest cost avoiders at the systemic level.[9]

B. Move Two: Decision—From Prevention to BDM

In a later article on strict liability, Calabresi and Hirschoff proposed a refinement: liability should be assigned to the party "in the best position to make the cost–benefit analysis between accident costs and accident-avoidance costs and to act on that decision." This formulation shifts the focus From "who can physically prevent the accident?" To "who should be making, and acting on, the optimization decision?"

---

[8] G. Calabresi THE COSTS OF ACCIDENTS – A LEGAL AND ECONOMIC PERSPECTIVE, 26–31, 135–173 (1970); Roberto Pardolesi & Bruno Tassone, *supra* note 5, at 11–12; Guido Calabresi, *The Decision for Accidents: An Approach to Nonfault Allocation of Costs*, 78 Harv. L. Rev. 713, 714–715 (1965).

[9] Calabresi *Id*, 719–730; Emanuela Carbonara, Alice Guerra & Francesco Parisi, *Sharing Residual Liability: The Cheapest Cost Avoider Revisited*, 45 J. LEGAL STUD. 173, 178–180, 191–194 (2016).

Subsequent scholarship has highlighted this as the move from a pure CCA perspective to a "best positioned decision-maker" test, especially relevant when prevention requires information-rich, technical, or systemic judgments rather than simple, observable acts of care.[10]

The Calabresi–Hirschoff refinement pushes the analysis from physical control over a specific accident to informational control over a class of accidents. Their BDM test asks which actor is best placed to gather information about risks and precautions, to perform the cost–benefit comparison, and to translate that comparison into changes in activity levels, technologies, and contractual arrangements. In many simple, bilateral accidents the CCA and the BDM will coincide, but Calabresi and Hirschoff already anticipated settings where design choices, information flows, and institutional structures are so complex that the party who could have grabbed the last clear chance is not the one who should be internalizing the full optimization problem.[11]

Later scholars have used this perspective to support selective strict liability or hybrid regimes in which responsibility is shifted "upstream" to manufacturers, platform operators, or other institutional actors that can redesign processes and technologies for many users at once. This literature also highlights a tension that is central for AI: sometimes the actor best placed to identify the efficient combination of precautions is not the actor best placed to implement them, which forces courts to choose between deterring bad decisions and inducing concrete behavioral change. The *Best Algorithmic Risk Governor* framework can be presented as an explicit attempt to operationalize the BDM idea for algorithmic ecosystems characterized by opacity, continuous updating, and multi-layered value chains.[12]

---

[10] G. Calabresi & J. T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 YALE LAW JOURNAL 1055, 1060–1064 (1972); For further analysis, see Helmut Koziol ed., BASIC QUESTIONS OF TORT LAW FROM A COMPARATIVE PERSPECTIVE 438–456 (Jan Sramek Verlag 2015); See also Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 11–12.

[11] Yuval Sinai & Benjamin Shmueli, *Calabresi's and Maimonides's Tort Law Theories – A Comparative Analysis and a Preliminary Sketch of a Modern Model of Differential Pluralistic Tort Liability Based on the Two Theories*, 26 YALE J.L. & HUMAN. 101 (2013). This article argues that by uncovering utilitarian-economic foundations in Maimonides's tort theory and placing them in dialogue with Calabresi's cheapest cost avoider and best decision maker doctrines (and with Posner's negligence theory), one can develop a modern, differential and pluralistic model of tort liability that integrates efficiency and justice by allocating strict or fault-based liability according to the type of risk-creating activity.

[12] For a critique of the theory, see Megan L. Richardson, *Revisiting Strict Product Liability: Taking Law and Economics Further*, 35 OSGOODE HALL L.J. 195 (1997). This essay argues that Dewees, Duff and Trebilcock's empirical and economic critique of strict product liability does not preclude a strong case for such liability, because a refined version of Calabresi and Hirschoff's cheaper cost avoider test – one that incorporates information costs, loss-

C. Move Three: Calculation—Hand, Posner, and the Economics of Negligence

Learned Hand's formula in *Carroll Towing*[13]—negligence where $B < PL$—has become the canonical translation of negligence into marginal cost–benefit terms. Law-and-economics scholars, above all Posner[14], read the formula as an implicit efficiency test: a defendant is negligent when an additional unit of precaution would have reduced the expected social loss by more than its cost. On this view, negligence liability induces actors to choose an efficient level of care at which marginal prevention costs equal marginal reductions in expected accident losses. Crucially, however, because a non-negligent actor escapes liability, a negligence rule fails to optimally regulate activity levels—a distortion that, as economic theory notes, is typically cured only under strict liability, where residual losses are internalized and reflected in market prices.[15]

At the same time, a rich critical literature emphasizes that courts rarely apply a literal Hand calculus and that real-world fact-finders often treat B, P, and L in moral or distributive terms rather than as purely economic inputs. This suggests that the Hand formula is best understood as a family of standards rather than a precise algorithm: it authorizes judges and juries to reason in terms of avoidable risk and reasonable precautions, but it leaves open which risks count and how to weigh catastrophic low-probability harms, dignitary interests, or structural inequalities. For AI, this opens space to argue that Hand-style negligence analysis should be anchored at the level of BARGs, not frontline users, because the most meaningful choices about B, P, and L are made when models are designed, trained, and integrated—ong before a physician runs a scan or a driver presses "engage".[16]

D. Move Four: Decoupling—Why AI Splits CCA from BDM (and Makes Room for BARG)

spreading and corrective justice – may yield a more certain and efficient liability rule than the negligence regime they favour.

[13] *United States v. Carroll Towing Co.*, 159 F.2d 169 (2d Cir. 1947).

[14] R. A. POSNER, ECONOMIC ANALYSIS OF LAW (5th ed. 1998); W. M. LANDES & R. A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW (1987).

[15] Eugênio Battesini, *Incremental Learned Hand Standard, Degrees of Negligence and Allocation of Damages: A Comparative Tort Law and Economics Approach*, 8 RJLB 1249, 1250–1251, 1258–1260 (2022); Yotam Kaplan & Maytal Gilboa, *The Other Hand Formula: Explaining Gain-Based Liability* 1, 9–12 (2021); Richard W. Wright, *Hand, Posner, and the Myth of the "Hand Formula"*, 4 THEORETICAL INQUIRIES IN L. 145, 146–150 (2003).

[16] Christopher Brett Jaeger, *The Hand Formula's Unequal Inputs*, 135 YALE L.J. 461, 465–481 (2025); See also generally Jeonghyun Kim, *Revisiting the Learned Hand Formula and Economic Analysis of Negligence*, 169 J. INST. & THEORETICAL ECON. 407 (2013).

In many classic accident settings—two drivers, a manufacturer and a consumer, a landowner and a visitor—the CCA and the "best cost–benefit decision-maker" often coincide. The party who can cheaply prevent the accident (by driving carefully, installing guards, or designing a safer product) is usually also the party well placed to evaluate the costs and benefits of precautions. In these relatively simple, bilateral interactions, the same actor typically controls both the relevant behavior and the relevant information: the driver knows how much effort safe driving requires, the manufacturer knows the marginal cost of redesigning a product, and the landowner can estimate the expense of making premises safer. As a result, assigning liability to the CCA effectively channels incentives to the actor who is also best situated to perform the cost–benefit analysis that tort law implicitly demands.

AI decision-making environments break this convenient alignment. Algorithmic systems are developed, trained, deployed, updated, and monitored by different actors who operate at different points in time and at different levels of abstraction. The human end-user—whether a physician relying on a diagnostic tool, a driver "supervising" an automated vehicle, or a caseworker using a risk-scoring system – often appears, at the level of a single incident, to be the CCA: she can double-check the output, apply common sense, or refuse to follow an AI recommendation. Yet she is not the actor best placed to govern systemic algorithmic risk across cases. She does not design the model architecture, set the decision thresholds, choose the training data, determine the feedback loop, or decide how errors are distributed across groups and contexts.

Conversely, the entities that are in fact best positioned to engage in meaningful cost–benefit analysis of algorithmic precautions—AI developers, platform operators, large institutional deployers, and sometimes regulators – frequently do not look like the CCA in any particular accident. Their contribution to a specific harm is distributed, probabilistic, and mediated through code, updates, and data pipelines. They may never interact with the injured party, never see the specific decision, and never appear in the immediate "who-could-have-prevented-this?" narrative that traditional tort analysis often privileges. Nonetheless, these upstream actors are precisely the ones who can observe patterns across thousands or millions of decisions, compare alternative designs, and internalize the long-run costs of false positives, false negatives, and biased error distributions.

The preliminary claim, then, is that while in simple, non-algorithmic accidents the CCA and the best cost–benefit decision-maker frequently converge, AI pushes them apart. Focusing exclusively on the apparent CCA at the point of harm risks misallocating responsibility to local human operators and under-incentivizing those who actually govern algorithmic risk at scale. As the next sections show, the actor who looks like the cheapest cost avoider in a single incident is often not the actor who is best placed to govern systemic algorithmic risk across cases.[17]

III. The Alignment Breaks—Why AI Rewires Tort Law's Liability Map

A. Rewiring Point One: The Actor Stack—A Multi Layered Value Chain

In AI, the "actor" is rarely a single person or firm. What looks like one decision at the point of harm is usually the end product of an actor stack—a multi layered value chain in which distinct entities make distinct risk shaping choices at distinct moments. Tort law, by contrast, is built to tell a cleaner story: identify the relevant actor, specify the duty, evaluate breach against a standard of care, and connect the conduct to the injury. The actor stack disrupts that narrative. Control is layered. Knowledge is layered. So is the capacity to prevent harm. And once those layers are separated, the familiar question—"who should have done more?"—cannot be answered by looking only at the last human in the room.

Start with model developers and vendors. They do not merely "build" a system; they write the system's risk profile into its DNA. They choose architectures, curate training data, tune parameters, and hard-wire defaults—confidence thresholds, escalation rules, retraining cadence—that silently govern downstream behavior. In effect, they preselect which errors will be common, which will be rare, and which will be tolerated. Long before any end-user encounters a concrete case, the developer has already defined the boundaries of what the system will treat as normal, exceptional, and ignorable.

Then comes the platform and cloud layer. Providers offer the infrastructure and tooling that make deployment feasible, and increasingly they supply pre-trained foundation models or large generative systems that downstream actors adopt as building blocks rather than bespoke

[17] Katherine Drabiak, *Leveraging Law and Ethics to Promote Safe and Reliable AI/ML in Healthcare*, 2 FRONT. NUCL. MED. 1, 4–7, 10–14 (2022); Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 5–11; George Maliha et al., *supra* note 1, at 630–639.

products.[18] This layer is often treated as background, but it is not neutral. Platforms can determine what is measurable (logging and monitoring), what is auditable (access to model behavior and change histories), and what can be constrained (rate limits, filters, safety tooling). Those design choices shape not only safety outcomes, but the practical possibility of proving negligence and correcting it. A world in which the relevant evidence is systematically unavailable—or structurally expensive to obtain—is a world in which the standard tort inquiry is tilted before it begins.

Institutional users—hospitals, banks, insurers, ride sharing and logistics platforms—sit at the integration layer, where "operations" becomes governance.[19] They decide which system to procure, where to place it in the workflow, how strongly to frame its outputs, and how much discretion humans will realistically have. A simple configuration choice can rewrite the risk map: a hospital may lower an alert threshold to reduce "noise" route low risk outputs into a slow queue or compress review time under productivity pressures. The model may not change at all, yet the institution's integration can convert a recommendation into an action-forcing directive—or, just as consequentially, an inaction-forcing default. This is the layer where organizational incentives, staffing levels, and compliance pressures translate algorithmic output into lived harm.

Finally, the frontline user appears—the physician, the driver, the loan officer, the HR staffer, the consumer—closest in time and space to the injury. That proximity makes the end-user the easiest defendant to name and the easiest story to tell. But the actor stack makes that convenience misleading. The end-user's "choice" is frequently constrained by upstream interface design, institutional protocol, and the thinness of information available to evaluate reliability. Tort law's impulse to locate responsibility at the last human touchpoint risks confusing visibility with control.

Responsibility is therefore distributed across multiple private and public actors whose decisions are separated in time, space, and expertise.[20] The point is not merely descriptive ("there are many

---

[18] Chen Chen et al., *Trustworthy, Responsible, and Safe AI: A Comprehensive Architectural Framework for AI Safety with Challenges and Mitigations*, ARXIV:2408.12935 1, 2–3, 10 (2025); Philipp Hacker et al., *Regulating ChatGPT and other Large Generative AI Models*, in PROC. ACM CONF. FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 1112, 1112–1116 (2023).

[19] Muhammad Uzair, *supra* note 2, at 12–13 ; George Maliha et al., *supra* note 1, 630–637; Katherine Drabiak, *supra* note 17, at 10–14.

[20] Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *Attributing Responsibility in AI-Induced Incidents: A Computational Reflective Equilibrium Framework for Accountability*, ARXIV , 1–4 (2025); Gabriel Lima & Meeyoung Cha, *Responsible AI and Its Stakeholders*, ARXIV:2004.11434 1, 2–3 (2020); Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, 5–8.

actors"). It is normative and doctrinal: when control and knowledge are layered, tort law must rethink how it assigns duty, assesses breach, and understands causation across a value chain. The actor stack is the first structural reason why AI rewires the liability map—and it sets up the radical information asymmetries that follow.

B. Rewiring Point Two: The Black Box—Radical Information Asymmetries

AI architectures, training pipelines, and performance characteristics are often opaque even to sophisticated institutional users, let alone to frontline professionals and lay users.[21] This opacity is not a mere inconvenience; it is a structural feature of contemporary algorithmic systems that reshapes how (and whether) tort law can do its ordinary work. In classic negligence settings, the law assumes that key actors can *appreciate* the relevant risk, *select* among precautions, and *explain* their choices in a way that courts can evaluate against a standard of reasonable care. The black box disrupts each of those assumptions at once. It disconnects the point of harm from the point of meaningful knowledge, and it converts what looks like a simple "error" into an evidentiary and governance problem: who actually knows enough to calibrate safety, and who can prove what, when something goes wrong?

Developers and platforms typically control documentation, logs, test suites, performance metrics, and outcome data at scale.[22] That control matters because the most legally salient facts in an AI-driven accident—error rates in the relevant subpopulation, known failure modes, model drift, the conditions under which performance degrades, the distribution of false negatives versus false positives, the existence (or absence) of post-deployment monitoring—are rarely visible from the outside. In practice, the information needed to perform a Hand-style "B versus PL" analysis is often *locked upstream*, and it remains locked precisely when litigation begins to ask for it. The result is a predictable distortion: courts can easily scrutinize what the last human actor did in the room, but they may never see the upstream design and governance choices that set the baseline risk in the first place. When the relevant safety knowledge is treated as proprietary (or is simply

[21] Clara Cestonaro et al., *supra* note 1, at 3–4, 9; Adriano Koshiyama et al., *Towards Algorithm Auditing: Managing Legal*, *Ethical and Technological Risks of AI, ML and Associated Algorithms*, 11 R. SOC. OPEN SCI. 1, 4–6, 13–14 (2024); George Maliha et al, *supra* note 1, at 638.
[22] Adriano Koshiyama et al., *Id*, at 3–6.

not recorded in a usable way), the tort system's ordinary fact-finding process becomes structurally tilted toward the visible and away from the causally important.

End-users operate interfaces that may provide limited explanations, coarse risk scores, or blunt recommendations, with minimal ability to interrogate or recalibrate the underlying model.[23] This means that the "human in the loop" is frequently asked—by designers, institutions, and later by plaintiffs – to supply the very thing the interface withholds: a reasoned, context-sensitive assessment of reliability. Yet the user's practical capacity to do so is often thin. A clinician sees a confidence score without the model's calibration data; a driver is told to monitor a system whose failures are rare but catastrophic; a loan officer is given a binary output without access to the model's feature space, training distribution, or threshold logic. The end-user's discretion is therefore not just institutionally constrained (as the actor-stack analysis explains), but informationally hollowed out. What looks like an opportunity to "double-check" is often a demand to validate a system that is intentionally non-interrogable at the point of use.

These asymmetries create what has been described as "black box" liability and responsibility gaps in medical AI and other fields, where no single frontline actor can meaningfully understand, let alone optimize, system-wide risk.[24] The gap is doctrinal as well as practical. Duty and breach become harder to specify when the content of "reasonable care" depends on facts controlled by another actor; causation becomes harder to prove when the mechanism of error is inaccessible; and failure to warn analysis becomes unstable when the most important warnings would require disclosures that vendors resist or institutions never demand. The black box thus amplifies the very divergence this article emphasizes: the apparent CCA at the point of harm may be asked to carry legal responsibility, even though the BDM—and, more precisely, the BARG—is located upstream where the system's risks can actually be observed, quantified, and reduced across cases. In short, radical information asymmetries do not merely complicate tort adjudication; they help explain why liability rules that fixate on the last human decision systematically misallocate incentives in AI ecosystems.

[23] Clara Cestonaro et al., *supra* note 1, at 2–4; Katherine Drabiak, *supra* note 17, at 3–5.
[24] Benjamin H. Lang et al., *Responsibility Gaps and Black Box Healthcare AI: Shared Responsibilization as a Solution*, 2 DIGIT. SOC. 52, 55–56, 59–61 (2023); George Maliha et al, *supra* note 1, 637–639; Clara Cestonaro et al., *Id*, 3–4, 9–10.

Crucially, this is not a static problem. Once the informational baseline is skewed—once the model's behavior, limitations, and change history are not meaningfully transparent to those who rely upon it—the legal system is primed to misread algorithmic harm as a sequence of isolated "bad calls," rather than as a governance failure in a socio-technical system. That is the bridge to the next rewiring point: the moving target. When systems update, retrain, and drift, the black box does not stay in one shape long enough for ordinary liability narratives to stabilize—making information asymmetry not only radical, but continuously renewed.

C. Rewiring Point Three: The Moving Target—Dynamic Systems and Feedback Loops

Unlike static products, many AI systems are not "finished" when they are shipped, installed, or first deployed. Their risk profile is not a stable attribute that can be assessed once and then treated as fixed. Instead, the system's behavior is often a moving target—because the model changes, the environment changes, the data changes, and, crucially, the system itself changes the data environment in which it operates. That dynamism matters for tort law because negligence doctrine is structured around relatively stable objects of evaluation: a design, a warning, a practice, or a professional decision. In AI ecosystems, those anchors drift.

1. Dynamic Adaptation and Environmental Shifts

Models are regularly updated or retrained, sometimes continuously via online learning. A central feature of modern AI deployment is that performance is expected to be maintained through iterative updating. Models are patched, thresholds are recalibrated, retraining data is refreshed, and "improvements" are pushed in cycles that resemble software development more than classic product manufacture.

Even where a model is not literally learning online, it is often subject to "concept drift" and "distribution shift"—the world that generated the training data is not the world in which the model now operates. A diagnostic model trained on one hospital's imaging pipeline may degrade when scanners, protocols, patient demographics, or prevalence rates change; a driving model may fail when weather, signage, road geometry, or fleet composition shifts; a credit model may become brittle as macroeconomic conditions change or consumer behavior adapts. The moving-target problem is therefore not simply that models get updated—it is that "reasonable care" cannot be

evaluated by looking only at a single frozen snapshot of model performance at time t_0, when the harm occurs at time t_1 in a materially different risk environment.

2. Feedback Loops and Self-Creating Systems

AI systems are often subject to feedback loops, where the system's own outputs influence future data (e.g., predictive policing, credit approvals, and hiring pipelines).[25] These feedback loops intensify the moving-target problem because they make the system partially self-creating. When an algorithm's outputs shape what gets observed, recorded, and later treated as "ground truth," the system can lock in its own assumptions—sometimes amplifying error, sometimes amplifying inequality, and often doing both while appearing to improve by internal metrics.

If police deployment decisions concentrate surveillance in particular neighborhoods, the resulting arrest data can "confirm" the model's premise that those neighborhoods are higher risk. If a lender denies applicants predicted to default, the model may never learn whether those applicants would in fact have repaid—creating selective labels and an evidentiary gap that looks like predictive success but is partly a byproduct of the institution's own denial policy. If an employer's screening model narrows the applicant pool, the workforce data that later "validates" the model is a downstream artifact of the model's own gatekeeping. In these settings, the system does not merely predict the world; it helps produce the world it then claims to measure. Tort law's ordinary intuition—evaluate the defendant's conduct against an external reality of risks—becomes harder to apply when the defendant's system is actively reshaping that reality.

3. The Problem of Scale

AI is typically deployed at scale, so that small design choices (e.g., risk thresholds, loss functions) propagate into thousands or millions of decisions. Scale transforms marginal technical parameters into population-level governance decisions. A minor shift in a confidence threshold, an alert suppression rule, a loss function that privileges one kind of error over another, or a UX choice that frames the output as a "recommendation" versus an "instruction" can translate into system-wide changes in false negatives and false positives.

---

[25] Aurora S. Zhang & Anette E. Hosoi, *Structural Interventions and the Dynamics of Inequality*, in PROCEEDINGS OF THE ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1014, 1015–1016 (2024); Madalina Busuioc, *supra* note 7, at 826–827; Solon Barocas & Andrew D. Selbst, *supra* note 3, 682–687.

In classic negligence narratives, the accident is the salient unit: a discrete event calling for a post hoc evaluation of whether the defendant should have been more careful in that moment. In AI systems, by contrast, the salient unit is often the policy encoded in the system—a policy that is executed repeatedly, at speed, and in ways that no single frontline actor can meaningfully audit decision-by-decision. Once the system is scaled, the tort system's preference for individualized stories can systematically miss the more important question: who set the parameters that governed risk across the entire population of decisions?

4. Structural Patterns of Harm

Harms may emerge not as isolated "accidents" but as repeated, structurally generated patterns—for example, systematic under-diagnosis in underrepresented subpopulations, or recurring disparate impact in algorithmic lending and hiring.[26] This is the core doctrinal pressure point. Traditional tort analysis is comfortable with rare malfunctions, one-off mistakes, and localized departures from reasonable care. But dynamic AI systems can generate harms through stable patterns that are produced by an evolving system operating under stable institutional incentives.

In medicine, the harm may not look like "the system broke" but rather "the system is consistently less sensitive for certain patients," especially where training data, calibration choices, or workflow integration make errors more likely for underrepresented groups. In employment and credit, the harm may be a recurring disparate impact that is statistically predictable yet individually deniable—each single denial can be defended as "reasonable reliance" on a score, while the system as a whole operates as a structural sorting mechanism that predictably burdens the same communities. The moving-target character of AI makes this worse: even when an issue is detected, updates can change the system's behavior before courts, regulators, or injured parties can stabilize the factual record.

5. Doctrinal Implications for Tort Law

For tort law, the moving-target and feedback-loop dynamics do not merely "complicate proof." They disrupt how duty, breach, and causation are conceptually organized. A regime built around

[26] Solon Barocas & Andrew D. Selbst, *Id*, 684–687; Savina D. Kim, Stefan Lessmann, Galina Andreeva & Michael Rovatsos, *Fair Models in Credit: Intersectional Discrimination and the Amplification of Inequity*, ARXIV:2308.02680V1, 12–16 (2023); Clara Cestonaro et al., *supra* note 1, at 3, 9.

one-time design choices and static warnings is poorly matched to systems whose safety depends on continuous monitoring, version control, incident response, and retraining governance. When the relevant "precaution" is not a discrete act by a frontline user but an upstream practice—ongoing auditing, drift detection, rollback capability, human-override realism testing, and post-deployment evaluation across subpopulations—the identity of the legally relevant decision-maker shifts.

This is precisely where the divergence between the apparent CCA (Cheapest Cost Avoider) and the BDM (Best Decision Maker) becomes practically consequential: the actor who can most cheaply prevent this instance of harm (often the last human in the loop) is frequently not the actor who can most cheaply and effectively govern the system's evolving risk over time.

In that sense, the moving target is not just a technical feature of modern AI. It is a structural reason why tort law's liability map must be rewired around the actors who can manage dynamics: the entities who control updating, monitoring, deployment constraints, feedback-loop mitigation, and institutional integration. Those are quintessential BARG functions. And once we see AI risk as dynamic and self-reinforcing, the next step follows naturally: tort law must learn to treat algorithmic harm less as a collection of isolated accidents and more as a set of systemically produced patterns—precisely the shift taken up in the next rewiring point.

D. Rewiring Point Four: From Accidents to Patterns—Systemic vs. Individual Risk

Traditional tort paradigms often imagine a discrete accident between an injurer and a victim: a bounded event, a localized lapse, and a post hoc inquiry into whether a specific actor failed to take a reasonable precaution in that moment. Algorithmic systems, by contrast, tend to produce risk as a feature of the system's repeated operation—across many users, contexts, and time periods. In this setting, the legally salient unit is often not "the accident", but the policy encoded in model design, thresholds, workflows, and update practices—policies that can quietly shift the distribution of harms at scale. AI systems therefore generate systemic risk:

1. Structural harms, such as discrimination, exclusion, and loss of privacy or autonomy, often emerge not from a single deviant output but from stable patterns of design and data use—choices about objectives, feature selection, labeling practices, and representativeness that

predictably burden certain groups or contexts even when each individual decision is facially "routine".[27]

2. Causation is often diffused across the socio-technical chain: no single decision looks clearly wrongful in isolation, and no single human actor can be said to have "caused" the harm in the classic but-for sense, yet the aggregate pattern is both foreseeable and socially costly. This is the familiar paradox of pattern-based injustice: each instance is deniable, while the distribution is systematic.[28]

In such environments, effective governance turns on the capacity to detect and respond to patterns—through population-level performance monitoring, auditing for disparate impact, stress-testing and drift detection, and the ability to revise thresholds, interfaces, and deployment constraints. The actors who can see and alter patterns across many cases—those with access to comprehensive data, auditing tools, and meaningful deployment levers—are therefore central to risk reduction in a way that frontline actors rarely can be.[29]

This shift from accidents to patterns pressures tort doctrine on multiple fronts. Standards of care calibrated for one-off mishaps may miss harms that are "reasonable" in any single case but unreasonable in their cumulative incidence or distribution. Similarly, evidence and causation doctrines that privilege individualized narratives may underweight statistical proof of recurring failure modes, feedback-loop amplification, or systematically skewed error rates.

Once risk is understood as systemic, liability design must follow the points of systemic control: where risk is measured, where it can be recalibrated, and where changes propagate across users. This reframes the tort question from assigning blame for an isolated incident to assigning responsibility for governing an evolving risk-generating system.

E. Rewiring Point Five: The Liability Map – Doctrinal Implications for Tort Law

Point Four reframed algorithmic harms as patterned and systemic rather than isolated mishaps. Point Five translates that shift into doctrine: once the relevant "risk" is a moving, feedback-driven

[27] Madalina Busuioc, *supra* note 7, at 826–827; Xukang Wang et al., *supra* note 3, at 3–4; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 677–692.
[28] Aurora S. Zhang & Anette E. Hosoi, *supra* note 25, at 1015–1017; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 1–3.
[29] Adriano Koshiyama et al., *supra* note 21, at 4–6; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *Id*, at 2, 7–9; Madalina Busuioc, *supra* note 7, at 826–827.

system, the liability inquiry must map onto the actors and institutions with genuine governance leverage over that system—design choices, deployment constraints, monitoring, and iterative updates.

Doctrinally, this pushes tort law away from a purely event-based perspective (a single breach causing a single injury) and toward a governance-based perspective. Courts still rely on familiar tools—duty, breach, causation, and reasonableness—but those tools must be applied with attention to upstream design and training decisions, ongoing post-deployment control, and the allocation of informational and organizational capacity to detect risk patterns before they crystallize into harm.

In practical terms, the "liability map" is a mapping between legal responsibility and institutional control. It asks where the system's risk thresholds are set, who decides when to retrain or rollback, who can meaningfully audit outcomes, and who can internalize the costs of harm through pricing, insurance, and enterprise-level precautions. Where those levers sit upstream, doctrines that default to blaming the most visible downstream actor will misallocate incentives.

That is why an exclusively backward-looking question—who could have prevented this accident tomorrow morning?—is radically incomplete. It invites courts to focus on the last human in the chain, even when that person lacks the authority, information, or time to override a system that is calibrated elsewhere.

Put differently, focusing only on the immediate encounter risks treating downstream discretion as control. In AI settings, the key preventive acts often occur earlier—during system design, calibration, documentation, deployment constraints, and post-deployment monitoring—where risk patterns can be detected and corrected.

Tort law must also ask: ***Who is best positioned to govern systemic algorithmic risk—to monitor, understand, and redesign the system in response to emerging harms?***

Rewiring the liability map therefore has concrete doctrinal consequences. First, it reframes duty and breach around governance capacity: the relevant standard of care may include reasonable auditing, robust logging, meaningful human-override pathways, and timely updating or retraining when performance drift is foreseeable. Second, it changes how courts should think about causation. Where harm arises from aggregate system behavior, the evidentiary focus often shifts from pinpointing a single "but-for" decision to establishing that the system's configuration and

oversight regime made the harm materially more likely—and that a feasible governance intervention would have reduced that risk.

Third, it informs the choice among negligence, product liability, and enterprise-liability-style doctrines. Depending on the context, traditional negligence may under-incentivize prevention because the party with the best information is not the party facing the lawsuit. Doctrines that better track informational asymmetry and cost internalization—whether through design-defect analysis, failure-to-warn framed as failure-to-disclose model limits, or institutional liability for deployment policies—can help align legal incentives with the locus of control.

Finally, it clarifies how tort law should interact with public regulation. Compliance may be relevant evidence of reasonableness, but it cannot substitute for governance when regulation lags, is incomplete, or is gamed through box-checking. Precisely because algorithmic systems are moving targets, tort doctrine remains a complementary mechanism for enforcing ongoing risk governance rather than one-time certification.

Against that backdrop, the next section offers concrete examples showing how a surface-level "cheapest cost avoider" (CCA) intuition often points courts toward the wrong defendant, while the true best cost–benefit decision-maker (BDM)—and the actor who can function as an effective BARG—sits upstream.

This is where the divergence between the apparent CCA and the true best cost–benefit decision-maker (BDM) in AI contexts becomes stark—and why the identity of the BARG is typically determined by system-level governance leverage rather than by proximity to the injurer.

IV. The Frontline Fallacy: When CCA Intuition Misfires

A. Medical AI: The "Independent Judgment" Myth

In contemporary medico-legal discourse, clinicians are routinely positioned as the primary bearers of responsibility when AI-assisted diagnosis or treatment contributes to patient harm. Courts, regulators, and professional guidelines alike tend to emphasize the physician's duty to "exercise independent judgment," framing algorithmic tools as mere aids that must never displace human

decision-making.[30] On this view, the physician remains fully accountable for the ultimate clinical decision, even when that decision is heavily shaped by algorithmic output.

From a superficial CCA perspective, this allocation of responsibility appears intuitively plausible. The physician is physically present at the point of care, occupies a well-defined professional role, and fits comfortably within existing malpractice narratives. In theory, the clinician can disregard an algorithmic recommendation, order additional tests, consult a colleague, or rely on clinical intuition. Courts can easily reconstruct this moment of choice, and negligence doctrine is well equipped to ask whether a reasonable physician should have done more in that encounter.

Yet this intuition collapses once the analysis shifts from the single clinical encounter to the structure of algorithmic decision-making in medicine. The most consequential cost—benefit trade-offs that shape diagnostic risk are rarely made at the bedside. Instead, they are embedded upstream in the design, calibration, and deployment of medical AI systems. Choices about sensitivity versus specificity, acceptable false-negative rates, confidence thresholds, and escalation protocols are typically determined long before a physician encounters a particular patient.[31] These choices define not only the system's error profile, but also the kinds of clinical vigilance that will appear "reasonable" or "excessive" ex post.

Developers and vendors play a central role in this process. Through decisions about model architecture, training data, loss functions, and calibration strategies, they effectively encode normative judgments about which errors matter most and which risks are tolerable. A diagnostic system optimized to minimize false positives will, by design, increase false negatives; one calibrated to reduce unnecessary biopsies may systematically under-detect malignancies in certain subpopulations. These are not incidental technical details. They are policy choices with predictable clinical consequences, replicated across thousands of cases.[32]

Healthcare institutions add another critical layer of risk governance. Hospitals and health systems decide which AI tools to procure, how to configure them, and how tightly to integrate them into

---

[30] Maroudas, Vasileios P., *Fault-Based Liability for Medical Malpractice in the Age of Artificial Intelligence: A Comparative Analysis of German and Greek Medical Liability Law in View of the Challenges Posed by AI Systems*, 57 REV. EUR. & COMP. L. 135, 145–151 (2024); Aagaard Lise, *Artificial Intelligence Decision Support Systems and Liability for Medical Injuries*, 9 J. RES. PHARM. PRACT.125, 126–127 (2020); George Maliha et al., *supra* note 1, at 632–633; Clara Cestonaro et al., *supra* note 1, at 4–5.

[31] Aagaard Lise, *Id*, at 125–126; George Maliha et al., *Id*, at 632–639; Clara Cestonaro et al., *Id*, at 2–4.

[32] Aagaard Lise., *Id*, at 125–127; George Maliha et al., *Id*, at 632–633, 638–641; Clara Cestonaro et al., *Id*, at 3–4, 9.

clinical workflows. Interface design, alert fatigue mitigation, default settings, and documentation practices all shape how clinicians actually experience and rely upon algorithmic output.[33] A system presented as a "recommendation" may, in practice, function as an action-forcing directive when time pressure, staffing constraints, and institutional protocols converge. Conversely, a nominally advisory tool can become an inaction-forcing default when low-risk outputs are routed away from meaningful review. None of these governance choices are controlled by the individual physician facing a patient.

Empirical and doctrinal analyses of medical AI liability increasingly recognize this mismatch. Systematic reviews of AI-related malpractice risks show that existing legal frameworks tend to overload physicians with responsibility while under-deterring developers and institutions, despite the fact that systemic risk is overwhelmingly determined upstream.[34] The physician is asked to supply "independent judgment" precisely where meaningful independence is structurally undermined by information asymmetries, opaque model behavior, and institutional reliance on algorithmic outputs.

In this context, the physician may look like the CCA at the level of a single adverse outcome, but she is not the actor best positioned to perform the relevant cost—benefit analysis about algorithmic risk. She cannot recalibrate the model, retrain it on more representative data, audit its performance across populations, or modify its integration into clinical workflows. Those capacities lie with developers, vendors, and healthcare institutions—the actors who can reduce risk not just for one patient, but for all future patients simultaneously.

Medical AI therefore exemplifies the core claim of this Article: in algorithmic environments, the apparent CCA at the point of harm often diverges sharply from the true best cost—benefit decision-maker. Treating the clinician as the primary locus of liability rests on the "independent judgment" myth—the assumption that frontline professionals retain meaningful control over risks that are in fact governed elsewhere. Once that myth is exposed, responsibility must be reoriented toward the actors who actually set the baseline level of algorithmic risk and who can most

[33] Deimantė Rimkutė, *AI and Liability in Medicine: The Case of Assistive-Diagnostic AI*, 16 BALTIC JOURNAL OF LAW & POLITICS 64, 70–71, 79 (2023); Aagaard Lise., *Id*, at 125–126; Katherine Drabiak, *supra* note 17, at 10–13.
[34] Deimantė Rimkutė, *Id*, at 65–67; George Maliha et al., *supra* note 1, at 629–634; Clara Cestonaro et al., *supra* note 1, at 1, 4, 9–10.

efficiently alter it. In medical AI, those actors are the system's developers and the institutions that deploy and govern it.[35]

The same structural illusion reappears beyond the hospital walls: just as clinicians are told to exercise "independent judgment" over opaque diagnostic systems, drivers of semi-autonomous vehicles are cast as ever-vigilant supervisors—formally responsible, yet practically deprived of meaningful control over risks that are engineered elsewhere.

B. Autonomous and Semi-Autonomous Vehicles: The "Ready-to-Intervene" Myth

In the context of autonomous and semi-autonomous vehicles, contemporary liability frameworks often rely—explicitly or implicitly—on what may be called the "ready-to-intervene" assumption. Under this model, responsibility is anchored in the figure of the human driver who is formally designated as a supervisor: the system may control steering, acceleration, braking, and navigation, but the human operator is expected to remain vigilant, attentive, and capable of taking over instantaneously when the system encounters an unexpected scenario.

At first glance, this allocation of responsibility appears compatible with classical Cheapest Cost Avoider logic. The driver is physically present, directly connected to the vehicle's controls, and—at least in theory—capable of preventing harm by braking, steering, or disengaging automation. Courts and commentators therefore often treat the driver as the natural locus of duty: the last human in the loop, and thus the last chance to avert catastrophe.

Yet a growing body of empirical research in human automation interaction fundamentally undermines this premise. Continuous, high-quality monitoring of a highly reliable automated system is not merely difficult; it is cognitively and psychologically unrealistic.[36] Decades of research on vigilance, automation complacency, and skill degradation show that humans are systematically ill-suited to act as passive supervisors of systems that fail rarely, unpredictably, and at machine speed. The more reliable the system appears during ordinary operation, the more

[35] George Maliha et al., *Id*, at 630–640; Katherine Drabiak, *supra* note 17, at 3–6, 9–14; Deimantė Rimkutė, *Id*, at 65–71.
[36] Alice Guerra, Francesco Parisi & Daniel Pi, *Liability for Robots I: Legal Challenges*, 18 J. INST. ECON. 331, 332–333 (2022); A. Feder Cooper & Karen Levy, *Fast or Accurate? Governing Conflicting Goals in Highly Autonomous Vehicles*, 20 COLO. TECH. L.J. 221, 227–235 (2022); Muhammad Uzair, *supra* note 2, at 5–7.

human attention decays; the rarer the intervention opportunities, the slower and less effective the human response becomes when intervention is suddenly demanded.

The result is a structural mismatch between legal expectation and human capability. The driver is formally tasked with monitoring, but the system is designed in a way that predictably erodes the very vigilance that monitoring requires. Reaction times in takeover scenarios routinely exceed the temporal window in which avoidance is physically possible, especially when failures involve perception errors, misclassification of objects, or complex traffic dynamics that unfold faster than human situational awareness can recover. In such settings, the "ready-to-intervene" driver is less a genuine safeguard than a legal fiction.

By contrast, manufacturers and software providers occupy a radically different position in the risk architecture of autonomous driving systems. They determine how control is shared between human and machine, how and when control is transferred back to the driver, and how clearly takeover requests are communicated.[37] They design the sensor fusion logic, object-classification thresholds, and decision policies that govern how the vehicle prioritizes speed, comfort, and safety.[38] These actors also control the system's learning and updating processes, enabling them to respond to incidents by deploying software updates across entire fleets, redesigning user interfaces, recalibrating thresholds, or modifying default behaviors.[39]

Crucially, these upstream actors are capable of learning from accidents in a way that individual drivers cannot. A single driver experiences one crash; a manufacturer observes patterns across thousands or millions of miles. A driver cannot redesign the perception stack or adjust the system's tolerance for uncertainty; a manufacturer can. From a cost–benefit perspective, the manufacturer is therefore vastly better positioned to evaluate the trade-offs between false positives and false negatives, between earlier alerts and driver overload, and between aggressive automation and conservative fallback strategies.

---

[37] Xuan Di, Xu Chen & Eric Talley, *Liability Design for Autonomous Vehicles and Human-Driven Vehicles: A Hierarchical Game-Theoretic Approach*, 118 TRANSPORTATION RESEARCH PART C: EMERGING TECHNOLOGIES 1, 3–6 (2020); Muhammad Uzair, *Id*, at 5–7, 14–15.
[38] A. Feder Cooper & Karen Levy, *supra* note 36, at 232–235; Xuan Di, Xu Chen & Eric Talley, *Id*, at 1–6.
[39] Jack Boeglin, *The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation*, 17 YALE J.L. & TECH. 171, 198–201 (2015); Xu Chen & Xuan Di, *supra* note 2, at 418–420.

Legal scholarship has increasingly recognized that traditional tort doctrines struggle to account for this asymmetry. Attempts to anchor liability in driver negligence rest on an implicit assumption that the driver is meaningfully capable of preventing the harm at reasonable cost. But when supervision itself is the weak link—when the system is engineered in a way that predictably defeats sustained human vigilance—the driver is not the cheapest cost avoider in any meaningful sense. Nor is the driver the best positioned decision-maker about system-level risk. That role belongs to the entities that design, train, deploy, and update the automated driving system in the first place.[40]

From the perspective advanced in this article, semi-autonomous driving thus exemplifies a broader pattern: the apparent CCA at the point of harm diverges sharply from the actor who actually governs risk at scale. Treating the driver as the primary bearer of liability not only misallocates responsibility, but also distorts incentives. It encourages manufacturers to externalize systemic design risk onto individual users, while providing weak legal pressure to redesign interfaces, takeover protocols, and automation boundaries in ways that reflect real human limitations.

Once this misalignment is acknowledged, the normative implication follows naturally. The best cost—benefit decision-maker in semi-autonomous vehicle systems is not the human monitor who is structurally set up to fail, but the manufacturer or software provider who can redesign the system to reduce the probability and severity of harm across an entire fleet. In BARG terms, the manufacturer is the actor best positioned to gather information about failures, evaluate trade-offs among alternative designs, and implement changes that propagate system-wide. Assigning liability accordingly does not absolve drivers of all responsibility, but it rejects the myth that human supervision can function as a robust fail-safe in environments engineered for machine control.

The collapse of the "ready-to-intervene" narrative in semi-autonomous driving thus exposes a broader structural pattern: across AI-mediated domains, tort law repeatedly assigns responsibility to human actors who appear to retain formal control, while the true governance of risk—through design choices, thresholds, and system architecture—resides elsewhere, a dynamic that becomes even more pronounced in algorithmic decision-making systems governing credit, insurance, and employment.

[40] Shi Rui, supra note 2, ta 159–162; Alice Guerra, Francesco Parisi & Daniel Pi, *supra* note **Error! Reference source not found.**, at 332–335; Muhammad Uzair, *supra* note 2; Xuan Di, Xu Chen & Eric Talley, *supra* note 37, at 2–6.

C. Algorithmic Credit, Insurance, and Employment: The "Neutral Score" Myth

In domains such as credit scoring, insurance underwriting, and algorithmic hiring, machine-learning systems are routinely framed as instruments of objectivity. By translating complex personal histories into numerical scores, these systems promise to replace subjective human judgment with neutral, data-driven assessment. Yet a substantial interdisciplinary literature has demonstrated that algorithmic scoring systems can encode, reproduce, and even amplify existing social hierarchies and structural inequalities, even when protected attributes are formally excluded from the model.[41]

From a surface-level tort perspective, the frontline clerk, loan officer, or HR professional applying the algorithmic output may appear to be the Cheapest Cost Avoider. In principle, this actor could question the score, request additional information, or deviate from the automated recommendation. This intuition fits comfortably within traditional negligence narratives: the human decision-maker is visible, proximate to the harm, and seemingly capable of exercising discretion at low cost.

In practice, however, this discretion is often more illusory than real. Institutional policies frequently instruct employees not to deviate from algorithmic outputs in the name of consistency, auditability, and the reduction of "human bias".[42] Deviations from automated recommendations may trigger internal scrutiny, disciplinary action, or accusations of arbitrariness, while adherence to the score is treated as compliance with objective procedure. The result is a familiar inversion: the human operator bears formal responsibility for the decision, but lacks meaningful authority to alter the risk calculus embedded in the system.

Meanwhile, the core cost–benefit decisions that shape discriminatory risk are made upstream, by institutions and vendors that design, train, and deploy the scoring systems. These actors select training datasets, define objective functions, and choose how to trade off accuracy, profit, and fairness across populations.[43] They decide whether, and how, to impose fairness constraints; which

---

[41] Lauri Kai, *supra* note 3, at 3–4, 12–14; Nicholas Schmidt & Bryce Stephens, *An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination*, arXiv 130, 130, 134–138 (2019); Xukang Wang et al., *supra* note 3, at 1–5; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 673–675, 691–692.
[42] Natalie Sheard, *supra* note 3, at 627–629, 633–635; Xukang Wang et al., *Id*, at 2, 5–7.
[43] Neil Menghani, Edward McFowland III & Daniel B. Neill, *Insufficiently Justified Disparate Impact: A New Criterion for Subgroup Fairness*, ARXIV:2306.11181 1, 1–2, 7 (2023); Greta Coraglia et al., *Evaluating AI Fairness in*

metrics count as acceptable performance; and what levels of disparate impact are tolerated as the price of efficiency.[44] Crucially, they also control access to portfolio-level and population-level data—the only vantage point from which systemic disparate impact can be detected, measured, and addressed.[45]

Extensive work in antidiscrimination law and algorithmic fairness has shown that persistent inequality in automated credit, insurance, and employment decisions is rarely the product of isolated frontline choices.[46] Rather, it emerges from design and deployment decisions that structure how risk is defined, measured, and distributed across groups. Treating the individual clerk or HR officer as the Cheapest Cost Avoider in this setting trivializes structural injustice by collapsing system-level governance failures into individualized moments of application.

From a cost–benefit perspective, the actor best positioned to govern algorithmic discrimination is the institution or vendor that sets the system's goals, fairness parameters, data governance practices, and override policies.[47] These actors can recalibrate thresholds, retrain models on more representative data, audit outcomes across populations, and redesign workflows to mitigate disparate impact at scale. By contrast, the frontline employee can at most alter the outcome of a single case, often at personal or professional risk, and without access to the information necessary to assess systemic effects.

The "neutral score" myth thus mirrors the broader pattern identified throughout this Article. The apparent CCA at the point of decision is not the actor who governs algorithmic risk in any meaningful sense. Liability regimes that fixate on downstream human application misdirect incentives, shielding those who define and control the system's distributive consequences while exposing individual employees to responsibility for harms they neither designed nor can effectively prevent.

Credit Scoring with the BRIO Tool, ARXIV:2406.03292 1, 2–3, 7, 15–16 (2024); Lauri Kai, *supra* note 3, at 2–4, 10–14, 20–28.

[44] Michael Feldman, Sorelle A. Friedler, John Moeller, Carlos Scheidegger & Suresh Venkatasubramanian, *Certifying and Removing Disparate Impact*, ARXIV:1412.3756 1, 2–3 (2015); Xukang Wang et al., *supra* note 3, at 2–4; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 688–692.

[45] Xukang Wang et al., *Id*, at 7–8; Solon Barocas & Andrew D. Selbst, *Id*, at 711–714.

[46] Lauri Kai, *supra* note 3, at 2–7; Natalie Sheard, *supra* note 3, at 625–627; Solon Barocas & Andrew D. Selbst, *Id*, at 673–675 ; Xukang Wang et al., *Id*, at 3–4.

[47] Solon Barocas & Andrew D. Selbst, *Id*, at 677–692; Xukang Wang et al., *Id*, at 1, 3–7; Lauri Kai, *Id*, at 12–14, 26–28.

Algorithmic scoring in credit, insurance, and employment thus exposes a familiar liability illusion: the law assigns responsibility to the last human decision-maker while ignoring where algorithmic risk is actually designed, calibrated, and governed. Downstream discretion is treated as control, even when it is procedurally discouraged and informationally empty. This misalignment is not confined to scoring systems. It resurfaces even more starkly in debates over generative AI, where users are portrayed as autonomous risk creators based on prompts and publication choices, while the architecture that shapes and constrains harmful outputs is built and controlled upstream. The next subsection examines this displacement of responsibility through the lens of generative AI and the "user-control" myth.

D. Generative AI and Content Harms: The "User-Control" Myth

Generative AI systems can produce a wide range of content-related harms, including defamation, copyright infringement, privacy violations, incitement, and other legally cognizable injuries. Much of the emerging legal and policy discourse frames these harms through the lens of user behavior. On this account, the end user who crafts the prompt or republishes the output is treated as the primary locus of responsibility: the user could have chosen different prompts, exercised greater restraint, or refrained from dissemination altogether.[48] From a traditional tort perspective, this framing is intuitively attractive. The user appears to be the CCA, exercising direct control at the moment closest to the harm.

Yet this intuition rests on a deeply misleading understanding of how generative AI systems actually generate, constrain, and propagate risk. A growing body of technical and governance-oriented scholarship demonstrates that content risk in generative AI is centrally shaped by platform-level design and deployment choices, including the structure of guardrails, filtering mechanisms, logging and monitoring practices, access tiers, and safety fine-tuning regimes.[49] At the same time, research on adversarial prompting and so-called "jailbreaks" shows that even well-

[48] Richard J. Tong et al., *A First-Principles Based Risk Assessment Framework and the IEEE P3396 Standard*, ARXIV:2504.00091 1, 2–6 (2025); Laura Weidinger et al., *Sociotechnical Safety Evaluation of Generative AI Systems*, ARXIV:2310.11986 1, 6–12 (2023); Susan Hao et al., *supra* note 4, at 2–4.
[49] Laura Weidinger et al., *Id*, at 6–11; Susan Hao et al., *Id*, at 1–5.

intentioned user conduct cannot reliably neutralize content risks when system-level constraints are brittle, incomplete, or strategically bypassable.[50]

These limitations are not accidental. Generative AI models are designed to be flexible, general-purpose, and responsive to a wide variety of inputs. As a result, platform providers face persistent trade-offs between expressive capacity, usability, and safety. The management of these trade-offs occurs upstream, through decisions about training data, reinforcement learning objectives, refusal behaviors, and post-deployment oversight. Empirical work on generative AI safety therefore emphasizes the centrality of systematic risk assessments, red-teaming, incident reporting, and iterative patching at the model and platform level.[51] These practices, when present, shape the baseline level of harm for all users simultaneously; when absent or under-resourced, they leave downstream actors exposed to risks they cannot meaningfully diagnose or mitigate.

By contrast, the user's preventive capacity is narrow and episodic. A user may avoid causing harm in a particular interaction, but lacks the ability to recalibrate default behaviors, redesign safety layers, or deploy updates that reduce risk across the system as a whole. In tort terms, the user's control is real but case-specific, while the platform's control is systemic. Treating the former as the primary focus of liability therefore misdirects incentives. It encourages after-the-fact blame of individual prompting or publication choices while undercutting investment in "safety by design" measures that only platform-level actors can implement.[52]

Once this structural asymmetry is made explicit, the identity of the relevant BDM becomes clearer. The most consequential cost—benefit decisions in generative AI concern how much harmful content is tolerable, which categories trigger refusals, how aggressively models generalize from training data, and how quickly vulnerabilities are detected and patched. These are not decisions that can be made meaningfully by end users. They require access to population-level data, technical expertise, and the capacity to implement changes that propagate across millions of interactions. In most generative AI ecosystems, these capacities reside with model developers and

[50] Banerjee, Somnath et al., *How (Un)Ethical Are Instruction-Centric Responses of LLMs? Unveiling the Vulnerabilities of Safety Guardrails to Harmful Queries*, ARXIV 193, 193–199 (2024); Federico Bianchi & James Zou, *Large Language Models Are Vulnerable to Bait-and-Switch Attacks for Generating Harmful Content*, ARXIV 1, 1–2 (2024).
[51] Richard J. Tong et al., *supra* note 48, at 1–5; Chen Chen et al., *supra* note 18, at 6–14.
[52] Ssan Hao et al., *supra* note 4, at 2–3; Laura Weidinger et al., *supra* note 48, at 7–12 ; Richard J. Tong et al., *Id*, at 2–6.

platform providers, sometimes in conjunction with large institutional deployers that integrate generative systems into consumer-facing products or services.[53]

From the perspective advanced in this Article, the "user-control" narrative therefore exemplifies a recurring liability illusion. Downstream discretion is treated as if it were genuine governance over risk, while the upstream actors who design, calibrate, and continuously update the system remain comparatively insulated from tort-based accountability. The result is a familiar misalignment: liability signals are sent to actors who cannot efficiently reduce future harm, while those who can redesign the system to alter the distribution of risk face weaker legal pressure to do so.

This misalignment does not require absolving users of all responsibility. Users may still bear duties related to intentional misuse, republication, or reliance in particular contexts. But focusing tort liability primarily on user behavior misconceives where meaningful prevention occurs in generative AI systems. The actor best positioned to evaluate cost–benefit trade-offs and to implement system level safeguards is typically the platform or model provider—the actor who functions, in practice, as the Best Algorithmic Risk Governor (BARG) for generative content risks.

Seen in this light, generative AI does not present an isolated anomaly in tort law, but rather a particularly clear manifestation of a broader structural error. Once again, liability intuition gravitates toward the last human actor in the chain, mistaking formal discretion for genuine control over risk. What appears as user choice at the surface is in fact tightly channeled by upstream architectures, defaults, and governance decisions that shape harmful outcomes across users and contexts. This displacement of responsibility is not unique to generative content systems. It recurs across AI-mediated domains whenever human involvement is preserved largely as a legal placeholder rather than as a realistic locus of risk governance. The next subsection distills this recurring error into a general insight—the "last human" pattern—and explains why it poses a foundational challenge to tort law's traditional liability framework.

E. Interim Insight: The "Last Human" Pattern

[53] Philipp Hacker et al., *supra* note 18, at 1115–1117; Laura Weidinger et al., *Id*, at 7–11; Richard J. Tong et al., *Id*, at 3–5.

The preceding sections examined distinct doctrinal arenas—medical AI, semi-autonomous vehicles, algorithmic scoring in credit and employment, and generative AI content systems. Although these domains differ technologically and institutionally, they expose a common structural feature of AI-mediated harm.

Tort intuition repeatedly converges on the same figure: the last human in the chain.

In each setting, the actor temporally and spatially closest to the injury—the physician reading the scan, the driver supervising the vehicle, the loan officer applying the score, the user crafting the prompt—appears, at first glance, to be the CCA. This actor occupies the final decision node, fits comfortably within existing negligence narratives, and offers courts a familiar anchor for duty, breach, and causation analysis. The legal story is straightforward: *a human saw the output, could have done more, and did not.*

But once the analysis shifts from the isolated incident to the architecture of algorithmic risk production, this intuition begins to fail.

Across AI ecosystems, the most consequential cost—benefit trade-offs are rarely made at the point of application. They are embedded upstream in decisions about model architecture, training data, objective functions, confidence thresholds, escalation rules, interface defaults, workflow integration, and post-deployment monitoring and updating. These choices determine the system's baseline error profile—which errors are frequent, which are rare, and which are structurally obscured. They also determine how much room for meaningful human intervention actually exists downstream. The frontline actor's "discretion" is therefore often constrained by design, channeled by institutional protocol, and informationally hollowed out by opacity.

This produces a systematic divergence that recurs across domains:

The apparent CCA is often the last manual link in the chain.
The true best cost–benefit decision-maker—the actor positioned to govern algorithmic risk systemically—is typically located upstream, where architectures, datasets, thresholds, and governance processes are shaped.[54]

[54] Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 13–15; Muhammad Uzair, *supra* note 2, at 12–15; Laura Weidinger et al., *supra* note 48, at 7–9, 22–25; George Maliha et al, *supra* note 1, at 630–634, 637–641; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 677–692, 715–722.

This is the "Last Human" pattern: tort law's liability intuition gravitates toward the most visible human actor, even when that actor lacks meaningful control over the system-level parameters that structure risk across cases. Visibility is mistaken for governance. Proximity to harm is conflated with capacity to reduce future harm.

Doctrinally, this pattern reveals a structural mismatch between classical CCA reasoning and AI-mediated harm. In many traditional accident settings, the CCA and the Calabresi–Hirschoff BDM coincide. AI systems fracture that alignment. The actor who could, in principle, avert *this* instance of harm at the last moment is often not the actor who can most efficiently redesign the system to reduce the class of harms from which this instance emerged.

The consequence is not merely analytical. If courts continue to map liability reflexively onto the last human in the loop, they risk over-deterring actors with limited systemic leverage while under-deterring those who actually shape the distribution of algorithmic risk. Tort law would then function less as an incentive mechanism for safer design and more as a mechanism of scapegoating.

This interim insight therefore necessitates a new organizing concept—one that tracks governance capacity over algorithmic risk, rather than mere proximity to injury. The next section introduces that concept: the BARG—the actor best positioned to gather information, perform the relevant cost–benefit analysis, and implement system-level changes that alter the baseline level of risk for many users at once.

V. Re-Centering Tort Law: The Actor Who Actually Governs Algorithmic Risk

A. Identifying the Real Risk Governor

Once AI-related harm is understood as the product of layered architectures, radical information asymmetries, dynamic updating, and pattern-based risk, the central tort inquiry must be reformulated at a more structural level. The relevant question is no longer merely who could have prevented *this* injury at the last moment, but rather who governs the system that makes injuries of this type systematically more or less likely across cases. The shift is from episodic control to systemic governance, from the moment of harm to the architecture of risk production. The BARG concept provides the analytic framework for performing that shift within tort doctrine.

The BARG is the actor (or set of actors) in an AI ecosystem best positioned to gather and interpret information about algorithmic risks and harm patterns across users and contexts; to evaluate cost—benefit trade-offs among competing design, deployment, and governance options; and to implement system-level changes—technical or organizational—that modify the baseline level of algorithmic risk for many users at once. Each component of this definition is critical. Information-gathering matters because algorithmic risk is statistical and distributed rather than episodic. Cost–benefit evaluation matters because AI safety decisions are rarely binary; they involve continuous trade-offs among accuracy, robustness, fairness, and operational constraints. System-level implementation capacity matters because only interventions that propagate across users can meaningfully alter the overall distribution of harm rather than merely shifting outcomes in isolated instances.

The BARG framework therefore treats algorithmic risk as a governance problem embedded in socio-technical systems rather than as a series of disconnected human mistakes. It recognizes that meaningful control over AI-related harm depends on actors who operate at the level of model design, data governance, deployment architecture, monitoring, and updating practices—domains where the structure of risk is defined before any particular user encounters the system. By centering these actors, the BARG concept translates tort law's concern with efficient accident prevention into the language of algorithmic system governance, preserving the economic logic of liability while adapting it to environments characterized by scale, opacity, and continuous change.[55]

B. Why BARG Is More Than BDM Applied to AI

At this point, a fair objection must be confronted directly. If the BARG framework builds on Calabresi and Hirschoff's best decision-maker analysis, is it doing anything more than giving the BDM concept a new label for the age of artificial intelligence? The answer is yes—but the claim requires precision. BARG does not purport to replace the BDM framework, nor does it reject the economic logic from which BDM emerged. Rather, BARG operationalizes that logic under conditions that the classical framework did not have to confront in systematic form: layered

[55] Gabriel Lima & Meeyoung Cha, *supra* note 20, at 1–3; Chen Chen et al., *supra* note 18, at 1–14, 67–70; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 1–4, 7–9; Madalina Busuioc, *supra* note 7, at 828–834.

algorithmic production, radical information asymmetries, continuous updating, feedback loops, and pattern-based harm at scale.

The BDM test asks the right foundational question: who is best positioned to make the relevant cost–benefit decision and to act upon it? In simple or moderately complex accident settings, that inquiry may be sufficient. The court can ask which party had superior information, superior capacity to compare accident costs and avoidance costs, and superior ability to adjust behavior accordingly. In such settings, the identity of the BDM can often be inferred from ordinary indicia of control: who designed the product, who operated the activity, who selected the precaution, or who could have altered the relevant conduct at reasonable cost. But AI systems alter the structure of the inquiry. They do not merely add technical complexity to familiar tort problems; they change where risk is generated, where it can be observed, and where it can be reduced.

BARG is therefore not a competing theory of accident-cost allocation, but a more specific governance-oriented instantiation of BDM for algorithmic systems. Its analytic contribution lies in shifting the inquiry from decisional capacity in the abstract to risk-governance capacity in a particular socio-technical ecosystem. In AI-mediated environments, the relevant cost–benefit decision is rarely a single discrete choice made at the point of harm. It is usually embedded in a series of upstream and ongoing decisions: what data to use, how to train the model, which loss function to privilege, how to calibrate thresholds, how to design the interface, how to structure human oversight, how to monitor performance after deployment, and when to update, retrain, restrict, or withdraw the system. These are not merely technical implementation details. They are the institutional locations at which the accident-cost calculus is actually performed.

This is why BARG adds something that BDM alone leaves underdeveloped. BDM identifies the type of actor tort law should care about: the actor capable of making and acting upon the relevant cost–benefit judgment. BARG identifies the kinds of capacities that matter when that judgment concerns algorithmic risk: access to training and performance data, control over model architecture, ability to observe outcomes across a population of cases, capacity to detect drift or disparate error patterns, authority to change deployment conditions, and ability to propagate safety improvements across users. Put differently, BARG translates BDM’s general economic question into operational markers suited to algorithmic environments. It asks not only who is best placed to decide, but who is best placed to govern the system through which risk is repeatedly produced.

The distinction also matters because BARG can generate different liability outcomes than a court might reach under an undifferentiated BDM analysis. A general BDM inquiry may still be pulled toward the actor who appears to exercise professional or operational judgment in the individual case: the physician who accepted an AI-generated diagnosis, the driver who failed to resume control, the loan officer who relied on a proprietary score, or the content moderator who applied a platform's automated classification. Each of these actors appears to make a "decision." Each is proximate to the injury. Each may be described, at least formally, as having the last opportunity to avoid harm. But BARG directs attention away from formal decisional moments and toward the actor that controls the risk architecture that made those moments legally and practically meaningful. In many AI cases, that actor will be upstream: the developer, platform provider, or institutional deployer that controls system design, monitoring, updating, and integration.

This does not mean that BARG always selects the same defendant, or that it mechanically assigns liability to the most technologically sophisticated actor. The point is functional, not categorical. A model developer may be the BARG where the relevant risk arises from training data, architecture, calibration, or known model limitations. An institutional deployer may be the BARG where the risk arises from workflow integration, staff incentives, override policies, alert thresholds, or failure to monitor local performance. A platform provider may be the BARG where it controls logging, guardrails, access rules, safety tooling, or system-wide updates. In some cases, algorithmic risk will be jointly governed, and more than one actor may perform BARG-like functions. But even in those cases, the framework clarifies the relevant inquiry: liability should track functional governance over the risk, not merely proximity to the injury or formal participation in the final decision.

BARG also functions as a doctrinal tool rather than merely a descriptive label. It can guide courts in applying familiar tort concepts—duty, breach, causation, and the choice between negligence, product liability, and institutional liability—without requiring courts to invent an entirely new cause of action. At the duty stage, BARG helps identify which actors have sufficient control over algorithmic risk to justify legally cognizable obligations of care. At the breach stage, it focuses the standard of care on governance practices such as validation, documentation, monitoring, updating, warnings, escalation pathways, and realistic human-override design. At the causation stage, it helps courts see that the legally relevant cause may lie not in a single downstream act, but in a system configuration that materially increased a class of foreseeable harms. And at the remedy and

deterrence stages, it channels liability toward the actors capable of changing the system for future cases, rather than merely punishing the last human link in the chain.

In this sense, BARG is best understood as a bridge between economic tort theory and algorithmic governance. It preserves the central insight of CCA and BDM analysis—that liability should be assigned in a way that induces the actor best positioned to reduce accident costs to do so. But it recognizes that, in AI ecosystems, accident costs are often shaped by governance decisions that are distributed, technical, dynamic, and invisible at the point of injury. The legal challenge is therefore not simply to find the cheapest cost avoider or even the best abstract decision-maker. It is to identify the actor with the institutional capacity to observe, evaluate, and alter the algorithmic system that generates risk across cases.

That is the conceptual work BARG performs. It does not claim that BDM was wrong; it claims that BDM requires specification when the "decision" at issue is no longer a discrete human choice but an ongoing governance process embedded in software, data, organizational routines, and platform architecture. BARG is that specification. It turns the BDM insight into a usable liability framework for AI by asking where algorithmic risk is actually governed—and by insisting that tort law's incentives should be directed there.

C. Functional Markers of Algorithmic Control

To operationalize the BARG framework, courts require criteria that move beyond formal labels and toward functional indicators of algorithmic control. The central question is not who is nominally responsible for an AI system, but which actor (or actors) possesses the practical capacity to shape the system's risk profile across cases. The following functional markers identify where meaningful governance over algorithmic risk actually resides.

To make this inquiry more concrete, courts can treat BARG identification as a functional, multi-factor test. The factors below are not mechanically dispositive; rather, they indicate where algorithmic risk is actually designed, observed, updated, and internalized. The greater an actor's control over the high-weight factors, the stronger the case for treating that actor as the relevant BARG.

Table 1. A Multi-Factor Test for Identifying the Best Algorithmic Risk Governor

Factor	Relative Weight	Doctrinal Significance	Typical Indication of BARG Status
Control over model architecture and system design	High	Identifies the actor that defines the system's baseline risk profile before any downstream use occurs.	The actor selects or controls model architecture, training pipelines, parameters, thresholds, interfaces, or core safety features.
Control over training data and data governance	High	Shows who shapes the informational foundation from which algorithmic errors, biases, and blind spots emerge.	The actor selects, curates, labels, validates, cleans, excludes, or updates the data on which the system depends.
Access to outcome data and population-level observability	High	Determines who can detect recurring error patterns, disparate impacts, drift, or systematic underperformance across cases.	The actor possesses logs, performance metrics, incident reports, post-deployment outcomes, feedback data, or auditing capacity across users or populations.
Ability to update, retrain, recalibrate, suspend, or withdraw the system	High	Captures who can respond to liability signals by reducing future risk at scale rather than merely correcting isolated outcomes.	The actor can deploy patches, change thresholds, retrain the model, alter guardrails, rollback versions, restrict use, or remove the system from deployment.
Control over deployment conditions and workflow integration	High/Medium	Identifies who determines how algorithmic output is translated into real-world decisions and whether human oversight is meaningful or merely formal.	The actor controls procurement, configuration, escalation rules, override policies, alert settings, user training, staffing incentives, or institutional reliance on the system.
Capacity to monitor, audit, and document ongoing performance	High/Medium	Links BARG status to the ability to maintain reasonable post-deployment	The actor conducts or controls validation, auditing, documentation, red-teaming, drift

		governance over a dynamic system.	detection, incident response, or periodic review.
Ability to spread costs and internalize residual risk	Medium	Reflects traditional tort-law concerns with efficient cost allocation, insurance, pricing, and enterprise-level risk management.	The actor can insure, price risk into the service, distribute losses across users or customers, or invest in system-wide precautions.
Contractual or practical authority over other actors in the AI value chain	Medium	Shows whether the actor can impose safety obligations on developers, vendors, deployers, or users even without direct technical control.	The actor can require documentation, audits, safety standards, data access, compliance reports, indemnification, or contractual safeguards.
Proximity to the immediate harm	Low	Helps explain why the last human actor should not automatically be treated as the BARG merely because she is closest to the injury.	The actor is present at the point of harm but lacks meaningful control over design, data, monitoring, updating, or system-level risk reduction.
Formal human discretion at the point of use	Low	Distinguishes nominal decision-making from genuine governance capacity.	The actor can accept, reject, or question an output in a single case, but cannot alter the system's risk architecture or observe patterns across cases.

This table should be read functionally rather than formally. No single factor is necessary in every case, and no factor is always sufficient on its own. The inquiry is cumulative: an actor that controls design, data, observability, updating, and deployment conditions will ordinarily be a strong BARG candidate, even if it is distant from the immediate injury. Conversely, an actor who is proximate to the harm but lacks access to system-level information, updating authority, or meaningful control over deployment should not be treated as the BARG merely because she is the last human in the loop. The table therefore translates the BARG concept into a doctrinally usable

test: responsibility should track functional governance over algorithmic risk, not formal proximity to an individual accident.

1. Design control

The most salient marker of algorithmic control is authority over system design. Actors who determine model architecture, training pipelines, parameter tuning, confidence thresholds, user interfaces, and integration into organizational workflows effectively define the system's baseline error distribution and its interaction with human users.[56] These design choices encode *ex ante* judgments about acceptable trade-offs between false positives and false negatives, sensitivity and specificity, automation and discretion. Because such decisions are made upstream and propagate across all downstream uses, design control is a strong indicator of BARG status.

2. Data access and observability

Algorithmic risk is statistical and pattern-based rather than episodic. Accordingly, the ability to observe system behavior at scale is a second critical marker. Actors with access to logs, performance metrics, post-deployment outcome data, and feedback across populations are uniquely positioned to detect systematic failure modes, bias, drift, or feedback-loop amplification.[57]By contrast, actors limited to isolated encounters with the system lack the informational basis required for meaningful cost—benefit analysis of algorithmic precautions. Control over observability therefore maps closely onto control over risk governance.

3. Systemic leverage

A third marker concerns the capacity to implement changes that propagate broadly rather than locally. Actors who can deploy updates, patches, configuration changes, policy modifications, or infrastructural constraints can alter the system's behavior for many or all users simultaneously.[58] This capacity distinguishes governance from discretion: the former reshapes the risk environment itself, while the latter merely affects outcomes in individual cases. Systemic leverage is thus central to identifying who can actually respond to liability signals by reducing future harm.

4. Economic capacity and incentives.

Finally, tort law has long emphasized the importance of assigning liability to actors who can

---

[56] Muhammad Uzair, *supra* note 2, at 5–6, 12–13; Weisz et al., *supra* note 4, at 5–9; George Maliha et al, *supra* note 1, at 633–634.

[57] Savina D. Kim, Stefan Lessmann, Galina Andreeva & Michael Rovatsos, *supra* note 26, at 10–15; Adriano Koshiyama et al., *supra* note 21, at 4–6; Madalina Busuioc, *supra* note 7, at 826, 829–834; Xukang Wang et al., *supra* note 3, at 6–7.

[58] Aagaard Lise, *supra* note 30, at 125–126; Laura Weidinger et al., *supra* note 48. At 7–9, 19–20, 22–23, 27–28; Philipp Hacker et al., *supra* note 18, at 1115–1116, 1118–1120; Adriano Koshiyama et al., *Id*, at 4–7.

internalize the costs of precaution and residual harm. In algorithmic environments, this consideration takes on renewed importance. Actors with the financial capacity to invest in safer design, monitoring, and governance—and to spread residual risk through pricing or insurance—are better positioned to respond efficiently to liability incentives without undermining socially valuable innovation.[59] Economic capacity therefore functions as both an efficiency filter and a realism constraint on BARG identification.

Taken together, these functional markers underscore that BARG status is not tied to formal proximity to the harm or to nominal human involvement at the point of decision. It tracks control over the architecture, data, and update mechanisms that generate risk across cases. Importantly, the BARG is not necessarily a single entity. In many AI ecosystems, a primary BARG (such as a model developer or platform provider) will coexist with secondary BARGs (such as large institutional deployers) whose integration and governance choices materially shape system-level risk.

The BARG is not necessarily a single entity; in many cases, a *primary BARG* (e.g. the model developer) will coexist with *secondary BARGs* (e.g. a large institutional deployer or platform provider).

These functional markers serve a deliberately pragmatic role. They translate the abstract insight that algorithmic risk is governed upstream into criteria that courts can actually apply when allocating liability. Rather than asking who touched the harm or who formally retained discretion at the point of decision, the markers redirect attention to where algorithmic risk is designed, observed, and recalibrated over time. In doing so, they provide a bridge between the economic logic of the best cost–benefit decision-maker and concrete doctrinal analysis.

The sections that follow build on this framework to show how traditional tort doctrines—negligence, product liability, and institutional responsibility—can be reinterpreted once liability is anchored in functional control over algorithmic systems rather than episodic human involvement. Read through the lens of these markers, familiar doctrines no longer ask whether the last human actor behaved reasonably in isolation, but whether the relevant risk governor exercised reasonable

[59] Emanuela Carbonara, Alice Guerra & Francesco Parisi, *supra* note 9, at 173–176, 190–191; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 8–11; Roberto Pardolesi & Bruno Tassone, *supra* note 5, 11–12, 29–31.

care in designing, monitoring, and updating a system whose errors predictably propagate across cases.

D. From Cheapest Cost Avoider to Risk Governor

The conceptual move from the CCA to the BARG does not abandon the economic logic of tort law, but rather exposes its limits in contemporary technological settings. Classical CCA analysis was designed for environments in which accident prevention is relatively localized and where the actor best positioned to avert harm in a given instance is also the actor who can efficiently internalize the costs of precaution. In such contexts, identifying the CCA provides a workable proxy for allocating responsibility in a way that minimizes the social costs of accidents.

Yet even within Calabresi's own framework, this proxy was never meant to be exhaustive. Where harm prevention depends not on simple, observable precautions but on complex judgments about design, information, and system-wide trade-offs, the identity of the cheapest cost avoider becomes increasingly indeterminate. This concern motivated Calabresi and Hirschoff's refinement of the analysis through the BDM concept, which shifts the focus from physical prevention to decisional capacity. The relevant inquiry becomes who is best situated to perform the cost—benefit analysis between accident costs and accident-avoidance costs and to act upon that analysis in a meaningful way. Seen in this light, the BDM framework already anticipates settings in which responsibility should attach not to the last actor in the causal chain, but to the actor who governs the conditions under which risk is produced and managed. The BARG concept builds directly on this insight, extending it to algorithmic environments in which risk is systemic, probabilistic, and generated through layered socio-technical systems rather than isolated human acts.[60]

Algorithmic systems make this extension necessary. In AI-mediated environments, the actor who appears to be the CCA in any individual incident—often a frontline human user—is frequently not the actor who governs risk in a meaningful sense. While such users may retain nominal discretion at the point of application, they typically lack control over the design choices, training data, thresholds, deployment architecture, and updating practices that determine the system's baseline error profile. By contrast, upstream actors—such as developers, platform operators, and large institutional deployers—are positioned to observe algorithmic behavior across cases, to evaluate

[60] *See* generally John C.P. Goldberg, *supra* note 5; Roberto Pardolesi & Bruno Tassone, *supra* note 5, at 11–12.

trade-offs among alternative configurations, and to implement changes that propagate across many decisions simultaneously. In these settings, the capacity to reduce harm lies less in last-moment intervention and more in the governance of the system that repeatedly generates risk. The divergence between the apparent CCA and the actor who actually governs algorithmic risk is therefore not accidental, but structural.[61]

Understood in this way, the Best Algorithmic Risk Governor is best seen as an algorithmic specification of the BDM rather than as a departure from established tort theory. It captures the intuition that, where risk is produced by complex and scalable systems, responsibility should follow governance capacity rather than proximity to harm. The shift from CCA to BARG thus reframes tort law's efficiency inquiry without altering its core commitment: allocating liability so as to incentivize those actors who are best positioned to reduce the social costs of harm.

This conceptual shift from accident prevention to risk governance sets the stage for examining how algorithmic systems systematically decouple apparent control from actual control—and why that decoupling repeatedly misleads tort law's traditional liability intuitions.

E. Limits and Objections: Over-Deterrence, Operational Knowledge, and Shared Governance

The BARG framework is not without limits. Indeed, if it were read too aggressively, it could reproduce a familiar problem in tort law: a concept designed to improve deterrence might become a vehicle for excessive liability. Three objections therefore require direct attention before the doctrinal implications are developed. First, would shifting liability toward developers, platforms, and institutional deployers over-deter socially valuable AI innovation? Second, does the framework understate the operational knowledge possessed by hospitals, banks, employers, and other institutional users—knowledge that may be unavailable to model developers? Third, if algorithmic risk is often jointly produced by developers, deployers, and users, why should tort law search for a single BARG rather than apportion responsibility among multiple actors?

The over-deterrence objection is serious, but it rests on a misunderstanding of what BARG requires. The framework does not impose strict, automatic, or enterprise-wide liability for every

[61] Laura Weidinger et al., *supra* note 48, at 7–12; George Maliha et al, *supra* note 1, at 630–635; Muhammad Uzair, *supra* note 2, at 1–8; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 674–692.

harm involving an AI system. Nor does it treat technological sophistication as a sufficient basis for responsibility. BARG identifies the actor best positioned to govern the relevant risk: the actor with access to the information, design levers, monitoring capacity, updating authority, and institutional ability to reduce future harms across cases. Liability should attach to failures of reasonable algorithmic governance, not to the mere fact that an AI system was used or that an AI-related harm occurred.

This distinction matters because innovation can be chilled not only by excessive liability, but also by poorly targeted liability. A regime that places primary responsibility on frontline users—physicians, drivers, loan officers, or employees—may discourage adoption in high-value settings while doing little to improve system safety. Those actors often cannot redesign the model, audit population-level error patterns, recalibrate thresholds, or update the system. By contrast, liability directed toward genuine BARGs creates incentives for safer innovation: better documentation, stronger validation, more realistic human-override design, continuous monitoring, incident response, and timely updating. The point is not to punish AI development, but to distinguish responsible governance from ungoverned deployment.

Over-deterrence concerns can also be addressed through doctrinal calibration. Courts need not treat BARG status as conclusive of breach. An actor may be the relevant risk governor and still demonstrate reasonable care. Evidence of meaningful pre-deployment testing, post-deployment monitoring, documentation, auditing, version control, incident response, and good-faith recalibration should matter in determining breach. Similarly, carefully designed regulatory safe harbors may be appropriate where an actor can show not only formal compliance, but sustained governance practices capable of detecting and correcting emerging risks. In this sense, the BARG framework is compatible with innovation-protective doctrines, so long as protection follows demonstrated governance performance rather than mere certification or formal labeling.

The second objection is that information asymmetry cuts both ways. Developers and platform providers often control model architecture, training data, documentation, logs, and update mechanisms. But hospitals, banks, insurers, employers, and other institutional deployers may possess operational knowledge that developers lack. A hospital may know how a diagnostic system is actually integrated into emergency-room workflows; a bank may know how loan officers are instructed to rely on a score; an employer may know how an automated screening tool

interacts with productivity pressures, applicant pools, or internal compliance incentives. In these contexts, the developer may understand the model, but the deployer understands the environment in which the model becomes legally consequential.

This objection does not defeat BARG; it refines it. BARG is a functional inquiry, not a categorical preference for upstream developers. Where the relevant risk arises primarily from model design, training data, calibration, or known technical limitations, the developer or platform provider will often be the primary BARG. Where the risk arises from procurement choices, workflow integration, override policies, alert thresholds, staffing pressures, user training, or failure to monitor local performance, the institutional deployer may be the primary or at least a secondary BARG. The framework therefore accommodates the fact that different actors possess different forms of knowledge. The legally relevant question is not who knows everything, but who controls the particular knowledge and intervention points necessary to reduce the relevant risk.

This is especially important in high-stakes institutional settings. A developer may provide a model with disclosed limitations, but a hospital may deploy it in a setting for which it was not validated, suppress alerts to reduce noise, or instruct clinicians to treat low-risk outputs as presumptively reliable. A credit-scoring vendor may design a model, but a bank may determine the threshold at which human review is bypassed or discouraged. An employer may purchase an automated screening tool but decide how heavily to weight its output, whether to audit disparate impact, and whether employees are permitted to override the recommendation. In such cases, the institutional deployer is not a passive consumer of technology. It is an active governor of algorithmic risk.

The third objection concerns shared governance. Many AI harms are not produced by one actor alone. They emerge from the interaction of model developers, platform providers, institutional deployers, and sometimes users. A model may be poorly documented by the developer, inadequately configured by the deployer, and uncritically followed by the end-user. If BARG were understood to require one exclusive defendant, it would oversimplify the very socio-technical complexity the framework is designed to capture.

The better reading is different. BARG need not always be singular. In some cases, there will be one dominant risk governor; in others, there will be primary and secondary BARGs; and in still others, responsibility should be apportioned among several actors according to their respective governance functions. A developer may be responsible for architecture, training data, calibration,

and warnings. A platform may be responsible for logging, guardrails, access controls, and system-wide updates. An institutional deployer may be responsible for procurement, local validation, workflow integration, override policy, monitoring, and escalation practices. A frontline user may remain responsible where she possesses meaningful information and realistic discretion in the particular case. But proximity alone should not transform the user into the principal risk governor.

Shared governance therefore does not undermine the BARG framework. It explains why the framework must remain functional and comparative. The task is not to search metaphysically for "the" actor behind an AI harm, but to map the functions through which risk was designed, observed, deployed, and controlled. Tort law already has tools for dealing with multiple responsible actors, including comparative fault, contribution, indemnity, and joint or several liability where appropriate. BARG does not eliminate those tools. It helps courts decide how to use them by identifying which actors exercised which forms of governance over the risk-generating system.

These objections ultimately sharpen the theory rather than weaken it. BARG is not a command to impose maximal liability on upstream actors. It is a method for aligning responsibility with practical governance capacity. It guards against over-deterrence by tying liability to unreasonable governance failures rather than to AI use as such. It accounts for two-sided information asymmetries by recognizing that institutional deployers may sometimes be primary risk governors. And it accommodates shared governance by allowing courts to identify primary and secondary BARGs and apportion responsibility according to functional control. Properly understood, the framework does not simplify algorithmic ecosystems into a single defendant; it gives tort law a structured way to see where meaningful control over algorithmic risk actually resides.

F. Stop Asking Who Touched the Harm—Ask Who Governs the System

At the point where algorithmic harm materializes, tort law's instinctive move is still to ask a familiar and deceptively simple question: *who touched the harm?* Who made the last decision, who relied on the output, who could have intervened at the final moment before injury occurred? This instinct reflects a deep structural feature of negligence doctrine, which has long been organized around discrete encounters, identifiable human actors, and localized failures of care.

Yet in AI-mediated environments, this question is increasingly orthogonal to the problem tort law purports to solve. The decisive determinants of harm are rarely located at the moment of application. They are embedded upstream—in system architecture, training data, model calibration, deployment design, and post-deployment governance. Focusing liability analysis on the last human touchpoint therefore risks mistaking proximity for control, and visibility for governance.

Once harms are produced by socio-technical systems that operate at scale, update dynamically, and distribute risk across thousands or millions of decisions, the analytically prior question must shift. The relevant inquiry is not *who touched the harm*, but *who governs the system that makes harms of this type systematically more or less likely*. Tort law, if it is to remain an instrument of efficient accident prevention, must reorient its liability lens toward actors with genuine system-level control.

This shift follows directly from the logic of the best cost–benefit decision-maker. In algorithmic environments, meaningful cost–benefit analysis does not occur at the point of use. It occurs where actors can observe aggregate performance, detect patterns of failure, and compare alternative designs and governance strategies. Those actors—developers, platform operators, and large institutional deployers—are uniquely positioned to gather and interpret information about algorithmic risks across populations, contexts, and time.[62] They can observe error distributions, model drift, feedback effects, and disparate impact patterns that remain invisible to frontline users operating case by case.

Crucially, these upstream actors also make the baseline risk decisions in the first place. Choices about training data composition, loss functions, confidence thresholds, escalation rules, and workflow integration determine which errors will be common, which will be rare, and which will be structurally obscured. These design and deployment choices encode normative judgments about acceptable risk long before any particular harm occurs.[63] By the time a physician, driver, clerk, or user encounters an algorithmic output, the system's error profile has already been largely fixed.

---

[62] Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 2–3. 6–7; Adriano Koshiyama et al., *supra* note 21, at 2–5, 15–17, 26–29; Madalina Busuioc, *supra* note 7, at 828–829, 831–834 .

[63] George Maliha et al, *supra* note 1, at 632–634, 638–641; Muhammad Uzair, *supra* note 2, at 5–8, 10–11, 14–16, 24; Lauri Kai, *supra* note 3, at 11–14, 20–22, 30–33.

Finally, tort law's efficiency rationale has always been forward-looking. Liability is justified not merely as retrospective blame, but as a mechanism for inducing future risk reduction. That mechanism operates only if liability is placed on actors who can respond to legal signals by redesigning systems rather than merely altering isolated behavior. Actors with the capacity to update models, recalibrate thresholds, modify interfaces, and deploy fixes across an entire user base are precisely those who can respond most efficiently to liability pressure by reducing future harms at scale.[64]

Seen in this light, the familiar focus on the last human in the loop is not simply incomplete—it is systematically misleading. The actor who appears to be the cheapest cost avoider in a single incident is often not the actor who governs algorithmic risk across incidents. Tort law's traditional fixation on the moment of harm therefore risks over-deterring downstream users while under-deterring the upstream entities that actually shape the system's risk architecture.

The central claim of this section is thus straightforward: in AI contexts, tort law must pivot from an event-based inquiry to a governance-based inquiry. The proper focal point of liability is the actor—or constellation of actors—who controls the informational, architectural, and organizational levers through which algorithmic risk is produced and managed. Asking who governs the system, rather than who touched the harm, is not a normative luxury. It is a doctrinal necessity if tort law is to align responsibility with real capacity for risk reduction in algorithmic societies.

VI. From Accidents to Architecture: Tort Law in the Age of Algorithmic Risk

A. Negligence Revisited: Reasonable Care in Algorithmic Systems

Negligence doctrine has long revolved around a deceptively simple inquiry: did the defendant fail to exercise reasonable care under the circumstances? In economic terms, this inquiry is often glossed through a Hand-style cost—benefit analysis, asking whether the burden of additional precautions would have been lower than the expected reduction in accident costs. In traditional,

[64] Emanuela Carbonara, Alice Guerra & Francesco Parisi, *supra* note 9, at 174–176, 191–194; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 8–13; Roberto Pardolesi & Bruno Tassone, *supra* note 5, at 11–12, 15–16, 25, 29, 31–32.

non-algorithmic settings, this framework typically directs courts toward the actor closest to the injury—the person who acted, failed to act, or could have intervened at the decisive moment.

In algorithmic systems, however, this intuitive mapping between reasonable care and proximity to harm systematically breaks down. The content of "reasonable care" cannot be specified by focusing solely on the last human decision-maker, because the most consequential cost–benefit decisions that shape algorithmic risk are made elsewhere: upstream, *ex ante*, and at scale. As a result, negligence analysis in AI contexts must be reinterpreted through the lens of the BARG—the actor best positioned to evaluate and govern algorithmic risk across cases rather than merely react to it in isolated incidents.

For **developers and vendors**, reasonable care encompasses a suite of system-level governance practices rather than isolated technical competence. It includes robust training and validation procedures; careful selection and documentation of training data; bias, robustness, and stress testing across relevant subpopulations; version control and change tracking; and meaningful mechanisms for post-deployment monitoring, incident response, and recall. These practices are not ancillary to negligence analysis—they define the baseline risk profile of the system itself. Choices about model architecture, loss functions, thresholds, and retraining cadence embed normative judgments about which errors are acceptable and which harms are tolerable, and they predictably shape outcomes across thousands or millions of downstream decisions.[65]

For **institutional deployers**—such as hospitals, banks, insurers, platforms, and large employers—reasonable care lies not in blind reliance on vendor assurances, but in active governance of algorithmic integration. This includes due diligence in system selection; scrutiny of documented limitations and known failure modes; ongoing auditing of performance in real-world conditions; clear and realistic human-override policies; and training for frontline staff that reflects actual system behavior rather than formal disclaimers. Crucially, institutional choices about workflow integration, alert thresholds, default settings, and escalation procedures can transform an ostensibly advisory tool into an action-forcing or inaction-forcing mechanism. Negligence analysis

[65] Katherine Drabiak, *supra* note 17, at 4–8, 11–14; Miriam Buiten, Alexandre de Streel & Martin Peitz, *Id*, at 2–8, 13, 15–16; Adriano Koshiyama et al., *supra* note 21, at 2–7, 15–17, 28; George Maliha et al, *supra* note 1, at 632–634, 638–641.

must therefore attend to how institutional configuration decisions materially shape risk, even when the underlying model remains unchanged.[66]

By contrast, for **frontline users**—physicians, drivers, loan officers, moderators, or individual consumers—reasonable care should be defined more narrowly. It may include appropriate reliance on AI outputs given the information reasonably available, adherence to institutional protocols, and communication with affected individuals. But it should not be stretched to encompass system-level governance responsibilities that users neither control nor meaningfully understand. Treating frontline actors as if they bear responsibility for model calibration, data representativeness, or post-deployment monitoring mistakes formal discretion for genuine control and converts negligence doctrine into a vehicle for scapegoating rather than efficient risk reduction.[67]

Reframed in this way, negligence doctrine no longer asks simply whether the last human actor behaved reasonably in isolation. Instead, it asks whether each relevant actor exercised reasonable care commensurate with its governance capacity over algorithmic risk. This reinterpretation aligns negligence analysis with the Calabresi—Hirschoff insight that liability should follow the best cost–benefit decision-maker, rather than reflexively attaching to the most visible human at the point of harm. In algorithmic systems, reasonable care is thus inseparable from system design, deployment architecture, and ongoing risk governance—and negligence law must evolve accordingly.

B. Product and Strict Liability After Software: When Algorithms Are the Defect

BARG is not meant to displace modern product liability. Rather, it explains when and why product-liability reasoning should be directed toward upstream algorithmic risk governors. In many AI cases, existing doctrines of design defect, failure to warn, enterprise liability, and strict liability already provide the doctrinal vocabulary for shifting responsibility toward actors that design, market, deploy, or profit from scalable risk-generating systems. The difficulty is not that product liability lacks relevance. The difficulty is that, in algorithmic ecosystems, courts must identify which actor actually performs the product-like governance function: defining the system's

[66] Deimantė Rimkutė, *supra* note 33, at 66–71; Clara Cestonaro et al., *supra* note 1, at 3–5, 8–10; Madalina Busuioc, *supra* note 7, at 828–834; Xukang Wang et al., *supra* note 3, 3–7, 9–10.
[67] Natalie Sheard, *supra* note 3, at 625–630, 633–638; Clara Cestonaro et al., *Id*, at 3–4, 6–10; George Maliha et al, *supra* note 1, at 632–634, 638–640.

baseline risk profile, controlling the data and architecture, monitoring performance, issuing warnings, updating the system, and spreading residual losses.

Put differently, product liability supplies the doctrinal form; BARG supplies the sorting principle. Product doctrine asks whether a product was defectively designed, inadequately warned against, or placed into the stream of commerce under conditions that justify enterprise responsibility. BARG helps courts determine where that inquiry should attach when the "product" is not a static object but a dynamic software-and-data system distributed across developers, platforms, vendors, and institutional deployers. The point is therefore not to replace product liability with a new label, but to prevent product-liability analysis from becoming trapped by formal categories or by the visible downstream user. When algorithmic design choices generate systemic risk, product-liability reasoning should follow the actor with functional control over those choices.

Debates over whether AI systems should be characterized as "products" or "services" have intensified, particularly in light of emerging European initiatives and AI-specific liability proposals. Yet from a BARG-oriented perspective, this formal categorization question is less important than a functional one: where is systemic design risk located, and which actor is best positioned to internalize and govern it?[68]

When an AI system operates as a standardized, scalable artifact—such as a medical diagnostic device, an autonomous driving control system, or a prepackaged credit-scoring model—the analogy to classic product liability becomes analytically powerful. In such contexts, strict or product liability directed at the BARG can serve its traditional efficiency function: internalizing systemic design risk in the hands of the actor who defines the system's baseline error profile and who can spread residual losses across users through pricing and insurance. Assigning liability at that level does not merely compensate victims; it incentivizes upstream redesign, recalibration, and safer deployment across all future cases.[69]

---

[68] J.K.C. Kingston, *Artificial Intelligence and Legal Liability*, in *Artificial Intelligence and Law*, 5–7 (Springer 2016); Sundaraparipurnan N. & Mark Potkewitz, *A Risk-Based Approach to Assessing Liability Risk for AI-Driven Harms Considering EU Liability Directive*, 3–8 (2024); also Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 4–6, 12–13; Aagaard Lise, *supra* note 30, at 125–126.

[69] Sundaraparipurnan N. & Mark Potkewitz, *Id*, at 4–8, 10–14; J.K.C. Kingston, *Id*, at 5–7, 8–11; Aagaard Lise, *Id*, at 125–126.

Within this framework, traditional product doctrines—especially design-defect and failure-to-warn analysis—require reinterpretation rather than abandonment. In algorithmic environments, "design" encompasses not only physical components, but model architecture, training data composition, feature selection, objective functions, risk thresholds, and robustness testing practices. A poorly calibrated confidence threshold, systematically unrepresentative training data, or the absence of stress testing across relevant subpopulations may constitute the functional equivalent of a design defect. Similarly, the failure to disclose known limitations, degradation risks, or population-specific performance gaps can be reframed as a failure-to-warn in a data-centric context. These doctrinal tools are already available; what changes is the object of scrutiny.[70]

The dynamic character of many AI systems complicates but does not undermine the product-liability analogy. Unlike static manufactured goods, algorithmic systems are frequently updated, retrained, and recalibrated after deployment. Strict liability in this setting may therefore need to be coupled with ongoing duties of monitoring, updating, and communicating material changes in system behavior. Where risk evolves over time—through model drift, feedback loops, or distribution shifts—the relevant "defect" may lie not in the original release alone, but in inadequate post-deployment governance. Product liability, properly adapted, can accommodate this dynamism by recognizing continuing obligations tied to control over updates and system performance.[71]

In many practical settings, the BARG will be identifiable as a manufacturer, software provider, or integrated platform that exercises design authority, data control, and systemic leverage over deployment conditions. Existing product-liability frameworks therefore offer a natural starting point for allocating responsibility—provided they are recalibrated to address software- and data-centric harms rather than purely physical malfunctions. The key insight is that when algorithms themselves embed the risk-generating design choices, the algorithm can be the defect in a legally meaningful sense, and liability should follow the actor who governs that design.[72]

---

[70] K.C. Kingston, *Id*, at 5–11; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 4–8, 15–16; Aagaard Lise, *Id*, at 125–126.
[71] Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 5–6, 10–14; Katherine Drabiak, *supra* note 17, at 8–9, 12–14; George Maliha et al, *supra* note 1, at 638–641.
[72] Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 4–8, 15–16; Aagaard Lise, *supra* note 30, at 125–126; J.K.C. Kingston, *supra* note 68, at 5–7, 10–12.

C. Who Is the Employer When the Algorithm Decides? Institutional and Vicarious Liability Reframed

As algorithmic systems increasingly organize and structure everyday professional activity, the traditional premise underlying employment-based liability—that employees function as autonomous decision-makers operating under human supervision—becomes progressively unstable. Across a growing range of domains, frontline actors no longer originate the relevant judgment but instead implement outputs generated elsewhere: by models designed, calibrated, and continuously updated at institutional or upstream levels. Physicians rely upon diagnostic scoring systems, loan officers operationalize automated credit determinations, and platform workers execute routing or pricing directives embedded in software architectures. In such environments, human actors frequently operate less as independent evaluators and more as executors of algorithmic instructions.[73]

This transformation exposes a structural tension within classical doctrines of institutional and vicarious liability. Respondeat superior presumes a hierarchical chain of supervision in which responsibility flows upward from identifiable employee negligence. Yet algorithmic governance often redistributes functional authority away from human supervisors and toward socio-technical systems embedded within institutional workflows. When decision thresholds, escalation pathways, and performance expectations are encoded directly into interfaces or automated feedback mechanisms, the meaningful locus of supervision may no longer correspond to formal managerial hierarchy.

Courts should therefore reconsider the tendency to conceptualize institutions merely as derivative bearers of liability for employee misconduct. Hospitals, digital platforms, insurers, and large employers increasingly operate as primary BARGs, exercising system-level authority over how algorithmic risk is generated, distributed, and constrained. Their responsibility should not depend exclusively upon identifying downstream negligence by individual staff members, but rather upon evaluating institutional governance choices concerning procurement, configuration, monitoring, and integration of algorithmic systems into operational decision-making.

[73] Madalina Busuioc, *supra* note 7, at 828–830, 832–833; Natalie Sheard, *supra* note 3, at 618, 622–628, 633–635.

In certain contexts, the algorithm itself functionally assumes characteristics traditionally associated with supervision. Decision architectures may constrain discretion through confidence thresholds, automated escalation rules, or performance metrics that discipline deviation from algorithmic output. A clinician instructed not to override diagnostic confidence scores, or a platform worker evaluated through automated productivity analytics, operates within a chain of control that is algorithmic rather than purely hierarchical. These developments raise a doctrinal question insufficiently addressed by existing vicarious liability frameworks: whether responsibility should continue to follow formal employment relationships alone, or instead track the locus of algorithmic control that meaningfully structures conduct.[74]

Reframing institutional liability around the BARG restores coherence to tort doctrine in algorithmic environments. When an institution elects to embed and rely upon AI systems as integral components of its workflow, it assumes governance authority over a risk-generating socio-technical environment. That authority encompasses decisions regarding vendor selection, calibration choices, override discretion, auditing practices, and post-deployment monitoring—decisions that shape risk distributions across thousands of interactions rather than a single encounter. Liability aligned with those governance capacities therefore directs legal incentives toward actors capable of implementing precaution at scale.

From a cost–benefit perspective, institutional actors uniquely possess the informational and organizational capacity necessary to govern algorithmic risk. They control access to aggregated performance data, maintain contractual leverage over developers and vendors, and possess the operational ability to suspend deployment, recalibrate thresholds, or redesign workflows in response to emerging harms. Unlike individual employees, they can transform lessons drawn from isolated incidents into system-wide preventive redesign affecting entire populations of users. Recognizing institutions as primary BARGs thus does not merely expand liability exposure; it aligns tort responsibility with the actors capable of reducing systemic risk.

Accordingly, when institutions choose to operationalize AI systems within decision-making processes, tort law should impose obligations commensurate with that system-level control. Liability grounded in BARG principles reflects not a doctrinal rupture but a principled extension

[74] Pratik Shukla, *Vicarious Liability or Liability for the Acts of Others in Tort: A Comparative Perspective*, 5 INT'L J. for Multidisciplinary Rsch. 1, 4–6 (2023); Madalina Busuioc, *Id*, at 825. 828, 832–833.

of institutional responsibility into algorithmically mediated environments, ensuring that responsibility follows governance authority rather than formal proximity to the moment of harm.[75]

D. Tort Law and AI Regulation: Complementarity, Not Displacement

If algorithmic harm is best understood as a problem of ongoing governance rather than episodic misconduct, then the question confronting tort law is not whether regulation replaces liability, but how the two systems should interact once responsibility has been relocated upstream along the liability map. The rapid emergence of dedicated AI regulatory regimes—most prominently the European Union's AI Act, the proposed AI Liability Directive, and an expanding body of sector-specific governance guidance—reflects a parallel institutional recognition that algorithmic risk must be managed ex ante through structured obligations imposed across the AI value chain, particularly for systems classified as "high risk".[76] In this respect, contemporary regulation increasingly targets the same institutional actors identified by the BARG framework: developers, large deployers, and platform operators whose design, calibration, and monitoring decisions shape risk distributions long before individual injuries materialize.

The expansion of regulatory governance, however, does not render tort law redundant. Rather, it clarifies tort law's distinctive institutional role. Regulatory regimes establish baseline safety through standardized mechanisms—conformity assessments, documentation duties, transparency requirements, and monitoring protocols—designed to prevent foreseeable harms before deployment. Tort law operates differently. It evaluates institutional competence under conditions of uncertainty and hindsight, responding to failure modes that regulation cannot fully anticipate, to deployment environments that evolve beyond certification assumptions, and to organizational incentives that convert nominal safeguards into ineffective practice.

From a BARG perspective, treating regulation as a substitute for tort liability would therefore risk reproducing the very distortions identified throughout this Article. Regulatory compliance may inform the content of reasonable care, but it cannot automatically negate it. Algorithmic systems are dynamic socio-technical arrangements rather than static products. Performance drift, feedback

[75] Aagaard Lise, *supra* note 30, at 125–126; Natalie Sheard, *supra* note 3, at 620–622, 633–635; George Maliha et al, *supra* note 1, at 630–634, 637–639.
[76] Philipp Hacker et al., *supra* note 18, at 1118–1120; Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 4–6, 10–14 ; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 2–6, 17–18.

loops, changing data environments, and workflow pressures may generate foreseeable risks even where formal regulatory benchmarks have been satisfied. If compliance were treated as a categorical shield, minimum regulatory standards would risk becoming ceilings on responsibility rather than floors for continuous governance.

At the same time, the institutional orientation of contemporary AI regulation provides an opportunity to calibrate tort incentives more precisely. Where actors functioning as Best Algorithmic Risk Governors demonstrate sustained governance capacity—through transparent risk assessments, meaningful auditing, continuous monitoring, and credible incident-reporting mechanisms—carefully structured safe harbors may be normatively justified.[77] Such protections can mitigate excessive deterrence that might otherwise discourage socially beneficial innovation or incentivize withdrawal from high-stakes applications where experimentation is socially valuable. Crucially, however, safe harbors should attach not to formal compliance alone, but to demonstrable institutional practices capable of detecting and correcting emerging risks across populations and over time. In this sense, protection follows governance performance rather than certification status.

The converse implication follows with equal force. Regulatory violations by BARGs should weigh heavily in negligence and strict-liability analysis, not merely as evidence of noncompliance but as markers of governance failure by actors uniquely positioned to internalize and operationalize regulatory knowledge.[78] Obligations directed toward developers, platforms, and institutional deployers function as operational signals about how algorithmic risk must be monitored, documented, and mitigated. When actors possessing system-level visibility and intervention capacity disregard those signals, the resulting lapse reflects not an isolated mistake but a breakdown in institutional competence at the precise point where tort law expects optimization decisions to occur.

Aligning tort doctrine and regulatory governance around BARGs therefore mitigates two symmetrical dangers that increasingly characterize AI ecosystems. On one side lies the over-deterrence of frontline users—physicians, drivers, clerks, and other end-users—who appear

[77] Chen Chen et al., *supra* note 18, at 13–14, 52–53; Adriano Koshiyama et al., *supra* note 21, at 2–7, 27–29; Richard J. Tong et al., *supra* note 48, at 4–6, 8.
[78] Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 4–6, 10–14; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, 2–5, 8–11.

proximate to harm yet lack meaningful informational or organizational control over system risk. On the other lies the under-deterrence of upstream decision-makers whose design choices, deployment constraints, and update practices structure harm across populations and across time. Complementarity between tort and regulation offers a path between these extremes. Properly aligned, regulation establishes the vocabulary of risk governance, while tort law supplies the adaptive pressure that ensures those obligations remain operational rather than symbolic. Accountability thus follows governance leverage rather than mere proximity to injury, allowing tort law to function not as a rival to AI regulation but as its dynamic partner in shaping responsible algorithmic systems.

E. Insuring the Algorithmic Society: Risk Distribution as Governance Architecture

Insurance has long occupied a central place in Calabresi's taxonomy of accident costs, particularly in the domain of secondary cost spreading. Yet in AI-driven environments, insurance performs a function that exceeds classical loss distribution. It becomes a governance multiplier—an institutional mechanism that translates tort liability into continuous oversight of socio-technical systems.

Insurance markets are already reacting to AI-related risks in healthcare, autonomous vehicles, financial scoring systems, and cyber infrastructures.[79] What distinguishes algorithmic risk, however, is not simply technological novelty but its structural character. Harm is produced not episodically but through scalable architectures of design, deployment, and iterative updating. Under these conditions, insuring individual users is analytically incomplete. The economically salient question becomes: who governs the system that generates risk across cases?

The BARG framework supplies the answer. When underwriting aligns with Best Algorithmic Risk Governors, insurance follows governance leverage rather than formal role. This alignment enables three interlocking functions.

First, pricing becomes informational governance. Premium differentiation tied to auditing capacity, drift detection protocols, incident reporting systems, and population-level monitoring transforms actuarial calculation into an incentive for institutional competence. Actors capable of

[79] Jack Boeglin, *supra* note 39, at 176, 193–194, 197–198, 200–202; Katherine Drabiak, *supra* note 17 at, 1–2, 8–11, 13–14; George Maliha et al, *supra* note 1; Sundaraparipurnan N. & Mark Potkewitz, *Id*, at 1–2, 10–14.

observing and recalibrating systemic risk face financial signals that reward sustained governance rather than minimal compliance.

Second, AI-specific liability products can embed structured governance incentives directly into coverage design. Policies directed toward developers, platforms, and large institutional deployers may condition favorable terms on demonstrable oversight practices—independent algorithm audits, documentation of threshold calibration, fairness testing across subpopulations, and transparent retraining procedures.[80] In this way, insurance operates as a private enforcement layer reinforcing tort doctrine's orientation toward upstream risk control.

Third, insurers increasingly function as cross-institutional observers of failure patterns. By aggregating claims experience across sectors, insurers may detect recurring error modes—bias amplification, model drift, inadequate override design—that remain locally invisible.[81] This aggregation capacity positions insurers as systemic risk intermediaries rather than passive indemnifiers.

Routing premiums and loss exposure through BARGs therefore does more than distribute loss. It embeds tort law's cost-internalization logic within a feedback loop of monitoring, pricing, and redesign. Where liability identifies the proper governor of risk, insurance sustains that identification over time. In dynamic algorithmic systems—where harm evolves through updates, feedback loops, and scaling effects—this sustained governance function is indispensable.

Properly structured, insurance thus serves as a bridge between episodic adjudication and continuous institutional adaptation. It amplifies tort law's preventive logic by linking financial exposure to demonstrable governance capacity. In the algorithmic society, insurance is not merely a mechanism of compensation; it is an architecture of accountability.

VII. Conclusion: Rewiring Tort Law for Algorithmic Governance

A. The BARG Turn: Restating the Core Move

Artificial intelligence exposes a growing tension at the center of modern tort theory. Continued reliance on the Cheapest Cost Avoider paradigm as the primary organizing principle of liability

[80] Adriano Koshiyama et al., *supra* note 21, at 3–7, 22–26; Richard J. Tong et al., *supra* note 48, at 2–6, 8.
[81] Susan Hao et al., *supra* note 4, at 3–8; Adriano Koshiyama et al., *Id*, at 4–7, 5–17, 22, 28.

risks directing judicial attention toward actors who are visible at the moment of harm but structurally incapable of governing the risks that produced it. The CCA framework emerged in relatively bounded accident settings in which prevention capacity and decision-making authority frequently converged in a single actor. Algorithmic environments disrupt that alignment. Harm increasingly arises from layered socio-technical systems in which design, training, deployment, monitoring, and institutional integration are temporally and organizationally dispersed. In such settings, focusing liability analysis on the last human in the loop encourages courts to fixate on proximity rather than governance, overlooking the complex, systemic, and data-driven structures through which algorithmic risk is actually produced and can be most efficiently reduced.[82]

The persistence of this downstream focus remains doctrinally understandable. Physicians, drivers, clerks, and other frontline users fit comfortably within familiar negligence narratives: they occupy identifiable roles, exercise apparent discretion, and stand closest to the injury. Yet visibility does not reliably track control. In algorithmic ecosystems, the most consequential safety decisions are embedded upstream—in architectural design choices, data curation practices, calibration thresholds, interface constraints, and institutional deployment policies that silently shape how risk is distributed across populations and over time.

To remain economically rational and normatively credible under these conditions, tort law must adjust its analytic lens. The relevant inquiry is no longer merely who could have prevented a particular accident at the lowest immediate cost, but which actor is best positioned to gather information about algorithmic risk, evaluate competing precautionary strategies, and implement interventions capable of reducing harm across cases. This move does not abandon the economic logic underlying the CCA or the best decision-maker framework. Rather, it extends that logic to environments characterized by opacity, scale, and continuous updating.

The Best Algorithmic Risk Governor captures this shift. The BARG is the actor best positioned to observe systemic risk patterns, internalize long-term accident costs, and recalibrate socio-technical systems in response to emerging harms. Directing liability toward such actors allows tort law to preserve its deterrence and loss-allocation functions while avoiding the systematic over-attribution

[82] Clara Cestonaro et al., *supra* note 1, at 9–10; Muhammad Uzair, *supra* note 2, at 1–3, 12–15; Solon Barocas & Andrew D. Selbst, *supra* note 3, at 671–672, 677–692; Laura Weidinger et al., *supra* note 48, at 6–12, 17–20; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 5–11; George Maliha et al, *supra* note 1, at 630–640.

of responsibility to downstream human actors whose apparent discretion masks structural constraint—and whose proximity to harm too often substitutes for genuine governance authority.

B. Liability as Governance Infrastructure: The Normative Payoff

Reorienting tort liability around the BARG yields an important normative payoff. Once liability is aligned with the actors who actually govern algorithmic risk, tort law ceases to function merely as a retrospective device for allocating losses and instead begins to operate as governance infrastructure for AI systems.

First, locating liability at the level of BARGs strengthens incentives for safe AI design and deployment. When responsibility attaches to the actors who control model architecture, training data, deployment parameters, and institutional governance structures, legal incentives are directed toward the points where systemic risk is created and can most efficiently be reduced. Developers, platforms, and large institutional deployers are thus induced to internalize the long-term costs of algorithmic failures and to invest in robustness, fairness, transparency, and meaningful post-deployment oversight. Liability in this configuration does not merely respond to harm; it structures the incentive environment within which algorithmic systems are designed and governed.[83]

Second, the BARG framework prevents the systematic over-attribution of responsibility to frontline human actors. Physicians, drivers, clerks, moderators, and similar operators often appear—at the level of a single incident—to be the cheapest cost avoiders. Yet their apparent control is frequently superficial. They neither design the model nor determine its training data, calibration thresholds, or institutional deployment constraints, and they typically lack the information necessary to evaluate the system's reliability across contexts. Treating these actors as the primary locus of liability therefore misdirects legal incentives while producing little systemic improvement. A BARG-oriented approach instead aligns liability with those actors who possess the informational and organizational capacity to govern algorithmic risk across cases.[84]

---

[83] Miriam Buiten, Alexandre de Streel & Martin Peitz, *Id*, at 8–11; George Maliha et al, *Id*, at 630–637; Adriano Koshiyama et al., *supra* note 21, at 2–6, 10–16, 21–28; Madalina Busuioc, *supra* note 7, at 829–834.
[84] Muhammad Uzair, *supra* note 2, at 1–8, 12–17; Natalie Sheard, *supra* note 3, at 621–635; Clara Cestonaro et al., *supra* note 1, at 2–4, 9–10; George Maliha et al, *Id*, at 630–635, 638–639.

Third, focusing liability on BARGs improves the legal system's ability to address harms that are structural rather than episodic. Algorithmic discrimination illustrates this dynamic clearly. Disparate outcomes in credit scoring, hiring, insurance underwriting, and other algorithmic domains rarely arise from isolated deviations in individual decision-making; they emerge from patterns embedded in data selection, model objectives, and deployment practices. Effective mitigation therefore depends on population-level monitoring, auditing, and recalibration—functions typically performed by developers, platforms, and institutional deployers rather than by individual decision-makers at the point of application. Locating liability at the level of algorithmic governance thus enables tort law to respond to systemic harms while encouraging institutions to detect and correct discriminatory patterns before they crystallize into widespread injury.[85]

The broader implication is that artificial intelligence need not be treated as an external disruption to tort doctrine. Properly structured, tort liability can shape the trajectory of AI integration itself. By directing legal incentives toward actors capable of governing algorithmic systems, tort law helps steer technological development toward configurations that are safer, more accountable, and more consistent with public values.

C. The Next Map: Evidence, Institutions, and Multi-BARG Futures

Calabresi's original insight—that accident law should focus on the actor best positioned to reduce the costs of harm—was never meant to be a static formula. It was a method for adapting legal responsibility to changing technological and institutional conditions. The BARG framework continues that project. By redirecting attention toward the Best Algorithmic Risk Governor, it offers a way to translate the logic of the cheapest cost avoider into the architecture of contemporary algorithmic systems. Yet identifying the relevant BARG is only the starting point. The next challenge is institutional and doctrinal: adapting procedural mechanisms, evidentiary rules, and responsibility doctrines so that tort law can operate effectively in technologically complex environments. Several directions for further inquiry follow from this shift.

One such direction concerns the problem of mass harms. Many AI-related injuries do not arise as isolated accidents but as systemic effects experienced simultaneously by large populations.

---

[85] Savina D. Kim, Stefan Lessmann, Galina Andreeva & Michael Rovatsos, *supra* note 26, at 2–4, 11–13; Xukang Wang et al., *supra* note 3, at 1–7, 9–10; Solon Barocas & Andrew D. Selbst, *supra* note 3, 671–672, 677–692, 717–719; Madalina Busuioc, *supra* note 7, at 825–826, 831–834.

Algorithmic systems used in domains such as credit scoring, employment screening, pricing, and content moderation can produce patterns of harm that are individually modest yet collectively substantial. In such circumstances, individualized litigation risks obscuring the structural nature of the harm and diffusing accountability. BARG-based liability therefore raises an important institutional question: how should responsibility be operationalized where algorithmic harms are widely distributed? Mechanisms such as class actions, collective redress procedures, and forms of public enforcement may become essential complements to the BARG framework, enabling courts and regulators to identify and discipline the relevant risk governors at scale rather than through fragmented individual claims.[86]

A second direction concerns the role of the state. Governments increasingly deploy AI systems in areas including policing, welfare administration, taxation, and immigration control. These deployments raise distinctive questions about how the BARG concept should operate in the public sector. In some cases, the relevant BARG may be the public authority that decides to adopt and structure the system; in others, it may lie with private actors responsible for design, training data, or operational architecture. Applying BARG analysis in this domain inevitably intersects with doctrines of public law, including sovereign immunity, administrative accountability, and constitutional limits on governmental power. Clarifying when public bodies themselves should be treated as BARGs—and how responsibility should be distributed between public institutions and their technological partners—remains a central challenge for the emerging law of algorithmic governance.[87]

A third direction concerns proof and causation. The epistemic structure of many AI systems complicates the operation of traditional evidentiary doctrines. Black-box models, probabilistic decision processes, and complex technological supply chains may make it difficult for injured parties to establish how a particular algorithmic output produced a legally cognizable harm. Without doctrinal adaptation, this informational asymmetry risks undermining deterrence by insulating the relevant BARG from effective liability. Courts may therefore need to reconsider evidentiary frameworks in ways that align informational access with responsibility. Instruments

[86] Aurora S. Zhang & Anette E. Hosoi, *supra* note 25, at 1014–1018, 1020–1023; Xukang Wang et al., *Id*, at 1–2, 4–8, 10; Solon Barocas & Andrew D. Selbst, *Id*, at 673–675, 684–687, 691–693, 701–714; Madalina Busuioc, *Id*, at 825–827, 823–833 .

[87] Annette Zimmermann & Chad Lee-Stronach, *Proceed with Caution*, 52 CAN. J. PHIL. 6, 6–9, 19–22 (2022); Madalina Busuioc, *Id,* at 826–830, 832–834.

such as reversed burdens of proof, evidentiary presumptions, and duties of disclosure imposed on actors controlling the relevant technological infrastructure may help restore that alignment. Integrating the BARG framework with such evidentiary innovations represents an important frontier for both scholarship and doctrinal development.[88]

Finally, many AI systems operate within layered technological ecosystems involving multiple actors distributed across global supply chains. Model developers, platform operators, downstream deployers, and regulators may each exercise partial control over the creation and management of algorithmic risk. In such environments, responsibility may not reside in a single actor but in a constellation of overlapping BARGs. The challenge, therefore, is not merely to identify the actor best positioned to govern risk, but to develop doctrines capable of allocating responsibility among multiple such actors operating at different stages of the technological pipeline. Crafting coherent principles for these multi-BARG environments—particularly in cross-border contexts—will be essential as AI systems become increasingly integrated into global infrastructures of production and governance.[89]

These lines of inquiry ultimately return tort theory to the broader ambition that animated Calabresi's work. The goal was never to freeze accident law within a single doctrinal rule, but to equip courts and lawmakers with a conceptual vocabulary capable of evolving alongside technological change. In the age of artificial intelligence, that vocabulary must now include the Best Algorithmic Risk Governor.

If courts accept this invitation, tort law will do more than assign liability after algorithmic harms occur. It will help structure the incentives that shape how AI systems are designed, deployed, and governed in the first place. By directing responsibility toward the actors best positioned to anticipate and control algorithmic risks, the BARG framework can help build a legal and

---

[88] David Fernández Llorca, Vicky Charisi, Ronan Hamon, Ignacio Sánchez & Emilia Gómez, *Liability Regimes in the Age of AI: A Use-Case Driven Analysis of the Burden of Proof*, 76 J. ARTIFICIAL INTELLIGENCE RSCH. 613, 616–617, 620–622, 630–631 (2023); Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *supra* note 20, at 1–4, 7–9; Clara Cestonaro et al., *supra* note 1, at 3–4, 9–10; Sundaraparipurnan N. & Mark Potkewitz, *supra* note 68, at 1, 3. 5–6, 9, 12–13.

[89] Gabriel Lima & Meeyoung Cha, *supra* note 20, at 2–3; Yunfei Ge, Ya-Ting Yang & Quanyan Zhu, *Id*, at 1–3, 7–8; Philipp Hacker et al., *supra* note 18, at 1115–1117, 1120; Miriam Buiten, Alexandre de Streel & Martin Peitz, *supra* note 5, at 5–6, 11–13.

economic environment in which the development of AI is continuously recalibrated in light of human safety, fairness, and dignity.

In the algorithmic age, the central question of accident law remains the one Calabresi posed half a century ago: who is best positioned to govern risk? The BARG framework offers a contemporary answer.