

Research on the Criminal Imputation of Crimes Involving Level 3 Autonomous Vehicles

Kexin Wang

*Department of China West Normal University, Nanchong, China
3363489663@qq.com*

Abstract. The technical characteristics of Level 3 autonomous driving, namely "human-machine co-driving and dynamic takeover", have subverted the logic of criminal imputation in traditional traffic accident cases, and relevant accidents have highlighted core dilemmas in the application of criminal law. Specifically, at the level of liable subjects, current norms fail to clarify the standards of duty of care for drivers, automobile manufacturers and algorithm designers; algorithmic black boxes and data barriers cause difficulties in evidence production and responsibility shifting, and the qualification of artificial intelligence as a criminal subject remains controversial. At the level of subjective imputation, algorithmic black boxes invalidate the standard of concrete foreseeability of consequences under the traditional old fault theory, hindering the presumption of subjective culpability. At the level of causality, algorithmic black boxes block the attribution chain, and probabilistic decision-making breaks through traditional rules of causality determination. The criminal punishability of artificial intelligence should be clearly denied, the proportion of subject liability should be defined according to different scenarios, the standards of subjective imputation should be reconstructed with the new fault theory and the duty of care should be refined. Impediments to imputation should be resolved through rules such as the pre-duty of algorithm interpretability and the inversion of the burden of proof, so as to provide a feasible application approach for such cases.

Keywords: Level 3 autonomous driving, criminal imputation, algorithmic black box, risk criminal law, proportional causality

1. Introduction

The iterative upgrading of autonomous driving technology is profoundly reshaping the driving mode and liability logic of road traffic. As a key node of the shift of rights and responsibilities between humans and machines, Level 3 autonomous driving operates under the mode of "system-dominated driving with drivers on standby for takeover", which fundamentally conflicts with the traditional criminal imputation system for traffic accidents centered on drivers. At present, academic circles mostly conduct macro debates around issues such as the qualification of artificial intelligence as a criminal subject, the attribution of criminal liability for algorithms, and the distribution of risks in autonomous driving. There is a lack of refined research in line with judicial practice on specific imputation dilemmas caused by algorithmic black boxes in Level 3 scenarios, leading to frequent

dilemmas including unclear liability attribution and inconsistent judgment standards. Taking the obstacles to criminal imputation of crimes involving Level 3 autonomous vehicles as the core anchor point, this paper focuses on three key links: the definition of liable subjects, the identification of subjective fault and the proof of causality, clarifies the pain points of criminal imputation in such cases, and constructs an operable imputation path adapted to technical characteristics, so as to provide normative references for relevant judicial application.

2. Review of relevant regulations on Level 3 autonomous driving

Level 3 conditional autonomous driving has a fixed Operational Design Domain (ODD). Within this domain, the system can independently perform all dynamic driving tasks and assume corresponding driving responsibilities on its own. When the vehicle operates outside the ODD or the system issues a takeover request, the driver shall complete the transfer of rights and responsibilities within the legally prescribed time limit. This is an operation mode featuring dynamic transfer of rights and responsibilities between human and machine. Autonomous driving traffic accidents in recent years have exposed prominent legal dilemmas in the industry, such as the blurred technical boundaries of the system's Operational Design Domain and the unclear definition of the driver's statutory takeover obligation under the human-machine co-driving mode, providing important practical references for the construction and improvement of China's regulatory system for Level 3 autonomous driving. Specific cases include the Handan section case on the Beijing-Hong Kong-Macau Expressway and the Baofeng National Highway case in Henan Province.

A review of China's current regulations shows that the *General Technical Requirements for Autonomous Driving Systems of Intelligent Connected Vehicles*, issued by the State Administration for Market Regulation and the Standardization Administration of China in 2024, has formulated mandatory provisions addressing core risks exposed in practice. These provisions are precisely adapted to the technical characteristics of Level 3 autonomous driving, clearly requiring vehicles to mark the system's applicable scope in a conspicuous manner and clearly define the boundaries of their Operational Design Domain. Vehicles must reserve a takeover response time of no less than 5 seconds for the driver; if the driver fails to take over in a timely manner, the system must automatically implement a minimum risk strategy of decelerating and pulling over to the roadside. Meanwhile, vehicles are mandatorily required to retain complete operation data for no less than 30 days prior to an accident. This has anchored the compliance baseline for Level 3 autonomous driving from a technical source. Within the framework of national mandatory standards, Beijing and Wuhan have taken the lead in issuing special local legislations, forming implementable and replicable local regulatory models. Both the *Regulations on Autonomous Vehicles of Beijing Municipality* and the *Regulations on the Development and Promotion of Intelligent Connected Vehicles in Wuhan Municipality* closely follow the technical characteristics of Level 3 systems—that is, the system assumes dynamic driving responsibilities within the Operational Design Domain. They clearly stipulate that Level 3 vehicles must pass a multi-departmental joint safety assessment before being allowed on the road, strictly restrict the road sections and time periods in which they can travel, refine the driver's statutory takeover duties, and clarify the subjects responsible for the determination of traffic violations and traffic accidents. At the level of criminal regulation, China's current criminal law system, centered on Article 132 of the *Criminal Law of the People's Republic of China* concerning the crime of causing a traffic accident, has formed clear applicable rules in light of the technical characteristics of Level 3 autonomous driving. It is clarified that in Level 3 driving mode, the driver bears the statutory obligation to stand by for takeover. If the driver violates traffic and transportation management regulations by being distracted, failing to take over, or failing to

respond to a takeover request in a timely manner, thereby causing a major accident, the constitutive elements of the crime of causing a traffic accident are satisfied. Overall, China's current regulatory system for Level 3 autonomous driving has responded promptly to emerging issues in industrial practice, achieving synchronization between legislation and technological development. Nevertheless, it still suffers from problems such as inadequate operability of certain provisions and poor coordination between technical standards and legal norms, making it difficult to fully address the complex legal challenges brought about by the rapid development of the industry.

3. Imputation dilemmas of crimes involving Level 3 autonomous vehicles

Traffic accidents involving Level 3 autonomous driving present prominent identification dilemmas in the application of criminal law, mainly focusing on three aspects: liable subjects, subjective identification, and causality.

3.1. Dilemmas in the imputation of liable subjects

First, there are obvious opposing views in academic circles on whether artificial intelligence can be regarded as a subject of criminal liability. Some scholars hold that strong artificial intelligence possesses autonomous consciousness and independent capacity for action, and can effectively translate its rational thinking into practice. Like other legal subjects, it can carry out legal acts and produce legal effects, thus having the practical possibility of independently bearing legal liability [1]. Scholars supporting this view argue that legal personality can be granted to strong artificial intelligence through legal fiction, so as to construct a liability distribution system compatible with technological characteristics, properly allocate tort liability while taking into account the practical needs of technological innovation and industrial development. Opponents, however, maintain that artificial intelligence lacks learning and application supported by free will and is merely a mechanical process of input and output [2]. They further point out that the operation of artificial intelligence is essentially the mechanical execution of pre-set human codes, without free will or capacity for criminal responsibility in the criminal law sense. Granting it legal subject status will not only undermine the dual-subject imputation system of traditional criminal law but may also enable the real liable subjects to evade criminal prosecution, resulting in a liability vacuum.

Second, most Level 3 accidents involve multiple causes leading to one effect, and current laws fail to clarify the identification standards of duty of care for various subjects, making it difficult to implement liability. Accidents involving Level 3 autonomous vehicles usually result from the joint faults of multiple subjects, such as the driver's breach of takeover obligation, defects in the automobile manufacturer's product safety guarantee obligation, and the lack of research and development duty of care of algorithm designers. However, current traffic regulations and product liability norms do not clarify the boundaries, performance standards and identification rules of the duty of care of the three parties in light of the technical characteristics of Level 3 autonomous driving. Consequently, after an accident, all parties plead "no fault" and shift liability to each other, making it hard for criminal imputation to pinpoint specific subjects.

Finally, the dual effects of algorithmic black boxes and data barriers have plunged the proof and identification of criminal liability of producers and sellers into difficulty. On the one hand, algorithm defects are highly concealed and unexplainable. Even professional technical identification can hardly fully restore the complete logical chain of algorithmic decision-making, nor can it prove whether producers and sellers were aware of product defects but still allowed risks to occur. The core proof of the subjective elements of negligent crimes is difficult to complete, directly hindering

the progress of criminal imputation. On the other hand, core evidence such as algorithm operation logs, vehicle driving data and system decision-making parameters is absolutely controlled by technical controllers such as automobile manufacturers throughout the process of generation, storage, extraction and disclosure. Current laws do not specify their mandatory obligation to fully retain and disclose relevant data. As a result, in judicial practice, automobile manufacturers often refuse to provide core data on the grounds of trade secrets, and even tamper with or conceal unfavorable data, further increasing the difficulty of proof for the prosecution. Ultimately, the criminal liability of automobile manufacturers is hard to identify, and they may even evade criminal accountability.

3.2. Dilemmas in subjective identification

On the one hand, algorithmic black boxes invalidate the identification logic of traditional criminal law, namely the presumption of subjective intent through objective acts. In traditional criminal judicial practice, judicial organs can comprehensively presume the subjective intent or negligence of the actor through objective acts such as premeditation, means of conduct and post-incident disposal. The core premise of this rule is that the actor's objective acts are governed by his subjective will. However, in Level 3 autonomous driving scenarios, vehicle operation is the output of algorithmic decision-making rather than a direct reflection of the driver's subjective will. The complete process of algorithmic decision-making is completely obscured due to its black-box nature, which can hardly be fully traced by existing technical means. Furthermore, it is impossible to reverse-presume the subjective culpability of the driver, automobile manufacturer, algorithm designer and other subjects through vehicle trajectories or operational acts in accidents. The traditional subjective presumption rule loses its applicable foundation.

On the other hand, algorithmic black boxes make it difficult to apply the traditional foreseeability standard of negligent crimes, resulting in the failure of negligent imputation logic. In the imputation theory of negligent crimes, the duty to foresee consequences is the core criterion. Its theoretical basis lies in whether the actor has the realistic possibility of foreseeing the occurrence of harmful consequences—that is, the actor is obliged to foresee the occurrence of harmful consequences. If foreseeing is possible, the basis for imputation is established; otherwise, it is not [3]. Traditional negligent crimes take the actor's cognition and control of behavioral risks as the basis for judging foreseeability. The existence of algorithmic black boxes directly renders this standard completely inapplicable in Level 3 scenarios. For drivers, they only bear the obligation to stand by for takeover within the Operational Design Domain, cannot intervene in the internal decision-making process of algorithms, let alone foresee potential algorithm failures and sudden risks, thus lacking concrete foreseeability of accidents. For automobile manufacturers and algorithm designers, although they can recognize the abstract safety risks of autonomous driving technology, due to the autonomous learning characteristics and opacity of algorithmic black boxes, they cannot foresee specific accident scenarios and causal processes, and thus fail to meet the requirement of "concrete foreseeability" in traditional criminal law. In addition, the blurred boundary of rights and responsibilities in dynamic human-machine takeover makes it difficult to accurately distinguish whether the accident is caused by system technical defects or the driver's operational negligence after an accident. Ultimately, core subjective elements such as breach of duty of care and existence of fault cannot be effectively proved through judicial procedures, and the traditional negligent imputation logic falls into an overall failure dilemma.

3.3. Dilemmas in the identification of causality

First, the dual obstacles of algorithmic black boxes and data barriers prevent the formation of a complete closed loop of the accident causal chain, making causal identification extremely difficult. In traditional human driving scenarios, the causal chain of traffic accidents is clear and traceable. There is a direct and traceable objective connection between the driver's illegal operation and the accident result, and causal attribution can be completed through basic data and on-site traces. However, in Level 3 autonomous driving scenarios, the core driving decisions of the vehicle are made by algorithms. The algorithmic black box completely obscures the internal operational logic of decision-making, making it difficult to form a complete logical chain from data input and algorithm calculation to behavior output. Whether the algorithmic black box will break the causal relationship between the harmful result and relevant subjects such as designers or users directly affects the criminal imputation of algorithm designers, and is also a key issue in the criminal imputation of artificial intelligence technology [4]. Data barriers, from the perspective of evidence materials for causal proof, further aggravate the aforementioned dilemma of causal identification: core materials such as full training data supporting real-time algorithmic decision-making, dynamic operation parameters, and decision-making logs at accident nodes are all enclosed within the closed-loop system of technical parties. Even if basic driving data are obtained, it is difficult to acquire the full-link core data of algorithmic decision-making, making it impossible to complete closed-loop verification of the causal chain. Even if an objective correlation between system operation and the accident result can be proved, it is impossible to accurately distinguish the causes of the accident due to the invisible black-box logic and unobtainable core data, rendering the causal attribution process extremely complicated.

Second, the probabilistic output characteristic of algorithms is fundamentally in conflict with the identification standards of traditional criminal law causality. The operation of artificial intelligence is dynamic and continuous. It no longer relies entirely on pre-programmed deterministic calculation rules, but continuously adjusts internal parameters through autonomous learning of massive data and generates decision-making results through probabilistic reasoning. The identification of causality in traditional criminal law requires an inherent and necessary connection, or a connection of high probability, between the harmful act and the harmful result, with certainty and foreseeability of causality as the core. However, algorithmic decision-making in Level 3 autonomous driving is essentially a result of probabilistic trade-offs based on massive data, and the occurrence of harmful results is obviously contingent and uncertain. Even if the actor fulfills the corresponding duty of care, the possibility of accidents cannot be completely ruled out, making it difficult to meet the core elements of traditional causal identification and causing traditional causality rules to lose their applicable foundation.

Finally, the risk characteristics of autonomous driving are incompatible with traditional imputation models, creating a dilemma in the application of the risk criminal law theory. Traditional criminal law is centered on actual harmful consequences and allows imputation only after real damage occurs. Yet risks posed by artificial intelligence are latent, sudden and unpredictable; once risks turn into actual harm, irreparable losses will occur. Therefore, it is necessary to apply the risk criminal law theory to regulate risk-creating acts and adopt a risk prevention orientation, weakening the strict requirement for actual harmful results in traditional theories and focusing on the dangerousness of the act itself. However, the risk criminal law theory may lead to the fictionalization of causal judgment, potentially expanding the scope of criminal punishment and excessively intervening in reasonable risks. How to strike a balance between crime prevention and

human rights protection is a problem that must be solved in the identification of causality involving artificial intelligence.

4. Improvement of the imputation path for crimes involving Level 3 autonomous vehicles

In response to the above identification dilemmas in terms of subject, subjectivity and causality, this part puts forward improvement paths from corresponding dimensions respectively, so as to realize a systematic response to practical problems.

4.1. Clarifying the attribution of criminal liability of multiple subjects in different scenarios

First, the status of Level 3 autonomous driving systems as criminal subjects should be clearly denied. Level 3 autonomous driving is essentially a set of graphical symbols created by algorithms and computer programs, lacking realistic human physiological characteristics and capabilities, and should be regarded as the object of legal relations. Imposing liability on it can neither achieve the preventive and retributive purposes of criminal punishment, but also may allow the real liable subjects to evade criminal punishment and create a liability vacuum. On this basis, it is necessary to hierarchically define the proportion of criminal liability of multiple subjects in combination with typical accident scenarios of Level 3 autonomous driving.

Scenario 1: Liability Identification for Inherent System Design Defects Combined with Minor Negligence of the Driver

This type of scenario is the most common in Level 3 autonomous driving accidents. The accident occurs within the Operational Design Domain of the system, caused jointly by technical defects of the system itself and minor flaws in the driver's duty of care, which is a typical case of multiple causes leading to one effect. From the perspective of criminal law imputation logic, the core of identifying such accidents lies in defining the boundary of the duty of care between human and machine. The establishment of a negligent crime presupposes that the actor has the possibility of avoiding the result. There is an objective and universal rule in the reasonable physiological response time of human beings to sudden road risks, which serves as the core benchmark for defining the boundary of the driver's takeover obligation. If the takeover response time reserved by the system is shorter than the reasonable physiological interval, even if the driver remains highly focused throughout the process, there is no realistic possibility of effectively avoiding the accident, and the core liability of the accident shall not be transferred to the driver. The safety guarantee obligation borne by manufacturers and system developers is by no means limited to issuing takeover risk reminders; they shall also take the initiative to adopt risk reduction measures while reminding and fulfill the bottom-line safety obligation. Merely reminding of risks but allowing dangers to escalate constitutes a design defect in violation of the statutory safety guarantee obligation, and has a direct criminal law causality with the harmful result.

Therefore, judicial adjudication shall be comprehensively determined based on three core elements: first, the takeover time reserved by the system, which mainly judges whether the driver has a realistic possibility of effectively avoiding danger; second, the system's bottom-line safety measures, which mainly identifies whether the automobile manufacturer has fulfilled its statutory safety guarantee obligation and whether the product has design defects; third, the driver's actual takeover speed, which mainly examines the rationality of his conduct and the degree of fault under emergency conditions. Based on this, hierarchical adjudication rules shall be formed: if the system reserves less than 5 seconds for takeover and fails to adopt bottom-line measures, the automobile manufacturer shall bear about 70% of the primary liability, and the driver shall bear 30% of the

secondary liability due to minor negligence, with the specific proportion dynamically adjusted according to the core elements of the case.

Scenario 2: Liability Identification for Activating Autonomous Driving Functions Beyond the System's Operational Design Domain

This type of scenario is a frequent type of Level 3 autonomous driving accidents. The accident occurs in an inapplicable scenario outside the system's Operational Design Domain, caused jointly by the driver's serious fault of illegally activating the system and failing to fulfill the takeover obligation, as well as the manufacturer's flawed obligation of failing to clearly inform the boundary of the Operational Design Domain and failing to adopt risk prevention and control measures after the system goes beyond the scenario. The core of criminal liability identification in such scenarios lies in that the driver, as the final person responsible for vehicle safety, illegally uses the system in inapplicable scenarios and is seriously out of supervision, violating the statutory duty of care, which is the main reason for creating the core risk of the accident. The automobile manufacturer's failure to fulfill the obligation of informing the Operational Design Domain and disposing risks only creates conditions for the occurrence of the accident and shall bear secondary liability.

Therefore, judicial adjudication shall be comprehensively judged based on two core elements: first, the driver's subjective fault degree for using the system beyond the Operational Design Domain, which directly determines the primary and secondary division of liability; second, the performance of relevant obligations by the automobile manufacturer, which is used to define its fault proportion and liability share. At the basic imputation level, if the driver illegally activates the autonomous driving function beyond the system's Operational Design Domain and is seriously at fault, he shall bear 70% of the primary liability; if the automobile manufacturer fails to fulfill the obligation of clearly informing the Operational Design Domain in a prominent manner and fails to adopt risk disposal measures such as function degradation, it shall bear 30% of the secondary liability. At the dynamic imputation level, if the automobile manufacturer completely fails to fulfill its statutory obligations, both parties shall share the liability equally; if the accident is purely caused by system defects without any possibility of foresight by the driver, the automobile manufacturer shall bear full liability.

Scenario 3: Liability Identification When the Driver Is Objectively Unable to Take Over After the System Issues a Takeover Request

This type of scenario is a special case in the imputation of Level 3 autonomous driving accidents. The accident occurs within the system's Operational Design Domain. When the system issues a takeover request, the driver completely loses the ability to take over due to unforeseeable objective reasons, and the system only gives a risk reminder without starting the emergency bottom-line procedure, eventually causing the accident. The assumption of criminal liability presupposes that the actor has a realistic possibility of fulfilling statutory obligations. If the driver completely loses the ability to take over due to a sudden and unforeseen objective reason, he neither acts in violation of the duty of care nor has a realistic possibility of avoiding harmful results, lacks the basis for criminal imputation, and shall not bear primary liability. The automobile manufacturer's safety guarantee obligation shall cover such extremely special scenarios, and it is a part of its statutory obligation to foresee such risks and preset complete emergency disposal procedures. The system only reminds for takeover but fails to adopt any risk bottom-line measures, which constitutes a product design defect in failure to fully fulfill the safety guarantee obligation, has a direct criminal law causality with the occurrence of the accident, and shall bear primary liability.

Therefore, judicial adjudication shall be comprehensively determined based on three core elements: first, the foreseeability of the driver's loss of takeover ability, which mainly judges

whether there is the possibility of expectation for him to fulfill the takeover obligation; second, the completeness of the system's emergency bottom-line function, which mainly defines whether the automobile manufacturer violates the safety guarantee obligation; third, the performance of the driver's basic health duty of care before driving, which is mainly used to adjust the liability proportion. At the basic imputation level, if the driver suffers from a sudden and unforeseen illness and cannot take over, and the system has no emergency procedure, the automobile manufacturer shall bear 70%-80% of the primary liability, and the driver who fails to fulfill the health check obligation shall bear 20%-30% of the secondary liability. In terms of dynamic adjustment, if the driver can prove that he has fulfilled all health duties of care without any fault, his liability may be reduced to less than 10%; if the accident is purely caused by the failure of the system's emergency procedure, the automobile manufacturer shall bear full liability; if the driver still uses the system on the road despite knowing that he suffers from a disease unfit for driving, he shall bear full criminal liability.

4.2. Reconstruction of subjective identification criteria and clarification of obligations based on the new negligence theory

While algorithmic black boxes undermine the judgment of specific result foreseeability under the traditional old negligence theory, the new negligence theory focuses on the general cognition of behavioral risks. Its criteria for judging foreseeability can effectively resolve the problem of drivers evading liability by taking advantage of algorithmic black boxes. The traditional old negligence theory takes the specific foreseeability of harmful results as the prerequisite for identifying negligence. However, algorithmic black boxes inherent to artificial intelligence feature opaque and unexplainable reasoning processes, which hinder the application of traditional judgment standards for the duty of result foreseeability. Drivers may exploit this limitation to evade liabilities on the ground that the algorithmic black box makes it impossible to accurately predict the system's response to specific road conditions. In contrast, the new negligence theory argues that foreseeability only requires a general awareness of the dangerous nature of a given behavior. Specifically, actors are subjectively culpable as long as they recognize the unreasonableness and undue risks of their conduct, without needing to accurately predict specific harmful outcomes. This rule effectively prevents drivers from improper liability evasion. An actor can be deemed to have foreseeability if he realizes that his conduct violates precautionary norms. Even if the actor cannot predict the exact occurrence of an accident, the awareness that his behavior is "imprudent" or "beyond reasonable risk limits" suffices to establish foreseeability [5]. Accordingly, the existence of algorithmic black boxes cannot serve as a barrier to the judgment of foreseeability.

On this basis, it is necessary to further specify the duty of result foreseeability and duty of care of multiple liable subjects in Level 3 autonomous driving scenarios, and clarify the specific judging criteria for subjective criminal imputation:

First, drivers shall enhance their capacity to foresee potential risks and reduce the probability of accidents. To this end, drivers must fulfill four dynamic duties of care. These include the duty of alert activation to complete state transition within two seconds after the system releases preliminary warning signals, the duty of timely response to improve takeover efficiency upon receiving immediate takeover instructions, the duty of standardized operation compliant with vehicle dynamics principles, and the duty of stable control to restore vehicle safety by regulating deceleration.

Second, manufacturers shall foresee operational and takeover-related risks during system operation. They shall reduce risk occurrence through four-dimensional reminder mechanisms

including sound, light, vibration and head-up display (HUD) to trigger drivers' visual, auditory and tactile perception. Fully respecting human physiological and cognitive characteristics, manufacturers must reserve adequate response time for drivers. Within the total takeover request duration of five seconds, a four-second early warning interval achieves the highest score in fuzzy comprehensive evaluation, which greatly optimizes lane-changing efficiency and driving stability. Therefore, such a time configuration should be adopted as the default and optimal system setting to guarantee reasonable preparation time for drivers [6]. Manufacturers that fail to equip vehicles with a reasonable takeover reminder system shall be deemed foreseeable of harmful consequences and held corresponding liable.

Third, program designers shall foresee traffic accident risks stemming from algorithm design defects. The designers' duty of foreseeability shall be evaluated based on the prevailing technological standards [7]. In accordance with the dominance principle, program designers bear the duty of result foreseeability for traffic accidents caused by algorithm design defects of artificial intelligence systems. This duty is relatively lenient. It does not require designers to foresee detailed accident scenarios, but merely to recognize that algorithm design defects may potentially trigger traffic accidents.

4.3. Eliminating algorithmic black boxes to optimize the causal imputation path

To address the dilemma of causality identification caused by algorithmic black boxes, an imputation path that is both normative and operable should be established on the basis of adhering to the basic principles of traditional criminal law causality. The core is to remove imputation obstacles through four rules:

First, adopt the right to algorithmic interpretation in line with the principle of proportionality to solve the problem of causal tracing under algorithmic black boxes. The right to algorithmic interpretation does not require algorithm controllers to disclose source code, but focuses on explanation oriented toward understandability, covering data types, weights, data processing rules and other contents related to specific algorithmic decisions [8]. The establishment of the algorithmic interpretation obligation must abide by the principles of necessity and proportionality, so as to support judicial organs in reconstructing the causal chain without infringing intellectual property rights and trade secrets, and ensure that algorithmic decision-making processes are traceable and verifiable.

Second, adopt the inversion of the burden of proof and a third-party interdisciplinary identification system to overcome difficulties in adducing algorithm data and shortcomings in judicial technical review. Where the prosecution preliminarily proves a correlation between artificial intelligence and the harmful result, the technical controller shall bear the burden of proving that its product has no technical defects and that it has fully fulfilled its statutory duty of care. Legislation should explicitly stipulate that technical parties shall not refuse to produce evidence or evade criminal obligations on the ground of trade secrets. If they fail to provide complete algorithm data or offer reasonable explanations, courts may directly draw adverse factual findings against them. Meanwhile, a third-party interdisciplinary identification team composed of automotive engineering experts, algorithm engineers and legal scholars may be established [9]. Its expert opinions may directly serve as the core basis for determining causality, thereby constructing a judicial identification system for artificial intelligence and making up for the lack of technical review capacity of judicial organs.

Third, the risk criminal law theory should be applied restrictively under the principle of legality for crimes and punishments, so as to balance the prevention of technological risks and the protection

of human rights. Its application must take the principle of legality for crimes and punishments as an absolute bottom line, and be strictly limited to the scope of crimes explicitly stipulated in the current criminal law. Breaking through legal provisions by analogical application or creating new crimes shall be prohibited to prevent improper expansion of the scope of criminal punishment at the source. This means the focus of evaluation is no longer merely on whether the result occurred and whether the actor could have avoided it but failed to do so, but more on whether the actor, within the scope of his cognition and ability, managed risks in the interaction process in a normatively expected, prudent and responsible manner [10]. The core of criminal law evaluation is whether the actor, within his cognitive and capacity limits, has fulfilled the prudent duty of care required by legal norms and industrial standards to avoid potential risks caused by algorithmic black boxes. Even if no actual harmful consequence occurs, acts failing to fulfill such obligations should be subject to criminal regulation. Conversely, if the actor has fulfilled a reasonable and prudent duty of care and adopted feasible monitoring, auditing and risk control measures against algorithmic black boxes, he shall not be held criminally liable even if unforeseeable risks arise suddenly.

Fourth, introduce the theory of proportional causality to resolve the dilemma of traditional causality identification caused by algorithmic black boxes. Proportional causality refers to the rule that causality is established solely based on the probability of a causal link between conduct and damage, and the liability share is calculated according to the proportion of such probability. Proportionality is the core of proportional causality, and its determination requires examining the degree of causal possibility between each subject's conduct and the damage. The degree of possibility is essentially a quantitative assessment of risk [11]. Under the premise that algorithmic black boxes make the causal chain difficult to trace intuitively and the liability transmission path vague, the proportional value of causality should be comprehensively determined by considering factors such as the subject's duty of care, risk level, ability to control the algorithm, and the inexplicability of the algorithmic black box. This causality identification model abandons the traditional logic of necessary causality, can effectively respond to practical problems such as difficulties in proving causality and lack of imputation basis caused by algorithmic black boxes, is more compatible with the probabilistic technological characteristics and opaque operation reality of artificial intelligence, and provides a reasonable causal judgment path for criminal imputation in relevant cases.

5. Conclusion

Focusing on the criminal imputation dilemmas of Level 3 autonomous driving, this paper sorts out its technical characteristics and the current normative system, and disassembles three core imputation difficulties concerning the definition of liable subjects, the identification of subjective fault, and the proof of causality. It further proposes targeted improvement paths, including denying the criminal subject qualification of artificial intelligence, dividing liability by specific scenarios, reconstructing imputation criteria based on the new negligence theory, and breaking down obstacles to causality identification adapted to algorithmic characteristics. This study has certain limitations. It does not conduct an in-depth discussion on the imputation issues of higher-level autonomous driving, and lacks empirical research on the practical judicial application effect of the proposed rules. With the continuous iteration of autonomous driving technology in the future, the criminal regulation system needs to be updated synchronously to consolidate the legal bottom line for the sound development of the industry.

References

- [1] Tao, P. (2025). Strong artificial intelligence should be recognized as a new civil subject. *Journal of Tianzhong*, 40(4), 35–41.
- [2] Ren, Y., & Nie, R. (2024). Denying the qualification of strong artificial intelligence as a subject of criminal liability. *Social Science Forum*, (4), 132–141.
- [3] Chen, L., & Zhang, H. (2026). The transformation of the theory of criminal responsibility in the era of artificial intelligence: From the principle of culpability to the principle of tolerance. *Journal of Beijing Police College*. Advance online publication.
- [4] Chen, Q., & Sun, D. (2025). Attribution of negligence in autonomous vehicle traffic accidents: Challenges, approaches, and limitations. *Jiangxi Social Sciences*, 45(11), 154–165.
- [5] Liang, R. (2026). Reshaping the foreseeability of drivers and developers in human–machine co-driving. *Journal of Wuhan Technical College of Communications*, 28(1), 8–17.
- [6] Yu, H. (2026). On the foreseeability rule for tort liability involving Level 3 autonomous vehicles. In *Collected Studies on Smart Rule of Law (Vol. 2, Studies in Digital Law)*, pp. 53–65.
- [7] Hu, W. (2025). On criminal attribution for negligence by autonomous vehicle software designers. *Journal of Wuhan Technical College of Communications*, 27(4), 30–38, 90.
- [8] He, X. (2023). Justification and limitations of the right to algorithmic explanation. *Journal of Dalian University of Technology (Social Sciences)*, 44(3), 65–73.
- [9] Tan, S. (2025). Liability determination and reconstruction of the burden of proof under the algorithmic black box of autonomous driving. *Workers' Legal World*, (11), 78–80.
- [10] Chen, L. (2025). Conflicts and reconstruction of the criminal duty of care of large language model users under the algorithmic black box: Criminal law adaptation from the perspective of risk governance. *Journal of Political Science and Law*, 42(6), 51–58.
- [11] Lu, S. (2026). Identification and proof of causation in artificial intelligence tort liability. *Journal of Chongqing University of Posts and Telecommunications (Social Sciences Edition)*. Advance online publication.