

自动驾驶中人类驾驶员刑事责任认定的困境与破解

邵思语

西南民族大学, 四川 成都 610041

DOI:10.61369/SE.2026030038

摘 要 : 自动驾驶技术正经历从辅助驾驶到高阶自动驾驶的根本性变革。在这一技术变革中,人类驾驶员的角色从传统驾驶模式下的完全主导者,逐渐转变为智能驾驶系统运行过程中的监督者或纯粹的乘客,这一根本性转变对现有刑法归责体系构成严峻挑战。本文围绕人类驾驶员刑事责任认定面临的困境,构建了基于技术分级的差异化归责框架。并在该框架下探讨了智能驾驶系统的犯罪主体资格问题,根据不同自动化级别的技术特征,在传统驾驶员注意义务的基础上结合信赖原则重新界定人类驾驶员的义务边界。最后通过危险现实化理论有效破解主体多元化和算法黑箱导致的因果关系认定难题。以期增强刑法在技术迭代中的适应性与前瞻性,促进技术创新的同时维护社会平稳发展。

关 键 词 : 自动驾驶; 人类驾驶员; 刑事归责

The Dilemma and Solution Of Human Drivers' Criminal Responsibility in Automatic Driving

Shao Siyu

Southwest Minzu University, Chengdu, Sichuan 610041

Abstract : autonomous driving technology is undergoing a fundamental change from assisted driving to high-level autonomous driving. In this technological change, the role of human driver has gradually changed from the complete leader in the traditional driving mode to the supervisor or pure passenger in the operation of intelligent driving system. This fundamental change poses a severe challenge to the existing criminal liability system. Focusing on the dilemma faced by human drivers' criminal responsibility determination, this paper constructs a differentiated imputation framework based on technology classification. Under this framework, the subject qualification of intelligent driving system is discussed. According to the technical characteristics of different levels of automation, the obligation boundary of human drivers is redefined based on the traditional driver's duty of care and the principle of trust. Finally, we can effectively solve the problem of causality identification caused by subject diversification and algorithm black box through the theory of dangerous reality. In order to enhance the adaptability and foresight of criminal law in the technological iteration, promote technological innovation and maintain the stable development of society.

Keywords : automatic driving; human driver; criminal liability

一、问题的提出

随着自动驾驶汽车商业化变革的逐步落地,自动驾驶在未来出行方面必将成为主流趋势。在相关产品普及的过程中,人类驾驶员在混合驾驶或全自动驾驶模式中的责任归属问题日益成为关注焦点,抑或成为公众选择自动驾驶汽车的核心考量因素。法律层面的缺失往往会掣肘自动驾驶汽车的发展。从目前我国颁布的相关法律法规以及地方性法规、规章制度等来看,对自动驾驶车辆发生交通事故的民事责任等有所规定,但如何规制其刑事责任的确尚处于空白状态,亟需刑法层面予以回应。由此,本文旨在系统分析自动驾驶中人类驾驶员刑事责任认定的困境与出路,为相关立法与司法实践提供参考。

二、自动驾驶中人类驾驶员刑事责任的认定困境

(一) 技术分级下的归责层次化需求

自动驾驶汽车通过雷达、监控装置等多种传感器采集汽车外部环境信息,从而在功能层面替代了人类驾驶员的视觉感知还环节。随后,智能驾驶系统依托芯片或者算法对收集到的数据进行计算并生成控制策略,实现了人类驾驶员大脑的信息整合与处理职能。最终,经决策形成的指令被传导给具体的应用系统并执行,完成对人类驾驶员手脚操作的替代。可见,自动驾驶汽车与传统汽车的显著区别在于自动驾驶汽车的替代属性,它可以不同程度上接替人类驾驶员完成驾驶操作。在人机交互的自动驾驶的情境下,人类驾驶员并非汽车的唯一控制主体。在发生法益侵害

后果时，不能直接认定其刑事责任，而应根据智能驾驶系统对人类驾驶员的替代程度，认定其相应的责任范围。自动驾驶汽车对人类驾驶员的替代程度取决于其内部智能驾驶系统的智能化水平高低。在不同的技术级别中，智能驾驶系统对车辆的控制权存在差异，因此界定人类驾驶员的刑事责任范围时必须结合具体技术级别展开讨论，这又为人类驾驶员刑事责任的认定提出了新的挑战。

（二）对犯罪构成要件的冲击

1. 智能驾驶系统主体资格争议

由于智能驾驶系统的介入，驾驶活动涉及多个参与主体。因此，在认定人类驾驶员责任时要考虑到其他主体是否应当承担主责以及责任划分问题。从目前国际上对于自动驾驶汽车的法律规范、政策指引看，一般可以按照“谁掌控驾驶权谁负责”的基本思路来归责。相应地，掌握驾驶权的主体就是责任主体。在传统汽车事故中，掌控驾驶权只有人类驾驶员。但在智能驾驶的场景下，由于智能驾驶系统的介入程度越来越深，人类对驾驶权的掌控也随之减少。在L4、L5级场景下，智能驾驶系统成为驾驶权的实质掌控者，但认定智能驾驶系统的刑事责任要以其具有犯罪主体资格为前提。有学者认为人工智能技术得出的结论是基于海量数据拟合的概率结果，故自动驾驶所应用的人工智能缺乏自由意志，并不具备刑事责任能力，不具有犯罪主体资格^[1]。支持肯定说的学者认为在强人工智能时代的人工智能系统已然具备自主意识，针对人工智能在自我意识的支配下实施严重危害人类的行为，应将其作为刑事责任主体^[2]。对此议题学界尚未形成统一意见，然而犯罪主体范围不明确，也会导致刑事责任认定路径出现阻碍。

2. 认定因果关系存在障碍

因果关系是认定行为人对危害结果承担刑事责任的必要条件。在自动驾驶涉罪案件中，对因果关系的认定面临双重障碍。一方面，由于智能驾驶系统的介入使得责任主体呈现多元性，法益侵害结果往往是由多方主体相互作用形成，呈现“多因一果”的复杂结构。难以清晰区分直接结果或间接结果，亦难以精确评估各主体的原因力比例，这使传统刑法中的因果关系理论在适用时陷入困境。另一方面，更为根本的是，自动驾驶系统普遍采用的“端到端”方式接受传感器数据并输出车辆控制指令，该模型架构具有显著的算法黑箱特性，其内部决策逻辑缺乏可解释性，致使发生事故后难以通过技术溯源准确查明导致损害发生的具体环节或原因，从而动摇了因果关系认定的证据基础。

三、自动驾驶中人类驾驶员刑事责任认定困境的破解路径

（一）分级归责框架的构建

1. L2级及以下自动驾驶

L2级及以下自动驾驶系统仅作为执行特定操作的辅助工具，人类驾驶员仍全程主导各项驾驶操作。此种场景下，发生法益侵害的直接原因通常是驾驶员的人为操作或监控疏忽，人类驾驶员仍是唯一且明确的责任主体。在过失认定上，辅助驾驶犯罪的注意义务根据与传统车辆犯罪的注意义务根据是一致的，主要是交

通运输管理法规^[3]，无需因辅助技术的加入而改变原有注意义务标准。同时，由于驾驶行为与损害结果之间的因果链条清晰，自动驾驶技术介入不会使其中断或复杂化，因果关系的认定与传统交通事故无异。对此级别自动驾驶的人类驾驶员归责不存在上述困境，尚未动摇以驾驶员过错为中心的传统归责原则与责任认定框架，故本文不再赘述。

2. L3-L5级自动驾驶

在L3至L5级自动驾驶的不同阶段，人类驾驶员能否成为责任主体、其注意义务的具体内涵以及算法黑箱对因果关系判断的影响程度均存在显著差异。因此要解决上述责任认定困境须在不同技术级别下分类讨论。

（二）智能驾驶系统犯罪主体资格之否定

分析人工智能能否具备犯罪主体地位，主要是围绕其是否具备刑事责任能力和刑罚实施能力两个层面展开。关于刑事责任能力，即辨认和控制自己行为的能力。L3级驾驶系统因其智能化程度有限，缺乏独立意识和根据认知自主行为的能力，故不具有刑事责任能力。而L4、L5级智能驾驶系统通过独立的信息处理与自主决策，可实现完全自主的车辆运动控制及故障应对^[4]。此时的智能驾驶系统不单单只是传统意义上的工具而是具有自主意识而能够辨认和控制自己行为的实体，因而在理论上具有刑事责任能力。关于刑罚实施能力，部分学者提出对人工智能修改程序、删除数据、永久销毁等刑罚建议。对此，笔者认为该建议首先面临执行层面的现实障碍：当某一车辆的智能驾驶系统故障导致事故时，若要对该智能驾驶系统施以刑罚，处罚范围难以界定。是仅对涉事系统进行处罚，还是必须对基于同一设计原理的整批系统全部进行修正？若选择后者，不仅导致刑罚成本过高，还可能阻碍技术创新与产业发展。其次，刑罚的惩罚功能也难以实现。智能驾驶系统不具备生命体或法律拟制主体的感知能力，无法体会自由刑带来的痛苦，亦无独立财产可供剥夺。因此，传统刑罚难以对其产生实质的惩罚效果。最后，刑罚的预防功能也存在局限性。即便对涉事系统采取修改程序或永久销毁等措施，也只能防止该特定系统再次造成危害。对于采用相同设计原理，可能共享潜在缺陷的同批次其他系统，若未进行统一的处理，则无法达成期望的预防效果。程序修改本身也不能确保系统不再发生类似危害，“治标不治本”的潜在危险仍可能存在，其实际预防效果仍值得质疑。基于现有刑罚建议在处罚对象范围界定与刑罚功能实现上均存在障碍，故人工智能虽在理论上具备犯罪主体所需的刑事责任能力但在现行法律框架下仍暂不宜认定其犯罪主体资格。

（三）因果关系的厘清

自动驾驶参与主体多元化及智能驾驶系统的“算法黑箱”使因果关系更加错综复杂，刑法学传统的因果关系理论难适用于无人驾驶时代^[5]。根据刑法理论，只有必然的因果关系才是刑法中的因果关系，即危害结果与危害行为之间需存在内在、合乎规律的引起与被引起关系。然而在自动驾驶领域，内在、合乎规律的要求本就具有一定的抽象性，加之自动驾驶系统具有“黑箱”特性，难以认定危害结果与危害行为之间是否存在直接的引起关系，该理论的适用面临显著挑战。

由于 L3 级与 L4、L5 级中智能驾驶系统的参与程度不同，算法黑箱对因果关系判断的影响程度也存在显著差异。L3 级自动驾驶采用人机共驾模式，事故发生时需辨析危害结果究竟源于人类驾驶员违规运行系统、接管操作不当、算法严重缺陷抑或是算法在运行中产生的不可预见的异化。算法的“黑箱”特性使得追溯直接致害原因变得极为复杂。当多个主体的行为均与危害结果相关时，认定因果关系并区分必然因果关系和偶然因果关系将面临更大的障碍。在 L4 和 L5 级自动驾驶中，由于人类驾驶员完全退出驾驶操作，可排除因人类驾驶员接管操作不当导致的损害，事故归因的焦点得以简化，主要集中在系统自身。其核心问题在于判断损害是由产品缺陷导致，还是属于使用者违规运行系统的后果。

为破解上述因果关系认定困境，有必要引入“危险现实化”理论。该理论核心在于从事后看现实的因果经过，只有相应行为的危险在结果中得以现实化时，才具有因果关系^[6]。运用该理论须分两步进行分析。首先，要识别某行为所内含的特定危险，且该危险须为不被容许的危险。自动驾驶系统的设计者根据技术规范设计出相应的算法或程序的行为本身就是具有风险的行为，算法基于其技术原理以及学习的局限性可能会发生异化，导致系统故障乃至做出错误抉择并引发事故。尽管该行为潜藏着损害法益的危险，但该行为的目的是正当的，且客观上有益于社会。因此，算法异化的危险应当被认为是被容许的危险，不得将算法创设行为认定为造成事故的直接原因，从而追究设计者的责任。否则将严重制约自动驾驶算法的创新与更迭，阻碍自动驾驶技术的发展。在社会发展和科技创新等宏观利益面前，个人身体健康与生命安全的位阶在遵循相关行业标准、国家标准和法律法规前提下退居次席也是不得不承认的价值预设^[7]。第二步，在结果发生后，判断该损害是否可被评价为上述特定危险的现实实现。在 L3 级自动驾驶中，不同主体的行为都可能含有特定危险。如人类驾驶员未按规范使用自动驾驶汽车，包括出行前未检查系统性能以及在设计条件外使用，内含系统故障，引发事故的危险。人类驾驶员未及时、合理接管的行为，内含导致财产损失或人身伤亡的危险。智能驾驶系统设计者违反技术规范设计算法、程序的行为则内含导致系统故障，甚至引发人身或财产损害的危险。对智能驾驶系统设计者的违规行为的认定，主要是检验其在实验阶段是否存在算法缺陷，需要有关专门机构介入检验。若现实的危害结果被覆盖在特定行为内含的危险范围内，则可认定该行为与危害结果之间有因果关系。具体而言，若人类驾驶员出行前未检查系统性能以及在设计条件外使用自动驾驶导致事故发生，此种情况下，危险的现实化是由人类驾驶员未按规范使用自动驾驶汽车

参考文献

- [1] 康兰平、王紫函：《自动驾驶汽车交通犯罪中刑法规制研究》，《东南法学》2021 年第 1 期。
- [2] 朱凌珂：赋予强人工智能法律主体地位的路径与限度 [J]. 广东社会科学，2021，(05):240-253.
- [3] 姜盼盼，王润慧：自动驾驶汽车犯罪归责的困境与破解 [J]. 邢台学院报，2025，40(05):105-116.
- [4] 向鹏，卢娅：自动驾驶汽车刑事风险及归责路径研究 [J]. 安顺学院学报，2025，27 (05): 108-114.
- [5] 彭文华，熊浩宇：辅助驾驶与自动驾驶注意义务之比较 [J]. 江汉论坛，2024，(03):133-138.
- [6] 赵燃，李刚炎，胥军，等：基于自动驾驶的汽车智能制动系统架构设计及功能分级 [J]. 客车技术与研究，2024，46(04):1-7.DOI:10.15917/j.cnki.1006-3331.2024.04.008.
- [7] 付玉明：自动驾驶汽车事故的刑事归责与教义展开 [J]. 法学，2020，(09):135-152.
- [8] 宋满金：自动驾驶情形中注意义务的分配与边界 [J]. 司法警官职业教育究，2024，5(04):48-56.

车引起的，该行为与危害结果之间存在因果关系。若在智能驾驶系统因无法应对紧急情况而发出接管请求时，人类驾驶员未能及时接管或接管后操作不当导致事故发生，此时人类驾驶员的行为直接促使自身行为所内含的危险现实化，其作为或不作为与结果之间的因果关系应予认定。若因智能驾驶系统设计者未履行技术规范，致使系统存在设计缺陷，进而在紧急情况下发生故障，导致驾驶员客观上丧失及时干预的可能性，则危害结果应评价为设计者违规行为所内含危险的实现，其行为与结果之间存在因果关系。当情况更复杂一点，若系统设计者的违规行为与驾驶员的不当操作行为共同作用于事故的发生，即系统设计缺陷导致系统决策失误，而驾驶员未履行合理注意义务，未及时接管或接管后操作失误而未能避免损害。其中人类驾驶员的行为系介入因素，且介入因素异常。根据危险现实化理论，即使介入因素极其异常，但实行行为的危险性极高、介入因素对结果的贡献很小时，也应肯定因果关系的存在。系统设计者的违规设计行为创设了高度且紧迫的危险，其不仅显著增加了系统故障的发生几率，而且一旦故障发生，通常伴随着重大的人身或财产损害后果^[8]。在 L3 级自动驾驶设定的运行条件下，人类驾驶员对系统的正常运行具有合理的信赖。当因设计缺陷引发突发性系统故障时，驾驶员往往难以在极短时间内有效介入以避免危害结果的发生，欠缺结果避免的可能性，这意味着即便驾驶员存在接管延迟或操作不当，其行为对最终损害结果发生的推动作用也相对有限。因此，在此种高危险性、低回避可能性的情境下，应当肯定系统设计者的违规行为是危害结果发生的主要原因，其与危害果之间的刑法上的因果关系不应因驾驶员的介入行为而中断。

四、结语

人工智能技术的快速发展正强力推动自动驾驶领域迈向新的高度，目前高阶智能驾驶正处于从测试试点迈向商业化应用的关键时期，而未来必将是高阶自动驾驶汽车的时代。现行刑法在应对人类驾驶员角色转变带来的归责挑战时已明显显现出一定的滞后性。鉴于当前自动驾驶仍处于商业化初期，普及程度有限，且相关刑事司法案例尚不充足，在刑事责任认定规则的构建上，不宜急于建立独立的专门针对自动驾驶中人类驾驶员的罪刑体系。应先在现有法律框架下系统分析其刑事责任认定所遇困境，继而结合不同自动化级别的技术特征对既有规则进行灵活解释与适度发展，以渐进式调适的思路应对自动驾驶技术持续演进带来的法律挑战。