Legal Challenges in Attributing Responsibility for Autonomous Driving Accidents

Yuanqi Luo[1]*

[1] Jiangxi University of Finance and Economics, Jiangxi, China
* Correspondence: luoyq2001@163.com

https://doi.org/10.53104/curr.res.law.pract.2025.07004

Abstract: The rise of autonomous vehicles (AVs) presents unprecedented legal, ethical, and regulatory challenges to existing frameworks of liability and responsibility. Traditional legal doctrines, built around human agency and fault-based liability, are increasingly strained by the complexity and opacity of machine decision-making systems. This paper examines the fragmented distribution of responsibility across manufacturers, software developers, users, and regulators, and highlights the legal ambiguity that emerges when causality is shared across distributed technical systems. It explores the regulatory and conceptual gaps that hinder effective adjudication, the influence of public moral expectations on the legitimacy of liability frameworks, and the philosophical dilemmas involved in delegating moral judgment to algorithmic systems. Drawing on legal scholarship, empirical studies, and ethical theory, the paper argues for a multi-stakeholder approach to legal reform, one that incorporates hybrid liability models, institutional coordination, and participatory governance. It concludes by advocating for a reconceptualization of responsibility in the age of autonomous mobility—grounded in transparency, fairness, and normative clarity.

Keywords: autonomous vehicles; public trust

Received: 18 July 2025

Revised: 8 August 2025

Accepted: 11 August 2025

Published: 14 August 2025

Citation: Luo, Y. (2025). Legal Challenges in Attributing Responsibility for Autonomous Driving Accidents. *Current Research in Law & Practice, 3*(1), 41-54.

1. Introduction

The advent of autonomous vehicles (AVs) marks a watershed moment in the evolution of human mobility. In the span of little more than a decade, what once seemed confined to science fiction has emerged as a burgeoning reality: intelligent systems capable of operating complex machines without direct human intervention. These vehicles, powered by advanced sensors, deep-learning algorithms, and real-time data analytics, promise a revolution in traffic efficiency, safety optimization, and urban planning. They signify a shift not just in the mechanics of driving but in the very fabric of how transportation systems function. Yet this promise comes entangled with a deep and unresolved legal uncertainty: when an autonomous vehicle is involved in an accident, who is to blame?

Conventional legal frameworks are structured around the assumption that a human agent is in control. This presumption is so deeply embedded in tort and criminal law that it informs doctrines of negligence,

recklessness, and intent across jurisdictions. Human drivers are held accountable for speeding, running red lights, texting behind the wheel, and making poor decisions under stress. When an accident occurs, courts investigate driver behavior, level of attentiveness, and compliance with traffic laws to determine liability. The insurance system, in turn, aligns with this fault-based architecture. Even in cases of no-fault liability or strict liability, there remains an anchoring assumption: the human is the operative center of decision-making.

Autonomous vehicles disrupt this assumption at a foundational level. Their intelligence derives not from a conscious agent but from millions of lines of code operating across layered systems of perception, planning, and control. These vehicles "see" the world through LiDAR and radar, "decide" how to act through probabilistic models, and "execute" those decisions via actuators governed by software architectures. In high levels of autonomy—particularly Level 4 and 5, as defined by the Society of Automotive Engineers—the human occupant is not only relieved of control but may be completely removed from the decision loop. The accident, if one occurs, cannot be traced to human reaction time or poor judgment. Instead, it is a consequence of system architecture, edge-case misclassification, sensor failure, GPS inaccuracies, or flaws in the training dataset used in the vehicle's machine learning module.

This technical complexity introduces not just a factual challenge but a philosophical one. Responsibility, in the legal sense, is tethered to notions of volition, foreseeability, and causation. If the vehicle's decision-making process is opaque even to its developers—a phenomenon known as the "black box" problem—how can one ascertain intent or negligence? Can a software engineer be held liable for a decision the AV made in an unforeseen scenario months or years after deployment? Is the manufacturer responsible if the vehicle behaved exactly as programmed but produced an undesirable outcome? These are not merely hypothetical musings. They go to the core of what legal systems must now confront in practice.

The problem is amplified by the layered structure of modern autonomous systems. Unlike a single individual operating a vehicle, the AV comprises subsystems developed by different entities. The sensors may be produced by one company, the machine-learning model trained by another, the vehicle platform built by yet another, and the final integration performed by a separate autonomous mobility firm. Each of these components may function independently under different intellectual property regimes, safety standards, and corporate liability policies. When a crash occurs, disaggregating fault becomes an exercise in technical forensics and contractual analysis. Legal systems, which traditionally rely on proximate cause and binary fault attribution, find themselves in unfamiliar territory. The question is no longer who turned the wheel, but which layer of code, hardware, or network latency contributed to the failure—and whether this failure was reasonably preventable.

The challenge extends into the realm of public policy and governance. Lawmakers are tasked with constructing a regulatory framework that incentivizes innovation while protecting public safety. Overregulation may stifle technological progress, pushing developers to offshore jurisdictions with looser constraints. Underregulation, however, could result in significant harm to citizens and erosion of public trust. Finding this balance requires a nuanced understanding not only of AV technologies but of their evolving social context. The legal system must be adaptive, forward-looking, and capable of engaging with technical standards in a meaningful way. Yet, most courts and legislatures are not currently equipped with the institutional capacity to deal with these issues at the necessary level of granularity.

Further complicating the legal landscape is the shifting nature of human responsibility in human-machine interactions. In partially autonomous systems—such as those operating at Level 2 or Level 3—the human driver is still nominally in charge but may be lulled into complacency by the vehicle's high level of assistance. Studies have shown that when drivers are not actively engaged, their situational awareness deteriorates rapidly. When the vehicle

suddenly demands human intervention—say, in a sensor-occlusion scenario or when facing an unanticipated road hazard—the driver may not respond in time. This creates a paradox: the human is legally responsible, but functionally incapable of exercising control. Courts must then determine whether to treat such drivers as negligent or to adjust liability doctrines to reflect the psychological and cognitive limitations imposed by the vehicle's own level of autonomy.

This shifting boundary of control also raises insurance questions. Traditional auto insurance models are based on driver profiles: age, experience, driving history, and behavior patterns. Autonomous vehicles render these models obsolete. If the vehicle does most of the driving, should premiums be based on its performance record instead? Should liability insurance shift from private individuals to manufacturers and software developers? Should governments mandate universal accident compensation funds for AV incidents, akin to workers' compensation schemes in labor law? The insurance industry, like the legal system, must reconceptualize its risk models to reflect this new ecosystem.

One cannot overlook the sociocultural and ethical dimensions that underlie legal adaptation. Different societies have different tolerance levels for automation, varying perceptions of fairness in blame attribution, and diverse traditions of liability. In civil law systems, for instance, the role of codified statutes is paramount, and legislative updates may be required to incorporate AV-specific clauses. In common law jurisdictions, judicial interpretation plays a central role, and precedents must evolve case by case. Public opinion, too, influences legal change. If citizens broadly perceive AVs as safer, courts may be less inclined to impose strict liability. Conversely, high-profile accidents involving AVs may trigger populist backlash and reactionary legislation. Legal institutions operate within these shifting cultural currents.

What emerges from this picture is not merely a gap in doctrine but a conceptual vacuum. Autonomous vehicles confront us with a new kind of actor on the road: a non-human, non-conscious agent capable of real-time decision-making with physical consequences. This actor does not possess intent, cannot be punished or deterred, and cannot pay compensation out of moral obligation. The legal system, rooted in centuries of human-centered jurisprudence, must find new language to describe this presence. It must develop doctrines that assign responsibility not based solely on individual fault but on system-wide accountability. Only then can justice be done in a world where machines share our roads and our risks.

2. Fragmented Responsibility and Legal Ambiguity

The legal attribution of responsibility has historically relied on the existence of a clear, traceable human actor capable of decision-making and error. The premise underlying most tort and criminal liability frameworks is that a human agent, endowed with agency and control, can reasonably foresee the consequences of their actions. This legal scaffolding begins to collapse when applied to autonomous vehicles (AVs), which are defined by the absence or reduction of human control, and by their operation through a distributed system of inputs, outputs, and software-driven decision-making. The result is a landscape in which liability becomes a diffuse problem, fractured across a multitude of actors, each of whom may plausibly deny sole responsibility.

The structural complexity of AVs lies at the heart of this fragmentation. Modern autonomous driving systems are built atop multiple interdependent components that include real-time object recognition algorithms, high-definition mapping systems, LiDAR and radar sensors, deep-learning models trained on dynamic datasets, and actuators coordinated by embedded software. These components are often produced by different entities across a decentralized industrial supply chain. For example, one firm may manufacture the camera system, another may design the neural network architecture for perception and decision-making, while a third may integrate these systems into the vehicle's onboard control infrastructure. The final deployment may occur under the banner of a ride-hailing platform or an automobile

brand, which itself may not own or directly oversee the development of any of the component parts.

When an autonomous vehicle causes an accident, the causal chain that must be unraveled includes potential failures at any of these levels. A mislabeling of road signage by the vision model may result from inadequate training data. A failure to yield may arise from faulty decision logic in the software stack. A delayed braking response may stem from a miscalibration of sensor thresholds. Yet each of these failures may be invisible to the end-user or vehicle operator, who cannot reasonably detect or intervene in real time. In legal theory, assigning liability under such circumstances becomes significantly more challenging. Courts must navigate between doctrines of strict liability, negligence, and product defectiveness, none of which were originally designed to accommodate black-box systems whose internal reasoning processes are neither transparent nor predictable.

This challenge is compounded in semi-autonomous systems, where the boundaries of human and machine control blur. Level 3 vehicles, as defined by the Society of Automotive Engineers, allow the driver to disengage from active control in specific conditions, with the understanding that they must retake control upon system request. Yet real-world evidence increasingly shows that human drivers often fail to reengage in a timely or effective manner due to cognitive disengagement or overreliance on automation. In such scenarios, determining fault requires an assessment of whether the system provided an adequate warning, whether the driver was given sufficient time and contextual awareness to respond, and whether the expectation of human intervention is realistic given human attentional limitations. These questions are not easily answered by existing legal standards, which presume a binary distinction between operator and machine.

The Uber self-driving fatality in Arizona in 2018—although occurring outside China—demonstrated the profound difficulty of isolating legal responsibility in AV accidents. Investigations revealed a litany of contributing factors: a disabled emergency braking system, a distracted human safety driver, flawed software categorization of the pedestrian, and organizational oversight failures. While criminal charges were ultimately filed against the human safety driver, the corporate entities involved were shielded by the diffusion of responsibility across technical and organizational domains. This case, though often cited in global legal discourse, serves not as a clear precedent but as a warning: absent coherent legal frameworks, accidents involving AVs will continue to fall into grey areas of attribution where responsibility is both shared and diluted.

A key conceptual issue in this domain is the lack of a singular controlling mind. In traditional vehicular accidents, courts look for a "mens rea" or a legally relevant state of mind such as recklessness, negligence, or intent. Machines, of course, have no mind to speak of, and the developers of such systems rarely have specific intent concerning particular outcomes. Thus, the legal system is forced to pivot toward frameworks based on risk distribution and system-level accountability. Some scholars advocate treating AV manufacturers and software developers as "designers of risk environments," a position that aligns them with product liability standards. This approach draws from long-established principles in consumer protection law, where manufacturers are liable for defects in design, manufacture, or warnings, regardless of intent. Yet the application of product liability to software remains unsettled in many jurisdictions. The intangible nature of code complicates notions of defectiveness, and the iterative nature of software updates muddies the question of when liability attaches to a particular version of the codebase.

Academic literature has begun to engage these questions, emphasizing the epistemic and moral complexity of assigning responsibility in hybrid human-machine environments. Liu, Du, and Li (2021) argue that the misattribution of legal responsibility in AV contexts has not only psychological consequences for individual actors but also structural effects on regulatory legitimacy. Their study highlights how systems that superficially preserve the appearance of human control may in fact undermine moral

accountability by masking the underlying complexity of AV decision-making. When drivers are blamed for system failures they could neither foresee nor correct, public trust in legal processes and technological oversight is eroded (Liu et al., 2021).

Another strand of scholarship focuses on the necessity of legal harmonization and the construction of unified liability regimes that account for the systemic nature of AV technologies. Wei and Guo (2025), in their comparative study of liability frameworks in China and Germany, underscore the fragmented nature of current legal approaches and call for integrated regulatory instruments that transcend jurisdictional inconsistencies. They note that while both civil law and common law traditions recognize forms of vicarious and enterprise liability, neither has yet developed a mature framework for distributed systems where decision-making is non-linear, probabilistic, and emergent. The authors suggest that future legislation should be grounded in a systems-theoretic understanding of responsibility, where liability is allocated not on the basis of singular fault but in proportion to the actor's role in the causal architecture of the vehicle's operation (Wei & Guo, 2025).

The ambiguity surrounding AV responsibility also affects contractual relationships within the mobility ecosystem. Suppliers, manufacturers, and service providers must negotiate indemnification clauses that anticipate accident scenarios without clear legal precedent. In practice, this often leads to risk-averse behavior, with companies seeking to limit their exposure through layered contractual insulation and complex liability waivers. Yet such arrangements do little to address the needs of victims, who may face protracted litigation with multiple parties each disclaiming liability. The net result is a legal environment in which accountability is scattered and justice delayed.

From a policy perspective, fragmented responsibility raises critical questions about the future architecture of legal systems. Should regulators create new liability categories specific to artificial agents? Would it be appropriate to establish centralized compensation funds for AV-related harm, similar to vaccine injury compensation schemes, to decouple victim relief from the uncertainties of fault attribution? These proposals have gained traction in academic and legislative circles but have yet to materialize into concrete law. Any such move would require rethinking foundational legal concepts such as personhood, agency, and foreseeability, none of which currently accommodate the unique characteristics of autonomous systems.

The fragmented nature of responsibility in AV contexts is not merely a challenge of legal formalism but one that touches upon deeper epistemic and institutional deficiencies. The law, in its current form, is ill-suited to apportion blame in systems where causality is distributed, intent is absent, and transparency is limited. Without a coherent theory of machine-integrated responsibility, legal systems will continue to produce inconsistent and unsatisfying outcomes. This gap must be addressed not only through doctrinal innovation but through a broader reconceptualization of what it means to be responsible in the age of autonomous technologies.

3. Regulatory and Conceptual Gaps

The rapid evolution of autonomous vehicle (AV) technologies has outpaced the development of legal and regulatory frameworks across jurisdictions. While the promise of reduced traffic fatalities and enhanced transportation efficiency is compelling, the lack of legal clarity concerning responsibility for accidents involving AVs introduces significant regulatory uncertainty. This uncertainty is not confined to one legal tradition or national context; it spans civil law and common law systems alike, revealing the structural limitations of legal systems designed around human agency when faced with autonomous decision-making machines.

At the core of the regulatory vacuum is the incompatibility between traditional legal doctrines and the technical realities of autonomous driving. The principle of negligence, foundational to tort law, presumes that liability is attached to a breach of a duty of care by an individual who could reasonably foresee

the consequences of their actions. In the case of an AV operating in autonomous mode, there may be no direct human act or omission to evaluate. When a vehicle swerves unexpectedly or fails to recognize a pedestrian due to a misclassification in its machine vision system, the question becomes whether this malfunction constitutes negligence and, if so, whose negligence it is. The legal subject here is no longer a person but a technological composite whose behavior is emergent and, in some cases, non-deterministic. This undermines the doctrinal tools that courts use to draw causal inferences and assess fault.

Strict liability regimes, often invoked in product liability claims, offer one possible avenue for resolving these questions. Under strict liability, manufacturers may be held accountable for harm caused by defective products regardless of fault or intent. However, this doctrine, when applied to AVs, raises novel challenges. In traditional product liability cases, the defect is typically physical—such as a faulty brake pad or a malfunctioning airbag. With AVs, the defect may lie in software logic or in a failure of the system to generalize correctly from training data to real-world scenarios. Determining whether such an error constitutes a "defect" under legal definitions, or whether it was an unavoidable limitation of current AI technologies, requires a technical sophistication that many courts are not equipped to exercise. The opaqueness of AV decision-making processes—often referred to as the "black box" problem—further complicates matters by rendering it difficult to trace the internal causality of the system's actions.

This legal ambiguity is exacerbated by the tension between encouraging technological innovation and ensuring public safety. Legislators are often caught between two conflicting imperatives. On one hand, they are under pressure to promote the development and deployment of AVs as part of national strategies for smart transportation, industrial upgrading, and carbon reduction. On the other hand, they are responsible for protecting the rights and interests of citizens who may be exposed to new and uncertain risks. The absence of clearly defined liability rules may deter consumers from trusting AVs and companies from investing in long-term innovation. The result is a chilling effect that benefits neither public safety nor economic competitiveness.

Some legal scholars have proposed that autonomous vehicles require a sui generis regulatory regime, one that does not merely modify existing liability doctrines but rethinks them entirely. This argument stems from the recognition that AVs represent a new class of agents—non-human, non-conscious, but autonomous in operation—that current law is ill-prepared to categorize. Creating new categories of legal responsibility, such as "algorithmic accountability" or "systemic fault," would allow the law to assign liability based on the systemic features of AVs rather than attempting to fit them into existing human-centered categories. This approach would also support the development of standardized benchmarks for software safety, data integrity, and algorithmic transparency.

The comparative study by Wei and Guo (2025) illustrates how different legal traditions are struggling to respond to these challenges. In their analysis of Chinese and German liability regimes, the authors show that while both countries recognize the unique risks posed by AVs, their legal responses diverge significantly. Germany has adopted a more proactive legislative stance by introducing specific provisions for autonomous driving in its Road Traffic Act, including requirements for data recording and liability insurance that reflect the technical realities of AV operation. China, by contrast, has relied on general provisions in its Tort Liability Law and Road Traffic Safety Law, supplemented by local pilot regulations in cities such as Beijing and Shanghai. This fragmented regulatory landscape creates uncertainty not only for legal practitioners but also for manufacturers and developers seeking to operate in multiple jurisdictions (Wei & Guo).

The regulatory gap is not solely a legal issue; it is also deeply conceptual. Law is a normative system that relies on shared understandings of agency, fault, and causation. AVs challenge these foundations by introducing machines that act without intention, learn without explicit programming, and adapt to environments in ways that even their creators may not

fully predict. Legal theorists such as A. Hevelke and J. Nida-Rümelin have argued that holding users responsible for the decisions of autonomous systems is morally indefensible when those users neither understand nor control the mechanisms by which decisions are made (Hevelke & Nida-Rümelin, 2015). This position implies that traditional fault-based liability, whether criminal or civil, may be normatively inappropriate in the context of full autonomy.

The conceptual gap also extends to the issue of foreseeability. One of the key elements in establishing legal responsibility is the notion that the harmful outcome was foreseeable and therefore preventable. But AVs operate on probabilistic reasoning. Their behavior is governed by neural networks that process inputs and generate outputs based on statistical inference rather than deterministic rules. This makes it difficult to predict how the vehicle will behave in novel or edge-case scenarios, such as an unexpected road closure or an ambiguous pedestrian gesture. When an accident occurs, it is often unclear whether it was the result of a failure in system design or an unavoidable limitation of the technology. The legal system, which depends on hindsight evaluation and counterfactual reasoning, is ill-suited to adjudicate such uncertainty.

One potential solution is the integration of mandatory technical standards into the regulatory framework. By establishing baseline performance metrics—such as minimum detection rates, decision latency thresholds, and fail-safe response protocols—regulators can create a clearer standard of care against which AV behavior can be judged. This would align the legal evaluation of AV accidents with the operational metrics used by engineers and developers. It would also shift the emphasis from retrospective blame to prospective safety assurance. However, setting such standards is not trivial. It requires close collaboration between legal experts, engineers, ethicists, and policymakers. It also requires continuous updating, as the pace of technological change is relentless.

Another emerging strategy involves the use of event data recorders (EDRs) and black-box systems to capture the operational state of the vehicle before and during an accident. These devices can provide vital evidence for reconstructing accidents and attributing fault. Some jurisdictions have already mandated their inclusion in autonomous vehicles, and the data they collect may play a critical role in future litigation and regulation. Yet the use of EDRs raises concerns about data privacy, ownership, and admissibility in court. The tension between transparency and confidentiality is a recurring theme in the regulation of AVs, reflecting the broader dilemma of how to balance individual rights with collective safety.

Cross-border regulatory coordination also presents a formidable challenge. As AVs are often developed by multinational corporations and tested across different legal environments, the absence of harmonized international standards creates regulatory arbitrage and enforcement gaps. Vehicles that meet the safety requirements in one country may fall short in another. Discrepancies in liability laws, data governance, and technical certifications hinder both consumer confidence and commercial scalability. Efforts by international bodies such as the United Nations Economic Commission for Europe (UNECE) to develop uniform regulations for automated driving systems have made some progress, but implementation remains uneven. Legal convergence in this area will require not just intergovernmental agreements but also shared philosophical commitments about the role of law in managing technological risk.

Finally, the conceptual reworking of liability must be accompanied by institutional reform. Courts may need to develop specialized technical benches or expert panels capable of understanding the nuances of machine learning and autonomous systems. Regulatory agencies must be endowed with sufficient authority and resources to monitor compliance, enforce standards, and update rules in light of new developments. Law schools and continuing legal education programs must prepare future practitioners to operate in a world where legal questions increasingly intersect with software design, data ethics, and systems engineering. Without these institutional adaptations, even the most forward-

looking regulations will remain difficult to implement and enforce.

The transition to autonomous mobility represents not just a technological shift but a jurisprudential crisis. The conceptual and regulatory frameworks that once provided legal certainty on the roads are being unmoored by the rise of non-human agents whose actions defy conventional theories of control and fault. Bridging these gaps will require a comprehensive reimagining of legal responsibility, grounded in interdisciplinary collaboration and an openness to normative innovation. Until such frameworks are developed, the law will continue to lag behind the machine, leaving society exposed to both legal uncertainty and moral ambiguity.

4. The Role of Moral and Social Expectations

Legal systems do not operate in isolation from the societies they serve. They are grounded in culturally and historically contingent expectations about fairness, accountability, and justice. Laws derive their legitimacy not only from formal authority but also from their resonance with widely held moral intuitions. When technology radically alters the structure of responsibility, the gap between legal determinations and public perceptions of right and wrong can widen, sometimes dramatically. Autonomous vehicles present precisely such a disruption, generating a new and uncertain space in which legal doctrines struggle to keep pace with evolving social attitudes.

Public intuitions about responsibility in road accidents are shaped by centuries of interaction with human drivers. When an accident occurs, people instinctively seek a human cause—a lapse of attention, a reckless maneuver, a failure to yield. This pattern of attribution is deeply ingrained, reinforced by insurance procedures, police reports, courtroom protocols, and everyday conversations. The shift to autonomous systems destabilizes this pattern. Machines lack consciousness and moral intent; they do not learn in the human sense, nor do they experience guilt, remorse, or care. Their actions are the outcome of algorithms, probability distributions, and sensor inputs. As such, the kinds of explanations they provide do not align with the psychological expectations most people bring to questions of blame and accountability.

Zhai, Wang, and Liu (2024) explored this mismatch in a controlled experimental setting. They presented over 2,600 participants with vignettes describing accidents involving fully automated vehicles. Even when the scenario clearly established that the vehicle was in control and the human occupant had no opportunity to intervene, respondents continued to assign significant moral and legal responsibility to the human "driver" (Zhai et al., 2024). This finding reveals not just a cognitive bias but a moral heuristic: people expect that someone, not something, should be held responsible. Responsibility in this sense is not just about causation; it is about personhood, intention, and the ethical structure of social relationships.

This cognitive and emotional expectation poses serious challenges for the legal regulation of AVs. If courts or legislatures adopt liability rules that diverge too sharply from public moral judgment, the result may be perceived as unjust, even when such rules are technically defensible. This perception, in turn, could erode trust in both the legal system and the technology itself. Public acceptance is not merely a matter of engineering; it is deeply bound up with questions of moral legitimacy. Trust in AVs requires not only confidence in their safety but also reassurance that when harm does occur, responsibility will be fairly assigned and justice meaningfully pursued.

The ethical complexity of AV decision-making is most evident in the so-called "trolley problem" scenarios, where the vehicle must choose between two harmful outcomes. These edge cases, although statistically rare, have become central to the moral discourse surrounding AVs. When a human driver makes a split-second decision that harms one person to save another, legal systems often treat it as a tragic accident without criminal liability. But when an AV makes such a decision, even if the outcome is statistically optimal, the absence of human agency transforms it into a profound moral dilemma. The public demands to know who programmed the algorithm, whose values were embedded in the decision matrix, and why a

machine was allowed to decide life and death outcomes at all.

These concerns are not purely hypothetical. In a study by Li, Zhao, and Malle (2016), participants were asked to evaluate different AV accident scenarios that required sacrificing one life to save many. The responses showed deep ambivalence. Many supported utilitarian decision rules in theory but rejected them when applied to AVs they themselves might ride in. This inconsistency—supporting a rule in the abstract while rejecting it in practice—reflects the fraught terrain of moral decision-making in automated contexts (Li et al., 2016). Ethical preferences become unstable when the agent is a machine, and the consequences directly affect the self.

This tension points to a broader phenomenon: the asymmetry in how people judge human and machine agency. Machines are often held to higher moral and safety standards than humans, even though they are not capable of moral reasoning. This is partly due to the perception that machines are controllable, programmable, and predictable—qualities that invite higher expectations. At the same time, people are less forgiving of machine failure, perceiving it not as a lapse but as a design flaw. When an AV causes harm, the public does not treat it as an accident in the traditional sense. Instead, it becomes evidence of technological overreach, systemic failure, or corporate negligence.

This asymmetry creates a dilemma for designers and regulators. On one hand, AVs must be trusted to make decisions in unpredictable environments. On the other hand, they must operate within a normative framework that aligns with public values, many of which are underdeveloped or contested. Embedding ethics into AV design—whether through ethical programming, value-sensitive engineering, or human-in-the-loop protocols—requires clarity about which values are at stake and whose values they are. Yet public opinion is often fragmented, inconsistent, and culturally variable.

Moral expectations around AVs also intersect with broader societal narratives about automation, control, and accountability. In societies with strong collectivist traditions, there may be greater tolerance for system-level solutions to harm, such as no-fault compensation schemes or collective liability models. In societies that emphasize individual responsibility and legal formalism, there may be greater insistence on identifying a culpable party. These differences influence how new technologies are received and what kinds of legal reforms are politically and culturally acceptable.

The role of media in shaping moral expectations cannot be overlooked. High-profile AV accidents, even when statistically rare, generate intense public scrutiny. They are often framed in emotive and sensationalist terms, emphasizing loss, malfunction, and corporate irresponsibility. This framing reinforces the moral intuition that machines should not be in control and that when they are, someone should answer for the consequences. Legal responses shaped in the aftermath of such events may reflect not reasoned analysis but reactive sentiment, leading to regulations that are either too stringent or inadequately justified.

Educational and institutional strategies may help bridge the gap between legal doctrines and moral expectations. Public engagement initiatives, citizen juries, and participatory design processes can democratize the conversation around AV ethics and responsibility. By involving stakeholders early and transparently, regulators can foster a shared understanding of the trade-offs involved and the principles that should guide AV behavior. This, in turn, can inform legal reforms that resonate with both normative commitments and technological realities.

In the longer term, societal expectations may shift as exposure to AVs increases and familiarity breeds acceptance. Just as early resistance to elevators without operators eventually gave way to trust in automation, so too may public discomfort with driverless vehicles decline over time. But such shifts are not guaranteed, and they depend on a regulatory environment that both protects the public and respects their moral sensibilities. If the law fails to recognize the moral dimension of public attitudes, it

risks delegitimizing itself and undermining the social fabric it seeks to uphold.

Zhai et al.'s findings underscore the need for an interdisciplinary approach to AV regulation, one that combines legal analysis, psychological research, and ethical theory. Legal responsibility must not be decoupled from social legitimacy. Assigning blame in a way that is procedurally correct but morally unintelligible to the public may satisfy formal requirements without achieving justice. Conversely, aligning legal standards too closely with fluctuating public sentiment may jeopardize consistency and predictability. The challenge is to find a balance that honors both democratic accountability and principled legal reasoning.

This balance is especially delicate in transitional stages of automation, where partial autonomy coexists with residual human oversight. In these contexts, moral expectations become even more confused. Is the human driver expected to intervene instantly if the system errs? Is inaction tantamount to negligence, even when reaction time is insufficient? Should designers be held responsible for fostering false confidence in automation, or should users bear responsibility for misunderstanding the system's limits? Each of these questions implicates different layers of moral and legal judgment, none of which admit easy answers.

The psychological phenomenon of moral distancing also plays a role. When harm results from the actions of a machine, observers often experience a weakened sense of empathic engagement. The victim may be real, but the perpetrator is faceless. This emotional gap can lead to under-reaction or misdirected blame, distorting both public discourse and legal interpretation. Addressing this requires cultivating new forms of moral vocabulary capable of articulating responsibility in human-machine collectives.

Autonomous vehicles present a powerful test case for the integration of ethics, law, and technology. They force society to ask not just what can be automated but what should be. They challenge the assumption that legal responsibility can be cleanly assigned in systems characterized by distributed agency and emergent behavior. They demand new models of justice that can accommodate the absence of human intent and the presence of machine decision-making. Above all, they call for a recalibration of moral expectations to match the hybrid realities of 21st-century mobility.

5. Ethical and Philosophical Dilemmas

The ethical implications of autonomous vehicles challenge not only conventional legal categories but also foundational philosophical concepts related to agency, responsibility, and moral judgment. While law focuses on compliance, deterrence, and liability, ethics demands an account of what *ought* to be done, even in scenarios that fall outside the reach of statutory definitions. The automation of moral choice in AVs demands scrutiny because it represents a delegation of human judgment to non-human systems. In this delegation, the questions of accountability, intention, and the meaning of harm become increasingly opaque.

The classic illustration of this ethical opacity is the so-called "trolley problem," which forces a decision between harming one person or allowing harm to come to many. When transposed to the context of AVs, this dilemma becomes more than a thought experiment—it becomes a design decision encoded into the vehicle's decision-making architecture. An AV's programming may determine, for instance, whether to prioritize the safety of its occupants or that of pedestrians in unavoidable crash scenarios. Such decisions carry profound moral weight, yet are made preemptively, not in the heat of the moment, and by engineers and ethicists removed from the situation. The ethical accountability of these actors becomes a key concern, as they shape the normative framework within which machines will act.

Philosophical inquiry into responsibility has long emphasized the importance of intent, autonomy, and moral agency. Machines possess none of these qualities in the conventional sense. They do not have intentions, cannot reflect on their actions, and are not moral subjects. This creates a vacuum in moral responsibility. When an AV causes harm, society instinctively searches for an agent to blame, yet no such agent exists within the machine. The

responsibility must therefore be distributed across the socio-technical network that produced and deployed the system. This includes software developers, data scientists, corporate executives, regulators, and legislators. The ethical dilemma lies in the absence of a clear locus of intent, and the diffusion of agency across non-human actors and institutional structures.

Hevelke and Nida-Rümelin (2015) argue that traditional models of moral responsibility are inadequate in this context. They suggest that placing the burden of responsibility on end-users—those riding in or overseeing AVs—is morally unjustified when those users have no meaningful control over the vehicle's real-time decisions. Instead, responsibility should be understood collectively, grounded in institutional and systemic contributions to the decision-making framework (Hevelke & Nida-Rümelin). This ethical shift aligns with contemporary theories of distributed agency, which reject the idea that only individual actors can be moral agents. It also reflects a pragmatic recognition that harm can be the result of cumulative design choices made across different domains.

Designing moral behavior into AVs also raises questions about whose morality is encoded and whether that moral code is universally acceptable. Cultural variation in moral reasoning complicates any attempt to establish a standardized ethical protocol. A utilitarian logic of maximizing lives saved may be accepted in one jurisdiction but rejected in another where rights-based deontological ethics are dominant. Engineers are therefore tasked not just with technical implementation but with navigating the pluralism of moral worldviews. This burden is both philosophical and political, requiring legitimacy and public deliberation in how AVs are programmed to act.

Ethical dilemmas intensify as AVs begin to function within increasingly complex social environments. Decisions about lane changes, merging behavior, and pedestrian interaction all involve implicit moral norms. These micro-interactions, while less dramatic than trolley-like choices, shape public perceptions of fairness, empathy, and respect. A vehicle that always asserts right-of-way may be legally compliant but ethically hostile. A vehicle that is overly deferential may be seen as weak or unpredictable. Balancing assertiveness and caution requires moral sensitivity that machines cannot achieve autonomously. Designers and policymakers must fill this ethical gap with rules that are not only safe but socially intelligible.

Philosophy also plays a role in evaluating the broader societal effects of automation. There is an ethical question about the justice of replacing human labor and decision-making with machine alternatives. The deployment of AVs will affect millions of professional drivers, reshape public space, and redefine mobility access. Ethical analysis must ask not only what decisions AVs should make in emergencies, but what kind of society is being built through their adoption. The dilemmas are not limited to split-second crashes but extend to questions of equity, dignity, and collective risk distribution.

Addressing these dilemmas requires interdisciplinary engagement. Legal reasoning alone is insufficient. Philosophical ethics must inform regulatory frameworks, public consultations must supplement technical design, and educational programs must equip engineers with the tools to reflect on their normative responsibilities. The goal is not to eliminate all ethical risk—such a standard is impossible—but to ensure that decisions involving harm, fairness, and responsibility are made with transparency and justification. The moral legitimacy of AVs will depend not only on their safety records but on the public's confidence that their actions are grounded in values that reflect democratic deliberation and ethical reasoning.

6. Path Forward

Addressing the legal, ethical, and technical complexities associated with autonomous vehicles requires a coordinated, multi-actor effort. No single discipline or institution possesses the full capacity to define responsibility, ensure public safety, and preserve innovation within the rapidly shifting context of machine-driven mobility. The challenges involved cannot be solved by adapting existing

liability doctrines alone; they demand the construction of new legal frameworks that distribute accountability, define safety obligations, and incorporate public values.

One foundational step is the development of hybrid liability regimes. These models recognize that harm involving autonomous vehicles often results not from individual misconduct but from systemic and distributed causes. Under such regimes, responsibility can be shared among developers, manufacturers, platform operators, and possibly public infrastructure providers. Product liability may apply when harm results from flawed software design, hardware malfunction, or inadequate warnings. Enterprise liability is appropriate where a corporate entity oversees deployment and control, even if it delegates design to third-party suppliers. These overlapping liability layers shift the focus from pinpointing a singular guilty party to assigning responsibility across the network of actors who contribute to the operation of autonomous vehicles.

Legal clarity requires not just conceptual reform but enforceable technical standards. Regulatory authorities must define operational benchmarks—such as system reaction times, minimum fail-safe conditions, real-time override capabilities, and environmental adaptability thresholds. These standards should reflect engineering realities without locking innovation into outdated criteria. Legal definitions of system failure must also distinguish between reasonable limitations inherent in probabilistic learning models and negligence in design, training, or maintenance. Without such nuance, liability decisions risk being arbitrary or technologically incoherent.

Fairness to accident victims requires reforms that extend beyond tort litigation. In cases where fault is ambiguous or legally diffused, traditional court processes may delay compensation or leave claimants without effective remedies. Establishing **accident compensation schemes** tailored to AV-related harm could provide a reliable and impartial safety net. These schemes, supported through pooled industry contributions or public-private arrangements, would allow for timely compensation while preserving the option for parties to seek judicial resolution in cases of gross fault or systemic negligence.

No single regulatory body can cover the technical, ethical, and legal dimensions of autonomous mobility. Institutional coordination is necessary across transportation, consumer safety, insurance, and cybersecurity sectors. Inter-agency regulatory frameworks should share data, co-develop standards, and avoid conflicting mandates that create legal uncertainty for developers and users. Independent technical certification entities, responsible for verifying compliance with safety protocols and transparency obligations, should function autonomously from commercial stakeholders to avoid conflicts of interest.

Autonomous vehicles also raise questions that extend beyond national borders. International regulatory convergence is essential for global safety, trade interoperability, and legal predictability. Variations in liability rules, data-sharing norms, and safety verification standards can create incentives for jurisdiction shopping and technological evasion. Regional treaties or frameworks under multilateral institutions could establish shared minimum requirements, encouraging best practices while respecting domestic legal diversity.

Legal reform efforts must also remain sensitive to public values. Technocratic policy instruments will not be effective unless they are accepted as legitimate by the populations they affect. Structured public engagement mechanisms, such as ethics panels, consultative forums, and citizen advisory boards, provide spaces where ordinary people can deliberate on issues such as moral decision-making by machines, privacy in data collection, and expectations of human oversight. These engagements not only generate democratic legitimacy but help align technical priorities with evolving social norms.

Progress on AV governance requires more than legal innovation. It involves sustained effort among legislators, regulators, engineers, insurers, ethicists, and the public. Responsibility in a technologically mediated environment can no longer be assigned

along linear or individualistic lines. It must be redefined as a collective endeavor, embedded in institutional architecture and capable of adapting as autonomous systems become more embedded in everyday life. Only by building this kind of multi-stakeholder foundation can law and policy retain their relevance in an increasingly automated world.

7. Conclusion

The arrival of autonomous vehicles represents not just a technological inflection point but a profound legal and moral challenge. The very structure of responsibility, long rooted in human agency and intent, is being destabilized by the emergence of systems that operate without consciousness, that learn from data rather than instruction, and that often behave in ways opaque even to their creators. This transformation does not occur in a legal vacuum. It collides directly with centuries-old doctrines of fault, liability, and causation—doctrines that were designed to adjudicate conflicts between individuals, not between a person and a machine learning model embedded in a multilayered supply chain.

The fragmentation of responsibility across manufacturers, developers, data providers, operators, and end-users leaves existing legal frameworks stretched beyond their limits. Traditional tort and product liability doctrines rely on clear chains of causation and the presence of a culpable actor. AV accidents, by contrast, often emerge from diffuse system interactions. A software misjudgment, a failure in sensor calibration, or a gap in training data may result in harm, yet no individual actor may be provably negligent in a conventional sense. As technical complexity increases, the ability of courts to reconstruct fault using analog tools diminishes.

The conceptual vacuum this creates is not merely procedural—it is normative. A society must decide not just how responsibility can be assigned, but how it should be assigned. Law must balance the imperative to provide compensation and deterrence with the need to maintain technological progress and social trust. When the public perceives legal outcomes that are technically accurate but morally dissatisfying, the legitimacy of the entire framework begins to erode. Responsibility is as much about justice as it is about liability, and both must remain visible and comprehensible in the public sphere.

In response to this complexity, reform cannot be isolated to statute books or courtroom procedures. It must be structural. Legal frameworks must evolve to reflect systemic causality rather than individual failure. Hybrid liability regimes that combine aspects of fault-based and strict liability offer one pathway. Compensation mechanisms that bypass prolonged litigation in ambiguous cases can restore fairness to victims. Clear safety standards embedded in technical regulation, not legal abstractions, are needed to make responsibility tangible and enforceable. These reforms must also be designed with institutional infrastructure in mind: expert regulators, interdisciplinary courts, and international harmonization mechanisms are all prerequisites for a legal order capable of responding to autonomous mobility.

Yet legal reform alone is insufficient. Moral and psychological expectations shape how laws are received and obeyed. Public discomfort with machines making irreversible decisions, even if statistically optimal, cannot be dismissed as irrational. It must be met with design transparency, participatory rulemaking, and accountability structures that reflect collective values. Ethical questions about who programs decisions, how data is governed, and whose safety is prioritized must not be deferred to technical experts alone. Law must reclaim its place as the space in which these choices are openly debated and socially resolved.

A path forward requires shared responsibility. No single actor—not the manufacturer, not the regulator, not the user—can carry the weight of judgment alone. Responsibility must be shared, structured, and situated within a legal system flexible enough to adapt yet principled enough to preserve fairness. If law is to retain its relevance in the age of automation, it must not retreat in the face of complexity. It must respond with clarity, with moral seriousness, and with institutional imagination equal to the task.

References

Hevelke, A., Nida-Rümelin, J. (2015). Responsibility for Crashes of Autonomous Vehicles: An Ethical Analysis. *Sci Eng Ethics*, *21*, 619–630. https://doi.org/10.1007/s11948-014-9565-5

Li, J., Zhao, X., Cho, M., Ju, W. et al. (2016). From Trolley to Autonomous Vehicle: Perceptions of Responsibility and Moral Norms in Traffic Accidents with Self-Driving Cars. *SAE Technical Paper 2016-01-0164*, https://doi.org/10.4271/2016-01-0164

Liu, P., Du, M. & Li, T. (2021). Psychological consequences of legal responsibility misattribution associated with automated vehicles. *Ethics Inf Technol*, *23*, 763–776 https://doi.org/10.1007/s10676-021-09613-y

Wei, X., & Guo C. (2025). A Comparative Legal Study on the Attribution of Liability for Autonomous Driving Accidents in China and Germany. *Country, Area and Advanced Technology*, *1*(1). https://doi.org/10.37420/j.caatj.2025.003

Zhai, S., Wang, L., & Liu, P. (2024). Not in control, but liable? Attributing human responsibility for fully automated vehicle accidents. *Engineering*, *33*, 121-132. https://doi.org/10.1016/j.eng.2023.10.008