《学术界》(月刊)
总第276期,2021.5
ACADEMICS
No.5 May. 2021

# 自动驾驶汽车涉罪的归责难题及其解决方案〔*〕

姜　涛,刁永超
(南京师范大学　法学院,江苏　南京　210023)

〔**摘　要**〕自动驾驶汽车涉罪后会产生犯罪主体认定困难以及行为定性难题,结果归属问题也随之产生。在现有的刑法理论范式下,自动驾驶汽车涉罪之客观构成要件判定不清,是导致刑事归责认定困难之关键原因。对犯罪主体认定的争议不如转向更具现实处罚依据的结果归属之上,传统过失犯实行行为理论与因果关系理论存在不合时宜的一面。客观归责论可以弥补上述缺陷,即通过规范归责的范式将风险类型判定作为逻辑前提,借助正反向规则对不符合客观构成要件之事项进行双重限定排除。客观归责论所提供的范式从行为归责到结果归属,使个罪的客观构成要件更为清晰,也使上述难题能够得到有效解决。其中,以法律不容许的风险与法律容许的风险,合理区分生产者、研发者的责任,是客观归责论运用于自动驾驶汽车领域的关键所在。

〔**关键词**〕自动驾驶汽车;客观归责论;过失犯罪;人工智能;客观构成要件
DOI:10.3969/j.issn.1002-1698.2021.05.015

科技步伐的大幅跨越,已经将人们从互联网的"网络化"时代,推向"人工智能"时代。[1]人工智能正在以新时代科技的形象深刻地改变着人类的工作方式、生活方式和社会秩序,在促进社会生产力整体跃升和社会治理崭新变革的同时,也衍生出了诸多科技成果。[2]得益于这一领域的技术突破与高速发展,无人驾驶技术从天方夜谭变得触手可及,自动驾驶系统从弱人工智能向强人工智能的过渡,必然会推动着无人驾驶技术在未来的实现。因此,自动驾驶汽车无疑成为了近年来最前沿的科技热点。诚然,当科技应用于现实之时,自动驾驶汽车的确可以在不同程度上解决现如今比较严重的社会问题。比如,普通自动驾驶中的"定速巡航"系统,[3]可以有效地减少能源上的过度浪费、减轻司机疲劳感等,进而降低交通事故的发生率;又如,更高级别的无人驾驶汽车,其允许驾驶员执行驾驶以外的任务,在当今以生产力为导向的社会中更是吸引眼球,[4]因此,法律完全禁止自动驾驶汽车并不是一个可接受的解决方案。与此同时,自动驾驶汽车也会带来法律适用方面的难题。上述难题在刑法领域具体体现为,对于自动驾驶汽车涉罪时该如何

**作者简介**:姜涛,南京师范大学法学院教授,中国法治现代化研究院研究员,江苏高校区域法治协同创新中心研究员;刁永超,南京师范大学法学院研究生。

〔*〕本文系国家社会科学基金重点项目"刑法体系的合宪性控制研究"(17AFX018)的阶段性成果。

归责？传统刑法理论在解决自动驾驶汽车涉罪的归责问题时是否存在困境？造成这一困境的原因是什么？刑法理论又应该如何突破这一困境？这都是刑法教义学面临的新课题，正处于尚未解决的危机之中。本文拟对这些悬而未决的问题尝试解答，以求教于各位专家学者。

## 一、自动驾驶汽车涉罪的归责难题及思考论域的设定

自动驾驶汽车涉罪的归责作为一个全新命题，现有刑法教义学对此的讨论才刚刚开始，并没有达成共识。这一研究的前提是明确自动驾驶汽车涉罪的归责难题，并设定讨论的论域，以免刑法教义学的思考成为一种“无底盘的游戏”。

### （一）自动驾驶汽车涉罪的归责难题

20 世纪 80 年代以来，人工智能控制理论与技术在交通运输工程中的应用越来越广泛，[5]而无人驾驶汽车事故的风险不会是零。国外学者曾从理论上模拟讨论过自动驾驶汽车面临的五大困境，包括突然刹车失灵、汽车为了避免重大事故转向硬路肩等，若遇到乘客与行人的生命权冲突时，源于功利主义编程的解决方案，自动驾驶汽车必须编程为过马路，杀死行人，因为只有这种编程方式，乘客会保持安全，但这并不意味着杀人是合法的。[6]而现实中的案例是，2018 年 3 月，优步（Uber）技术公司的一辆自动驾驶汽车在行驶的过程中肇事，造成了全球首例致人死亡的事故。[7]这一事件给自动驾驶技术的发展当头棒喝，并将其推向了舆论的风口浪尖；反观国内，近年来，我国紧抓人工智能发展的重大战略机遇，于 2017 年发布了系列发展规划，[8]大力发展人工智能车联网产业。同时，百度、滴滴等相继推出自动驾驶汽车并上路实验。国家的政策支持加之科技的导向意味着，未来自动驾驶技术有望在我国飞速发展。然而，技术发展带来的问题也需要正视，刑法理论无法回避自动驾驶汽车上路肇事的问题。2016 年，京港澳高速河北邯郸段发生追尾事故，处于自动驾驶模式的特斯拉轿车当场损坏，作为自然人的驾驶员不幸身亡。2017 年在智能汽车设计大赛中，某代表队的无人驾驶车辆闪躲行人后，偏离主路冲过绿化带和非机动车道，最终撞上人行道上的树木。

以上案例已经暴露了自动驾驶汽车的安全隐患，这背后蕴藏着大量的法律规制难题，尤其重要的是，基于自动驾驶行为所产生的交通肇事行为的归责问题也无法避免。自动驾驶行为在发生危害结果之时，此类犯罪与传统刑法上的主体、行为及危害结果等构成要件内容会大有不同，比如，驾驶者与“肇事者”之间的分离、驾驶行为与“肇事行为”的分离等，从而导致刑法上的归责难题。再如，是否需要基于对于结果发生较为遥远的“条件”或者离奇的因果流程而让生产者或研发者为此负责，如果科学设想上与事实上的因果流程彼此之间有重大偏离的话，是否欠缺可归责性。诚然，导致自动驾驶汽车发生重大交通事故的肇事行为，可能存在故意或者过失两种罪过形态。在故意犯罪的情形下，无论强人工智能支持下的自动驾驶系统拥有多么高度的自主决定能力，其作为人类犯罪工具的本质地位始终无法被松动，并且在故意犯罪中，事实归因与规范归责两者相对较为清晰，即可以结合实行行为的性质，按照现行刑法规定的相关故意犯罪来认定行为人的刑事责任即可。

自动驾驶汽车真正的归责困境集中于过失犯罪的场合，正如有学者所指出，在人工智能时代，刑法适用中成为问题的是过失犯领域，往往会出现归责间隙。[9]与自动驾驶汽车肇事联系最为密切的罪名莫过于交通肇事罪，当传统汽车变为自动驾驶汽车后，中间穿插产品研发者与生产者责任的判断，导致归责主体、归责法理等发生变化。首先，根据《刑法》第 133 条之规定，[10]交通肇事罪的行为主体是一般主体，行为主体和责任主体应具有同一性。其次，本罪的实行行为是违反交通管理法规的行为，这里的实行行为应作具体认定。最后，严重的法益侵害结果要与违规

行为具有法律上的因果关联,即严重后果必须是由违章行为引起的,涉及应该判定谁的行为违章。本来,个罪的构成要件是清晰界定罪与非罪、此罪与彼罪的规范标准,当把交通肇事的案件事实与刑法有关交通肇事罪的规定对接起来的时候,个罪的构成要件提供了实现这种无缝对接的理论根据,有了这一理论根据的引导,司法裁判才不至于在具体操作上出现偏误。然而,当我们把自动驾驶汽车肇事的案件事实与交通肇事罪的构成要件关联起来时,就会发现对自动驾驶汽车肇事按照交通肇事罪定罪处罚存在一定的困难:(1)判断自动驾驶汽车涉罪的行为主体究竟是自动驾驶系统,还是作为自然人的驾驶员,行为主体与责任主体是否已经发生了分离。(2)何种行为是为法规范所不允许的实行行为。(3)在规范上判断危害结果应该归于何者。上述难题意味着当以交通肇事罪的构成要件涵摄自动驾驶汽车肇事之时,本来清晰的、明确的个罪之构成要件变得不够清晰了,由此导致归责难题。笔者认为,为何会带来这种难题,是因为自动驾驶系统的介入,使得作为自然人的驾驶员的驾驶角色逐渐消解,反之,自动驾驶汽车以及其幕后相关主体的影响日益强化,责任的归属或许已经发生转移,导致个罪之客观构成要件认定上的不清楚。

问题又在于,无论装载自动驾驶系统的汽车肇事后行为主体以及行为类型有多么特殊,其对公共安全这一法益造成的侵害是不可否认的,并且责任追究的强化,有利于将社会损害最小化。有鉴于此,理清自动驾驶汽车肇事后涉罪的客观构成要件,明确其归责法理、归责主体、归责范围、归责程度等,就是刑法学无法回避又亟待解决的时代难题。

(二)思考论域的设定

诚然,以上问题的解决会有不同逻辑主线,但如果在概念上任意发挥,便会形成无数岔路,进而造成业内自说自话或移花接木式“不同频”的虚假繁荣现象。为避免这一现象,本文根据自动驾驶系统的等级,将自动驾驶汽车分为三种类别,以期在确定讨论范围的前提下,展开对客观构成要件难题之阐述。

自动驾驶汽车作为人工智能领域的分支,必然具备人工智能的一切要件。如果说,人工智能是模拟和扩展人的智能的一门技术科学,根据智能机器人的“智能”程度,可以将其分为普通智能机器人、弱智能机器人和强智能机器人。[11]那么,我们也可以根据自动驾驶汽车的自动驾驶系统的“智能”程度,将其分为辅助驾驶型、半自动驾驶型及无人驾驶型三类。(1)“辅助驾驶型系统”。加减档变速、方向控制、启动与刹车等可以选择由自动驾驶系统自动控制,其余操作完全由人类驾驶员正常操控。作为普通人工智能发展阶段的产物,本质上人类驾驶员仍然对车辆的驾驶拥有天然的控制权。(2)“半自动驾驶型系统”。加减档变速、方向控制、启动与刹车等在一定情况下由自动驾驶系统操控,且由系统和驾驶员共同对驾驶环境进行监测,出现识别不能等情况时,系统发出请求信号,转由人类驾驶员接管驾驶。配备该类型自动驾驶系统的汽车在行驶的过程中,人类驾驶员退居二线,不必时刻参与驾驶,但是仍需要人类驾驶员保持警惕戒备状态随时接管车辆。(3)“无人驾驶型系统”。在一定路段甚至是全路段,所有操作全部由自动驾驶系统独立完成。全自动化的智能驾驶系统,解放了人类驾驶员的双手。以上可以分别对应国际自动机工程师学会(SAE)制定的自动驾驶汽车的分类标准,详见下页表1。

以上针对自动驾驶系统的分类,贴合其母体人工智能发展阶段的同时,又与国际自动驾驶汽车分类相吻合,更重要的是为自动驾驶汽车涉罪的客观构成要件难题的解决设定了统一的路标,使得不同的思维在同一概念下迸发出多彩的火光,从而消解自说自话、自我定义式的随意批判。将自动驾驶汽车类型化后,会得出辅助驾驶型和半自动驾驶型自动驾驶汽车在驾驶操作和对驾驶环境的监管方面仍然无法脱离人类驾驶员的驾驶

表 1　自动驾驶汽车的分类标准

| 对应特征 / 三分法 | SAE 等级 | 含义 | 人工智能水平 | 驾驶操作 | 驾驶环境监管 |
|---|---|---|---|---|---|
| 辅助驾驶型系统 | L1/L2 | 自动驾驶系统完成部分指定驾驶操作，其余由驾驶员完成 | 普通智能机器人阶段 | 驾驶员/系统 | 驾驶员 |
| 半自动驾驶型系统 | L3 | 自动驾驶系统单独完成驾驶操作，根据系统请求，驾驶员随时准备接管车辆 | 弱人工智能阶段 | 驾驶员/系统 | 驾驶员/系统 |
| 无人驾驶型系统 | L4/L5 | 自动驾驶系统单独完成所有驾驶操作，驾驶员驾驶地位虚化 | 强人工智能阶段 | 系统 | 系统 |

行为的结论,故在这两类自动驾驶汽车肇事后客观构成要件与传统交通肇事并无实质差别,下文将不作探讨。笔者将自动驾驶汽车框定在强人工智能阶段的无人驾驶领域,进而对自动驾驶汽车涉罪归责难题的成因进行阐明,并提出解决方案。

## 二、归责难题的成因:自动驾驶肇事涉罪之构成要件模糊

只有存在避免所禁止的结果或实现所要求的结果的可能性时,才能谈及行为的违法性及可非难性。[12]与传统交通肇事相比,自动驾驶肇事更多的是“技术故障”所导致的,驾驶者感官可知的结果并不属于其构成要件,生产者、研发者的过失才与驾驶者感官可知的结果具有关联。刑法对这种技术故障造成的法益侵害不得不管,但又面临着归责难题,因而对刑法教义学提出了全新挑战。归纳而言,造成自动驾驶肇事涉罪之归责难题的原因有二:一是自动驾驶肇事涉罪之构成要件主体及行为分离错位;二是自动驾驶肇事涉罪之因果关系模糊不清。

### (一)自动驾驶肇事涉罪之构成要件主体及行为分离错位

虽然强人工智能阶段下的自动驾驶汽车可以在一定路段甚至是全路段实现完全的自我驾驶,但自动驾驶汽车本身并不具有刑事主体地位。首先,强人工智能模仿了人类物种适应环境的一个基本特征,但是其仍然不具有法律人格。在刑法的语境中,人格与罪责是密切相关的,只有能够明辨是非、有意志自由的人,才会因选择做错事而受到归责。毕竟,归责的先决条件就是要有理解每一个选择所产生的后果的能力和自由选择的能力。其次,在某种程度上讲,人格与我们为自己设定目标并追求的能力有关,但目前来看,人工智能的这方面能力还是极其有限的,虽然它们可能拥有拓展和设置独立小目标进而达到总体目标的能力,但是这个更大的目标仍旧是由人类程序员或者是用户所设定的。[13]例如,自动驾驶系统可以在现场对交通作出判断与决定,但安全导航想要达到目的地的这个总体目标却是预先设计好的。最后,现代刑法及其所有进步的发展似乎都是人类为人类而设计的,只有人类的行为拥有构成刑事责任基础的这种能力。故而,应“刺破人工智能面纱”,[14]否定强人工智能的刑事主体地位,将行为主体与责任主体进行剥离,分别予以认定。

归责的核心在于判断一个事件能否被看作特定人的“犯行”,这种判断在不同案件、不同时代的标准并不相同。在传统汽车时代,当车辆发生交通事故致使法益受到侵害时,可以追责的主体无非是交通运输人员,其中,最为主要的便是机动车驾驶人员(人类)。但是,在强人工智能时代,因自动驾驶系统的介入使得人类驾驶员的驾驶地位逐步被架空,直至完全虚化,驾驶人的身份便无从认定。即使具备传统驾驶员的全部要件,自动驾驶汽车的使用者也不能被认定为传统意义上的驾驶人员,因为此时的角色定位已经由最初的“司机”变为了“乘客”。因此,自动驾驶汽车相关驾驶主体可以分为三类:(1)驾驶使用

者,即在自动驾驶的场景下,作为自然人的驾驶员相当于汽车的使用者。(2)驾驶行为者,即自动驾驶系统作为汽车的决策中枢,相当于操控汽车的行为者。(3)驾驶责任者,即自动驾驶系统背后的研发者与生产者。然而,当现实驾驶者并非作为自然人的驾驶员之时,要求自动驾驶汽车的使用者为自动驾驶系统的肇事行为负责,显然违背了现代刑法当中的责任主义原则。而要追究行为者的责任,要求自动驾驶汽车"自负其责",不仅于法无据,而且会动摇刑法甚至整个法律体系的归责基础。

问题的关键在于,否定自动驾驶系统的刑事主体地位,交通肇事罪的责任可能会转移到该系统背后的设计者、生产者以及销售者等主体之上,但这显然超出了交通肇事罪之构成要件的涵摄范围,即交通肇事罪的行为主体与自动驾驶汽车的潜在责任主体之间存在明显的错位。质言之,既然实施伤害"行为"的自动驾驶汽车既不具有刑事责任能力,又不具有刑罚适应能力,那么就不能将法益侵害结果看作是直接造成该结果的主体之作品,而需要其他主体为此担责。由此形成了行为主体与责任主体分裂的全新模式,即出现"直接损害主体不负责、其背后的设计者和生产者负责"的情况。[15]

涉罪主体的不清晰也决定了主体行为之不明朗。自动驾驶汽车上路肇事后,各个主体行为在何种程度上的偏离才属构成要件行为,需要在主体分离问题解决后,分别予以认定。也正是因为自动驾驶汽车所涉主体及行为的复杂性、分离性,如何在相关主体之中正确认定与合理划分归责对象,就成为了刑法理论与司法实践亟待解决的难题。

(二)自动驾驶肇事涉罪之因果关系模糊不清

如果行为主体与责任主体相互分离只是归责难题的外在表现,那法益侵害结果的归属模糊便直击问题的本质,[16]并在因果关系上得以集中体现。

在现行法律体系下,自动驾驶汽车根本无法对自己的行为给第三者造成的损害承担刑事责任,即依照现行刑法,所造成的法益侵害结果没有与之直接对应的追责对象。因此,就法律规范层面而言,只要自动驾驶汽车本身不能被归责,受害者和社会就会面临一种责任真空的状态。[17]但从规范社会、保护人民切身利益的需要出发,又难以将这种法益侵害行为视为意外事件而放任不顾,故归责对象亟需明晰。换言之,当自动驾驶汽车在既定的行驶程序下,对路况进行自主判断并在驾驶过程中造成无过错第三方的损害时,将法益侵害结果这笔"账"归属到何人的头上便成为了难题,问题之本在于清晰结果归属的判断模式,进而将行为主体与责任主体分离之问题予以"黏合"。

诚然,结果归属的判断模式必然囊括事实判断与价值评价两个维度。事实判断也即存在论层面的条件关系,价值评价即价值论层面的规范归责。一般而言,事实判断和价值评价整体上是两级位阶的关系,跃过经验层面的事实因果关系的位阶,直接跨入二级规范层面的价值评价位阶只会让规范成为空中楼阁,缺乏具体事实支撑,而无处适用;反之,只踏上第一层,而不进阶运用规范判断的范式,那也仅仅只是生活判断而非法规范判断。作为必要的重复,在现有的刑法规范体系下,自动驾驶汽车肇事致使无过错的第三者死亡的,这种过失最直接的便是与交通肇事罪相联系,可能被归责的主体有三类,以下分别考虑结果归属之问题。第一,讨论驾驶使用者,即作为自然人的驾驶员之责任。驾驶员丧失驾驶主体地位后,已经退出了交通肇事罪的刑事归责领域,即其因作为义务的阙如而不在刑法的涵摄范围之内,作为自然人的驾驶员本身没有实行行为,又何论因果关系及责任之有无。第二,讨论驾驶行为者,即自动驾驶系统之责任。考虑作为直接行为主体的自动驾驶系统是否应被归责并进行二分判断。在事实层面,如果没有自动驾驶汽车实施侵害行为,被害人法益便不会受到侵害,完全符合"无 P 则无 Q"的条件关系公式,但

是在规范层面,违章行为与法益侵害结果之间的“引起与被引起”关系却发生了偏移,正如前文所述,作为原因的违章行为无法归责于一个毫无人格可言的自动驾驶系统,故作为直接肇事主体的自动驾驶汽车无法成为结果归属下的责任主体。第三,讨论驾驶责任者,即自动驾驶系统的研发者及生产者之责任。反思自动驾驶汽车背后相关主体行为的原因性时会出现“算法黑箱”问题。

技术层面的“算法黑箱”已然改变了事实因果关系模型。自动驾驶汽车自主行驶的过程中,智能操作系统实时通过物联网大数据进行交通规则辨别及数据采集,以达到对车辆的控制。而自动驾驶系统从数据采集到数据重组,再到自我决策指令的输出,这一过程本身就是一个无法准确知其明细的技术性难题,也即从数据的“输入端”到决策的“输出端”,这一中间过程是如何具体运作的,就当前的技术水平而言仍然无法作出合理回应。自动驾驶技术领域“算法黑箱”的存在,使得作为自然人的驾驶员、受害人无法得知以及难以理解自动驾驶技术的内部运营逻辑或原理。[18]“算法黑箱”反映到刑法教义学层面,则体现为从过失的实行行为到法益侵害结果的发生存在一个未知的发展流程。由此带来的问题是,既然存在论层面的事实基础都未可知,那么作为桥梁的规范层面因果关系又如何去建立呢?事实上,技术层面的“算法黑箱”将传统的因果关系模型所对应的现实情况进行了重构,并且以上实行行为与构成要件结果之间因果链条走向的不清晰,也决定了结果应算作哪一主体之“作品”并不明确,同时,还包含着因为效用而容忍制造风险的情况,毕竟,就自动驾驶汽车而言,作为一项新技术,而具有社会必要性或社会有益性,制造风险在一定限度内是被容忍的,只有创设一个超过允许范围的风险——特别是因为没有遵守技术等规制并且因此发生了构成要件的结果时,这个结果才是可归责的。[19]这一无法绕过的技术难题,也为自动驾驶汽车涉罪的归责问题竖起了一座壁垒,带来归责难题。

总之,在自动驾驶场景下,从行为主体与责任主体的分离到结果归属的不明晰,都彰示着交通肇事罪之客观构成要件在自动驾驶情况下的理顺,已经成为了交通肇事罪归责的关键难题,亟需更有解释力的理论去解决。

## 三、传统理论的困境与客观归责论的提出

目光往返于事实与理论之间,追求两者的切合性,在科学面向与社会生活之间构建一种平衡才是解决问题的正确之钥。为此,本文在自动驾驶场景下着重对传统归责模式进行检思,并尝试将客观归责论架构予以基本介绍。

### (一)思维魔方:传统过失犯理论行为模式之检思

当传统交通领域幻化为自动驾驶场景之时,传统以注意义务为核心的过失犯归责模式将面临多维冲击,既有根植于过失理论框架的原生弊病,也有新兴科技所引发的规范适用障碍。突破之路选择之前,有必要对传统理论之于新事物无奈原因进行检思。对此,笔者将结果预见义务与结果回避义务作为检思轨线,结合自动驾驶场景展开如下讨论。

#### 1.“结果预见义务”为导向的过失行为模式

“注意义务是过失犯罪的核心要素”,[20]注意义务的界定是判断过失犯实行行为中最关键的一环,不同理论学说的观点不尽相同。其中,旧过失论及其修正理论就是从结果无价值论和心理责任论出发的一种以结果预见义务为导向的过失行为模式。“如果注意就能预见到结果,实际却没有预见而实施了行为让结果发生,因此受到非难。”[21]可见,预见可能性成为了旧过失论归责的考量工具,也成为了注意义务内容的规范性要素。然而,当人们逐步踏入风险社会之时,这种侧重于预见义务的归责逻辑愈发暴露出其原生理论不合时宜的一面。尤其是在医疗领域与交通领域,若注意义务的规范性要素仅仅为预见义务的违反,则带来的结果可能是归责范围的极度扩张。旧过失论过度重视预见义务,而忽

视行为层面,这几乎与出现了结果就要行为人负责的结果责任别无二致。面对以上批判,旧过失论尽管也进行了一定的修正,但是修正的理论依旧将过失实行行为的客观性判断依附于主观色彩强烈的预见可能性的判断之上,偏离了结果无价值的基本立场,不免陷入了自我构建的逻辑陷阱。如此一来,如若出现自动驾驶汽车肇事致人死亡,判断注意义务的违反便会从主观出发认为是其背后的研发生产人员应当预见这一系统操作失误却没有预见,而将结果责任直接安放在生产者及研发者之上,这使得注意义务丧失了判断实行行为性的价值,即使修正后的理论用具有法益侵害危险性的行为框定过失犯的实行行为性,但过失行为人行为的法益侵害性并非在明确犯意支配下而为,将故意犯的实行行为判定理论加持在过失犯之上,过失犯实行行为的价值又怎能得到实现?因此,以"结果预见义务"为导向的归责模式忽视了实行行为性,不当扩大作为自然人的驾驶员及相关技术人员的刑事归责风险,导致处罚范围的扩大化。问题在于,如果法律成为科技进步的绊脚石,将技术创新的活力抑制在萌芽之中,那么这对于正努力打破技术堡垒且追求高质量发展的中国无疑是不利的。

2."结果回避义务"为导向的过失行为模式

原生困境影响着旧过失论及其修正学说的运用,学界将眼光投向以"结果回避义务"为导向的过失行为模式。最典型的就是新过失论,"即便具有预见可能性,但倘若履行了结果回避义务,则不成立过失犯。"[22]由此,新过失论在预见可能性的基础之上增补了结果回避义务作为对结果预见义务的修正和限缩。[23]依循结果回避义务的逻辑演绎,在法益侵害结果发生之时,行为人应当预见而没有预见尚不成立过失,还需确定行为人是否采取适当措施回避结果的发生才能认定过失不法并加以非难。[24]不可否认,新旧过失论归责模式的调整符合了主观要素进军不法阶层的趋势。结果回避义务这一标准的增加,在一定程度上也对结果预见义务带来的过失犯罪的泛化进行了限制。然而,结果回避义务这一标准本身却流于模糊,究竟如何判断也成为该理论的原生弊病,并且与自动驾驶导致的交通肇事罪的认定不吻合。在自动驾驶场景下,因多方面主体的涉入及"算法黑箱"的存在导致结果回避义务的范围不明,此种情形下如何判断是否完全尽到了回避义务更是难上加难。自动驾驶汽车背后的设计者和生产者若在目前的科技水平之下尽到了合理的注意义务,即为履行了回避义务,但按照此种思路亦可推导出干脆一开始就不去生产研发,便可以避免自动驾驶汽车驾驶过程中所有偶然危险的发生的一种极端荒谬的结论。质言之,当过失的实行行为判断依赖于结果回避义务之时,就表明过失犯不法的确定很大程度上取决于行为人之社会行为的规范判断,这一标准的确定使得违法不仅受制于法益侵害结果的大小,还需要对行为进行社会规范上的再三思虑,然而行为所依凭的规范界限不明成为了此种归责模式的原生弊端。

在我国,以结果预见可能性为中心的过失犯理论一直占据通说地位,而后才出现结果预见可能性与结果回避义务两说对立之局面。[25]而注意义务的违反却是传统过失犯实行行为理论。在这种情况下,注意义务的规范性要素则由以上两说构成。如此便很难理清其与因果关系在判断上的差异,经常出现两者混同相互证明的逻辑演绎。可以说,传统的过失不法构成仍然需要判断注意义务违反与法益侵害结果之间的因果关系。而在传统因果关系理论的判断中却反复运用注意预见义务,不免在逻辑上陷入以行为违反注意义务而始,又以相当因果关系中的预见义务连接结果而终的循环论证怪圈,只会让过失犯不法构成要件的判定更为困难。以注意义务为核心的传统过失犯理论虽然能够在学术上基本阐明过失不法的构成要件,但其自身理论的如上界定争议,易与因果关系认定相互混淆而模糊了过失的不法要件,使其难以在自动驾驶汽车涉罪所代表的复杂过失犯罪有所运用。因此,针对自动

驾驶汽车涉罪的客观构成要件的诸多争议点,且基于以上传统归责理论的弊端,如果仍然援用以“结果预见可能性”作为逻辑起点的传统归责模式进行探讨,无异于一种“刻舟求剑”的固执做法。

(二)思维转换:传统因果关系理论存在不足

在传统过失犯不法构成中,除实行行为的判定外,无法绕过连接法益侵害结果的因果关系。这一桥梁不仅将行为与结果相连接,而且也是判定责任归属的唯一路径。因而理顺过失犯罪的因果关系是解决过失犯罪归责问题的第二个关键。而传统的因果关系理论主要有条件说、原因说、相当因果关系说以及普通法系的双层次因果关系理论等。其中,条件说与原因说属于传统因果关系单一论,而相当因果关系说与双层次因果关系理论则属于因果关系二元论。

1. 传统因果关系单一论

早期刑法受到自然行为论的影响,将纯粹的、物理的、存在的因果流程,不加任何修饰地直接套用到刑法上的因果关系判断当中,称为事实因果关系。[26]然而,实然层面的事实因果关系的判断是经验性的,因而存在如下困境:其一,在自然意义上无法得出实行行为与危害结果之间的关联时,则无法肯定刑事上的因果关系。例如,自动驾驶汽车因程序设计有误,其背后的生产者没有及时召回所售汽车而发生交通事故。通过经验上的判断无法肯定这种不作为与损害结果之间的因果关系,因而无法进行刑法上的归责认定。其二,根据经验的判断,具有因果关联的行为与结果,刑法在某些情况下会难以排除其因果关联性。[27]这会使得犯罪圈不当扩大。例如,产品的设计者在现有的技术水平下已经对自动驾驶系统的设计尽到了合理的注意义务,投入使用后却因算法错乱导致法益侵害结果。此时,设计者与事故的发生是具有经验上的因果关联的,但是将危害结果归责于设计者显然使得刑法的规制范围无限扩大,违背社会一般民众的法感情。因此,条件说等因果关系单一论学说呈现出“有归因而无归责”的一元格局。然而,因果关系之判断乃是实然与应然的统一、行为事实与规范判断的融合。在新康德主义的视域下,实然未必能当然地推导出应然,两者中间存在一种天然的隔板,理应予以先后的判断,不可混为一谈。即实然的事实因果关系只有经过应然的价值评价,才能转化为刑法上的因果关系。

2. 传统因果关系二元论

随着条件说等因果关系单一论受到日益严厉的批判,学界逐渐意识到刑法学作为规范学科,应将事实因果关系及时纳入自行建立的规范体系当中。[28]而后,将事实因果关系与法律因果关系相区分,提出了归因与归责的因果关系二元论的判断范式,通过规范论的视角对条件说等进行法律上的限缩,相当因果关系等理论也便应运而生。然而,这一理论有着天然的弊端。其一,相当性理论在排除异常因果流程之时大显身手,但在风险降低但仍具相当性的场合便难以有所作为,会将本不该当构成要件的行为宣誓其行为无价值,而后借助紧急避险等排除其违法性;其二,对于相当性的判断因缺乏明确性而难以得出一致性的结论。例如,自动驾驶汽车交通肇事后,其背后的设计者、生产者在何种情况下应该予以归责,这其中的相当性的判断依据无从得知;其三,如上文所述,相当因果关系在判断的过程中杂糅了注意义务理论,使得其内部出现了循环论证的理论怪圈,无意中提升了过失犯归责的判断难度。

因此,最为理性的解决方法便是跳出上述传统过失犯归责的窠臼,将目光转向客观归责论,从而为自动驾驶汽车肇事的教义学规制提供内在的理论支撑。

(三)客观归责:从行为归责到结果归属的阶层归责模式

客观归责论发轫于德国,伴随着工业化进程的加快以及风险社会加速扩张,其范围也覆盖到了整个客观构成要件理论。客观归责论“以规范的客观目的为客观归责的指导原则,分成制造风险、实现风险与构成要件效力范围三个阶段,确

立具体的阻却客观归责的规制”。[29]客观归责论是确定因果关系理论,抑或实质的构成要件理论,这在学说上存在争议,[30]本文坚持客观归责论是实质的构成要件理论,它对客观的构成要件符合性提出了一种新的解释,并以“制造法不允许的风险与实现法不允许的风险”为中心进行解释。

当中涉及的是构成要件行为理论和结果归责理论,从法秩序的角度通过具体的下位规则将客观构成要件精致化,解决的是行为人到行为再到结果三者之间客观行为不法的问题。如前所述,客观归责论的三个核心规则分别是:制造法不允许的风险、实现法不允许的风险和构成要件的效力范围,三者都在于填充“客观可归责性”的实质内涵,并呈现出具体的判断规制:第一,“制造法所不允许的风险”和“实现法所不允许的风险”这两个环节分别对应了构成要件行为与结果归属。前者主要解决实行行为的问题,[31]比如从实质上定义构成要件行为,后者主要是规范层面的因果关系理论,[32]以排除有因果流程相当性而不必受归责的行为。第二,客观归责的基准在于,一是行为的危险性,这依靠事前判断,二是危险之实现,这依靠事后判断。[33]第三,客观归责论认为归责的重心应落脚在客观行为与法规范提供的基准行为之间的期望偏离,以及由此制造并实现的法不容许之风险,[34]且明确区分了事实归因和规范归责。值得注意的是,客观归责论并没有去寻找统一的归责标准,而是将努力的方向放在搭建清晰的构成要件框架之上,并分门别类地将规则具体化,在归责问题日益复杂化的今天,或许是一个更加稳健的选择。[35]尽管客观归责论的下位规则并非其所独创,但是相比于传统过失理论,客观归责论最大的优势是完成了从行为归责到结果归属的理论架构及规则整合。其中,“制造法所不允许的风险”提供构成要件行为一个实质而共同的内涵,用“不被允许的风险”限制构成要件的效力范围,是构成要件理论上的一大进步,且“制造法不允许的风险与实现法不允许的风险”表述构成要件符合性的内涵,其最重要的意义在于客观定义行为不法。[36]

在存在论的维度,客观归责论是一种实行行为理论,对应客观归责论的第一环节即“行为制造法所不允许的风险”。质言之,制造法益风险的实质是违反行为规范,[37]由此,法不允许之风险成为了发动过失归责的直接依据。众所周知,故意犯罪实行行为由刑法条文予以明确规定,而过失犯罪因没有清晰且定型化的实行行为,在判断之时因其随意性似是而非的判断难保不出差池。即如果不以具体规则加以判断,作为因果关系起点的实行行为则语焉不详。[38]客观归责论用其第一阶层“行为制造法所不允许的风险”这一正面规则来明确实行行为,以行为人的举止行为与危害结果之间创设了法不允许的风险这一行为事实,作为过失犯注意义务偏离的现实依据,进而对过失实行行为进行框定。再加之其具体下位规则对日常行为进行反向排除。在故意行为中,利用“降低风险”等于没有制造风险来进行否定、“制造容许的风险”的行为没有造成法所反对的风险而否认实行行为;对于过失行为,则通过“义务违反关联性原理”说明义务违反与结果发生之间欠缺关联性,即因违反注意义务的行为所造成的结果,即使不违反注意义务而为,同样不可能完全避免出现这种结果的可能性的,[39]则不属于刑法上的实行行为。类似的对实行行为的限定标准不一一列举。

在规范论的维度,客观归责论也是一种结果归属理论,对应客观归责论的第二环节即“实现了法所不允许的风险”。结果犯中,在确定实行行为与构成要件结果后,还需审查构成要件结果是否是实行行为所涉风险的实现。这便是所谓的结果归属问题。[40]客观归责论对于认定“危害结果究竟应算到谁的头上”这一问题有着天然的优势,可以很好地解决行为与结果之间的关联性,进而决定结果归责。在我国传统的因果关系理论中,往往将事实判断与价值判断相混淆,完全将经验上的事实关联与规范评价糅为一体,甚至相互替代,呈现一种“你中有我、我中有你”的

判断态势,最终导致结果归属上出现偏差。反观客观归责论,其从法规范的视角探讨该风险是否在危害结果当中得以实现。可以说,在讨论规范归责之前,事实上的因果关系已经通过条件说等进行了判断,如果行为和结果之间有事实上的因果关系,那么才有必要用客观归责论来证成结果让行为人负责在法规范上是正当且有效的。其在思考逻辑上明确区分了事实归因与规范归责,在梳理案件事实以及判断责任归属方面拥有得天独厚的理论优势。

当然,客观归责论也在不同层面上遭受到了学者的质疑与批判。有学者认为客观归责论在讨论归责可能性之时,潜意识地肯定了因果关系的存在。[41] 例如,在雷电交加的暴风雨来临之际,劝人去树林里散步。客观归责论首先通过第一层次的判断,此时劝人散步的行为本就属于一种生活风险,而不为刑法所关注,即已经否认了其实行行为性,因而不纳入刑法的规制范围内。试问实行行为尚且不存在,又何谈在潜意识中肯定了因果关系的存在呢?该学者的质疑显然存在"盲人摸象"的嫌疑。回归到自动驾驶情境下,运用客观归责论的第一环节去判断过失的实行行为,恰好可以将行为人注意义务偏离的轨道进行重塑。例如,自动驾驶汽车肇事之时,作为自然人的驾驶员正在车中接打电话,难道要将肇事所致的第三人伤亡之结果与驾驶员接打电话的行为相关联吗?显然,在自动驾驶模式下,处于乘客角色的驾驶员接打电话的行为不属于刑法当中的实行行为,又何谈因果关系问题的讨论。还有学者认为,客观归责理论具有避免陷入杀人、伤害罪的实行行为判断困境的优势,但是客观归责论带来的最大问题是将构成要件行为的判断乃至限缩构成要件符合性的全部问题纳入其中,削弱了构成要件符合性概念所具有的罪刑法定主义机能。[42] 笔者认为并不如此,客观归责论的产生本就源于个罪实行行为的模糊,在过失犯等领域具有无法替代之用处。该学者既然认为客观归责论对部分犯罪实行行为定型化与构成要件的明确性具有积极作用,又怎能反过来以偏概全地认为其对构成要件机能进行了削弱呢?客观归责论在过失犯领域的运用,不但不会有违罪刑法定主义机能,反而能够适应新型社会涌现的大量问题。例如,现如今我们正处于一个风险急剧扩张的社会,不论交通运输、体育竞技、医疗卫生行业,还是工程作业等过失犯罪集中爆发的领域,因新兴科技的植入所带来的不可避免的风险仿佛一把把"达摩克利斯之剑",让公众的生活时刻处于一种不安当中。就像现如今无人驾驶汽车"算法黑箱"所导致的不可控风险,并不能被现在的社会公众所普遍接受。无论从政策学的视角或是法律视角,都应尽力维护人民的生活安宁,将有限的刑法资源投入到最迫切解决的社会问题当中去,便需要刑法在一定程度上有所扩张。客观归责论明确过失犯罪的构成要件,其利用自身的下位规则在客观构成要件方面合理地限制了刑法上风险的产出,它所要实现的目的在于限制、过滤客观构成要件中所包含的不法,具体通过风险概念及其规则体系限定人们由于其行为而招致的外部责任负担。[43] 恰好与自动驾驶汽车等风险领域的法益保护观念相互吻合。

客观归责论使个罪的构成要件更为清晰,也整合了具体的下位规则,为本文尝试运用该理论解决自动驾驶汽车过失肇事的归责难题提供了新路径。

## 四、难题的有效解决:客观归责论的精致化归责

如前所述,客观归责论在理顺案件事实及判断责任归属方面有着得天独厚的优势,即通过明晰实行行为与结果归属,将客观构成要件理论予以精致化,增强了理论研究的层次性与实务的可操作性。以下,本文将运用客观归责论逐一解决自动驾驶交通肇事带来的归责难题。

### (一)制造法不允许的风险:涉罪主体行为偏离注意义务轨迹之明晰

在客观归责论之下,进行过失犯行为判定的

前提是考察法所不允许的风险隶属于何人的风险支配领域,若行为人无法掌控因果流程的,则不能视为制造了法不允许的风险。如上文所述,与传统驾驶场景不同的是,自动驾驶场景下因自动驾驶系统的介入,使得对其风险支配领域的判断更为复杂化,实行行为偏离注意义务轨迹也显得杂乱无章。

1. 注意义务之明晰是刑事归责之基础

注意义务之明晰是实行行为偏离正轨后刑事归责之基础。自动驾驶汽车所涉主体众多,根据上文所作分类分别予以讨论。(1)驾驶使用者,即作为自然人的驾驶员之注意义务。车辆操控权从作为自然人的驾驶员手中剥离后,意味着人类驾驶员谨慎驾驶的注意义务也随之消灭。但在乘坐自动驾驶汽车前,其仍然有检查车辆是否适宜行驶的谨慎义务。[44](2)驾驶责任者,即自动驾驶系统背后的研发者与生产者之注意义务。自动驾驶系统虽然接替了作为自然人的驾驶员进行驾驶,从而对车辆进行实际控制。然而,自动驾驶汽车的此种高度自主性,驾驶者(使用者)对车辆施加的影响是十分有限的,因此其注意义务实质上是其背后的研发者与生产者之注意义务。

无人驾驶型自动驾驶汽车的运行完全靠车辆的自主驾驶,生产者的生产设计环节就如同基因一般决定着自动驾驶汽车的先天安全性能。根据《产品质量法》第 26 条之规定,[45]自动驾驶汽车的生产者负有车辆安全生产义务与安全保障义务。如果不符合安全生产标准,则自动驾驶汽车属于缺陷产品。而根据缺陷产品的类型,可以将自动驾驶汽车生产者的安全生产义务具体为两个维度:一是自动驾驶汽车的设计者维度。正是因为技术层面的支持使得设计者给予了自动驾驶系统自主判断的活力,保证系统运行安全且恰当是设计层面的基本义务。二是制造者维度。制造环节是一个庞大又复杂的过程,在生产、加工、组装等各个阶段都要时刻尽到安全生产的义务,经过反复试验确保自动驾驶汽车上路安全可靠、功能健全后才能投入市场。且在投入市场后,这两个维度的管理人应时刻保持警惕,设计者负担数据库系统更新义务,生产者负有产品警示与召回等义务。

自动驾驶汽车的后天安全性能取决于以上安全生产、保障义务的正确履行。投入市场流通后,自动驾驶汽车看似脱离了生产者的实际控制,但事实上仍受到生产者的实质影响,这也是自动驾驶汽车涉罪归责的前提与基础。然而,任何技术的成熟都不是一蹴而就的,自动驾驶汽车安全系数的提高,也是一个渐进的过程。基于此,若自动驾驶汽车涉罪时的系统缺陷超越了当下人类科学技术的认知水平,那么尽到合理注意义务的生产者,则因没有制造法所不允许的风险而不具有归责基础。基于以上对相关注意义务的梳理,随之而来的便是,判断行为偏离注意义务轨迹是否制造了法所不允许的风险,进而归于构成要件行为的问题。

2. 涉罪主体实行行为之分类建构

第一,作为自然人的驾驶员之实行行为明晰。在乘坐自动驾驶汽车前,作为自然人的驾驶员仍然有检查车辆是否适宜行驶的谨慎义务。类似的注意义务倘若作为自然人的驾驶员没有履行,导致自动驾驶过程中因系统故障扰乱正常行驶而造成法益侵害时,实行行为则是作为自然人的驾驶员因偏离了检查车辆适宜行驶的注意义务而创造的法所不允许的行为。质言之,车辆使用者对上述风险具有实际控制地位,即法所不允许的风险存在于其风险支配范围内,且该主体又没有合理履行支配义务,故而在此种情况下,跳出交通肇事罪的禁锢,转向过失致人重伤、死亡等犯罪,行为主体与责任主体便再次具有一致性。

第二,自动驾驶系统背后研发者与生产者之实行行为明晰。自动驾驶汽车上路行驶过程中,系统如何操控则很大程度上取决于其背后的技术支撑。表面上自动驾驶系统借助强大的数据算法可以对车辆进行自由控制,但实质上仍然取决于数据输入端对于障碍物、路况信息及交通法

规的正确录入与及时修补更新,其驾驶行为始终在人类理性与人格的延长线上,故其内部缺陷才是致使自动驾驶汽车发生重大交通事故的主要原因。基于此,自动驾驶汽车的研发者与生产者的注意义务也被框定在了以合理的技术设计自动驾驶系统的范围之内。上述技术要求便是研发者与生产者风险支配之领域,其拥有自动驾驶汽车的行车风险控制权限,发生交通事故而造成法益侵害时,实行行为是研发者与生产者对技术严密性把控不过关之注意义务的偏离,而此时的实行行为并非交通肇事罪之构成,而应隶属于产品责任类犯罪的构成要件。另外,如若发生交通事故时自动驾驶系统出现故障,但其他可以选择的行为具有同样的风险,那么也不能将研发者与生产者的行为看作实行行为。因为即使研发者与生产者履行技术严密性把控的注意义务,同样不可完全避免法益侵害结果出现的可能,那这一行为就不在构成要件的效力范围之内,因此不能定性为实行行为。

3. 涉罪主体分离困境之化解

自动驾驶场景下,面对行为主体与责任主体相互分离的异常样态,车辆发生重大事故就难以在交通肇事罪的范围内予以规制。既然作为实施法益侵害结果的自动驾驶汽车本身不能自负其责,那么自动驾驶车辆实控权的认定就成为了责任归属之关键突破点。根据以上论述,只要跳出交通肇事罪的构成要件,转向过失致人重伤、死亡罪,产品责任类犯罪(如生产、销售假冒伪劣产品罪),便可将行为主体与责任主体再次进行同一性整合。行为与结果的分离错位作为形式上的归责困境化解后,也为后续法益侵害结果的归属铺平了道路。

(二)实现法不允许的风险:自动驾驶涉罪二元因果关系之判断

“当谨慎义务的违反恰恰是侵害结果的客观上可归责的原因,它在构成要件上才是有意义的。”[46]简言之,制造了法所不允许的风险,只能彰示行为的规范违反性,但实现法不允许的风险之判断重心在于制造法不允许的风险的行为与法益侵害结果之间的规范关联。

1. 实然归因:事实因果关系之判断

自动驾驶汽车涉罪的归责难题很大程度上取决于上文所述的“算法黑箱”的存在,设计者输入端与自动驾驶系统输出端之间存在一个技术真空领域,以现有的技术无法清晰地了解在这之间机器提取了何种特征。而它的存在也成为了事实因果关系判断上的拦路虎。针对此种困境,有学者提出了“因果流程中间项”理论,认为如果认识到了该中间项,则说明可以预见到结果发生。[47]面对自动驾驶系统的“算法黑箱”难题,“因果流程中间项”的方案认为只需要预见到基本事实即可,从而去缓解事实因果不明的张力。笔者认为,该理论似乎没有意识到这原本是一个技术层面的存在论困境,又怎能通过意识层面的应然判断去架空实然理性呢?这种建立在新康德主义之下的实然与应然双时空的“唯我”主义思想,应该适时地向现象存在主义转变。为此,本文尝试将其事实层面的因果流程作不同情形的假设,以方便后续的规范归责讨论。一是研发者与生产者明知系统存在缺陷;二是研发者与生产者不明知系统存在缺陷。此处还可以分为三种不同类型:(1)自动驾驶系统的研发者与生产者研发生产的过程中没有发现程序错误且这种失误是可以避免的,投入市场后才得以显现且没有召回产品。(2)自动驾驶系统的研发者与生产者研发生产的过程中存在过失未发现程序错误,产品投入市场后由于第三人原因使得汽车肇事。(3)自动驾驶系统的研发者与生产者研发生产的汽车肇事,但是由于现有的技术水平下隐藏的系统程序错误无法避免。基于以上事实因果关系的假设,以下将进行刑法规范上的归责判断。

2. 应然归责:法律因果关系之评价

法律因果关系的判定建立在行为主体所为的实行行为之上。无论是霍金等科学家们的警告,还是各界人士的创意性思考都是超越现有生活事实的无限想象。无论人工智能主体地位肯

定说学者从立法拟制、独立意志角度的努力论证,还是否定说学者从罪刑法定、处罚不能维度的极力反对,都是建立在对未来不确定事项的基础之上的争论不休。与其整日构想人工智能具有独立意志而反控人类,[48]不如基于现实,从眼下最棘手的问题出发,去讨论自动驾驶汽车的研发者与生产者在何种范围内承担刑事责任。

在讨论研发者与生产者责任之前,还需排除一种情况。无论多么先进的传感器和控制软件,都不会完全没有造成或被损坏的风险,假若作为自然人的驾驶员(使用者)在特定情况下已经清晰预见自动驾驶系统可能因失灵而造成交通事故,或者任由未满 14 周岁的未成年人驾驶员操作,却视若无睹不采取任何措施,则应按照不作为承担相应的责任。例如,自动驾驶汽车的自然人驾驶员在乘坐汽车时多次在同一情况下发生故障,但恰好没有其他交通主体参与而未发生事故,此后驾驶员应在此种情况下主动接管驾驶或者排除故障。在自动驾驶系统控制车辆运行之时,之所以降低甚至免除作为自然人的驾驶员的注意义务,完全是建立在信赖原则的基础之上,自动驾驶系统的存在也是以解放人类双手为目的的,本质在于驾驶员的注意义务转移到了自动驾驶系统之上,一旦驾驶员知晓了故障的存在,那么对自动驾驶汽车日常维护以减少系统故障的注意义务便被违反,其不作为的行为已经偏离了合法行为基准,这体现了驾驶员对被害人生命或重大公私财产的漠视,其不但没有降低风险,反而增加了危险发生的概率,进而制造了法所不允许的风险。更为重要的是,驾驶员的不作为导致事故发生,这个因果流程是正常进行的,行为与结果之间存在常态关联,不存在因果流程的重大偏移。如果驾驶员履行了必要的注意义务,及时接管车辆或者对系统故障进行排除,则不会发生交通事故,因而具有结果避免可能性。可见,无论从事实的角度还是从规范视角上考虑,驾驶员此时所创设的风险都在事故发生中得以实现,因此,该损害结果应归属于驾驶员的行为。排除上述特殊情况下,仅需讨论在自动驾驶系统研发者及生产者过失的情况下的结果归属问题即可。[49]以下讨论分为三种情形:

第一种情形是自动驾驶系统的研发者与生产者研发生产的过程中没有发现程序错误,且这种失误是可以避免的,投入市场后才得以显现且没有召回产品。此时新的归责路径如下:首先,研发者与生产者制造了法所不允许的风险。研发者与生产者研发和生产的过程中没有尽到最大的注意义务去防止系统故障,严重威胁到产品使用者及无过错第三方的身体健康和生命安全,体现了研发者与生产者对他人人身、财产法益的漠视。即便不采取严格责任,其后续不但没有降低危险行为,反而以不召回的不作为增加了事故发生的概率,制造了法所不允许的风险。其次,研发者与生产者实现了法所不允许的风险。产品研发者与生产者疏忽大意所设计出的自动驾驶汽车上路因缺陷肇事,这个因果流程是正常进行的,设计行为与法益侵害结果之间存在常态的因果关联,不存在因果流程重大偏移的问题。[50]利用合义务替代行为可知,如果研发者与生产者履行了最大的注意义务或召回义务,那么生命危险及财产损失的后果将不会发生或减少到最小。相关法律法规对产品安全的保障,要求研发者与生产者履行安全生产与安全保障的义务。正是为了将新兴技术所带来的风险控制在合理的范围内,即使发生危险,履行了后续产品召回的义务也能及时排除或者降低,但此种情形下,研发者与生产者视国家法律于不顾,产品投入市场显现危害之时不履行及时召回的义务,这正是相关法律法规所要防止的风险,被害人的财产及生命法益恰好被这些规范的保护目的所覆盖。因此,从法规范的视角,研发者与生产者创设的风险在法益侵害结果生成的过程中得以实现。再次,贯穿于上述两点过程中的构成要件效力范围判断方面,自动驾驶汽车发生交通事故致使法益侵害结果发生,既不能由第三人负责,也不能由被害人自我答责。众所周知,消费者购买自动驾驶汽

车时,虽知道其本身具有一定风险,但决没有对发生损害的结果作出承诺。消费者不同意自动驾驶汽车存在系统缺陷而致使人身损害,更遑论是更大的生命及财产损害危险。此时定要揭开自动驾驶系统之面纱,将结果直接归属于研发者及生产者的行为。最后,有学者可能会认为这完全符合传统汽车召回模式,不涉及刑事责任问题,用其他社会治理的方法即可予以规制。然而,较一般人而言,自动驾驶汽车的研发者与生产者因其专业人员的身份,对人工智能行为所引发的风险具有更大的认识可能性,[51]且其基于自己的过失行为创设了一个风险,而这个风险恰好在其掌控范围内,那么针对这一风险就使得研发者与生产者居于由于该风险的先行行为而产生的保证人地位。先行行为的保证人义务所带来的风险恰好就是法所不允许的风险。即使产品投入市场一段时间,只要发现系统存在设计等问题,就有消除风险,也即立刻召回的保证人义务,如果违反义务进而导致危险在结果中得以实现,那么研发者与生产者依然要对结果负责,并为此承担相应的刑事责任。

第二种情形是自动驾驶系统的研发者与生产者在研发生产的过程中因工作失误未发现程序错误,产品投入市场后由于第三人故意操作的原因使得汽车肇事,当属于在行为人创造之法不容许风险的实现过程中,由于因果关系发生重大偏离,而不具有可归责性。首先,研发者与生产者制造了法不允许的风险。此时与第一种情况下的风险制造同理,研发者与生产者研发生产的过程中没有尽到最大的注意义务,由于工作失误导致程序错误未被发现,体现了其对他人生命及财产法益的漠视,研发者与生产者的行为偏离了注意义务的轨迹,使得自动驾驶汽车肇事风险升高,制造了本可避免的肇事风险,为法规范所不允许。其次,研发者与生产者虽制造了风险,但第三人故意操作之行为同样制造了另一个风险。这就使得研发者与生产者所制造的风险在实现的过程中因果关系发生了偏离,而在第三人故意操作行为与研发者、生产者不当生产的行为竞合下,汽车肇事致使法益侵害的结果应如何分配应谨慎思考。客观归责论本身可以运用“合义务替代行为”之方法去判断义务违反关联性的大小,即义务违反与结果发生之间的关联程度,进而帮助查明事实。如若仍然事实不清、证据不足,则利用存疑时有利于被告原则排除结果归责。而在这种自动驾驶汽车肇事的情况下,运用“合义务替代行为”之方法,假设研发者与生产者在生产的过程中完全遵循了生产研发义务,没有任何与法规范相背离的违规操作。如若第三人的故意操作行为足以使得肇事结果发生,则应切断研发者与生产者行为与肇事结果之间的关联,不应将结果归属于研发者与生产者。同样,如若第三人故意操作导致研发者与生产者之生产研发行为事实无法查明,则根据罪疑惟轻原则排除研发者与生产者之结果责任,切断因果关系不予归责。

第三种情形是自动驾驶系统的研发者与生产者研发生产的汽车肇事,但是由于现有的技术水平下隐藏的系统程序错误无法避免,当属于因为效用而容忍制造的风险,不具有可归责性。客观归责论本身包括“在允许性风险中排除归责”的子规则。客观归责理论的客观性格,明确体现在其所提出的判断公式中——只有当法益侵害结果,是行为人所创造之法不容许风险的实现的时候,才可以将结果归责给行为人。其中两个归责标准,无论是创造法不容许风险或者风险实现,都是从客观的角度来确认的。也就是说,对于结果归责而言,行为人必须创造出一个按照客观理智第三人之判断确实存在的非容许风险,并且此风险也必须以客观理智第三人得以理解的方式实现为侵害结果。而如此的风险判断,都和行为人的主观认知或态度无关。[52]反过来,如果行为人认识不到这一风险,但这一风险却在实践中实现了,如果以理智第三人之判断为标准,这种风险是一种需要容忍的风险,则不具有归责基础,比如,患者被送上手术台,患者就需要容忍符

合规范的手术失败带来的伤害。同理,在自动驾驶汽车技术发展中,容许的风险本质就是在法益保护与行为自由之间寻求平衡,当国民的行为自由更需要保护时,行为所致的法益侵害风险也会被法所允许。这种被容忍的风险确实存在,因为发展人工智能这样的高端科学技术,是目前全世界科学领域共同致力的目标之一,作为强人工智能的自动驾驶技术,是人类科技发展到一定阶段的结果,也是人类妥协与让步的结果,人类在享有这种科技进步带来的收益的同时,也需要容忍科技发展所带来的风险。如若不然,现有科技发展水平将会停滞不前。就像最初传统汽车上路行驶一样,本身就属于一种高风险的行为,对于"升级版"汽车的风险性更不可妄加制约。所以说这种风险已经被法律所允许,故在此时研发者与生产者的行为已经不再是实行行为,亦不能将结果归属于二者。这种情况恰好暴露了科技发展的原生弊病,也是一种需要容忍的风险。因此,只要研发者与生产者在研发与生产的过程中尽到了最大的注意义务,而只是基于现有的技术水平无法避免缺陷的产生,则生产研发行为不属于刑法中的实行行为,那么如果发生交通事故,刑法也应保持谦抑性原则,收起"严父"的面容不予规制。

自动驾驶汽车作为人工智能发展的标识产物给人们带来便利的同时,也会因为自动驾驶系统的介入而萌发出种种新型矛盾弊端。正是由于各种问题的产生,进而催生了刑法学者对自动驾驶汽车领域内刑事问题的纷繁争议。在人工智能不断向前发展的时代背景下,根据当前自动驾驶汽车的实际情况,在原有理论失范之时,就应寻求更合适的刑法教义学原理去明晰客观构成要件,去判断自动驾驶汽车肇事的实行行为,以规范归责将不属于刑法规制的行为排除,进而限定交通肇事罪的处罚范围,这大致可以作为理论方法的试金石。另外,应明确自动驾驶汽车所处的阶段,在相同阶段内去探讨结果归属抑或责任承担问题才会使话语同频,制造更灿烂的学术火花,进而以人工智能时代的刑法关切去有效解决自动驾驶汽车涉罪的归责难题,在保护或防范人民的生命、财产法益不受人工智能缺陷之侵蚀的同时,也不会阻碍自动驾驶技术的发展。长远地看,在这样新的归责理论体系下各种立场必然会达成一致。

---

**注释:**

〔1〕参见高铭暄、王红:《互联网+人工智能全新时代的刑事风险与犯罪类型化分析》,《暨南学报(哲学社会科学版)》2018年第9期。

〔2〕参见刘俊祥:《智能化时代中国政治学学科学术话语的创新》,《学术界》2020年第3期。

〔3〕"定速巡航"系统:按司机要求的速度开启后,不用踩油门踏板就使车辆以固定的速度行驶。采用了这种装置,当在高速公路上长时间行车时,司机不用再去控制油门踏板,减轻了疲劳,同时减少了不必要的车速变化,亦可以节省燃料。

〔4〕Gurney, Jeffrey K, "Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles", *Journal of Law, Technology & Policy*, Vol. 2013, Issue 2, Fall 2013, p. 255.

〔5〕张鸿涛等编:《物联网关键技术及系统应用》,北京:机械工业出版社,2017年,第236页。

〔6〕Ivó Coca - Vila, "Self - driving Cars in Dilemmatic Situations: An Approach Based on the Theory of Justification in Criminal Law", *Criminal Law and Philosophy*, Vol. 12, 2018, pp. 62 - 63.

〔7〕Stanila, Laura, "On the Necessity of Recognizing Artificial Intelligence as Subject to Criminal Law - The Case of Autonomous Vehicles", *Journal of Eastern - European Criminal Law*, Vol. 2019, Issue 2, 2019, p. 45.

〔8〕我国于2017年发布《新一代人工智能发展规划》和《国家车联网产业标准体系建设指南》,要求大力发展人工智能车联网产业,促进自动驾驶技术的进步。

〔9〕参见储陈城:《人工智能时代刑法归责的走向——以过失的归责间隙为中心的讨论》,《东方法学》2018年第3期。

〔10〕《刑法》第133条规定:"违反交通运输管理法规,因而发生重大事故,致人重伤、死亡或者使公私财产遭受重大损失的,处三年以下有期徒刑或者拘役;交通运输肇事后逃逸或者有其他特别恶劣情节的,处三年以上七年以下有期徒刑;因逃逸致人死亡的,处七年以上有期徒刑。"

〔11〕刘宪权:《涉人工智能犯罪中研发者主观罪过的认定》,《比较法研究》2019年第4期。

〔12〕参见〔德〕理查德·霍尼格:《因果关系与客观归责》,隗佳译,赵秉志、宋英辉主编:《当代德国刑事法研究》第1卷,北京:法律出版社,2017年,第70页。

〔13〕Dafni Lima, "Could AI Agents Be Held Criminally Liable:

Artificial Intelligence and the Challenges for Criminal Law", *South Carolina Law Review*, Vol. 69, Issue 3, Spring 2018, p. 685.

〔14〕参见袁曾:《人工智能有限法律人格审视》,《东方法学》2017 年第 5 期。

〔15〕参见姜涛、柏雪淳:《谁之责任与何种责任:人工智能的责任模式配置慎思》,《河南社会科学》2019 年第 4 期。

〔16〕广义的因果关系既包括因果关系本身,也包括结果归属,故结果归属模糊也体现为因果关系之模糊。

〔17〕Sabine Gless, "If Robots Cause Harm, Who Is to Blame: Self-Driving Cars and Criminal Liability", *New Criminal Law Review*, Vol. 19, Issue 3, Summer 2016, p. 432.

〔18〕参见许可:《人工智能的算法黑箱与数据正义》,《社会科学报》2018 年 3 月 29 日。

〔19〕参见〔德〕沃尔夫冈·弗里希:《客观结果归责理论的发展、基本路线与悬而未决的问题》,恽纯良译,赵秉志、宋英辉主编:《当代德国刑事法研究》第 1 卷,北京:法律出版社,2017 年,第 109 页。

〔20〕赵俊甫:《过失实行行为研究》,《中国刑事法杂志》2004 年第 6 期。

〔21〕〔日〕佐伯仁志:《刑法总论的思之道·乐之道》,于佳佳译,北京:中国政法大学出版社,2017 年,第 243 页。

〔22〕〔47〕参见〔日〕前田雅英:《刑法总论讲义(2008 年增修版)》,曾文科译,北京:北京大学出版社,2017 年,第 181 页。

〔23〕李川:《从注意义务违反到客观归责:医疗过失犯之理论嬗变与归责限缩》,《法学论坛》2014 年第 5 期。

〔24〕〔36〕参见陈子平:《刑法总论》,北京:中国人民大学出版社,2009 年,第 146-147、124 页。

〔25〕参见劳东燕:《过失犯中预见可能性理论的反思与重构》,《中外法学》2018 年第 2 期。

〔26〕Vgl Kindhäuser, *Zur Alternativstruktur des strafrechtlichen Kausalbegriffs*, ZIS, 2016, S. 574.

〔27〕〔28〕参见林钰雄:《新刑法总则》,北京:中国人民大学出版社,2009 年,第 119 页。

〔29〕〔33〕许玉秀:《当代刑法思潮》,北京:中国民主法制出版社,2005 年,第 476、503-504 页。

〔30〕关于这种争议,参见许玉秀:《当代刑法思潮》,北京:中国民主法制出版社,2005 年,第 481-484 页。

〔31〕第一层次的行为归责可以将过失犯罪实行行为定型化,并在教义学的视角下将以注意义务为核心的"意志归责"转向了制造法所不允许的风险之"规范归责"。实行行为是判断行为之违法性问题,是站在行为人当时的角度来看待行为的违法性,即属于事前判断。

〔32〕第二层次的结果归属则是"归因与归责"二分判断的因果关系理论。结果归属问题是在规范论层面以探讨创设的不容许风险是否在结果中得以实现,这是在危害结果发生后的事后判断,其主要任务是认定危害结果是否是行为人的"作品"。

〔34〕王霖:《自动驾驶场景下过失犯归责困境巡检与路径选择——以规范归责模式为视角》,《河北法学》2020 年第 3 期。

〔35〕参见劳东燕:《风险分配与刑法归责:因果关系理论的反思》,《政法论坛》2010 年第 6 期。

〔37〕周光权:《客观归责论在财产犯罪案件中的运用》,《比较法研究》2018 年第 3 期。

〔38〕参见周光权:《行为无价值论与客观归责理论》,《清华法学》2015 年第 1 期。

〔39〕参见周光权:《客观归责理论的方法论意义兼与刘艳红教授商榷》,《中外法学》2012 年第 2 期。

〔40〕参见劳东燕:《事实因果与刑法中的结果归责》,《中国法学》2015 年第 2 期。

〔41〕参见刘艳红:《客观归责理论:质疑与反思》,《中外法学》2011 年第 6 期。

〔42〕张明楷:《也谈客观归责理论兼与周光权、刘艳红教授商榷》,《中外法学》2013 年第 2 期。

〔43〕参见〔德〕乌尔斯·金德霍伊泽尔:《风险升高与风险降低》,陈璇译,《法律科学》2013 年第 4 期。

〔44〕譬如,行车前履行开启自动驾驶系统自检程序的注意义务,只有经过系统自检确认无故障之后才可正常行车。

〔45〕《产品质量法》第 26 条规定:"生产者应当对其生产的产品质量负责、产品质量应当符合下列要求:(一)不存在危及人身、财产安全的不合理的危险,有保障人体健康和人身、财产安全的国家标准、行业标准的,应当符合该标准。"

〔46〕参见孙运梁:《过失犯的客观归责:以结果避免可能性为中心》,《比较法研究》2017 年第 5 期。

〔48〕参见叶良芳:《人工智能是适格的刑事责任主体吗?》,《环球法律评论》2019 年第 4 期。

〔49〕首先,自动驾驶系统研发者及生产者故意设计生产程序故障汽车的概率凤毛麟角,少之又少。其次,如若出于故意,那么其归责路径也十分明确,制造法所不允许的风险与实现风险的判断没有疑问,结果当然归属于研发者或生产者,故本文在此不多赘述。

〔50〕参见孙运梁:《客观归责论在我国的本土化:立场选择与规则适用》,《法学》2019 年第 5 期。

〔51〕参见刘志强、方琨:《论人工智能行为法律因果关系认定》,《学术界》2018 年第 12 期。

〔52〕参见周漾沂:《从客观转向主观:对于刑法上结果归责理论的反省与重构》,《台大法学论丛》2014 年第 43 卷第 4 期。

〔责任编辑:邹秋淑〕