

自动驾驶紧急避险的刑事责任分析

王佳琪

广东工业大学法学院, 广东 广州

收稿日期: 2026年7月3日; 录用日期: 2026年8月4日; 发布日期: 2026年8月12日

摘要

自动驾驶技术的规模化落地, 使传统刑法理论中的“电车难题”从思想实验变为现实风险, 有关自动驾驶场景下紧急避险的刑事责任认定, 已成为刑法学必须回应的命题。其中, 针对危险源的防御性紧急避险, 因危险制造者的法益保护性降低, 在符合法定要件时可阻却违法; 针对无辜第三人的攻击性紧急避险, 尤其是生命权冲突的极端场景, 虽有驾驶员优先、数量功利主义为代表的生命权衡理论作为理论基础, 但其违背了生命权不得作为紧急避险损害客体的刑法基本原则。基于此, 本文认为自动驾驶算法预设的攻击性紧急避险行为, 在构成要件层面符合故意杀人罪等罪名, 且无法通过违法阻却事由出罪; 结合自动驾驶的技术特性以及不同应用场景, 可以分情况界定算法设计者、生产商、驾驶员的刑事责任边界, 实现刑法法益保护的规范目的。

关键词

自动驾驶, 紧急避险, 刑事责任, 算法归责

Analysis of Criminal Liability for Necessity in Autonomous Driving

Jiaqi Wang

School of Law, Guangdong University of Technology, Guangzhou Guangdong

Received: July 3, 2026; accepted: August 4, 2026; published: August 12, 2026

Abstract

The large-scale deployment of autonomous driving technology has transformed the “trolley problem”—a classic thought experiment in traditional criminal law theory—into a tangible real-world risk. Determining criminal liability for necessity in autonomous driving scenarios has thus become a core issue that criminal jurisprudence must address. Defensive necessity targeting the source of danger may negate wrongfulness when all statutory requirements are satisfied, given that the legal

interests of the party who created the danger warrant a reduced level of legal protection. As for offensive necessity that inflicts harm on innocent third parties, especially in extreme scenarios involving conflicts of the right to life, life-weighting theories such as the driver-first principle and quantitative utilitarianism are often cited as theoretical support. However, such theories run counter to the fundamental criminal law principle that the right to life shall not be the object of damage in necessity, and cannot provide valid justificatory grounds for these necessity acts. On this basis, this paper argues that preprogrammed offensive necessity embedded in autonomous driving algorithms meets the constitutive elements of crimes including intentional homicide, and cannot be exculpated through any justification defense. Taking into account the technical characteristics of autonomous driving and the specifics of different scenarios, the boundaries of criminal liability for algorithm designers, manufacturers and drivers can be delineated on a case-by-case basis, so as to realize the normative purpose of criminal law in protecting legal interests.

Keywords

Autonomous Driving, Necessity, Criminal Liability, Algorithmic Imputation

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着智能网联汽车技术的快速迭代, L4 级高度自动驾驶已在我国多地实现商业化运营, L5 级完全自动驾驶的技术落地亦进入倒计时。自动驾驶技术通过算法替代人类驾驶员的驾驶决策, 大幅降低了人类操作失误引发的交通事故, 但其无法从技术上完全排除极端碰撞场景的出现, 即当车辆陷入不可避免的碰撞风险, 只能在不同法益损害之间做出选择时, 算法的预设决策本质上符合刑法上的紧急避险行为[1]。但与人类驾驶员在紧急状态下的瞬时决策不同, 自动驾驶的避险决策是通过算法预先设定的, 而其中的算法逻辑不仅涉及紧急避险制度的传统教义学争议, 更引发了行为主体认定、归责路径重构、算法伦理与刑法规范协调等一系列新问题。

紧急避险制度的核心争议之一为能否以牺牲无辜第三人的生命保全其他生命[2]。在自动驾驶场景中, 这一命题从抽象的哲学思辨转化为具体的算法设计与责任认定问题, 即算法能否预设牺牲少数人生命保全多数人生命的决策? 优先保护车内驾驶员的算法设计是否具有刑法上的正当性? 预设违法避险算法的生产商、使用自动驾驶系统的驾驶员, 应分别承担何种刑事责任? 对此学界中各种学说分别站在不同的立场上对该问题做出诠释。生命权作为最高位阶的人格法益, 具有不可衡量、不可处分的属性, 这是我国刑法的一贯立场, 也为自动驾驶紧急避险的刑事规制提供了基本规范指引。在此价值指引下, 本文以我国《中华人民共和国刑法》(下文简称《刑法》)¹第 21 条关于紧急避险的规定为基础, 结合自动驾驶技术的特性, 对自动驾驶下紧急避险的正当性边界、构成要件该当性、刑事责任分配等问题进行思考, 并最终构建适配自动驾驶场景的刑事规制逻辑体系, 以期司法适用提供理论参考, 同时为自动驾驶技术的合规发展划定清晰的刑法红线。

2. 自动驾驶紧急避险的场景界定与核心刑事争议

自动驾驶紧急避险的刑事责任认定是以场景的类型化划分为逻辑前提。只有厘清不同避险场景的构

¹<https://flk.npc.gov.cn/detail?id=ff808181796a636a0179822a19640c92&fileId=&type=&title=%E4%B8%AD%E5%8D%8E%E4%BA%BA%E6%B0%91%E5%85%B1%E5%92%8C%E5%9B%BD%E5%88%91%E6%B3%95>

造差异，才能准确界定其刑事争议的核心焦点，进而展开针对性分析。

2.1. 自动驾驶紧急避险的场景类型化

根据避险行为指向的对象与危险来源的关系，可以将紧急避险分为防御性紧急避险与攻击性紧急避险，这一分类对自动驾驶场景同样适配，也是本文展开分析的基础框架。

防御性紧急避险，是指避险行为针对的是正在发生的危险来源本身，而非与危险无关的无辜第三人[3]。在自动驾驶场景中，该类避险主要表现为车辆因其他交通参与者违反交通规则的行为陷入碰撞风险，算法为避免损害发生，针对违规的危险源实施避险行为。如对向车辆违规逆行、非机动车突然闯入机动车道、行人违规横穿封闭高速公路等情形下，自动驾驶车辆为避免碰撞而采取的紧急制动、适度转向等行为，而造成了违规交通参与者的损害。此类避险的核心特征在于避险对象是危险的制造者与危险源，是其违规行为直接引发了碰撞风险，因此其法益的保护性低于守法的避险人。防御性紧急避险的正当性认定标准相较于攻击性紧急避险更为宽松，即便造成的损害与保全的法益基本相当，也存在阻却违法的空间，这一点在自动驾驶场景中同样适用。

攻击性紧急避险，是指为避免正在发生的危险，将损害转嫁给与危险来源完全无关的无辜第三人的行为。这是自动驾驶紧急避险的核心争议场景，也是“电车难题”的现实化体现[4]，根据侵害法益的不同衍生出两种攻击性紧急避险类型：其一，财产法益冲突的攻击性紧急避险，即为保全车辆内驾驶员的生命、健康权，或重大财产权，而损害无辜第三人的财产法益。车辆为避免与前方失控车辆相撞，转向撞向路边的护栏、无人商铺等财物。此类场景因涉及不同位阶的法益权衡，符合法益衡量原则的此类避险行为可直接阻却违法，因此在实务中几乎无裁判争议[5]。其二，生命权冲突的攻击性紧急避险，即车辆陷入不可避免的碰撞风险，只能在牺牲车内驾驶员生命与牺牲车外无辜行人生命之间或在不同数量的无辜主体生命之间做出选择。“电车难题”是其现实代表：车辆刹车失灵，直行将撞击前方5名行人，右转将撞击路边1名无辜行人，或驶向隔离带增加车内驾驶员死亡的可能。此类场景涉及生命权的权衡，是自动驾驶紧急避险刑事责任认定的难题，也是本文研究的核心问题。

2.2. 自动驾驶紧急避险的核心刑事争议

与人类驾驶员实施的传统紧急避险相比，自动驾驶紧急避险因“算法预设决策、多方主体参与、事前程序设定”的技术特性，产生了一系列全新的刑事争议。

首先，生命权冲突下紧急避险的违法阻却边界。学界存三大观点：一是肯定说，该说认为生命权的至高无上性仅具有抽象意义，在极端特殊场景下，生命的数量可作为法益衡量的标准，紧急避险可以成为牺牲少数人生命保全多数人生命的阻却事由[4]。该说同时主张对生命紧急避险设置严格的适用条件以避免功利主义的滥用。二是否定说，即坚守生命权不可权衡、不可量化的刑法底线，认为任何以牺牲无辜第三人生命为代价的避险行为都无法阻却违法[6]。该学说在充分肯定了生命价值的不可替代性的同时，明确不得通过牺牲无辜者的生命实现避险目的。三是折中说，该学说试图通过义务冲突理论、有限自治模式等来回避正当性争议。如主张由驾驶员事前选择避险模式，算法仅执行驾驶员的自主决策；或认为算法设计者面临两个不作为义务的冲突，履行其中一项义务即可阻却违法[7]。

其次，算法预设避险行为的行为主体界定。传统紧急避险的行为主体是实施避险行为的自然人，而自动驾驶避险决策的逻辑是由算法预先设定、系统再自动执行，在此情形下必然需要考虑“谁是避险行为的实施主体”这一前提性问题。对此，学界也存有以下观点：一是主张生产商、算法设计者是行为主体，认为算法的设计与编写直接决定了避险行为的模式，生产商对算法具有最终的控制权，因此预设避险算法的行为就是刑法上的实行行为，生产商是避险行为的核心主体[8]。二是主张驾驶员是行为主体，

该观点认为驾驶员是启动自动驾驶系统、选择使用具有预设避险算法的车辆的核心主体，其对算法下的避险行为具有利用与支配的能力，这与人类驾驶员亲自实施避险行为具有规范上的等价性，因此驾驶员是避险行为的主体[6]。三是主张生产商、驾驶员均为行为主体，该观点认为生产商的算法设计与驾驶员的系统使用对避险结果的发生具有共同作用，二者是避险行为的共同主体，应根据共犯原理分配责任。

最后，是不同主体的刑事责任分配规则。在明确避险行为的违法性与主体归属情形下，还需解决不同主体的刑事责任分配问题，预设违法避险算法的生产商，应承担故意犯罪还是过失犯罪的责任？驾驶员使用具有违法避险算法的车辆，能否以期待可能性缺失阻却责任？算法设计符合行业标准，但因技术局限导致避险不当的，能否以容许的风险排除归责？这些问题直接决定了自动驾驶紧急避险的刑事归责效果，也是司法实践中最需解决的现实问题。

3. 自动驾驶紧急避险正当性的检视

自动驾驶紧急避险的刑事责任认定，首先需解决正当性问题，即何种避险行为能够通过违法阻却事由出罪。对此，需以我国《刑法》第21条的规定为基础，结合防御性与攻击性紧急避险的类型差异而分别展开。

3.1. 防御性紧急避险的正当性认定

针对危险源的防御性紧急避险，在自动驾驶场景中具有完整的正当性基础，其教义学依据在于危险的制造者通过违规行为创设了法不容许的风险，其法益保护性相较于守法避险人显著降低，因此即便避险行为造成了危险制造者的损害，也符合紧急避险的法益衡量原则[3]。

我国《刑法》第21条规定紧急避险“造成的损害小于所避免的损害”，通说认为这是紧急避险的核心成立要件。在防御性紧急避险中，这一要件的判断并非简单的法益位阶对比，而需结合危险的来源、双方的过错程度、法益的保护性进行实质判断。在自动驾驶场景中，其他交通参与者的违规行为是碰撞风险的唯一来源，其不仅违反了道路交通安全法规，更创设了针对自动驾驶车辆与驾驶员的不法危险，此时合法性原则应当优先适用。避险行为造成了违规者的重伤甚至死亡，也不能简单认定为“造成的损害大于所避免的损害”，而应结合双方的过错程度进行实质判断。如自动驾驶车辆在封闭高速路段因制动距离不足而无法完全避免对突然闯出行人的伤害并造成死亡结果，应认定为防御性紧急避险阻却违法。这一结论不仅符合逻辑，也与我国道路交通的司法实践相契合。在人类驾驶员因避让违规行人而造成事故的案件中，司法机关通常会根据行人的过错程度减免驾驶员的责任，其背后的法理正是危险制造者的法益要保护性降低。需要强调的是，防御性紧急避险的正当性仅限于针对危险制造者本身。如果自动驾驶车辆为避让违规的交通参与者，转而撞击其他无辜的第三人，则不再属于防御性紧急避险，而转化为攻击性紧急避险，其正当性需另行判断。

3.2. 攻击性紧急避险的正当性否定

针对无辜第三人的攻击性紧急避险，尤其是生命权冲突的极端场景，无论基于何种理论，都不具备刑法上的正当化依据。其根本原因在于，此类行为违背了生命权不可量化、不得作为紧急避险损害客体的刑法核心原则，也不符合我国紧急避险制度的规范目的。

以修正功利主义为代表的生命权衡理论违背宪法基本原则。肯定说认为，抽象层面上所有生命价值平等，但在极端场景下，生命的数量可作为法益衡量的标准，牺牲少数人保全多数人符合“两害相权取其轻”的紧急避险制度本质。同时，该说主张通过“高偶然性、不得已、仅以数量为唯一衡量标准等”严格条件，限制生命紧急避险的适用范围[4]，但这一理论存在众多无法回避的难题。生命权的不可权衡性

是我国刑法的根本立场，生命数量的对比无法改变这一本质。我国刑法通说明确指出，生命是人格的基本要素，具有最高的、不可替代的价值，不能通过数量进行比较与衡量，因此以牺牲他人生命保全自己或多数人生命的行为，不能成立阻却违法的紧急避险[6]。这一立场的背后，是对人格尊严的绝对保护。每个生命都具有独立的、不可复制的价值，多个独立个体的生命价值并不天然高于单一个体的生命价值，反之如此。倘若将个体的生命价值加以量化，将得出“富人的生命价值高于穷人”“年轻人的生命价值高于老年人”的不当结论，最终瓦解“法律面前人人平等”的宪法原则。任何以“保全多数人”“综合考量生存概率”等名义对生命进行量化比较的算法逻辑，本质上是对生命权的功利化计算，与刑法保护人格法益的基本立场相悖。

生命权衡理论无法解决规范体系的逻辑矛盾。如果承认牺牲少数个体的生命保全多数个体生命的行为具有违法阻却性，则意味着无辜第三人对此负有容忍义务，不能对避险行为实施正当防卫。这一结论显然违背了基本的法感情与法理逻辑，遵守交通规定的行人在面对突发的驾驶事故时，该主体有权对车辆实施正当防卫，甚至可以通过造成车内人员伤亡的方式保护自己的生命；根据正当防卫的对象只能是不法侵害，这恰恰反证了此类避险行为的违法性。

修正功利主义的严格限制条件，无法改变其行为本质的不法性。肯定说试图通过“高偶然性、不得已、仅以数量为标准”等条件限制生命紧急避险的适用，但这些条件只能影响责任的认定，无法改变行为本身的不法属性。即便碰撞场景的发生概率极低，也不能让预设“牺牲无辜者生命”算法的行为合法化；即便避险行为是“不得已”的唯一选择，也只能说明行为人可能因期待可能性缺失阻却责任，而非行为本身合法。

除了功利主义理论，学界还提出了义务冲突理论、社会团结义务理论、被害人推定的承诺等学说，试图为自动驾驶攻击性紧急避险提供正当化依据[4]，但这些学说均存在一定的矛盾之处。义务冲突理论无法为攻击性紧急避险提供正当化基础。该说主张算法设计者既负有保护车内驾驶员安全的保证人义务，又负有不得侵害无辜第三人生命的不作为义务，二者形成了义务冲突，履行其中一项义务即可阻却违法。但该理论混淆了作为义务与不作为义务的位阶关系。当二者发生冲突时，禁止侵害他人的义务绝对优先，不存在所谓的义务冲突。更何况攻击性紧急避险中的无辜第三人，原本并未陷入任何危险，是算法的转向行为将其拖入了死亡风险。

社会团结义务理论无法证成牺牲生命的避险行为。社会团结义务是紧急避险制度的重要正当化依据，其核心是社会成员在危急时刻，应当在一定限度内容忍他人为保全重大法益而对自己造成的轻微损害。但社会团结义务明确限定不能延伸至生命权的牺牲，因为生命一旦丧失，便再无机会享受社会团结义务带来的任何利益。因此，要求无辜第三人容忍他人牺牲自己生命的避险行为，完全超出了社会团结义务的合理边界，也违背了该理论的基本逻辑。

被害人推定的承诺理论不具有适用空间。该说主张在“无知之幕”的遮蔽下，理性人事前会同意“牺牲少数人保全多数人”的算法规则，因为这能提高每个人的生存概率。但这一理论存在不足：其一，罗尔斯的“无知之幕”恰恰推导出理性人绝不会同意牺牲自己生命的规则。因为根据“最大最小值”原则，理性人会优先考虑最坏的结果，即自己成为被牺牲的少数人，而生命的丧失是无法弥补的终极损害；其二，我国刑法明确否定“承诺杀人”合法性。即便理性人事前可能做出此种承诺，得到被害人承诺的杀人行为依然构成故意杀人罪，更何况是根本无法证实的“推定的承诺”[9]。

3.3. 期待可能性理论下攻击性紧急避险的出罪

自动驾驶场景中针对无辜第三人生命的攻击性紧急避险，无论基于何种理论，其行为始终符合故意杀人罪的构成要件且具有不法性，不存有任何阻却违法的可能性。这并不意味着所有实施主体必然承担

刑事责任，应通过期待可能性理论判断阻却可能性。

根据期待可能性的核心理论，在紧急状态下无法期待行为人实施适法行为时，即便行为具有违法性，也可阻却责任[1]。在自动驾驶场景中，对于身处车内、生命面临紧迫、现实危险的驾驶员，法律无法期待其放弃自己的生命选择自我牺牲，因此其对避险结果的发生，可能因不具有期待可能性而阻却责任承担或承担避险过当的责任；但对于算法设计者、生产商而言，其在编写算法、预设避险规则时，并未面临任何紧迫的生命危险，具有完全的自由选择空间，法律完全可以期待其不预设牺牲无辜第三人生命的违法算法，因此无法以期待可能性阻却责任。这一区分，既坚守了生命权依法受最高保护的法律底线，又兼顾了紧急状态下行为人期待可能性的客观边界。

4. 自动驾驶紧急避险的构成要件与行为主体认定

自动驾驶紧急避险的刑事责任认定，以构成要件该当性的判断为前提。只有明确算法预设的避险行为是否该当刑法分则的构成要件，以及该行为归属于何种主体，才能进一步展开违法性与有责性的判断。

4.1. 算法预设避险行为的构成要件

自动驾驶算法预设的攻击性紧急避险行为完全符合故意杀人罪、故意伤害罪等罪名的构成要件，不存在构成要件层面的出罪空间[1]。从客观层面看，算法预设的避险行为与法益损害结果之间，具有刑法上的因果关系。一方面，算法的预设决策直接决定了车辆的避险行为，是法益损害结果发生的直接原因；另一方面虽有观点认为，自动驾驶的避险决策是技术计算的结果，并非人类的身体动作，因此不属于刑法上的行为[10]。但该观点忽略实行行为本质是“对法益的危险支配”。算法设计者与生产商通过编写程序、预设规则完全支配车辆避险行为的逻辑与法益损害的发展流程，其预设算法的行为与人类驾驶员在紧急状况下的决策具有规范上的等价性，属于刑法上的实行行为。

预设违法避险算法的主体对于侵害的法益主观上至少具有间接故意。算法设计者、生产商在编写“牺牲无辜第三人生命保全驾驶员”的算法时，明知该算法在极端场景下会造成无辜者的死亡结果，却依然将其植入车辆系统并投入市场，放任死亡结果的发生，完全符合间接故意的构成要件。即便生产商辩称其“不希望死亡结果发生”，也无法否定其对结果发生的可能性具有明确认知，因此成立间接故意而非过失。

从驾驶员的角度出发，对其“明知”的判断应当建立可操作的客观标准，具体可结合以下因素进行综合认定：一是车辆销售方是否通过书面告知、系统弹窗提示等方式，明确向驾驶员告知算法存在生命权衡的避险逻辑；二是驾驶员是否通过车辆说明书、操作培训等渠道知晓避险模式的具体选择规则与法律后果；三是驾驶员是否曾因该类避险算法的适用收到过警示或发生过相关事故；四是驾驶员是否主动开启、选择了优先保护车内人员的避险模式。根据上述标准，若驾驶员明知算法可能在极端场景下造成无辜者死亡结果，依然选择使用自动驾驶系统，即可认定其对结果至少持放任态度，具有间接故意。只有当驾驶员对算法的避险规则完全不知情，且已经尽到一般理性人的注意义务时，才能否定其主观故意。

4.2. 避险行为的主体归属与责任范围界定

运用“行为支配”原则界定自动驾驶紧急避险行为的实施主体对避险行为的支配力，进而明确其责任范围。

在完全自动驾驶场景中，生产商与算法设计者是避险行为的核心主体。该类主体对避险算法的设计、编写与植入具有最终的决定权，完全支配避险行为的实施与结果的发生。即便算法是在车辆行驶过程中自动执行的，也不能改变其事前预设行为的实行行为属性。正如德国学者黎安·沃尔娜所言，自动驾驶

的编程者与传统“电车难题”中的扳道工，在规范本质上并无差异，二者都通过事前的决策，决定了不同主体的生命走向，因此必须对其预设的行为承担刑事责任[8]。但需要区分：算法设计者是生产商的员工，其执行工作任务的法律后果由生产商承担。在刑事责任层面，算法设计者作为直接实行者，需对其编写违法算法的行为承担个人责任；生产商中直接负责的主管人员和其他直接责任人员则根据单位犯罪的规定承担相应责任。

在选择性自动驾驶的场景中，驾驶员与生产商可能同为核心主体。驾驶员并非是算法的编写人员，但其启动与选择使用行为是对避险算法的利用与认可，在规范层面可认定为避险行为的间接正犯或帮助犯。如果驾驶员明知车辆搭载违法算法，依然选择使用自动驾驶系统，则表明其具有利用该算法保护自己生命的故意，与生产商形成了共同犯罪的故意，应当对避险行为造成的结果承担共同犯罪的责任。如果驾驶员对算法的避险规则完全不知情，那么其对避险行为的发生不具有支配力，不具有主观故意，无需对避险结果承担刑事责任。如果生产商为驾驶员提供了避险模式的自主选择权，驾驶员选择相应的避险模式，其就是避险决策的最终决定者，应当对避险结果承担主要责任。

5. 自动驾驶紧急避险的刑事责任分配与归责路径

5.1. 符合法定要件的防御性紧急避险免责

当自动驾驶的避险行为属于针对危险源的防御性紧急避险，且符合我国《刑法》第21条规定的全部要件时，所有主体均无需承担刑事责任。具体而言，需满足以下条件：一为危险正在发生，且由其他交通参与者的违规行为直接引发；二为避险行为是不得已的唯一选择，不存在制动、减速等其他损害更小的替代方案；三为避险行为针对的是危险制造者本身，未损害无辜第三人的法益；四为避险行为没有超过必要限度，造成的损害与所避免的损害未形成显著失衡。需要注意的是，即便在防御性紧急避险中，也禁止为保护轻微的法益而造成危险制造者的生命损害。如果避险行为超过必要限度造成不应有的损害，属于避险过当，相关主体应当承担刑事责任，但应当减轻或者免除处罚。

5.2. 针对无辜第三人的攻击性紧急避险责任分配

作为避险行为核心支配者的生产商与算法设计者预设“牺牲无辜第三人生命”的避险算法，应当承担主要的刑事责任。首先，生产商明知其预设的算法在极端场景下具有造成无辜者死亡的可能仍将该算法运用于实际，该行为构成故意杀人罪的间接正犯，或与驾驶员构成共同犯罪。其次，如果生产商在算法设计中已尽到全部合规义务，未预设任何违法的避险规则，仅因技术局限、算法缺陷等不可预见的原因导致车辆在极端场景下实施不当避险行为，造成无辜者损害的，应当否定其主观故意；仅在存在过失的情况下，认定为过失致人死亡罪、过失致人重伤罪。如果生产商已尽到全部注意义务，损害结果是无法预见、无法避免的，应认定为意外事件，排除刑事责任的承担。最后，如因避险算法存在产品缺陷导致损害结果发生，生产商在主观明知情况下除承担民事产品责任外，还可能构成生产、销售不符合安全标准的产品罪，与故意杀人罪形成想象竞合，应从一重罪论处。

除传统单位犯罪与自然人责任路径外，可引入组织过失责任模式以适配现代企业的运作特征[11]。自动驾驶算法的设计、测试、上线是多部门协作、流程化决策的集体产物，若企业内部存在算法合规审查机制缺失、安全测试流程疏漏、风险防控义务未落实等组织性过错，导致违法避险算法被投入市场，无需追溯具体员工的主观过错，即可直接认定企业的组织过失责任，避免因具体行为人认定困难导致的责任落空，使刑事责任分配更贴合自动驾驶行业的实际运作模式。

驾驶员刑事责任的认定核心在于其对违法避险算法的认知与选择，以及期待可能性的判断。明知违法算法仍使用的情况下，驾驶员依然选择使用自动驾驶系统造成损害结果的，其与生产商构成故意杀人

罪的共同犯罪。但在责任层面，驾驶员在碰撞发生时，自身生命面临紧迫、现实的危险，依据期待可能性理论，法律无法苛求其放弃自身生命利益，在此情形下可阻却其责任，最终不予刑事处罚。需要严格限定的是，期待可能性的适用仅限于驾驶员自身生命面临现实危险的场景，不能泛化适用于财产损失、轻微身体伤害的避险场景。对于违法算法不知情的情况下，驾驶员仅正常使用自动驾驶功能，其对避险结果的发生既无故意也无过失，无需承担任何刑事责任。但如果生产商为驾驶员提供了避险模式的选择权，驾驶员主动选择优先保护车内人员生命利益的模式，则其是避险决策的最终决定者。即便存在期待可能性，也不能完全阻却责任，只能作为量刑情节予以从轻、减轻处罚。

6. 自动驾驶紧急避险刑事规制的完善

自动驾驶技术的发展，不能突破刑法的底线与公共安全的红线。刑法对自动驾驶紧急避险的规制，应在自然人紧急避险分析的基础上，从价值底线、归责体系、配套制度三个层面进行补充完善，以求实现技术创新与法益保护的平衡。

6.1. 明确生命权不可衡量的算法设计规则标准

通过立法、司法解释或行业规范，明确自动驾驶算法设计的三大禁止性规则。第一，绝对禁止预设针对无辜第三人生命的攻击性紧急避险算法，禁止以任何形式对生命权的量化；第二，绝对禁止基于年龄、性别、身体状况、社会地位等个体特征，对不同主体的生命价值进行差异化排序；第三，绝对禁止“驾驶员绝对优先”的算法设计，不得将车内驾驶员的生命安全置于车外无辜者的生命安全之上的优先地位。

明确自动驾驶算法设计的基本准则。在极端碰撞场景中，算法的唯一合法决策是采取紧急制动、减速等措施，最大限度降低碰撞造成的整体损害，不得通过转向等行为，将碰撞风险转嫁给无辜的第三人。即便转向可能保全更多的生命，也不得预设此类决策，绝不允许以牺牲无辜者生命为代价的“利益最大化”。

6.2. 明确分层化主体的责任边界

确立“生产商首要责任”原则。生产商是自动驾驶技术的开发者与算法的最终决定者，应当对算法的合法性承担首要责任。通过立法明确生产商负有算法合规审查义务、规范生产流程。在车辆投入市场前，必须对避险算法的合法性进行全面审查，确保其符合刑法与行业规范的要求。在车辆投入市场后，若政策明确相关的设计要求，生产商应无条件提供技术支持确保已售出车辆算法运行的合法性。若因生产商预设避险算法的违法性致使驾驶前端发生相应刑事案件，无论极端场景的发生概率高低，都应依法追究其刑事责任。

严格限定驾驶员责任的成立范围。驾驶员刑事责任的承担仅限于其明知算法违法仍主动使用，或自主选择违法避险模式的情形。对于普通驾驶员在正常使用自动驾驶功能时，因系统预设的违法算法造成的损害结果，应坚持无过错则无责任^[1]。同时应严格限制期待可能性理论的使用场景。仅在驾驶员自身生命面临紧迫危险的极端场景下，才能通过期待可能性阻却其责任，防止该理论的滥用。

明确自动驾驶分级的责任承担差异。对于 L2、L3 级的辅助自动驾驶，人类驾驶员负有实时监控车辆、随时接管的义务，若驾驶员有能力接管却未及时接管，导致不当避险行为发生的，驾驶员应当承担主要责任；对于 L4、L5 级的完全自动驾驶，驾驶员无接管义务，在此情形下避险行为的责任主要由生产商承担^[12]。

6.3. 技术与法律的协同治理

建立强制性的算法合规审查与备案制度。通过立法明确规定自动驾驶车辆的避险算法必须经过国家

指定机构的合规审查,审查通过后方可投入市场;同时,生产商需将避险算法的核心逻辑向监管部门备案,确保算法的可追溯、可审查。对于未通过合规审查、未备案的算法,禁止投入市场使用,从源头杜绝违法避险算法的出现[4]。

完善自动驾驶“黑匣子”数据记录制度。强制要求所有自动驾驶车辆安装符合国家标准的事态数据记录系统,完整记录车辆行驶过程中的环境感知、算法决策、车辆控制等全流程数据,且数据不可篡改、不可删除。该制度不仅能为交通事故的责任认定提供完整的证据支撑,也能为刑事责任的追究提供关键的证据保障。

推动自动驾驶紧急避险的行业标准制定。由行业主管部门牵头,联合法学界、汽车行业、人工智能领域的专家,制定自动驾驶紧急避险算法的行业标准,明确算法设计的合法边界、技术要求、测试规范,为生产商的算法设计提供明确的行为指引,也为司法机关的责任认定提供参考依据。

7. 结论

自动驾驶技术的发展对传统刑法中的紧急避险制度产生冲击,但并未动摇刑法的根本价值与核心原则。人的生命权始终是最高位阶、不可权衡、不可侵犯的根本法益,这是刑法对自动驾驶紧急避险进行规制的不可突破的底线。自动驾驶场景中的防御性紧急避险,因危险制造者的法益要保护性降低,在符合法定要件时可阻却违法;而针对无辜第三人生命的攻击性紧急避险,无论基于功利主义还是其他理论,都无法获得刑法上的正当化依据,其行为始终具有违法性。算法预设的违法避险行为,在构成要件层面符合故意杀人罪等罪名的构成要件,其刑事责任需根据各主体的行为支配力与主观过错进行分配。对自动驾驶技术的规制既不能因技术创新而突破法益保护的底线,也不能因过度保守而抑制技术的发展。坚守生命权依法受最高保护的核心原则,通过核心概念的精准解释构建清晰的责任边界,以完善的配套制度实现技术与法律的协同,才能让自动驾驶技术在法治的轨道上健康发展,最终实现技术与公共安全的协调统一。

参考文献

- [1] 魏超. 自动驾驶汽车“电车难题”伦理困境的刑法解答[J]. 政治与法律, 2025(6): 130-146.
- [2] 王钢. 对生命的紧急避险新论——生命数量权衡之否定[J]. 政治与法律, 2016(10): 95-108.
- [3] 陈璇. 紧急权: 体系建构、竞合适用与层级划分[J]. 中外法学, 2021, 33(1): 5-31.
- [4] 储陈城. 自动汽车程序设计中解决“电车难题”的刑法正当性[J]. 环球法律评论, 2018, 40(3): 82-99.
- [5] 张明楷. 刑法学(上)[M]. 第6版. 北京: 法律出版社, 2021: 267-279.
- [6] 高铭喧, 马克昌. 刑法学[M]. 第10版. 北京: 北京大学出版社, 2022: 178-185.
- [7] 张轩铭. 自动驾驶汽车应急算法的刑事风险与应对——基于有限自治模式的展开[J]. 华南理工大学学报(社会科学版), 2025, 27(5): 109-119.
- [8] 黎安·沃尔娜, 王德政. 自动驾驶汽车编程者的刑事责任——以规定参数进行紧急避险的角度[J]. 上海师范大学学报(哲学社会科学版), 2019, 48(6): 103-111.
- [9] Bonnefon, J.F., Shariff, A. and Rahwan, I. (2016) The Social Dilemma of Autonomous Vehicles. *Science*, 352, 1573-1576. <https://doi.org/10.1126/science.aaf2654>
- [10] 弗兰克·彼得·舒斯特, 李倩. 自动驾驶下生产商的刑事责任[J]. 苏州大学学报(法学版), 2021, 8(3): 41-50.
- [11] 耿佳宁. 单位固有刑事责任的提倡及其教义学形塑[J]. 中外法学, 2020, 32(6): 1489-1508.
- [12] 托马斯·魏根特, 樊文. 自动驾驶汽车的紧急避险权[J]. 中国应用法学, 2019(4): 174-185.