Categorical Identification of Criminal Liability for Semi-Autonomous Vehicles

Yihan Zhang *

School of Economic Law, Southwest University of Political Science and Law, Chongqing 500157, China

* Corresponding Author Email: z17749912105@163.com

Abstract. The widespread adoption of conditionally autonomous vehicles has rendered it imperative for China to regulate them within the criminal law framework. Confronted with challenges including the determination of criminal liability subjects, the assessment of subjective fault, and the delineation of causal relationships, the traditional model for identifying criminal liability in conventional vehicle cases is no longer applicable to autonomous vehicles, as autonomous driving systems cannot qualify as subjects of criminal liability. Drawing on the theory of objective imputation, the principle of reliance, and the objectives of industry development and legal regulation, this paper positively identifies the criminal liability of producers across the production and circulation stages, categorizes the duty of care for drivers into pre-driving and in-driving phases, and emphasizes the need to restrict such duties. From the perspective of criminal liability exemption, it negatively limits the scope of liability for relevant subjects by examining various potential natural or man-made exceptional circumstances. This approach achieves the dynamic transfer and multi-layered control of risks associated with conditionally autonomous vehicles between drivers and producers, ensuring the reasonable distribution of criminal liability among all stakeholders.

Keywords: Autonomous Driving, Criminal Liability, Duty of Care, Duty of Takeover, Exemption from Criminal Liability.

1. Introduction

In recent years, the deep integration of artificial intelligence (AI) technology, 5G technology, and the automotive industry has transformed autonomous vehicles from theoretical concepts into practical mobility solutions accessible to households worldwide. Since 2021, driven by the interplay of policy guidance, technological breakthroughs, and market demand, China's autonomous vehicle industry has entered an accelerated development phase. However, technological progress is a double-edged sword: while autonomous driving offers significant conveniences, it also poses non-negligible risks to personal and property safety. In response, China is expediting the establishment of a comprehensive regulatory system covering road testing, pilot applications, and market access for autonomous vehicles. At the state level, the draft amendment to the Road Traffic Safety Law has been included in the State Council's 2024 legislative plan, with multiple central ministries collaborating on the formulation of technical standards for intelligent connected vehicles. At the local level, numerous provinces and cities have actively explored and refined local regulations, such as the Beijing Autonomous Vehicle Regulations and the Shenzhen Special Economic Zone Regulations on the Administration of Intelligent Connected Vehicles.

According to China's Classification of Driving Automation Levels for Motor Vehicles, driving automation is categorized into Levels 0 to 5. Levels 0 to 2 correspond to assisted driving, Level 3 to conditional autonomous driving, and Levels 4 and 5 to highly autonomous and fully autonomous driving, respectively. In September 2025, the Ministry of Industry and Information Technology (MIIT) and other departments jointly issued the Work Plan for Stabilizing Growth in the Automotive Industry (2025-2026), explicitly stating that "conditional approval will be granted for the production access of Level 3 autonomous vehicles." The large-scale application of Level 3 autonomous vehicles has thus become a prominent trend in China's automotive industry development. Levels 0 to 2 autonomous driving technologies only assist human drivers, who retain ultimate control over the vehicle. Consequently, their criminal regulation is essentially identical to that of conventional human-driven

vehicles. Levels 4 to 5 autonomous technologies, while promising, are not yet feasible for rapid implementation, making their criminal regulation a non-urgent legal issue. In contrast, Level 3 autonomous driving serves as a critical milestone in achieving automated driving advancements and has garnered significant attention from all stakeholders. Therefore, this paper focuses on the study of criminal liability identification for semi-autonomous vehicles.

As defined in this paper, semi-autonomous driving refers to Level 3 conditional autonomous driving, wherein the system continuously performs all dynamic driving tasks under designed operating conditions, with the dynamic driving task fallback user (driver) executing a takeover when appropriate. In other words, the vehicle can operate autonomously without human intervention under specific conditions but requires the driver to be ready to take over at any time when necessary. Due to the involvement of autonomous driving technology and the uniqueness of the takeover duty in Level 3 autonomous driving, the identification of criminal liability for semi-autonomous vehicles presents numerous challenges, and existing criminal law contains significant gaps in regulating such cases. This paper first summarizes the challenges associated with identifying criminal liability for semi-autonomous vehicles, then positively identifies such liability, and finally discusses circumstances warranting exemption from criminal liability.

2. Challenges in Identifying Criminal Liability for Semi-Autonomous Vehicles

Compared to the identification of criminal liability for conventional vehicles, the determination of whether autonomous driving systems themselves can be held liable, the reasonable distribution of risks between autonomous systems (or their producers) and drivers, and the assessment of fault among relevant subjects constitute major challenges in the context of semi-autonomous vehicles, requiring further exploration.

2.1. Uncertainty in Identifying Liability Subjects

Under existing criminal law, criminal liability can only be borne by natural persons and legal persons. In conventional human-driven vehicle cases, the primary liability subject is the driver. Where accidents result from vehicle quality defects, manufacturers and sellers may be held liable for product liability. However, the introduction of autonomous driving technology raises the question of whether autonomous driving systems or the vehicles themselves can qualify as criminal liability subjects under criminal law. Currently, the mainstream view in Chinese academic circles regarding the criminal subject qualification of fully autonomous vehicles is the "negative theory." Scholars advocating this view argue that despite their high level of intelligence, autonomous driving systems lack independent consciousness and thus do not possess the cognitive and volitional capacities required for criminal liability, making them ineligible to bear criminal responsibility [1]. From the perspective of criminal law's purpose, classifying AI as a criminal liability subject would necessitate corresponding adjustments to the penalty system. However, the three penalties proposed by scholars—"data deletion, program modification, and permanent destruction"[2]—have no meaningful deterrent or corrective effect on non-sentient AI. Such measures merely satisfy utilitarian pursuits and contribute nothing to maintaining social order in the real world [3]. Notably, the purpose of developing AI is to enhance human well-being through technology; it is impermissible to treat autonomous vehicles, which serve as human transportation tools, as criminal subjects to prevent and punish their "crimes"[4].

This paper endorses the negative theory. Even fully autonomous driving systems are merely intelligent systems that simulate human brain functions, fundamentally differing from humans and lacking the essential elements required to be criminal liability subjects. Humans are emotional beings who actively or passively form relationships with others and the environment through social interaction. In contrast, AI "thinks" and "acts" based on massive amounts of data fed by programmers; its actions are determined not by consciousness but by inferences derived from such data, similar to mechanical behaviors performed by humans in a state of "loss of mental capacity," which are not punishable. If fully autonomous driving systems cannot be classified as criminal liability subjects,

semi-autonomous driving systems—with their lower level of automation—certainly cannot qualify either.

Having established that semi-autonomous driving systems cannot be held criminally liable, the key challenge lies in reasonably distributing criminal risks between drivers and producers in the context of the human-machine collaborative nature of semi-autonomous driving. The complexity of behavioral control inherent in human-machine collaboration makes the traditional driver-centric liability model inadequate in cases where accidents occur while the vehicle is under autonomous control or during the transition of control to the driver. Furthermore, imposing excessive liability on drivers in semi-autonomous driving would significantly dampen consumer demand, thereby hindering the long-term stable development of the autonomous vehicle industry. Consequently, the application of autonomous driving technology in the automotive industry has profoundly disrupted the traditional central liability model, presenting the major challenge of dynamically adjusting this model to achieve the reasonable distribution and transfer of risks and responsibilities between drivers and producers.

2.2. Difficulties in Assessing Subjective Fault

In conventional human-driven vehicles, drivers are influenced by numerous variables such as physiological functions and psychological emotions, leading to frequent traffic accidents. The application of autonomous driving technology can significantly reduce driving errors and accident rates, and the comfort it provides often fosters drivers' trust and reliance on the technology. However, this trust obscures the standards for assessing drivers' subjective fault in the event of accidents. In semi-autonomous driving, the primary challenge in assessing drivers' subjective fault lies in determining whether drivers are at fault when accidents occur before they receive a takeover request and assume actual control of the vehicle. For example, during autonomous driving, if the system fails to detect an obstacle due to a malfunction or other factors, and the driver—having not received a takeover request—reasonably believes the road conditions are safe or that the system will take effective measures, resulting in an accident, there is controversy regarding whether the driver is guilty of negligence. Holding the driver liable in such cases would require them to fulfill the same duty of care as in conventional driving, which contradicts the original intent of developing autonomous driving technology. Conversely, absolving the driver of liability would place primary responsibility on producers, but imposing such high technical requirements and duties of care on producers for semi-autonomous driving—with its lower level of automation and intelligence compared to fully autonomous driving—may be unfair and could inhibit technological innovation and development. Therefore, assessing drivers' subjective fault in semi-autonomous vehicle accidents is challenging and requires a comprehensive consideration of specific accident scenarios and the principle of fairness.

In semi-autonomous driving, where product defects in the design or manufacturing of autonomous vehicles cause property damage or personal injury to road users, producers may be held liable for product liability. Given the current state of technological development, producers often manufacture defective products due to technical or experiential deficiencies. Since negligent crimes are predicated on the breach of a duty of care, the key to determining whether a producer is negligent lies in defining the scope of their duty of care under specific circumstances [5]. The duty of care can be divided into the duty to foresee consequences and the duty to avoid consequences. Some scholars argue that due to the temporal and spatial separation between producers and specific autonomous driving accidents, producers' liability for negligence should be based on their failure to fulfill the duty to foresee consequences, which should be broader than that of conventional vehicle developers [6]. The determination of the scope of this duty to foresee consequences hinges on the identification of product "defects." According to the Product Quality Law, a "defect" refers to an unreasonable risk inherent in a product that endangers personal safety or the safety of others' property; where state or industry standards exist for a product, a defect also includes non-compliance with such standards. While China has initiated the development of multiple standards—such as the state standard General Technical Requirements for Autonomous Driving Systems of Intelligent Connected Vehicles, which sets

general technical requirements for autonomous driving systems—state and industry standards for the product quality of autonomous vehicles remain to be further clarified. Additionally, given the need for autonomous driving technology research and development and the unpredictability of algorithms, reasonable risks in autonomous driving systems should be permitted, but the definition of "permissible risks" requires further discussion.

2.3. Ambiguity in the Boundaries of Causal Relationship

Causal relationship in the criminal law context refers to a direct and necessary cause-and-effect link between the perpetrator's harmful act and the resulting harm. In cases involving semi-autonomous vehicle crimes, the identification of causal relationship faces numerous challenges, including determining whether a causal relationship exists, assessing liability in cases of multiple causes leading to a single effect, and addressing the "algorithm black box" problem.

Firstly, there is controversy regarding whether a causal relationship exists between producers and traffic accidents caused by product defects in autonomous vehicles. Some scholars argue that producers only have an abstract possibility of foreseeing harmful consequences and lack the ability to foresee or avoid specific negligent acts or their outcomes[7], thus negating a causal relationship between producers' manufacturing activities and the resulting harm. However, excluding producers as liability subjects in such cases would place full criminal liability on drivers, who would essentially be held accountable for uncontrollable product defects or technical vulnerabilities—an outcome that is clearly unreasonable. Furthermore, if producers have fulfilled their reasonable duty of care (i.e., their vehicles comply with all state and industry standards) but an unforeseeable defect causes an accident, the question arises as to whether producers should be held liable. Absolving producers of liability in such scenarios would leave criminal liability unassigned, denying victims fair compensation for their damages.

Secondly, the identification of multiple causes leading to a single effect in human-machine interaction is complex. Traffic accidents involving autonomous vehicles often result from multiple causes, with liability potentially shared among drivers, autonomous system producers, and other subjects. The existence of the takeover function in semi-autonomous driving exacerbates this complexity. For example, if an accident occurs while the vehicle is under autonomous control due to a obvious system malfunction, and the driver—out of overconfidence—fails to take effective measures, the question of how to allocate liability between the system and the driver arises. Similarly, if the autonomous system detects an emergency and issues a takeover request but provides insufficient time for the driver to respond effectively, resulting in an accident that could have been avoided with an earlier warning or a more timely response, determining the liability subject becomes problematic.

Thirdly, the "algorithm black box" problem poses a significant challenge. Factors contributing to autonomous vehicle accidents include not only human-controllable factors such as driver error and producer negligence but also non-human uncontrollable factors such as the algorithm black box. Due to the opacity and inexplicability of AI algorithms, no one can fully understand the features extracted by the system during decision-making. Even if an algorithm is designed in compliance with relevant legal and technical standards, the system may still produce unforeseeable harmful consequences during autonomous operation. There is controversy regarding whether such unforeseeable consequences constitute an abnormal intervening factor that breaks the causal chain or should be classified as an unforeseeable event. Improper regulation in this regard could create potential risks of lawbreakers exploiting legal loopholes. Given the extreme complexity of the algorithm black box, simplistic attribution is inappropriate, making the identification of criminal liability in such cases particularly challenging.

3. Categorical Identification of Criminal Liability for Semi-Autonomous Vehicles

The traditional model for distributing criminal liability in conventional vehicle cases is incompatible with semi-autonomous vehicles, necessitating the development of an adaptive liability framework. This section specifically identifies the criminal liability of producers and drivers across the production, circulation, and driving stages of semi-autonomous vehicles.

3.1. Identification of Producers' Criminal Liability

3.1.1. Criminal Liability During the Production Stage

In cases involving semi-autonomous vehicle crimes, the introduction of autonomous driving technology may render producers liable as criminal subjects. Some scholars argue that discussions regarding producers' duty of care are only necessary in the context of specific traffic accidents. However, producers cannot foresee or avoid specific traffic accidents during the design and manufacturing process, making it impossible to hold them liable for negligence [8]. This paper contends that producers bear an indirect duty to avoid consequences. Given the current immaturity of autonomous driving technology, producers can reasonably foresee potential harmful consequences arising from autonomous driving systems. Due to temporal and spatial separation, while producers cannot directly prevent accidents, they are obligated to take reasonable pre-emptive measures to indirectly avoid such consequences. According to the Opinions on Strengthening the Access Management of Intelligent Connected Vehicle Manufacturers and Products (hereinafter referred to as the Opinions) issued by the MIIT, producers are obligated to ensure the safe operation of vehicles under designed operating conditions. This includes ensuring that vehicles can automatically detect autonomous driving system failures, continuously meet designed operating conditions, and are equipped with autonomous driving data recording systems. Additionally, producers must fulfill a strict duty to inform, explicitly disclosing information such as vehicle performance, driver responsibilities, and the activation and deactivation mechanisms of autonomous functions to consumers. When necessary, producers should also provide driving skills and safe operation training to prevent foreseeable and preventable traffic accidents. Therefore, producers' criminal liability during the production stage is limited to breaches of the duty to ensure product quality and the duty to inform.

Building on the identification of producers' duties to ensure product quality and inform, this paper applies the theory of objective imputation to determine the causal relationship between producers' failure to fulfill these duties and the resulting harm. In today's risk society, individuals are often unable to predict future risks or harmful consequences, a phenomenon exacerbated by technological advancements. We must respond proactively to pervasive risks while allowing for reasonable risks to coexist, preventing technological stagnation or regression due to excessive regulation. Faced with the expanding scope of social risks, traditional theories such as the but-for test and the proximate cause theory focus on factual attribution but neglect normative evaluation, potentially leading to overcriminalization and the improper extension of causal chains. These theories are inadequate for objectively and reasonably attributing liability in a risk society. In contrast, the theory of objective imputation is grounded in objective facts and employs layered restrictions to regulate significant risks, making it well-suited to address risk distribution issues arising from autonomous driving technology.

When producers are unaware of defects in their manufactured autonomous vehicles, and such defects cause traffic accidents, how should the causal relationship be determined? According to the theory of objective imputation, the first step is to determine whether the producer created an unreasonable risk. It is important to note that the scope of "producers creating risks" should be negatively restricted by considering factors such as "reducing risks," "not creating risks," and "creating permissible risks." For example, inherent defects in autonomous driving technology that exist when vehicles are approved for market entry and put into circulation should be deemed "permissible risks," and producers should not be held criminally liable for accidents resulting from

such defects. The second step is to determine whether the risk was realized. During the autonomous driving phase, the vehicle is fully controlled by the system, and the driver's duty of care is reduced. If detecting and avoiding a defect exceeds the driver's duty of care, the accident caused by the product defect should be regarded as an extension of the system's actions, and the producer behind the system should be deemed to have realized the risk. At this stage, it is still necessary to determine whether the producer falls under circumstances such as "failure to realize the risk" or "failure to realize an impermissible risk." Finally, the realized risk must fall within the scope of protection of the criminal law provisions; otherwise, the producer cannot be held liable. Therefore, according to the theory of objective imputation, producers should be held liable for traffic accidents if they created and realized an impermissible risk that falls within the scope of criminal law regulation [9].

3.1.2. Criminal Liability During the Circulation Stage

If a producer fails to detect a system defect during the design and manufacturing process but discovers the defect after the vehicle has entered circulation and fails to take appropriate measures to prevent traffic accidents, should the producer be held liable? This paper argues that producers cannot avoid liability for traffic accidents merely by claiming to have fulfilled their duty of care during the production stage. For risk prevention purposes, producers and drivers must collaborate to implement multiple safeguards against potential risks both before and after vehicles enter circulation. Therefore, producers' performance of post-delivery remedial obligations during the circulation stage is an essential component of risk prevention.

From a practical perspective, such defects are not beyond the detection capabilities of current scientific and technological levels and thus do not fall within the scope of "permissible risks." Producers can identify previously undetected defects through various channels, such as issues discovered during the production of new vehicle batches, data feedback from traffic accidents, and consumer complaints. Therefore, when producers discover defects after vehicles have entered circulation, they must take post-delivery remedial measures to prevent traffic accidents. Additionally, compared to other relevant subjects, producers possess superior risk prevention and problem-solving capabilities, enabling them to more effectively identify the source of defects and control the realization of risks.

From a normative perspective, according to the Opinions, producers must establish self-inspection mechanisms. If significant issues such as over-the-air (OTA) update safety, driving assistance safety, or autonomous driving safety are identified in semi-autonomous vehicles, producers must immediately cease production of the relevant products and take corrective actions in accordance with the law. Thus, producers' performance of post-delivery remedial measures is not only an objective requirement based on practical considerations but also a legal obligation. Producers are responsible for ensuring the quality of the products they manufacture. When defects are discovered after vehicles have entered circulation, producers are obligated to take remedial actions such as ceasing production and recalling products. If producers knowingly fail to take proactive measures to address defects discovered during the circulation stage, they should be held liable for traffic accidents resulting from such defects.

3.2. Identification of Drivers' Criminal Liability

The type and scope of the driver's duty of care vary depending on the vehicle's operating status, the transfer of control during operation, and the level of automation of the autonomous vehicle. This section first qualitatively analyzes the driver's duty of care before and during driving, then quantitatively analyzes it to further substantively restrict the scope of the duty of care.

3.2.1. Types of Duty of Care

Before operating a semi-autonomous vehicle, the driver bears an inspection and update duty. In the conventional vehicle context, drivers are required to perform daily maintenance and regular inspections of their vehicles and check for obvious faults before driving. In the semi-autonomous vehicle context, due to the involvement of autonomous driving technology, drivers must not only

fulfill these traditional obligations but also promptly update and upgrade the autonomous driving system to prevent foreseeable risks of traffic accidents. Autonomous vehicles are the result of the application and iterative development of AI technology in the automotive industry. Producers continuously patch vulnerabilities and enhance functions based on issues identified during the design and manufacturing process, data generated by vehicles in market circulation, and feedback from autonomous vehicle traffic accidents[10]. On this basis, drivers must fulfill their duty to inspect and update the system before driving to improve driving safety and implement dual safeguards against the occurrence and realization of risks by both producers and drivers prior to vehicle operation.

During the operation of a semi-autonomous vehicle, the driver bears a duty to monitor or remain alert to the vehicle's operating status and road conditions. In emergency situations such as when the system issues a takeover request, the driver also bears a takeover duty. However, whether the driver is obligated to monitor, remain alert, or take over the vehicle varies depending on the specific phase of semi-autonomous driving.

During the autonomous system-led driving phase, the autonomous system independently performs all driving tasks. According to the Shenzhen Special Economic Zone Regulations on the Administration of Intelligent Connected Vehicles (hereinafter referred to as the Shenzhen Regulations), when a conditionally autonomous vehicle is operating in autonomous driving mode, the driver must be seated in the driver's seat and monitor the vehicle's status and road conditions. Thus, the driver is obligated to fulfill a monitoring duty in autonomous driving mode. In contrast, Germany's Road Traffic Act stipulates that drivers in Level 3 autonomous driving must fulfill an alertness duty. Therefore, there is controversy regarding whether drivers should bear a monitoring duty or an alertness duty during the autonomous driving phase. The monitoring duty requires the driver to actively observe the vehicle's status and road conditions, imposing a high standard of care. The alertness duty, by contrast, requires the driver to maintain a state of wakefulness and awareness during autonomous driving to enable timely takeover in emergency situations, imposing a lower standard of care. This paper argues that drivers only need to bear an alertness duty when the vehicle is under system control. While fulfilling a monitoring duty would more effectively prevent the occurrence and realization of risks, imposing such a high standard of care on drivers would contradict the original intent of developing autonomous driving technology. On one hand, the high requirements of the monitoring duty would significantly reduce consumers' willingness to purchase semi-autonomous vehicles for convenience. On the other hand, imposing a high standard of care on drivers might encourage producers to neglect product quality and technological innovation. Therefore, the adoption of the alertness duty is more appropriate.

During the driver-led driving phase, when the driver retains control of the vehicle, they should be held to the same standard of care as drivers of conventional vehicles. During the human-machine interactive driving phase, there is a takeover process from the moment the system issues a takeover request until the driver completes the takeover, during which the driver must fulfill a takeover duty. Due to the presence of numerous interfering factors during this takeover process, the timing of the driver's takeover duty requires further discussion[11]. Article 35 of the Shenzhen Regulations stipulates: "When an intelligent connected vehicle issues a takeover request or is in a state unsuitable for autonomous driving, the driver shall immediately take over the vehicle." Thus, the driver's takeover duty primarily arises in two scenarios: "when the system issues a takeover request" and "when the driver discovers that the vehicle is in a state unsuitable for autonomous driving." In the first scenario, the system continuously monitors the vehicle's status and road conditions and issues a takeover instruction when it is about to exceed the designed operating parameters, requiring the driver to passively fulfill the takeover duty. If the system fails to issue a takeover request or issues an erroneous request due to a malfunction, and the situation is not so obvious that a reasonable driver should have detected it or made the correct decision, the driver should not be held liable for accidents resulting from their failure to fulfill the takeover duty. In the second scenario, the driver must fulfill the aforementioned alertness duty and proactively take over the vehicle upon discovering that it is in a state unsuitable for autonomous driving. The definition of "discovering that the vehicle is in a state

unsuitable for autonomous driving" directly impacts the scope of the driver's obligation. Germany advocates that "the driver shall immediately take over the vehicle when they recognize or should recognize based on obvious circumstances that the autonomous driving function no longer meets the conditions for its intended use." This paper considers this an objectively reasonable requirement for drivers, effectively preventing the realization of risks, and recommends that China adopt this standard. Additionally, the vehicle's abnormal state refers to situations where "the autonomous driving function no longer meets the conditions for its intended use," which corresponds to the information disclosed by producers pursuant to their duty to inform. Therefore, if the driver fails to proactively take over the vehicle due to a failure to recognize such an abnormal state, it is reasonable to hold them at fault.

3.2.2. Restriction of the Duty of Care

Building on the identification of the types of the driver's duty of care, this section further restricts the scope of such duty. Some aspects of the duty of care exist only at the theoretical level; imposing all such duties as high-level legal obligations would improperly expand drivers' criminal liability. Therefore, further restricting the duty of care is essential to achieving a reasonable distribution of risks between drivers and producers.

As an example, during the autonomous driving phase, if the system fails to detect an obstacle due to a malfunction, and the driver—having not received a takeover request—reasonably believes the road conditions are safe or that the system will take effective measures, resulting in an accident without the driver taking any action, some scholars argue that drivers cannot be expected to continuously monitor the vehicle and road conditions during autonomous driving and should be allowed to reasonably trust the system, thus precluding a finding of negligence[12]. This paper contends that such a blanket exemption is inappropriate. Given the relatively low level of intelligence of semi-autonomous driving systems, drivers' reasonable trust in the system should be limited. If the driver observes obvious abnormalities in the autonomous driving system or perceives an imminent and urgent danger to which the vehicle fails to respond, the driver must immediately take over the vehicle and take measures to avoid an accident. Failure to fulfill this obligation constitutes reckless negligence.

Notably, the actual performance of the driver's takeover duty must also consider subjective factors such as the driver's takeover capability and objective factors affecting takeover effectiveness. Specifically, successfully taking over control of the vehicle within a short period and avoiding an accident during the human-machine interactive driving phase is not trivial, requiring the satisfaction of multiple conditions: "short time frame," "successful takeover," and "accident avoidance." Firstly, regarding the takeover time, when the system issues a takeover request or the driver detects an abnormality, the situation is often urgent, with an accident imminent. The driver must complete multiple steps within a short period, including receiving the takeover instruction, processing information, making decisions, and executing operations. Therefore, the system must allow sufficient reasonable time for the driver to complete the takeover when issuing a request. Meanwhile, as technology matures, the system should avoid issuing premature intervention requests to prevent drivers from ignoring such requests under the mistaken belief that the system is malfunctioning when no obvious abnormalities are observed. Secondly, regarding the driver's state during takeover, during the autonomous system-led driving phase, drivers may engage in non-driving activities. Different non-driving activities impose varying demands on the driver's senses and motor skills, affecting their reaction time and, consequently, their takeover capability. Therefore, relevant regulations should comprehensively consider the impact of non-driving activities on the performance of the takeover duty and restrict the scope of permissible non-driving activities. Finally, regarding accident avoidance, if the driver strictly complies with the system's takeover instruction or proactively takes over the vehicle when appropriate but still cannot avoid an accident, the driver should not be held liable for the accident.

4. Circumstances Warranting Exemption from Criminal Liability for Semi-Autonomous Vehicles

In addition to positively identifying the circumstances under which producers and drivers should be held criminally liable for semi-autonomous vehicles, it is necessary to explore specific exemption scenarios where relevant subjects are not liable or are exempt from criminal punishment from an exculpatory perspective, thereby substantively identifying criminal liability for semi-autonomous vehicles.

4.1. Exemption Scenarios for Producers

4.1.1. Accidents Caused by Product Defects Undetectable by Current Scientific and Technological Levels

When designing and manufacturing semi-autonomous vehicles, producers can only conduct relevant production activities based on existing scientific and technological levels and take necessary measures to avoid the occurrence and realization of controllable risks. However, due to the limitations of current scientific and technological levels, even with extensive road testing, producers cannot anticipate all potential road interaction outcomes and risks that may arise after vehicles enter circulation. Therefore, if producers strictly comply with relevant state or industry standards but unforeseeable risks still materialize, the law cannot impose an unreasonable burden by holding producers criminally liable[13]. If producers can prove that they fulfilled their reasonable duty of care under current scientific and technological conditions and that their products comply with relevant standards, they should not be deemed negligent during the production stage and should be exempt from criminal liability. Furthermore, scientific and technological progress is an ongoing process, and no technology is perfect. Autonomous driving technology inherently possesses unavoidable technical limitations and defects that cannot be overcome or effectively mitigated with current technological capabilities. If producers are aware of such inherent defects during manufacturing but take reasonable measures to provide risk warnings and implement preventive measures, and the products pass state approval and enter the market, such defects should be regarded as "permissible risks," and producers should be exempt from criminal liability for accidents resulting from these defects.

4.1.2. Accidents Caused by Drivers' Illegal Operations or Third Persons' Gross Negligence

Corresponding to producers' duty to inform, drivers must have a clear understanding of information such as the vehicle's equipment performance, the operation mechanism of the autonomous driving system, the procedures for activating or deactivating the system, and the requirements for fulfilling the takeover duty before driving. If an accident results from the driver's intentional or grossly negligent violation of relevant regulations—such as failing to perform required operations or engaging in illegal operations—producers should be exempt from criminal liability, provided they have strictly fulfilled their duty to inform or provided necessary driving training. Additionally, if an accident is entirely caused by a third person, producers shall not be held liable. For example, at a traffic light intersection, a pedestrian suddenly runs a red light when the vehicle is about to pass, leaving the autonomous system or driver insufficient time to take evasive action such as braking. In this case, the accident is caused by the pedestrian's gross negligence, and neither the producer nor the driver should be held liable.

4.1.3. Accidents Caused by Force Majeure

Force majeure constitutes a ground for excluding criminal liability under criminal law, referring to situations where the perpetrator foresees the occurrence of harmful consequences but is unable to prevent them due to objectively irresistible reasons. This principle also applies to semi-autonomous vehicles. Under extreme weather conditions such as heavy rain, dense fog, or sandstorms; natural disasters such as earthquakes, floods, or mudslides; or sudden road incidents such as road collapses, bridge failures, or falling objects, semi-autonomous vehicles may be unable to effectively avoid accidents within a very short period. Additionally, the introduction of autonomous driving technology

may give rise to special risks such as system hacking or signal interference. For example, vehicle systems may be maliciously hacked and controlled, or solar storms may disrupt the Earth's magnetic field, causing signal deviations in semi-autonomous vehicles. The aforementioned natural or man-made events are beyond human control, and even with maximum effort, accidents cannot be prevented. Therefore, producers and drivers are not at fault subjectively and may be exempt from or have their criminal liability reduced if they have fulfilled their reasonable duty of care.

4.2. Exemption Scenarios for Drivers

4.2.1. Accidents Caused by Failure to Take Over the Vehicle in a Timely and Effective Manner

During the human-machine interactive driving phase of semi-autonomous vehicles, factors such as whether the system issues a takeover instruction, when the instruction is issued, and how it is communicated all affect the driver's ability to complete the takeover and, consequently, the occurrence of accidents. Drivers shall not be held criminally liable for failing to successfully take over the vehicle in the following circumstances: First, if the system fails to detect abnormalities in the vehicle's operating status or surrounding environment due to its own defects, and such abnormalities are not so obvious that a reasonable driver should have detected them, the driver should be excluded from criminal liability for the traffic accident. Second, if the time provided by the system for the driver to take over after issuing a request is shorter than the time required for the driver to complete the takeover and avoid the accident, the driver lacks the capacity to take over in a timely manner and should not be held liable. Third, regarding the method of issuing the takeover request, systems typically use auditory alerts to notify drivers of the need to take over. If the system fails to issue an alert or the alert is too faint for the driver to perceive when an abnormality is detected, leading to an accident, the driver should not be held liable. To address this, producers can design multi-sensory takeover request mechanisms to mitigate the risk of failure or insufficient notice from a single mechanism. For example, installing vibration devices in the driver's seat or olfactory devices in the vehicle that emit a pungent but non-harmful odor.

In addition to system-related factors, drivers may be unable to take over the vehicle in a timely manner due to non-subjective personal factors. If the system issues a takeover request at a time when the driver is experiencing a sudden illness or has reached their physical limit, preventing them from taking over within the normally reasonable time frame and resulting in an accident, the driver lacks the capacity to take over and should not be held criminally liable. Finally, if the system issues the takeover request in a timely and appropriate manner, and the driver strictly complies with the request but an accident still occurs due to extremely rare and unforeseeable factors, the driver's actions objectively caused the harm, but they are not at fault, and the outcome exceeds the cognitive and predictive capabilities of a reasonable driver. In such cases, the incident should be classified as an unforeseeable event, and the driver should be exempt from criminal liability.

4.2.2. Accidents Caused by Reasonable Trust in the System

Based on the principle of reliance in criminal law, an actor who reasonably believes that another person will act or refrain from acting in a certain manner in accordance with social norms shall not be held liable if the other person's unreasonable conduct causes harmful consequences. In other words, an actor may not be obligated to avoid consequences even if they have the possibility of foreseeing them. This principle prevents the improper expansion of the duty to foresee, reasonably limits the scope of the actor's duty of care, and achieves a dynamic balance between the actor's freedom and the protection of the other person's rights and interests [14]. In the context of conventional vehicles, drivers are entitled to reasonably believe that other motorists and pedestrians will comply with traffic rules during normal driving. If other road users suddenly violate traffic rules, drivers are generally not held liable for negligence.

The principle of reasonable reliance also applies between drivers and autonomous driving systems in semi-autonomous vehicles. From the driver's perspective, the vehicle operating in autonomous

driving mode is analogous to other road users, and the bond between them is the shared obligation to comply with semi-autonomous vehicle operation and safety rules. Drivers are entitled to reasonably believe that the system will operate in accordance with pre-set rules. Therefore, drivers shall not be held liable for negligence if they have fulfilled their own duty of care. During the autonomous system-led driving phase, drivers only need to fulfill a low-level alertness duty and are not required to maintain the same level of constant attention as in conventional driving. If a system malfunction or road abnormality is not so obvious as to attract the driver's attention, the driver is entitled to reasonably believe that the system or road conditions are normal and is not obligated to proactively take over the vehicle, and thus shall not be held liable for any resulting accident. However, if the system or road abnormality is within the scope of the driver's duty of care, the driver cannot rely on the principle of reasonable trust in the system to claim exemption from liability.

4.2.3. Accidents Caused by the System's "Algorithm Black Box"

In general, AI algorithms exhibit strong randomness and opacity in data feature extraction, making the "algorithm black box" problem inevitable [15]. Although semi-autonomous vehicles have a much lower level of intelligence than fully autonomous vehicles, the possibility of "autonomous decision-making" during operation cannot be ruled out. In cases where neither producers nor drivers can foresee or avoid harmful consequences caused by the algorithm black box, the question of who should bear liability arises. Typically, producers cannot foresee whether or when the algorithm black box phenomenon will occur during research and development, nor can they take effective measures to avoid it, and thus should be excluded from criminal liability. Meanwhile, holding drivers—who are neither at fault nor have engaged in any harmful conduct—liable would be inconsistent with existing legal norms. The simultaneous exclusion of producers and drivers from criminal liability leaves victims' legitimate rights and interests unprotected [16]. However, in reality, the rapid development of explainable AI technology in recent years has made it possible to overcome the technical barriers posed by the algorithm black box [17]. Even in the presence of an algorithm black box, professional identification institutions can investigate the operating principles of autonomous vehicles through post-hoc transparency methods such as code analysis to determine whether producers are liable for manufacturing defects.

Furthermore, even if the algorithm black box problem cannot be fully resolved at the technical level, legal techniques can be employed to determine the causal relationship in criminal liability identification. Some scholars have proposed a "reverse elimination" rule for this purpose. Specifically, after conducting an overall assessment of the accident's consequences and relevant indicators, all other impossible causes are eliminated to confirm the existence of a causal relationship. This method effectively prevents subjects from using the algorithm black box as a pretext to avoid liability and avoids the creation of artificial legal loopholes [18]. Therefore, the occurrence of the algorithm black box does not automatically exempt producers from criminal liability. If post-hoc transparency techniques or the reverse elimination rule can establish a causal relationship between the producer and the accident, the producer should be held criminally liable. In contrast, drivers—who have no substantive connection to the system's design, research and development, or performance—may claim exemption from criminal liability for accidents caused by the algorithm black box.

5. Conclusion

Autonomous vehicles represent the application of AI technology in the automotive industry, bringing unprecedented conveniences to travel while posing significant risks to road safety. With the mass market entry of semi-autonomous vehicles, criminal law must proactively address various risks arising throughout the "full life cycle" of semi-autonomous vehicles and identify criminal conduct falling within its regulatory scope. Simultaneously, while acknowledging the convenience and limitations of autonomous driving technology, criminal liability should be excluded from certain scenarios through the establishment of exemption provisions. This approach enables the autonomous vehicle industry to achieve technological innovation and development within the boundaries of the

law, protects the legitimate rights and interests of consumers and other road users to a certain extent, and ultimately promotes the coordinated development of legal regulation, industry progress, and consumer welfare.

References

[1] Zhuang Yonglian, Huang Jingping, Gao Yandong, et al. Artificial Intelligence and the Future of Criminal Rule of Law [J]. People's Procuratorate, 2018, (01): 41 - 48.

[2] Liu Xianquan. "Internal Troubles" and "External Threats" in the Age of Artificial Intelligence and Criminal Liability [J]. Oriental Law, 2018, (01): 134 - 142.

[3] Xu Ximing. A New Perspective on Criminal Liability for Traffic Accidents Involving Autonomous Vehicles [J]. Journal of Chongqing University of Posts and Telecommunications (Social Science Edition), 2020, 32 (05): 44 - 50.

[4] Pi Yong. Basic Issues Concerning the Criminal Rule of Law for Artificial Intelligence [J]. Journal of Comparative Law, 2018, (05): 149 - 166.

[5] Zhao Sicong. On the Criminal Risks of Autonomous Driving and Their Responses [J]. Northern Legal Science, 2022, 16 (02): 132 - 142.

[6] Lin Yujia. An Analysis of Negligent Crimes in Autonomous Driving Accidents [J]. Journal of Chongqing University (Social Science Edition), 2024, 30 (06): 238 - 249.

[7] Zhou Mingchuan. On the Criminal Liability for Traffic Accidents Involving Autonomous Vehicles [J]. Journal of Shanghai Jiao Tong University (Philosophy and Social Sciences Edition), 2019, 27 (01): 36 - 43.

[8] Jiang Su. The Challenges of Autonomous Vehicles to the Law [J]. China Law Review, 2018, (02): 180 - 189.

[9] Yao Wanqin. The Theory of Objective Imputation and the Attribution of Criminal Liability for Traffic Accidents Involving Autonomous Driving [J]. Journal of Dalian University of Technology (Social Sciences Edition), 2023, 44 (06): 93 - 101.

[10] Huang Chenchen. The Division and Undertaking of Criminal Liability in Traffic Accidents Involving Autonomous Vehicles [J]. Journal of Dalian Maritime University (Social Science Edition), 2021, 20 (03): 40 - 47.

[11] Cai Xian. The Takeover Duty and Its Criminal Attribution in the Human-Machine Co-Driving Model [J]. Journal of Soochow University (Law Edition), 2023, 10 (03): 27 - 40.

[12] Huang Bo. The Criminal Law Response to Traffic Accidents Involving Autonomous Vehicles [J]. Tianfu New Discussion, 2018, (04): 112 - 118.

[13] Wang Ying. How is Law Possible? ——A Legal Perspective on Risk Scenarios of Autonomous Driving Technology[J]. Law and Social Development, 2019, 25 (06): 99 - 112.

[14] Cai Xian. The Development of the Attribution System for Negligent Crimes in Autonomous Driving[J]. Journal of Comparative Law, 2023, (04): 65 - 81.

[15] Chen Jinghui. The Legal Nature of Algorithms: Speech, Trade Secrets, or Due Process? [J]. Journal of Comparative Law, 2020, (02): 120 - 132.

[16] Long Min. The Identification and Distribution of Criminal Liability for Traffic Accidents Involving Autonomous Driving [J]. Journal of East China University of Political Science and Law, 2018, 21 (06): 77 - 82.

[17] A Jincheng. A Hierarchical Theory of Algorithm Transparency [J]. Chinese Journal of Law, 2023, 45 (02): 52 - 66.

[18] Shi Jieyang. The Demarcation of Criminal Liability Between Drivers and Producers in Human-Machine Interaction Scenarios [J]. Huxiang Forum, 2025, 38 (05): 112 - 124.