Research on the Legal Allocation of Criminal Responsibility in Autonomous Vehicle Accidents

Zezheng Liu

Tianjin Hexi Prison, Tianjin City, China

1405167640@qq.com

Keywords: Autonomous Vehicles; Traffic Accidents; Criminal Liability; Liability Allocation; Negligence Liability; Product Liability; Algorithmic Explainability; Duty of Care; Causality; Legal Regulation

Abstract: The rapid development of autonomous driving technology has revolutionized traditional transportation while continuously challenging the conventional criminal liability allocation framework. This article focuses on the core dilemma of distributing criminal liability in accidents involving autonomous vehicles. Research reveals severe challenges in addressing this issue, manifested in the following aspects. To ensure effective implementation, the following recommendations are proposed: Legislative revisions: Amend the Criminal Law and Road Traffic Safety Law to enact practical regulations tailored to specific scenarios. Procedural adjustments. This study aims to provide new theoretical foundations and perspectives for criminal liability allocation in autonomous vehicle accidents, contributing to the harmonious advancement of AI innovation, technological progress, and societal safety.

1. Introduction

Artificial intelligence is driving the fourth industrial revolution. This revolution is transforming human society. Autonomous vehicles represent one of its critical applications. These vehicles have moved beyond laboratories and testing grounds onto public roads. Level 3 (L3) vehicles are gradually being deployed and commercialized, while Level 4 (L4) and Level 5 (L5) technologies continue to advance[1]. This transformation enhances travel convenience and reduces human driving errors, yet it creates significant challenges in defining and allocating legal liability when accidents occur.

Traditional traffic criminal law centers on the "human driver." It establishes offenses such as the crime of causing traffic casualties and dangerous driving based on the driver's intent or negligence, requiring clear causation between actions and outcomes. However, autonomous driving technology disrupts this foundation: At L3, the driver transitions to a supervisory role. At L4/L5, occupants become passengers or users, fundamentally altering their responsibilities. Multiple parties—vehicle designers, manufacturers, programmers, owners, and operators—now contribute to driving. Their actions or product flaws may cause accidents, raising debates over whether algorithms themselves can bear liability.

1.1 Difficulty in assigning responsibility

The shift from driver-centric to multi-stakeholder liability complicates accountability.

1.2 Challenges in proving fault

Define the "reasonable duty of care" for drivers (especially at L3) in complex human-machine interactions.

Provine negligence when accidents stem from system design flaws, algorithmic errors, or sensor failures.

Traditional methods struggle to pierce the "technological black box" to determine whether developers or manufacturers acted negligently or intentionally.

DOI: 10.25236/etmhs.2025.002

1.3 Unclear causation

Autonomous accidents often result from intertwined factors—human operation, hardware/software malfunctions, algorithmic issues, environmental conditions, or third-party acts (e.g., hacking). Algorithmic decision-making resembles a "black box," making it hard to pinpoint exact causes or establish criminal causation between specific actions and outcomes[2].

Truly responsible parties evade punishment. Drivers with limited capability face disproportionate penalties. This not only threatens individual rights and judicial fairness but also risks stifling technological innovation or compromising public safety due to ambiguous liability rules.

2. Literature review

Scholars worldwide have recognized the legal liability challenges posed by autonomous vehicles and are actively conducting research. Existing studies primarily focus on the following areas[3]:

Research on liability allocation: Significant work has been conducted in this domain. Studies concentrate on product liability, adjustments to compensation rules, and reforms to insurance systems, aiming to achieve more efficient and equitable loss distribution.

Regulatory and policy research: Numerous studies address governmental management aspects such as testing standards, safety requirements, data governance, and ethical guidelines for autonomous driving.

Fundamental criminal liability discussions: Scholars have begun examining core issues including the expansion of liable parties, updates to causation theories, and evolving duty-of-care standards. Novel concepts like "human-centric approaches," "expanded product liability," and "algorithmic accountability" have been proposed.

However, dedicated research on detailed criminal liability allocation remains insufficient. Current gaps include:

Fragmented frameworks: Most studies address single liable parties (e.g., users or manufacturers) or isolated issues (e.g., causation or duty of care), failing to holistically analyze interactions between different stakeholders, automation levels (L3 vs. L4/L5), and accident causes.

Theoretical-practical disconnect: Proposed innovations often diverge significantly from existing criminal law frameworks. They lack practical interpretation methods within China's legal system or concrete legislative amendment proposals, with insufficient discussion on flexibly applying traditional doctrines (e.g., negligence, causation, corporate crime) to autonomous accidents[4].

Absence of operational guidelines: Research fails to provide clear, actionable standards for judges to determine fault thresholds, establish causation, or define specific duty-of-care requirements—such as "reasonable supervision" standards at L3 or "adequate testing" boundaries for manufacturers. Critical evidentiary elements like "algorithmic black boxes" and vehicle event data remain underexplored.

Limited China-specific analysis: Studies seldom address how China’s Criminal Law, Road Traffic Safety Law, and relevant judicial interpretations should respond to autonomous driving challenges. Research rarely integrates China's technological development stage and judicial characteristics to formulate context-appropriate liability rules.

Therefore, this study addresses the core question: How can China's current criminal law framework establish a fair, reasonable, predictable, and operationally viable method for allocating criminal liability in severe traffic accidents involving autonomous vehicles (particularly L3-L5)? This method must clearly delineate liability boundaries among drivers, manufacturers, vehicle operators, and other responsible parties.

3. Methodology

This study employs the following integrated methodologies:

Involves extensive collection, systematization, and critical analysis of domestic and international academic works, journal articles, and legal regulations. This establishes fundamental theoretical foundations while capturing cutting-edge developments[5].

Examines experiences and lessons from autonomous driving pioneers (e.g., Germany, U.S., Japan) regarding accident liability determination, legislative developments (e.g., Germany's amended Road Traffic Act), judicial practices, and theoretical discourse. Valuable insights are distilled and adapted to China's socio-legal context.

Selects landmark autonomous accident cases (e.g., the 2018 Uber self-driving fatality in Arizona, U.S.) for detailed examination. Judicial precedents are analyzed alongside theoretical frameworks to construct this study's proposed liability architecture.

Conducts theoretical analysis of current Criminal Law provisions and relevant offenses—including the crime of causing traffic casualties, crime of negligently causing serious accident, and crime of manufacturing/selling non-compliant products[6]. Actionable refinement measures and legislative recommendations are formulated.

4. Result

4.1 Ambiguity in Identifying Liable Parties

Autonomous driving refers to technology enabling vehicles to perform driving tasks without human intervention. Current automation levels (L0–L5) are defined as follows: L0 – No Automation: Full driver control without assistance. L1 – Driver Assistance: Single-function aids (e.g., adaptive cruise control), requiring continuous driver monitoring. L2 – Partial Automation: Combined acceleration/steering control (e.g., lane centering), with driver supervision and readiness to intervene. L3 – Conditional Automation: Full task control under specific conditions; driver must resume control when requested. L4 – High Automation: Full autonomy within operational design domains (e.g., geo-fenced areas), no driver intervention needed. L5 – Full Automation: Unrestricted autonomy in all scenarios. This study focuses on criminal liability dilemmas for L3+ systems.

- Can Autonomous Vehicles Bear Criminal Liability?

Criminal law fundamentally aims to punish offenders and protect society, emphasizing retribution. Vehicles lack independent consciousness—a subjective capacity for experience (e.g., perceiving red lights as warnings). Current "weak AI" technology executes specific tasks (e.g., driving) without cross-domain cognition or self-awareness. Holding vehicles criminally liable would undermine the retributive purpose of criminal law.

- Can Human Drivers Bear Criminal Liability?

Relevant offenses include causing traffic casualties and dangerous driving . Traditional liability requires "actual vehicle control" [1]. Under L3+, drivers become "fallback-ready users" [2] who typically perform no driving acts. Merely observing road conditions does not constitute "control"—defined as volitional command over physical acts (acts/omissions). Legally, L3 drivers resemble passengers. This conflicts with traffic offense statutes requiring active control. Judicial precedent (e.g., the 2018 Uber fatality in Arizona [3]) declined to prosecute safety operators for phone use while monitoring, highlighting liability ambiguities.

- Can Manufacturers Bear Criminal Liability?

Manufacturers (legal entities) potentially fall under corporate crime doctrines. Algorithm development—critical to L3+ safety—involves software engineers, hardware teams, and testers. Corporate liability requires identifying "directly responsible personnel," yet technical defects often stem from collective decisions [4]. The crime of negligently causing serious accident demands "safety violations during production." Algorithm development constitutes scientific creation rather than "production" (i.e., transforming resources into tangible goods/services). Thus, prosecuting manufacturers under this statute faces legal incongruity.

- Can Operators Bear Criminal Liability?

Operator liability involves debates over supervisory negligence vs. managerial negligence. Key issues include: Whether failing to update systems constitutes criminal omission (Art. 15). Operators exert less control than traditional transport firms: California DMV reports show 31% remote control over critical hardware (vs. 85% for conventional companies). NTSB data indicates traditional firms bear liability in 73% of L3 accidents, while operators account for only 22%—reflecting judicial recognition of diminished control, corroborated by Tsinghua University experiments.

Liability allocation remains unclear in multi-party accidents due to absent criminal law standards.

4.2 Complexity in Establishing Subjective Fault and Causation

The core challenge lies in the fundamental conflict between intelligent driving technology and traditional criminal law doctrines of subjective fault. Criminal liability hinges on accurately determining the actor's mental state—whether intentional ("knowing violation") or negligent ("foreseeable yet unanticipated"). However, when decision-making shifts from humans to algorithm-controlled systems, assessing subjective fault becomes problematic. Concurrently, criminal causation requires proving direct causation between harmful conduct and damage—a link often disputed in autonomous accidents.

- The "Black Box" Dilemma

AI research consensus holds that even designers cannot fully reconstruct algorithmic decisions post-accident. Choices emerge from complex computations of massive data, lacking human-like transparent reasoning. When failures trace to hidden training data flaws, hardware anomalies, or unforeseen model interactions, proving foreseeability of harm or breach of duty of care becomes nearly impossible. Judges struggle to establish that individual developers: Could foresee specific failure modes from billions of code interactions.

- Evolving Technical Standards and Temporal Misalignment

Negligence liability relies on industry norms and knowledge at the time of conduct. Yet autonomous driving safety thresholds constantly shift: yesterday's "acceptable risk" may today be deemed an "obvious flaw" due to new research or accident data. This creates a judicial dilemma:

Strict adherence to contemporary standards may fail to punish risks later deemed foreseeable.

Applying retroactive knowledge violates the principle against ex post facto liability.

OTA updates further complicate defining the precise "time of conduct."

4.3 Autonomous Technology's Inherent Flaws and Criminal Law Scrutiny

While promising safer transportation, autonomous systems harbor technical vulnerabilities in core algorithms, sensors, decision logic, and environmental interactions. When these flaws trigger severe accidents (e.g., fatalities or major property damage), they inevitably invoke criminal law, creating unprecedented legal challenges. Technological defects now form a perilous bridge between innovation and criminal liability.

- Perception & Recognition Failures

Sensors (cameras, LiDAR, radar) may fail due to weather, glare, obstructions, aging, or miscalibration, causing "blind spots" or misjudgment (e.g., unidentifed pedestrians). Liability ambiguities arise: design defect (manufacturer), sensor failure (supplier/maintainer), or unforeseeable interference (act of God)?

- Decision & Planning Flaws

The most complex risk area: Ethical dilemma mishandling: E.g., flawed responses to unavoidable collisions (trolley problem scenarios). Could preset logic violating societal ethics incur designer/manufacturer liability (e.g., negligent homicide)? Rule-conflict errors: Incorrect trade-offs between traffic rules and safety optimization (e.g., emergency evasion violating laws). Prediction failures: Faulty behavior models of pedestrians/other vehicles. Software bugs: Code errors triggering unintended dangerous acts.

- Execution & Control Failures

Mechanical/electronic faults in steer-by-wire, brake-by-wire, or throttle systems may derail correct decisions, causing loss of control.

- Cybersecurity Vulnerabilities

Connected vehicles risk hacker manipulation (data tampering, remote control, system crashes), raising liability questions for manufacturers with inadequate security.

- Human-Machine Interaction & Takeover Failures

Inadequate/alerts or driver distraction/incompetence may prevent effective L3 takeover. This involves concurrent liability for interface design flaws and driver negligence.

4.4 Dispelling Legal Ambiguities

By amending the Criminal Law or enacting an Autonomous Driving Act, explicitly deny legal subjectivity to autonomous systems. Liability must rest with natural persons, legal entities, or unincorporated organizations—responsibility follows accountability.

Users face criminal liability only when: Failing to take over within a reasonable time after a lawful, clear takeover request. Their takeover action directly causes the accident.L4-L5: Driver transitions fully to passenger. Designers/manufacturers/commercial operators become primary criminal liable parties.

Establish Algorithmic Safety Guardianship Obligations for manufacturers/developers: Mandate compliance with national Safety of the Intended Functionality (SOTIF) and cybersecurity standards across design, testing, deployment, and updates via the Product Quality Law, Standardization Law, or Autonomous Driving Act. Coverage must include known/foreseeable edge cases. Legally require safety redundancy and minimal risk strategies (e.g., safe stop upon system failure/operational domain exceedance). Ensure algorithmic transparency: Legislate requirements for developers to reconstruct accident decision logic, dismantling the "black box." Mandate recording system states, sensor data, decisions, and human interactions pre-accident. Criminalize intentional concealment of defects or delayed updates causing accidents. Strengthen defect liability: Legally enforce disclosure obligations for system limitations and proper usage. Hold manufacturers criminally liable for misleading publicity or inadequate warnings causing misuse. Introduce the crime of major autonomous system safety violations: Penalize manufacturers (≤3 years imprisonment) for knowingly releasing defective systems (e.g., AEB false activation rates exceeding standards) causing severe injury/death. Enact an Autonomous Driving Act (modeled on Germany's §1b): Require L3+ vehicles to have EDRs ("black boxes") and data interfaces. Impose post-accident data provision duties; refusal triggers presumed fault liability.

Finally, define operators' System Operation Safeguard Duties. Ensure airworthy vehicle status. Legally mandate regular hardware maintenance/checks. Require software updates to latest certified versions. Impose liability for unauthorized modifications/delayed updates causing accidents. Establish remote monitoring & emergency intervention: Require 24/7 operational centers for L4/L5 fleets. Criminalize negligence in remote intervention exacerbating accidents. Standardize user training & qualification:Legally obligate operators to train users (especially L3 fallback-ready users) on system limits/takeover protocols.

4.5 Advancing Autonomous Driving Technology

Build comprehensive sensing systems. Single sensors (prone to blind spots) are insufficient. Fuse. LiDAR (penetrates rain/fog; precise shape detection). Cameras (color/texture detail). Radar (all-weather reliability, e.g., 4D radar measuring distance, speed, height, and motion vectors). HD maps & RTK positioning. This creates "super eyes" resilient to diverse conditions. Note: Tesla added radar to Model S/X after camera-only limitations surfaced.

Prioritize intelligent yet explainable algorithms. Train decision systems (e.g., deep reinforcement learning) on real/simulated data to handle complex scenarios at human/expert levels. Combat the "black box" via interpretable AI—making decision logic auditable to build trust. Implement shadow mode: Systems run invisibly during human driving, learning from real-world scenarios without vehicle control. Leverage V2X (vehicle-to-everything) communication: Enable real-time data exchange between vehicles/infrastructure to detect beyond-line-of-sight risks (e.g., preempting intersection collisions), mitigating ethical dilemmas like the trolley problem.

Ensure precision with fail-safe architectures. Use by-wire systems (steering/braking/throttle) for electronic precision exceeding human reflexes. Adopt dual redundancy: Duplicate critical components (power, controllers, actuators) to maintain functionality during failures. Enforce functional safety standards (e.g., ISO 26262) for rigorous hardware/software validation.

Accelerate validation via simulation platforms. Deploy cloud-based large-scale simulations to recreate millions of edge cases (e.g., pedestrians emerging in heavy rain)—unfeasible in physical testing. Example: Waymo’s 20 billion simulated miles vs. 20 million real-world miles. Combine virtual/physical testing to maximize system robustness and minimize liability risks.

5. Conclusion

The power of technology is gradually maturing, and the development of artificial intelligence is constantly driving autonomous driving from a dream to a reality. However, the road to safety for autonomous driving is long and arduous. In addition to continuing to deepen research in the field of science and technology and striving for breakthroughs in technology, it is also necessary to provide legal support for the development of autonomous driving. Through this study, we have identified difficulties in criminal liability allocation for autonomous driving accidents. At the technical level, there are deficiencies in perception and recognition, decision-making and planning, execution and control, network security vulnerabilities, and failures in human-machine interaction and takeover; At the legal level, there is ambiguity in the identification of the subject, which needs to be analyzed and broken through respectively from the autonomous vehicle itself, the driver, the manufacturer and the platform operator. The subjective responsibility and causality are also complex, and the subjective fault theory and traditional causality are challenged. In the field of technology, we strive for excellence in perception, decision-making, execution, and verification to solve technical problems; In the legal field, we start from the constituent elements of criminal law to improve the application conditions of various charges in order to address the problems that may arise in future autonomous driving. We distinguish the obligations and responsibilities that each subject should bear, such as the driver's duty of care and immediate takeover obligation, the manufacturer's algorithm safety obligation, and the obligation to be responsible for product defects. Finally, from a subjective perspective, we strive to break the causal relationship and overcome legal mysteries. I believe that with continuous efforts to upgrade technology and improve laws, autonomous driving will become a reliable partner in the future. Every safe arrival in the future is a silent commitment of technology and law!

References

[1] Zhang Mingkai. Criminal Law (6th ed.) [M]. Beijing: Law Press China, 2021: 230-245. (Fundamental Theories of Criminal Law)

[2] SAE International. Taxonomy and Definitions for Terms Related to Driving Automation Systems (J3016_202104) [S]. Warrendale: SAE, 2021.

[3] Huang Chengyan, Zha Xiaoyun, Ding Qunyan, et al. Data augmentation and rule-guided large Language Model for Power Grid Legal Defense Document Generation [J]. Journal of National University of Defense Technology,2025,47(04):180-188.

[4] Wang Ying. Rethinking the Framework of Algorithmic Tort Liability in Criminal Law [J]. Science of Law, 2023, 41(1): 112-125

[5] Fang Juan. Innovative Paths of Community Legal Aid in the Context of Digitalization [J]. Legal Review,2025,(14):91-93

[6] Koopman, P., & Wagner, M. Challenges in Autonomous Vehicle Testing and Validation. SAE International Journal of Transportation, 2017